

# INTELLIGENT ANALYSIS OF BIOMEDICAL IMAGING DATA FOR PRECISION MEDICINE

EDITED BY: Kuanquan Wang, Shuo Li, Xiu Ying Wang, Jun Feng and  
Yong Xu

PUBLISHED IN: *Frontiers in Medicine*



# frontiers

## Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-83250-438-3

DOI 10.3389/978-2-83250-438-3

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [frontiersin.org/about/contact](https://frontiersin.org/about/contact)

# INTELLIGENT ANALYSIS OF BIOMEDICAL IMAGING DATA FOR PRECISION MEDICINE

Topic Editors:

**Kuanquan Wang**, Harbin Institute of Technology, China

**Shuo Li**, Western University, Canada

**Xiu Ying Wang**, The University of Sydney, Australia

**Jun Feng**, Northwest University, China

**Yong Xu**, Harbin Institute of Technology, Shenzhen, China

**Citation:** Wang, K., Li, S., Wang, X. Y., Feng, J., Xu, Y., eds. (2022). Intelligent Analysis of Biomedical Imaging Data for Precision Medicine. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-438-3

# Table of Contents

- 05 Editorial: Intelligent Analysis of Biomedical Imaging Data for Precision Medicine**  
Kuanquan Wang, Shuo Li, Xiuying Wang, Jun Feng and Yong Xu
- 08 Acute Angle of Multilobulated Contours Improves the Risk Classification of Thymomas**  
Xiaowei Han, Song Luo, Bing Liu, Yue Chen, Wenwen Gao, Yige Wang, Xiuxiu Liu, Hongwei Yu, Longjiang Zhang and Guolin Ma
- 18 Differentiation of Brain Abscess From Cystic Glioma Using Conventional MRI Based on Deep Transfer Learning Features and Hand-Crafted Radiomics Features**  
Linlin Bo, Zijian Zhang, Zekun Jiang, Chao Yang, Pu Huang, Tingyin Chen, Yifan Wang, Gang Yu, Xiao Tan, Quan Cheng, Dengwang Li and Zhixiong Liu
- 30 Pyramid-Net: Intra-layer Pyramid-Scale Feature Aggregation Network for Retinal Vessel Segmentation**  
Jiawei Zhang, Yanchun Zhang, Hailong Qiu, Wen Xie, Zeyang Yao, Haiyun Yuan, Qianjun Jia, Tianchen Wang, Yiyu Shi, Meiping Huang, Jian Zhuang and Xiaowei Xu
- 42 Red Blood Cell Classification Based on Attention Residual Feature Pyramid Network**  
Weiqing Song, Pu Huang, Jing Wang, Yajuan Shen, Jian Zhang, Zhiming Lu, Dengwang Li and Danhua Liu
- 54 A Robust Training Method for Pathological Cellular Detector via Spatial Loss Calibration**  
Hansheng Li, Yuxin Kang, Wentao Yang, Zhuoyue Wu, Xiaoshuang Shi, Feihong Liu, Jianye Liu, Lingyu Hu, Qian Ma, Lei Cui, Jun Feng and Lin Yang
- 66 Automatic Segmentation of Novel Coronavirus Pneumonia Lesions in CT Images Utilizing Deep-Supervised Ensemble Learning Network**  
Yuanyuan Peng, Zixu Zhang, Hongbin Tu and Xiong Li
- 78 MPMR: Multi-Scale Feature and Probability Map for Melanoma Recognition**  
Dong Zhang, Hongcheng Han, Shaoyi Du, Longfei Zhu, Jing Yang, Xijing Wang, Lin Wang and Meifeng Xu
- 88 Three-Dimensional Liver Image Segmentation Using Generative Adversarial Networks Based on Feature Restoration**  
Runnan He, Shiqi Xu, Yashu Liu, Qince Li, Yang Liu, Na Zhao, Yongfeng Yuan and Henggui Zhang
- 105 Deep Learning in Prostate Cancer Diagnosis Using Multiparametric Magnetic Resonance Imaging With Whole-Mount Histopathology Referenced Delineations**  
Danyan Li, Xiaowei Han, Jie Gao, Qing Zhang, Haibo Yang, Shu Liao, Hongqian Guo and Bing Zhang
- 114 U-shaped GAN for Semi-Supervised Learning and Unsupervised Domain Adaptation in High Resolution Chest Radiograph Segmentation**  
Hongyu Wang, Hong Gu, Pan Qin and Jia Wang

- 126 ***Cost-Sensitive Uncertainty Hypergraph Learning for Identification of Lymph Node Involvement With CT Imaging***  
Qianli Ma, Jielong Yan, Jun Zhang, Qiduo Yu, Yue Zhao, Chaoyang Liang and Donglin Di
- 135 ***Progression-Free Survival Prediction in Small Cell Lung Cancer Based on Radiomics Analysis of Contrast-Enhanced CT***  
Ningxin Chen, Ruikun Li, Mengmeng Jiang, Yixian Guo, Jiejun Chen, Dazhen Sun, Lisheng Wang and Xiuzhong Yao
- 145 ***Feasibility Study of Intelligent Three-Dimensional Accurate Liver Reconstruction Technology Based on MRI Data***  
Shaodong Cao, Huan Li, Suyu Dong and Zhenxuan Gao
- 152 ***RDHCformer: Fusing ResDCN and Transformers for Fetal Head Circumference Automatic Measurement in 2D Ultrasound Images***  
Chaoran Yang, Shanshan Liao, Zeyu Yang, Jiaqi Guo, Zhichao Zhang, Yingjian Yang, Yingwei Guo, Shaowei Yin, Caixia Liu and Yan Kang
- 162 ***PET and CT Image Fusion of Lung Cancer With Siamese Pyramid Fusion Network***  
Ning Xiao, Wanting Yang, Yan Qiang, Juanjuan Zhao, Rui Hao, Jianhong Lian and Shuo Li
- 173 ***Early COPD Risk Decision for Adults Aged From 40 to 79 Years Based on Lung Radiomics Features***  
Yingjian Yang, Wei Li, Yingwei Guo, Yang Liu, Qiang Li, Kai Yang, Shicong Wang, Nanrong Zeng, Wenxin Duan, Ziran Chen, Huai Chen, Xian Li, Wei Zhao, Rongchang Chen and Yan Kang
- 188 ***A Brain Tumor Image Segmentation Method Based on Quantum Entanglement and Wormhole Behaved Particle Swarm Optimization***  
Tianchi Zhang, Jing Zhang, Teng Xue and Mohammad Hasanur Rashid
- 207 ***Knee Bone and Cartilage Segmentation Based on a 3D Deep Neural Network Using Adversarial Loss for Prior Shape Constraint***  
Hao Chen, Na Zhao, Tao Tan, Yan Kang, Chuanqi Sun, Guoxi Xie, Nico Verdonchot and André Sprengers



## OPEN ACCESS

## EDITED AND REVIEWED BY

Alice Chen,  
National Cancer Institute (NCI),  
United States

## \*CORRESPONDENCE

Kuanquan Wang  
wangkq@ieee.org

## SPECIALTY SECTION

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

RECEIVED 12 August 2022

ACCEPTED 25 August 2022

PUBLISHED 26 September 2022

## CITATION

Wang K, Li S, Wang X, Feng J and Xu Y  
(2022) Editorial: Intelligent analysis of  
biomedical imaging data for precision  
medicine. *Front. Med.* 9:1017751.  
doi: 10.3389/fmed.2022.1017751

## COPYRIGHT

© 2022 Wang, Li, Wang, Feng and Xu.  
This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License](#)  
(CC BY). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Editorial: Intelligent analysis of biomedical imaging data for precision medicine

Kuanquan Wang<sup>1\*</sup>, Shuo Li<sup>2</sup>, Xiuying Wang<sup>3</sup>, Jun Feng<sup>4</sup> and Yong Xu<sup>1</sup>

<sup>1</sup>Faculty of Computing, Harbin Institute of Technology, Harbin, China, <sup>2</sup>Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, United States, <sup>3</sup>Faculty of Engineering, School of Computer Science, The University of Sydney, Sydney, NSW, Australia, <sup>4</sup>School of Information Science and Technology, Northwest University, Xi'an, China

## KEYWORDS

biomedical imaging, intelligent analysis, classification, segmentation (image processing), deep learning—artificial neural network

## Editorial on the Research Topic

### Intelligent analysis of biomedical imaging data for precision medicine

Biomedical imaging, such as X-ray, ultrasound, computed tomography (CT), magnetic resonance (MR), positron emission tomography (PET), and microscopic imaging, has been widely applied in clinical practices due to its capabilities in depicting physical anatomy and revealing functional and biochemical process of the human body. Biomedical images are becoming indispensable for more accurate diagnosis and precision medicine including pre-/intraoperative planning, survival prediction, and evaluation of the therapeutic response. However, the accuracy of the diagnostic decisions often heavily relies on the experience of radiologists, and thereafter inevitably is subject to inter-/intra-operator variations. Besides, due to the limitations of imaging techniques, the image artifacts, inhomogeneity in intensity, and low contrast between tissues also impede the accurate diagnosis. Thus, there is a demand to design intelligent analysis methods of biomedical imaging data for more efficient, objective, and effective precision medicine.

Intelligent analysis methods, based on biomedical imaging data, are designed to provide quantitative and qualitative evaluation for auxiliary diagnosis. In this Research Topic, the intelligent analysis methods are used to segment and recognize the region of interest (ROI) in different organs and tissues (Zhang T. et al., Peng et al., Ma et al., Li D. et al., Bo et al., Chen H. et al., Zhang D. et al., Zhang J. et al., He et al., Wang et al.), evaluate the histological risk (Han et al., Li D. et al., Bo et al., Yang Y. et al., Xiao et al.), locate, count and classify cells (Li H. et al., Song et al.), reconstruct and visualize the three-dimensional model (Cao et al.), predict the progression-free survival (Chen N. et al.) and measure organs (Yang C. et al.).

For evaluation of histological risk types, Han et al. introduced the acute angle between adjacent lobulations as a new quantitative indicator for the prediction of the risk classification on thymomas. In their method, the least absolute shrinkage and selection operator (LASSO) was adopted to make the feature selection and the individualized imaging nomogram was used to evaluate the prediction ability of the selected feature. The authors verified that the acute angle was significantly associated with the risk classification ( $p < 0.05$ ). For improving the diagnostic ability of prostate cancer, Li D. et al. divided the deep learning model into three parts: prostate gland segmentation, classification, and prostate cancer area segmentation. First of all, the prostate gland segmentation network was implemented to acquire the mask of the gland, based on which, the mask was moreover cropped on ADC, DWI, and T2WI sequences. Secondly, the prostate classification network was executed to determine whether the cropped area contained prostate cancer. Once the gland was abnormal, the prostate cancer area segmentation network was finally applied to segment the lesion area. The experimental results showed that all the accuracy, precision, and sensitivity had obtained an evident improvement. Bo et al. combined deep transfer learning with hand-crafted radionics features for classification between brain abscess and cystic glioma, which achieved higher accuracy than deep learning or hand-crafted based model. To understand the relationship between the chronic obstructive pulmonary disease (COPD) risks and age, Yang Y. et al. separated the subjects by COPD stages. Within each group, the age data is divided into eight equal intervals. Then, the survival Cox model was created on lung radionics features to estimate the risk probability of COPD. The evaluation metrics area under curve showed the excellent performance of the proposed model. To enrich the image and provide more details for the clinical diagnosis of lung cancer, Xiao et al. proposed a siamese pyramid fusion network for PET images and CT images fusion (PET-CT) to simultaneously display the metabolic and anatomical information. In order to validate the improvement of image fusion, the authors used five classification methods (multilayer perceptron, support vector classifier, random forest, K-nearest neighbor, and naive Bayes classifier) for training and testing on PET, CT, and PET-CT. The experimental data demonstrated that all the models based on PET-CT data obtained an evident improvement compared to PET and CT data.

For brain tumor segmentation, Zhang T. et al. creatively applied the wormhole theory on quantum-behaved particle swarm optimization (QPSO) for dealing with smeared and irregular shapes in medical image segmentation tasks. Even with the low contrast and high inhomogeneity in the medical image, the proposed QWPSO method recovered the contours of the tumor well which were consistent with the ground truth. In recent years, coronavirus disease 19 (COVID-19) has been spread around the world. Peng et al. designed an ensemble model consisting of deep supervised learning

networks (DeepLab V3+, U-Net, PAN, and FPN) for COVID-19 lesion segmentation. The proposed ensemble model achieved a better segmentation result than manual segmentation as 0.7279 in IOU metric and 92.4604 in Hausdorff distance metric. As the lymph nodes are highly relative to lung adenocarcinoma, it is necessary to identify lymph nodes from CT to make a better diagnosis and treatment. To identify the lymph node accurately, Ma et al. constructed a cost-sensitive uncertainty hypergraph learning (CSUHL) scheme. On the one hand, both epistemic and aleatoric uncertainty were adopted to improve the quality of pathological representation. A new hypergraph-based learning scheme was used to reconsider the correlation between samples to generate high-order representations. On the other hand, the scheme was devised to capture the cost sensitivity of negative samples and assign more weights to the lymph node. The loss function let the model focus on the patient with lung adenocarcinoma. In the experiment phase, the proposed model obtained the highest prediction accuracy in 0.95238 among state-of-the-art methods. It is evidently better than the method BC-GNN which has the second highest prediction accuracy in 0.91667.

Intelligent medical image segmentation has also become important in the orthopedic area. As the manual segmentation for knee bone and cartilage is tedious and subjective, Chen H. et al. chose a three-dimensional deep neural network (nnU-Net) as their baseline model, and the adversarial loss was selected to provide the prior shape constraints and expand the contextual receptive field for resampled volume segmentation. The proposed method was assessed on the public dataset SKI10 and achieved a score of more than 76 in the validation phase. It was proven that the method can either extract the area of healthy bone and cartilage accurately or the pathological cases. Similar to research in Chen H. et al., He et al. introduced an adversarial learning scheme into 3D U-Net. Additionally, a Squeeze and Excitation (SE) module was added to increase the weight of relevant features and decrease the weight of irrelevant features for liver segmentation. As for the malignant melanoma recognition tasks in a whole-slide image (WSI) with huge size, Zhang D. et al. first broke up the WSI into several patches to relieve the computational burdens. Then, the location information, predicted categories, as well as confident probabilities, were combined to acquire the recognition result of malignant melanoma. Thirdly, as the pathological features appear on different scales, a multi-scale feature fusion architecture was designed to enrich the lesion features. Meanwhile, for irregularly shaped lesion areas, the deformable convolution style residual blocks and channel attention mechanism were constructed to focus on the essential features and reduce the influence of noise.

For intelligent medical image analysis methods, segmenting the thin structure from low contrast and ambiguous images is still a great challenge. Zhang J. et al. attempted to fuse

multi-scale features within each layer called intra-layer pyramid-scale aggregation blocks (IPABs). The blocks generated two associations at both high and low scales. Besides, the pyramid skip connection and deep pyramid supervision were used for further enhancement. The performance of the proposed block was verified in three public datasets (DRIVE, STARE, and CHASE-DB1). The experimental data showed that the proposed block can effectively extract thin vessels and outperformed the current state-of-the-art methods. As biomedical images often come from different domains, directly applying the same intelligent analysis model to different domains may cause poor prediction results. To relieve this problem, Wang et al. put forward to model of the semi-supervised learning approach and unsupervised domain adaptation approach into the same framework. And generative adversarial network changed to predict the label for each pixel for leveraging the annotated and unannotated data in the segmentation task here.

Intelligent diagnosis of a pathological image often demands counting the number of positive cells to estimate the illness state. Li H. et al. selected the two-stage feature pyramid network as the baseline model. The anchor-based model predicted the categories and refined the anchors several times which is more suitable for cell counting tasks. Some researchers created an intelligent analysis method for predicting progress-free survival (PFS) in patients with cancer. For example, Chen N. et al. explored radiomics signatures from the contrast-enhanced CT images to predict the PFS of a patient with small cell lung cancer. At first, the image features, including shape, intensity, texture, and so on, were calculated from each tumor area. The univariate prognostic ability of features was estimated by Cox proportional hazard (CPH) model. Then, the variation inflation factor (VIF) method was performed to remove the redundant features. Random survival forests (RSFs) were further applied to simplify the features. Only the features with a high importance score could be selected to form the radiomics signature. Based on selected radiomics and clinical features, the prognostic model was created. Finally, 11 radiomics features are selected and the model can predict the PFS with high accuracy.

After reviewing this Research Topic, we conclude that the intelligent analysis of biomedical image data for precision medicine can improve the diagnosis procedure in a more efficient and scientific way. The featured researchers are devoted to obtaining more accurate diagnosis results for specialized clinical problems, and the experiment results show that their method is positive examples of precision medicine, which indicates that the methods could be assisted in making personalized clinical plans for diagnosis in the future.

## Author contributions

KW wrote the overall framework of the paper and summarized the contents of the edited articles. SL, XW, JF, and YX summarized the contents of their respective edited articles. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the National Natural Science Foundation of China under Grant 62272135.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



# Acute Angle of Multilobulated Contours Improves the Risk Classification of Thymomas

Xiaowei Han<sup>1†</sup>, Song Luo<sup>2†</sup>, Bing Liu<sup>3</sup>, Yue Chen<sup>3</sup>, Wenwen Gao<sup>3</sup>, Yige Wang<sup>3</sup>, Xiuxiu Liu<sup>3</sup>, Hongwei Yu<sup>3</sup>, Longjiang Zhang<sup>2</sup> and Guolin Ma<sup>3\*</sup>

<sup>1</sup> Department of Radiology, The Affiliated Drum Tower Hospital of Nanjing University Medical School, Nanjing, China,

<sup>2</sup> Department of Medical Imaging, Jinling Hospital, Medical School of Nanjing University, Nanjing, China, <sup>3</sup> Department of Radiology, China-Japan Friendship Hospital, Beijing, China

## OPEN ACCESS

### Edited by:

Kuanquan Wang,  
Harbin Institute of Technology, China

### Reviewed by:

Xiyang Liu,  
Xidian University, China  
Wen Cheng,  
Harbin Medical University Cancer  
Hospital, China

### \*Correspondence:

Guolin Ma  
maguolin1007@qq.com

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

Received: 20 July 2021

Accepted: 30 August 2021

Published: 29 September 2021

### Citation:

Han X, Luo S, Liu B, Chen Y, Gao W,  
Wang Y, Liu X, Yu H, Zhang L and  
Ma G (2021) Acute Angle of  
Multilobulated Contours Improves the  
Risk Classification of Thymomas.  
Front. Med. 8:744587.  
doi: 10.3389/fmed.2021.744587

**Background:** Computed tomography plays an important role in the identification and characterization of thymomas. It has been mainly used during preoperative evaluation for clinical staging. However, the reliable prediction of histological risk types of thymomas based on CT imaging features requires further study. In this study, we developed and validated a nomogram based on CT imaging and included new indices for individualized preoperative prediction of the risk classification of thymomas.

**Methods:** We conducted a retrospective, multicenter study that included 229 patients from two Chinese medical centers. All the patients underwent cross-sectional CT imaging within 2 weeks before surgery. The results of pathological assessments were retrieved from existing reports of the excised lesions. The tumor perimeter that contacted the lung (TPCL) was evaluated and a new quantitative indicator, the acute angle (AA) formed by adjacent lobulations, was measured. Two predictive models of risk classification were created using the least absolute shrinkage and selection operator (LASSO) method in a training cohort for features selection. The model with a smaller Akaike information criterion was then used to create an individualized imaging nomogram, which we evaluated regarding its prediction ability and clinical utility.

**Results:** A new CT imaging-based model incorporating AA was developed and validated, which had improved predictive performance during risk classification of thymomas when compared with a model using traditional imaging predictors. The new imaging nomogram with AA demonstrated its clinical utility by decision curve analysis.

**Conclusions:** Acute angle can improve the performance of a CT-based predictive model during the preoperative risk classification of thymomas and should be considered a new imaging marker for the evaluation and treatment of patients with thymomas. On the contrary, TPCL is not useful as a predictor for the risk classification of thymomas in this study.

**Keywords:** thymoma, computed tomography, risk classification, predictive model, nomogram

## INTRODUCTION

Thymomas are rare primary thymic epithelial neoplasms, and they account for <1% of all adult malignancies (1). These tumors are often located in the anterior mediastinum and have the potential for local invasion (1, 2). Traditionally, thymomas are usually divided into invasive (Masaoka stage III/IV) and non-invasive (Masaoka stage I/II) lesions according to the Masaoka–Koga clinical staging system (3, 4). Thymomas can also be histologically classified as A, AB, B1, B2, or B3 according to the WHO classification system (revised version of 2015), based on the morphology of epithelial cells and the ratio of lymphocytes to epithelial cells (5). WHO types B2 and B3 are typically considered to be more invasive and are associated with lower survival rates than types A, AB, and B1. Therefore, thymomas can be divided into a low-risk group (types A, AB, and B1) and a high-risk group (types B2 and B3) (6). CT plays an important role in the identification and characterization of thymomas. This imaging technique has been mainly used during preoperative evaluation for clinical staging (4). However, the reliable prediction of histological risk types of thymomas based on CT imaging features still needs further exploration (7, 8).

Previous studies have focused on the relationship between CT imaging findings and the WHO histological classification (9–15). For instance, one study reported that some features (contours, heterogeneous enhancement, infiltration of surrounding fat and lung, and node enlargement) are significantly associated with the WHO classification categories (13). Another study reported that the histological features of aggressive thymomas were significantly correlated with decreased doubling time (DT) and increased growth when DT was evaluated retrospectively and dynamically (14). However, these studies were based exclusively on the evaluation of inter-group CT imaging feature differences and did not include the development of models for classification prediction. Furthermore, other studies have explored the relationship between the tumor perimeter contacting the lung (TPCL) with postoperative pleural recurrence on preoperative CT findings (16, 17). One study developed a more objective and quantitative method, with which the authors measured the angle formed by adjacent lobulations to predict lung invasion by thymomas (18). In this previous study, the authors found that adjacent lung invasion can be precisely predicted by the multilobulated aspect of the thymoma when it includes at least one acute angle (AA). However, the relationship between these quantitative imaging indicators and the WHO histological classification has been rarely reported.

Until now, the classification of thymomas from preoperative CT imaging has mostly employed traditional morphological indicators, without reproducible individualized prediction models or the inclusion of objective and quantitative indicators (19–22). Therefore, in this study, we sought to investigate whether TPCL and an AA formed by adjacent lobulations could constitute quantitative and reliable predictors of thymoma classification. We hypothesized that these new quantitative imaging indicators could be used as independent factors in the development of a predictive model for the risk categories included in the WHO histological classification system of

thymomas. Moreover, we also hypothesized that this new model would outperform a model that only includes traditional morphological indicators. The goal was to develop and validate an imaging nomogram to be used in individualized prediction of the risk classifications of thymomas preoperatively using non-invasive data and with minimal demand on patients.

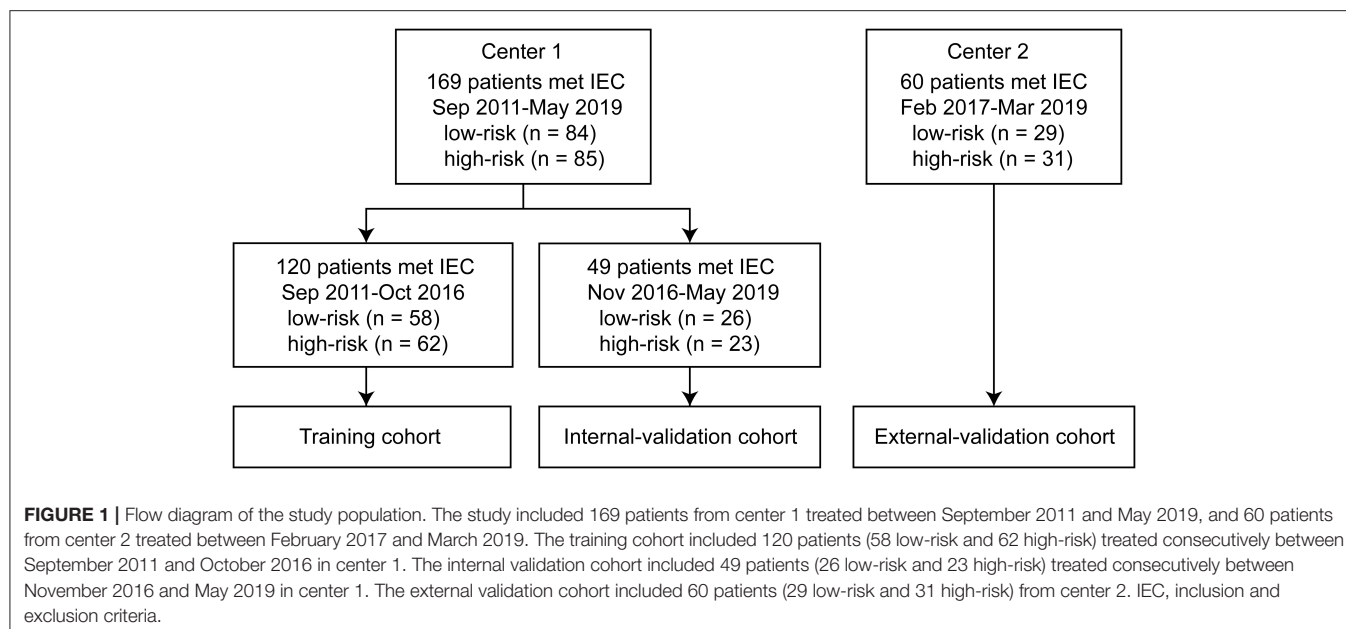
## METHODS

### Clinical Samples

We conducted a retrospective multicenter study that included 229 patients. Inclusion criteria were (i) thymoma diagnosed by postoperative pathological examination, (ii) contrast-enhanced CT examination performed, and (iii) CT imaging performed within 2 weeks before surgery. Exclusion criteria were (i) CT imaging performed after preoperative neoadjuvant chemotherapy; (ii) myasthenia gravis, hormone therapy, or other treatment options; (iii) CT artifacts that affected the assessment of the lesions; and (iv) recurrent anterior mediastinal mass after thymectomy. We included 169 patients treated between September 2011 and May 2019 in center 1 and 60 patients treated between February 2017 and March 2019 in center 2. We divided the patients into training, internal validation, and external validation cohorts. The training cohort included 120 patients (58 low-risk and 62 high-risk) treated consecutively between September 2011 and October 2016 in center 1. The internal validation cohort contained 49 patients (26 low-risk and 23 high-risk) treated consecutively between November 2016 and May 2019 in center 1. The external validation cohort included 60 patients (29 low-risk and 31 high-risk) from center 2 (**Figure 1**). We retrieved related clinical information on age, sex, symptoms, myasthenia gravis, clinical stages, and histological classification from the surgical records and pathological reports in the medical records database. Ethical approval was obtained from the institutional review boards of both center 1 (China-Japan Friendship Hospital in China) and center 2 (Jinling Hospital, Medical School of Nanjing University); the need for informed consent was waived because of the retrospective nature of this study.

### Computed Tomography Imaging

All the patients underwent CT imaging within 2 weeks before surgery and obtained the pathological examination results. They also underwent preoperative cross section spiral CT scanning examinations. CT images were obtained with a variety of scanners, namely, 16-row multi-detector CT (MDCT) (Aquilion; Toshiba, Tokyo, Japan), 320-row MDCT (Aquilion TM ONE; Toshiba, Tokyo, Japan), and 256-row MDCT (Revolution; GE Healthcare, Chicago, IL, United States) in center 1 and dual-source CT (Somatom Definition; Siemens Healthineers, Erlangen, Germany), 128-row MDCT (Somatom Perspective, Siemens Healthineers, Erlangen, Germany), and second-generation dual-source CT (Somatom Flash; Siemens Healthineers, Erlangen, Germany) in center 2. All images were obtained when the patients were in a supine position with suspended inspiration. An intravenously administered contrast medium was used in all the patients. The images were



reconstructed from both the mediastinal (window width, 400–450 HU; window level, 20–50 HU) and lung windows (window width, 1,000–1,500 HU; window level, –650 to –750 HU). Images with 5-mm slice thickness after reconstruction were used for evaluation, and original images with 0.6–1.25-mm slice thickness were available, when necessary, in all cases.

## Computed Tomography Imaging Interpretation and Pathological Examination Findings

All the CT images were reviewed using a picture archiving and communications system (PACS; GE Healthcare, Chicago, IL, United States) and were retrospectively reviewed by two radiologists with 10 years of experience, who were all blinded to the clinical details of the patients and pathological findings at the time of image interpretation. Where differences occurred, a third chest tumor radiologist with 23 years of experience addressed the differences for the final decision. The image interpretation criteria used standard reporting terms defined by the International Thymic Malignancy Interest Group (ITMIG) for anterior mediastinal masses suspected to be thymoma (23). Evaluated CT features included the following data about the primary mass and its surrounding structures: lesion location (tumors in the anterior mediastinum were classified into centrally located, right-sided, and left-sided lesions; any tumors that were located around or on the line running through the sternum were considered to be central); size in the x, y, and z axes; contour (smooth, single-lobulated, or irregular multilobulated); internal density (homogenous or heterogeneous); calcifications (without calcification, single, or multiple calcifications); infiltration of surrounding fat; tumor abutment  $\geq 50\%$  or  $<50\%$  of an adjacent mediastinal structure; and direct vascular endoluminal invasion. The following information regarding the surrounding structures was also included: adjacent lung abnormalities, pleural effusion

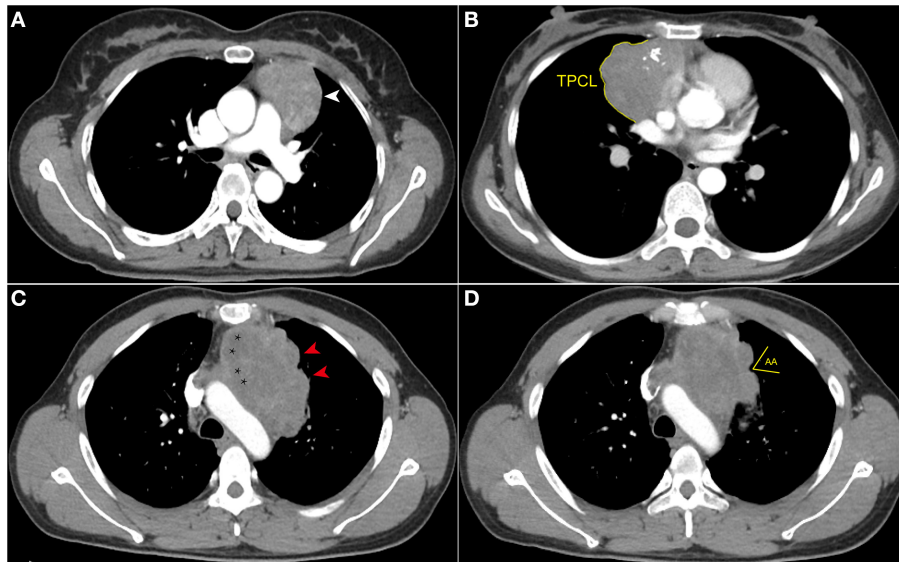
(without, unilateral, or bilateral), mediastinal lymph node enlargement ( $>1$  cm in short axis on an axial image), and phrenic nerve involvement (consistent with elevated hemidiaphragm). Any differences in findings were resolved on a consensual basis.

The thymomas were classified according to the 2015 revised WHO histology classification (5), which is mainly based on the morphology of epithelial cells and the ratio of lymphocytes to epithelial cells. When a tumor showed multiple histological components, it was classified based on the predominant component. The thymomas were divided into low-risk (types A, AB, and B1) and high-risk (types B2 and B3) subgroups, because types B2 and B3 are considered to be more malignant than type A, AB, or B1 (3, 6). A modified Masaoka–Koga stage was obtained by reviewing surgical records and pathological reports (24).

## Measurement of the New Quantitative Indices and Evaluation of Their Consistency

The maximum tumor diameter was measured at the largest section of the tumor on axial images (size\_max) by the same two radiologists who reviewed the CT images. The TPCL was measured by manually drawing the surface of the tumor contour adjacent to the lung, at the two maximum adjacent sections of the tumor shown on axial images. The average TPCL was then calculated as the final value. When the tumor contacted the bilateral mediastinal pleura, only the side overhanging the pleural cavity was measured to measure the TPCL. In multilobulated thymomas interfacing with the lung, the smallest angle formed by adjacent lobulations was uniformly measured twice on lung windows. The average angle was then calculated and classified as an AA or an obtuse angle (OA) (Figure 2). All tumors with smooth contours were counted as OA (18).

Intraclass correlation coefficients were used to determine intra- and inter-observer agreement in the measurement of new



**FIGURE 2 |** Computed tomography imaging evaluation of contour, internal density, and measurement of the new index. **(A)** Axial contrast-enhanced chest CT image obtained at the level of the pulmonary trunk demonstrated a thymoma with smooth contour (white arrowhead) in a 42-year-old woman (WHO classification AB). **(B)** Axial contrast-enhanced chest CT image showed a thymoma (WHO classification B1) in a 54-year-old woman. The multi-lobulated contour and internal calcifications were seen, and the tumor perimeter that contacted the lung (TPCL) was measured along the margin between tumor and adjacent lung (yellow arc line). **(C)** Axial contrast-enhanced chest CT image obtained at the level of the aortic arch demonstrated a thymoma (WHO classification B2) in a 49-year-old man who complained of chest distension and chest pain. The tumor has several obtuse angles formed by adjacent lobulations (red arrowhead), and low-density cystic or necrotic areas are indicated by stars. **(D)** In the same patient as **(C)**, the acute angle (AA) formed by adjacent lobulations can be seen in the upper slices.

indices. A pool of 50 patients was randomly selected from the cohorts, including 25 low-risk and 25 high-risk patients. For inter-observer agreement, two radiologists independently identified the cross-sectional images of the largest tumor area and measured the TPCL and the smallest angle in these patients at the same time. The intraclass correlation coefficients (ICCs) were calculated and analyzed between measurements. To evaluate intra-observer agreement, the TPCL and AA were all measured twice for each patient by each radiologist within 1 month, and the ICCs were separately calculated and analyzed.

## Feature Selection and Building of the Models

For imaging features defined by ITMIG terms and the new quantitative indices, regularized multivariate logistic regression with the least absolute shrinkage and selection operator (LASSO) penalty method was applied to the training cohort to reduce overfitting or any type of bias in feature selection (25). The selected features were then weighted by their respective coefficients in the regression equation formula as follows:

$$y = \beta_0 + \sum_{s=1}^n \beta_s X_s + \varepsilon \quad (1)$$

where  $y$  is 1 for patients with high-risk thymoma and 0 for low-risk patients,  $\beta_0$  is the constant term,  $n$  is the number of features used in the model,  $\beta_i$  ( $i = 0, 1, 2, \dots, n$ ) is the model parameter

of coefficient,  $X_j$  ( $j = 0, 1, 2, \dots, n$ ) is the feature, and  $\varepsilon$  is the error term.

The LASSO criteria for selecting parameters based on minimizing the value of the following cost equation:

$$\sum_{i=1}^N \left( y_i - \sum_{j=1}^n X_{ij} \beta_j - \beta_0 \right)^2 + \lambda \sum_{j=1}^n |\beta_j| \quad (2)$$

where  $N$  is the number of patients,  $y_i$  is the outcome labels of patient  $i$ ,  $n$  is the number of features,  $X_{ij}$  is the  $j$ th feature of the  $i$ th patient,  $\beta_i$  ( $i = 0, 1, 2, \dots, n$ ) is the model parameter of coefficient,  $\beta_0$  is the constant term, and  $\lambda$  is the regularization parameter.

Least absolute shrinkage and selection operator method is a shrinkage and selection method for linear regression (26). It aims to minimize the sum of squares of residual errors (MSE) under the condition that the sum of absolute values of a regression coefficient that is less than a constant is deleted during variable selection. We selected the optimal value of  $\lambda$  by leave-one-out cross-validation. We considered  $\lambda$  optimal if it minimized MSE and maximized the area under the receiver operating characteristic curve (AUC) in the training cohort. To test the robustness of the final number of features included in the model, we repeated the feature selection procedure at one SE of the optimal  $\lambda$  value (lambda 0.1se).

The model, including the new quantitative index, was compared with a model without this new index using the corresponding Akaike information criterion (AIC) values for

each model and their related statistical tests to estimate model complexity and data fitting performance. Furthermore, DeLong tests were performed to compare all pairs of receiver operating characteristic curve (ROC) (27).

## Construction and Evaluation of the Nomogram

We constructed a nomogram based on CT imaging features with the new quantitative index and applied it to predict the risk classification of thymomas. We assessed the accuracy of the nomogram using ROC curves. We then calculated AUCs and compared them between the training cohort and the two validation cohorts by DeLong tests (27). We also determined sensitivity and specificity.

We assessed the calibration of the nomogram using calibration curves and unreliability (U) statistics. We also conducted a decision curve analysis to evaluate the clinical utility of the nomogram by quantifying the net benefit of its use at different threshold probabilities in the validation datasets.

## Statistical Analysis

Statistical analysis was conducted using SPSS (version 23.0; IBM, Armonk, NY, United States), R (version 3.5.0; R Foundation, Vienna, Austria), STATA (version 15.0; StataCorp, College Station TX, United States), and MATLAB (version 2013a; Mathworks, Natick, MA, United States). A two-sided  $p < 0.05$  was used as a threshold for statistical significance.

## RESULTS

### Clinical Information and Imaging Features

The clinical characteristics of the patients are summarized in **Supplementary Table 1**. We found that imaging features, namely, size\_max, contour, density, calcification, and the new index, AA, were significantly associated with risk classification in the training cohort ( $p < 0.05$ ). Only AA was significantly associated with risk classification in the internal validation cohort ( $p < 0.05$ ). The inter-group statistical results of features, such as density, calcification, and pleural effusion, approached statistical difference. Size\_max, density, and AA were significantly associated with risk classification in the external validation training cohort ( $p < 0.05$ ); contour and calcification trended toward significance.

### Evaluation of the Consistency of the Measurement of the New Quantitative Indices

We found that the measurement of both TPCL and AA presented good inter- and intra-observer agreements. The ICCs were 0.9 for TPCL and 0.85 for AA between the two radiologists. For intra-observer agreement, the ICCs were 0.93 for TPCL and 0.88 for AA in one radiologist, and 0.92 for TPCL and 0.89 for AA in the other.

### Feature Selection and Model Development

From all the imaging features, 23 (with AA) were reduced to three potential predictors based on data from the 120

patients of the training cohort (8:1 ratio; **Figures 3A,B**). These three features were retained with non-zero coefficients in the LASSO logistic regression model with a minimum  $\lambda$  of 0.1052 and then used in the regression equation to build the first model. The result demonstrated that TPCL was ineligible for the risk classification of thymoma (11:1 ratio; **Figure 3**). Without considering the AA factor, 22 features were reduced to two predictors with a minimum  $\lambda$  of 0.1155 based on data from the training cohort. These two features were used in the regression equation to build the second model (**Supplementary Table 2**). We found significant differences between the two models for both AIC and AUC (**Supplementary Table 3; Figure 4**).

## Construction and Evaluation of an Individualized Nomogram

We used the first model with the smaller AIC to construct an individualized imaging nomogram incorporating three independent predictors: contour, density, and AA (**Figure 5**). The classification accuracies were 77.5, 73.47, and 70% in the training, internal validation, and external validation cohorts, respectively. The sensitivities were 74.19, 60.87, and 70.97%, and the specificities were 81.03, 84.62, and 68.97% in the training, internal validation, and external validation cohorts, respectively (**Supplementary Table 4; Figure 4**).

## Validation of Individualized Nomogram

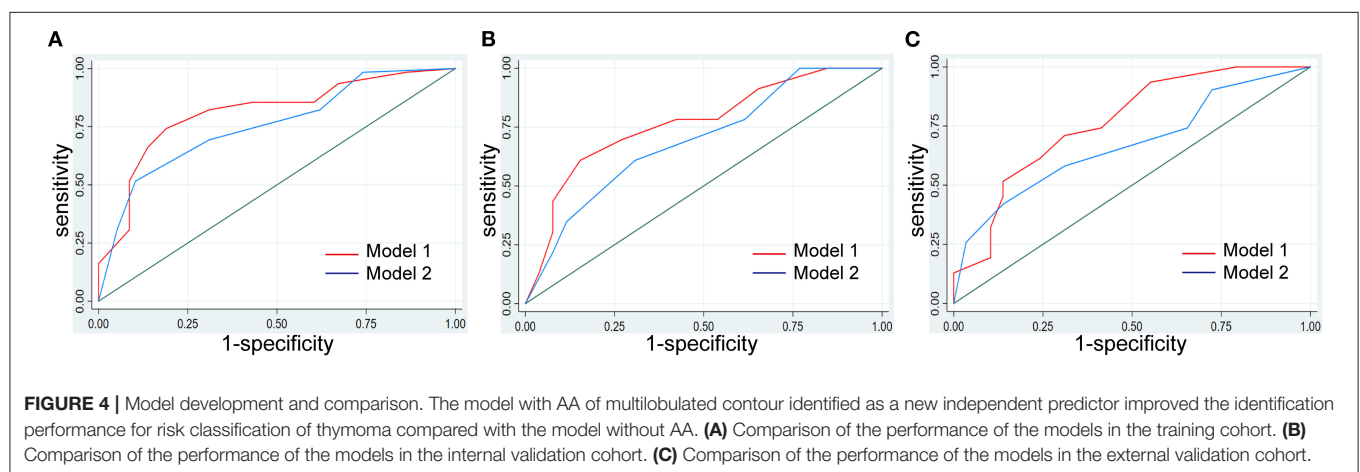
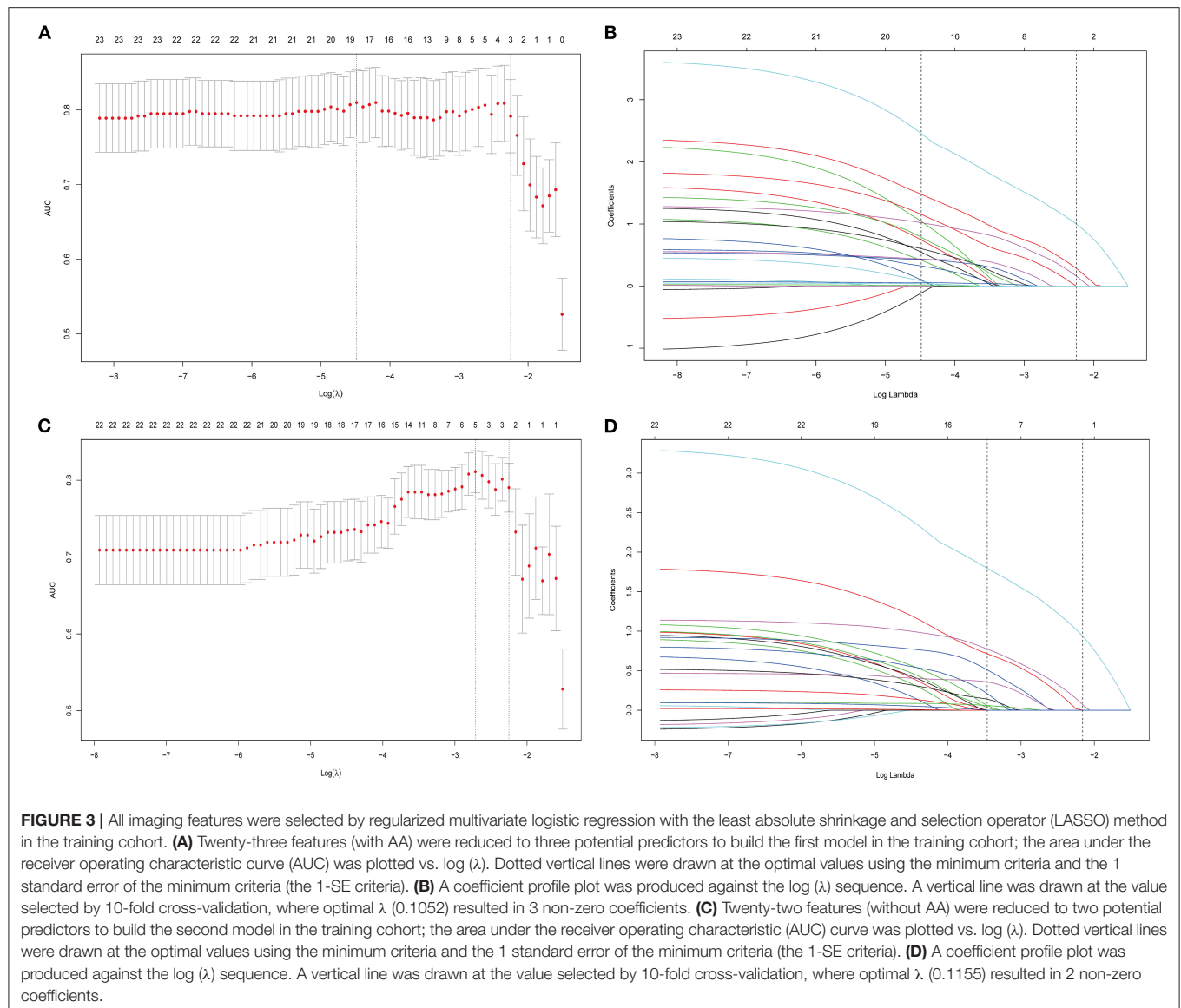
We found that calibrations for the probability of risk classification of thymomas were good in the three cohorts (training,  $p = 0.062$ ; internal validation,  $p = 0.267$ ; external validation,  $p = 0.14$ ). The C-index of the nomogram for the prediction of lymph node status was 0.811 (95% CI, 0.731 to 0.889), 0.766 (95% CI, 0.63 to 0.902), and 0.765 (95% CI, 0.644 to 0.886) in the training, internal validation, and external validation cohorts, respectively (**Supplementary Table 2; Supplementary Figure 1**).

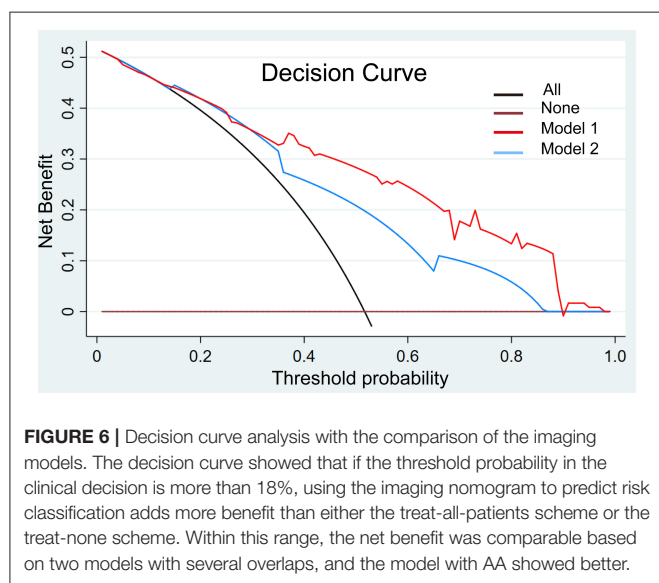
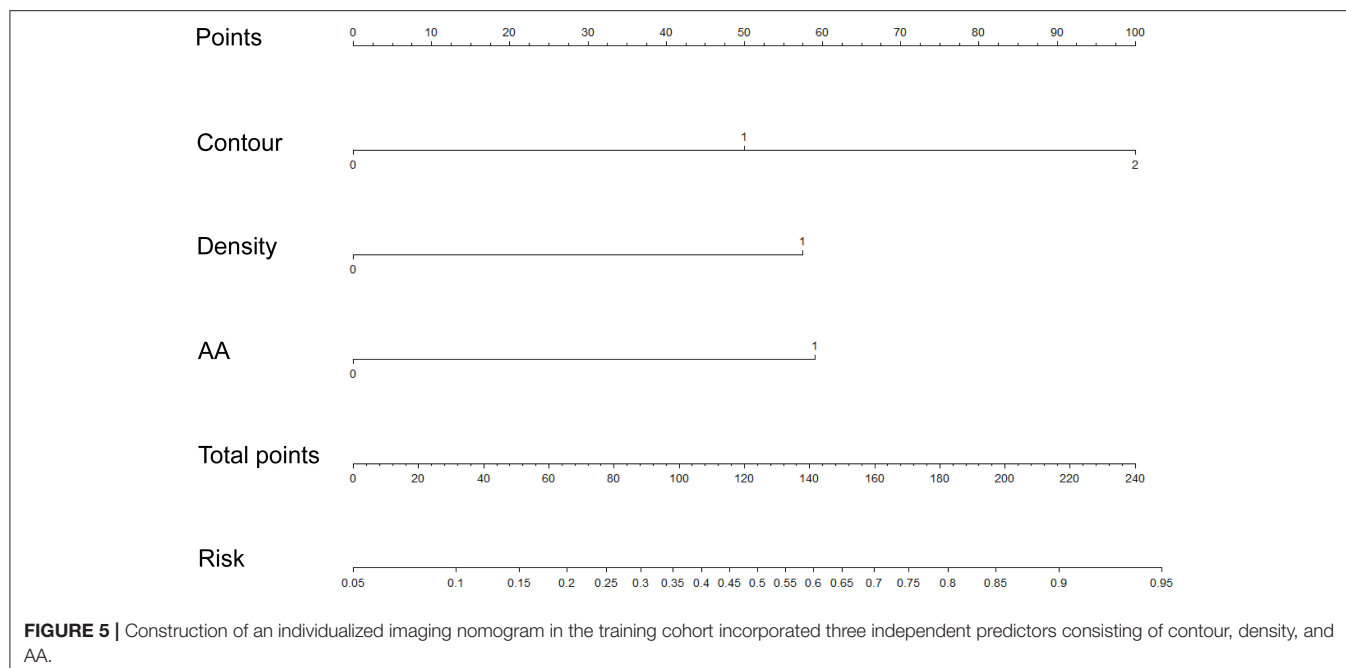
## Clinical Use

We present the decision curve analysis for the first and second models in **Figure 6**. The decision curve showed that if the threshold probability in the clinical decision is higher than 18%, using the radiomics nomogram to predict risk classification performs better than either the treat-all-patients or the treat-none schemes. Within this range, the net benefit was comparable between the two models with several overlaps.

## DISCUSSION

In this study, we developed and validated a diagnostic CT-based predictive model for individualized preoperative risk classification in patients with thymomas. The new model incorporates contour, density, and the new quantitative index AA. The new model successfully stratified the histological grading of tumors according to their risk classifications. We established a nomogram that can facilitate the clinical evaluation and treatment of thymomas.





In this study, all the patients underwent preoperative cross-sectional spiral CT examinations using the same body position and scanning methods. However, we should highlight that the CT images were obtained with different types of scanners in the two centers. This is a common limitation in multicenter studies, and we have performed the following measures to improve the consistency of the results. First, all the patients underwent preoperative cross-section spiral CT examinations. They were scanned in the same body position and same methods, although the CT images were obtained with different types of scanners. All the images were reconstructed with the same parameters and methods, thus ensuring consistent evaluation

and measurement of image quality and reducing errors in imaging processes among the different CT scanners (28, 29). In addition, the research focused on the imaging features and quantified indicators based on CT from the perspective of clinical diagnosis. Therefore, any error mainly comes from the experience of the radiologists and measurements. We introduced two new key quantitative indices, TPCL and AA (16–18) and showed that they can be consistently measured. Indeed, the results of the consistency analysis by the ICC evaluation of inter- and intra-observer agreements showed that the consistencies of both TPCL and AA measurements were satisfactory. The use of standardized techniques in this study minimized errors related to image assessment, assuring the quality of these new quantitative indices (30).

During the construction of the first and second models in this study, the candidate imaging features were reduced to three or two potential predictors using the LASSO method to obtain a subset of features with optimal stability and accuracy (25, 26). The LASSO regression algorithm is a regularized feature selection method with which all independent variables are processed simultaneously based on variance trade-off. Several previous studies have demonstrated the advantages of using LASSO to select parameters when constructing predictive models (25, 26, 31, 32). However, we are not aware of any study that has used this method for histological grading of thymomas. Tumor size, contours, internal density, calcification, and the new index AA all showed significant differences between low- and high-risk thymomas in this study. However, tumor size, TPCL, and calcification were not selected as independent risk factors in the final regression model. These results match those of previous studies in the field, although we should note that CT

imaging was shown to not differentiate between histological subtypes of thymomas in these previous studies. Of note, Johnson et al. reported an association between CT imaging and pathologic assessments of the tumor response after neoadjuvant treatment (21). Tumor size and calcification have been found to be related to type B3 thymoma in previous studies; however, we could not replicate these findings (33).

By comparing AICs and AUCs between the two models, we show that the new model that included AA has better predictive performance. The AIC is a standard method for estimating associations between model complexity and data-fitting performance (34). In this study, we used the AIC to examine whether the inclusion of three predictors in the first model would counteract the performance benefits by bringing additional complexity costs. We found that this is not the case. In fact, the differences between the two models in the AIC pinpointed that the benefits of including an extra predictor outweighed the costs caused by the increased complexity. We also compared the predictive performance of the two models by testing for differences between their AUCs by DeLong tests. We found that the new model with AA performed better in the identification of risk classifications in thymomas in all the three cohorts, but we determined that TPCL is ineligible for risk classification of thymoma. The presentation of an AA in tumors indicates a high-risk pathological type with high invasion potential. Therefore, we speculate that AA, being an independent predictor of risk category, is closely related to this pathological category of thymomas (18).

Thymomas are heterogeneous tumors, and their histopathology grading is associated with multiple imaging features (8, 35–37). Choe et al. reported that the histological features of aggressive tumors were significantly correlated with decreased DT and increased growth (14). In addition, Green et al. found that the presence of an AA between lobulations in multilobulated thymomas can predict lung invasion with satisfactory accuracy (18). Until now, there has been no study reporting the reliable identification of histological types of thymomas based on CT imaging features (36). We aimed to bridge this gap by constructing and validating a nomogram incorporating new quantitative imaging variables. The nomogram presents a good performance in generating individualized probabilities of the risk classification of thymomas. The results demonstrated good generalizability and provided a tool that can be used by both physicians and patients to perform preoperative individualized prediction of the risk classification of thymomas, following the current trends toward personalized medicine.

The introduction of the model in clinical practice has the potential to benefit both the diagnosis and treatment of patients with thymomas. We comprehensively evaluated the discrimination and calibration of the risk classification predictions and demonstrate the reliability of the results of the model. However, it is also important to consider whether the nomogram-assisted decisions would improve clinical outcomes. We performed a decision curve analysis to examine the clinical consequences of the tool based on different threshold probabilities and net benefits. We implemented this analysis by

calculating the proportion of true positives minus the proportion of false positives weighted by the relative harm of false-positive and false-negative results. We found that for threshold probabilities higher than 18%, the use of the nomogram to predict risk classification outperforms both the treat-all-patients and the treat-none schemes. This analysis demonstrates the clinical utility of the tool.

This study has some limitations. First, the incidence rate of thymoma was very low. It is very precious and time-consuming to obtain patients with complete clinical, imaging (qualified contrast-enhanced CT), and definite pathology results. Second, we relied on data collected retrospectively; 120 and 49 patients were used for model training and internal validation in this study, respectively, while only 60 patients were included for external validation. Compared with the training set and internal validation set, the proportion of the two risk types was relatively balanced and may support the results of this study. Third, the CT scanners used in this study and their parameters were different, and the diaphragm evaluation was based on the coronal reconstruction of images. However, we suggest that these parameters are unlikely to have strongly affected the evaluation of parameters that can be assessed and measured using CT images. Using high-dimensional data from a quantitative analysis of tumor volume and radiomics could improve the predictive models.

## CONCLUSIONS

In conclusion, we developed and validated a new CT-based model incorporating AA that can improve predictive performance during the risk classification of thymomas, when compared with traditional imaging predictors. By decision curve analysis, we demonstrated the clinical utility of the tool. AA should be considered as a new imaging marker for the evaluation and treatment of patients with thymomas.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

## ETHICS STATEMENT

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Ethical approval was obtained from the institutional review boards and the informed consents were waived from the patients for the retrospective nature of this study.

## AUTHOR CONTRIBUTIONS

XH and SL acquired, analyzed, explained the data, and drafted the manuscript. HY, YC, BL, and WG analyzed and explained the imaging data. YW and XL acquired the clinical information and revised the manuscript. LZ and GM designed the study and

revised the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was funded by Beijing science and technology planning project (Z211100003521009).

## ACKNOWLEDGMENTS

We sincerely thank Prof. Chen Huang (Department of Pathology, China-Japan Friendship Hospital) for retrieving the pathological data of the patients in this study, Dr. Liwen Zhang (CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences) for improving the quality of the plots, and Lu Zhang (Department of Science and Education, Shangluo Central Hospital, Shanxi Province, People's Republic of China) for polishing the revised edition of the manuscript.

## REFERENCES

- Marom EM. Advances in thymoma imaging. *J Thorac Imaging*. (2013) 28:69–80. doi: 10.1097/RTI.0b013e31828609a0
- Marom EM. Imaging thymoma. *J Thorac Oncol*. (2010) 5(10Suppl.4):S296–303. doi: 10.1097/JTO.0b013e3181f209ca
- Klimiec E, Quirke M, Leite MI, Hilton-Jones D. Thymus imaging in myasthenia gravis: the relevance in clinical practice. *Muscle Nerve*. (2018) 9:26096. doi: 10.1002/mus.26096
- Carter BW, Benveniste MF, Madan R, Godoy MC, Groot PM, Truong MT, et al. IASLC/ITMIG staging system and lymph node map for thymic epithelial neoplasms. *Radiographics*. (2017) 37:758–76. doi: 10.1148/rg.2017160096
- Marx A, Chan JK, Coindre JM, Detterbeck F, Girard N, Harris NL, et al. The 2015 world health organization classification of tumors of the thymus: continuity and changes. *J Thorac Oncol*. (2015) 10:1383–95. doi: 10.1097/JTO.0000000000000654
- Jeong YJ, Lee KS, Kim J, Shim YM, Han J, Kwon OJ. Does CT of thymic epithelial tumors enable us to differentiate histologic subtypes and predict prognosis? *Am J Roentgenol*. (2004) 183:283–9. doi: 10.2214/ajr.183.2.1830283
- Priola AM, Priola SM, Giraudo MT, Gned D, Fornari A, Ferrero B. Diffusion-weighted magnetic resonance imaging of thymoma: ability of the Apparent Diffusion Coefficient in predicting the World Health Organization (WHO) classification and the Masaoka-Koga staging system and its prognostic significance on disease-free survival. *Eur Radiol*. (2016) 26:2126–38. doi: 10.1007/s00330-015-4031-6
- Padda SK, Terrone D, Tian L, Khuong A, Neal JW, Riess JW, et al. Computed tomography features associated with the eighth edition TNM stage classification for thymic epithelial tumors. *J Thorac Imaging*. (2018) 33:176–83. doi: 10.1097/RTI.0000000000000310
- Yanagawa M, Tomiyama N. Prediction of thymoma histology and stage by radiographic criteria. *Thorac Surg Clin*. (2011) 21:1–12. doi: 10.1016/j.thorsurg.2010.08.008
- Shen Y, Ye J, Fang W, Zhang Y, Ye X, Ma Y. Efficacy of computed tomography features in predicting stage III thymic tumors. *Oncol Lett*. (2017) 13:29–36. doi: 10.3892/ol.2016.5429
- Hayes SA, Huang J, Golia Pernicka J, Cunningham J, Zheng J, Moskowitz CS. Radiographic predictors of resectability in thymic carcinoma. *Ann Thorac Surg*. (2018) 106:242–8. doi: 10.1016/j.athoracsur.2018.02.019
- Hayes SA, Huang J, Plodkowski AJ, Katzen J, Zheng J, Moskowitz CS, et al. Preoperative computed tomography findings predict surgical resectability of thymoma. *J Thorac Oncol*. (2014) 9:1023–30. doi: 10.1097/JTO.0000000000000204

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2021.744587/full#supplementary-material>

**Supplementary Figure 1 |** Calibration curves of the imaging nomogram in the (A) training cohort, (B) internal validation cohort, and (C) external validation cohorts. The y-axis represents the actual results of risk classifications. The x-axis represents the predicted results. The diagonal gray line represents a perfect prediction by an ideal model. The fit double arcs represent allowable errors in unreliability ( $U$ ) statistic. The dotted line represents the performance of the nomogram, of which a closer fit to the arc line represents a better prediction. Good calibrations were observed for the probability of risk classification of thymomas in the training ( $P = 0.062$ ), internal validation ( $P = 0.267$ ), and external validation cohorts ( $P = 0.14$ ).

**Supplementary Table 1 |** Clinical information and imaging features of patients in the training and validation cohorts.

**Supplementary Table 2 |** Risk factors for imaging features models.

**Supplementary Table 3 |** Evaluation of the imaging features models among cohorts.

**Supplementary Table 4 |** Evaluation of individualized nomogram among cohorts.

- Zhao Y, Chen H, Shi J, Fan L, Hu D, Zhao H. The correlation of morphological features of chest computed tomographic scans with clinical characteristics of thymoma. *Eur J Cardiothorac Surg*. (2015) 48:698–704. doi: 10.1093/ejcts/ezu475
- Choe J, Lee SM, Lim S, Choi SH, Kim N, Do KH, et al. Doubling time of thymic epithelial tumors on CT: correlation with histological subtype. *Eur Radiol*. (2017) 27:4030–6. doi: 10.1007/s00330-017-4795-y
- Luo H, Xie S, Ma C, Zhang W, Tschöpe C, Fa X, et al. Correlation between thymus radiology and myasthenia gravis in clinical practice. *Front Neurol*. (2019) 9:1173. doi: 10.3389/fneur.2018.01173
- Kato T, Iwano S, Taniguchi T, Kawaguchi K, Fukui T, Ishiguro F, et al. The contact length between the tumor contour and the lung on computed tomography is a risk factor for pleural recurrence after complete resection of thymoma. *Gen Thorac Cardiovasc Surg*. (2015) 63:343–8. doi: 10.1007/s11748-015-0525-z
- Do YW, Lee HJ, Narm KS, Jung HS, Lee JG, Kim DJ, et al. Tumor perimeter and lobulation as predictors of pleural recurrence in patients with resected thymoma. *Lung Cancer*. (2016) 98:79–83. doi: 10.1016/j.lungcan.2016.05.017
- Green DB, Eliades S, Legasto AC, Askin G, Port JL, Gruden JF. Multilobulated thymoma with an acute angle: a new predictor of lung invasion. *Eur Radiol*. (2019) 29:4555–62. doi: 10.1007/s00330-019-06059-1
- Ackman JB, Verzosa S, Kovach AE, Louissaint A Jr, Lanuti M, Wright CD, et al. High rate of unnecessary thymectomy and its cause. Can computed tomography distinguish thymoma, lymphoma, thymic hyperplasia, and thymic cysts? *Eur J Radiol*. (2015) 84:524–33. doi: 10.1016/j.ejrad.2014.11.042
- Khandelwal A, Sholl LM, Araki T, Ramaiya NH, Hatabu H, Nishino M. Patterns of metastasis and recurrence in thymic epithelial tumors: longitudinal imaging review in correlation with histological subtypes. *Clin Radiol*. (2016) 71:1010–7. doi: 10.1016/j.crad.2016.05.007
- Johnson GB, Aubry MC, Yi ES, Koo CW, Jenkins SM, Garces YI, et al. Radiologic response to neoadjuvant treatment predicts histologic response in thymic epithelial tumors. *J Thorac Oncol*. (2017) 12:354–67. doi: 10.1016/j.jtho.2016.09.118
- Okumura M, Yoshino I, Yano M, Watanabe SI, Tsuboi M, Yoshida K, et al. Tumour size determines both recurrence-free survival and disease-specific survival after surgical treatment for thymoma. *Eur J Cardiothorac Surg*. (2019) 56:174–81. doi: 10.1093/ejcts/ezz001
- Marom EM, Rosado-de-Christenson ML, Bruzzi JF, Hara M, Sonett JR, Ketani L. Standard report terms for chest computed tomography reports of anterior mediastinal masses suspicious for thymoma. *J Thorac Oncol*. (2011) 6:S1717–23. doi: 10.1097/JTO.0b013e31821e8cd6

24. Detterbeck FC, Nicholson AG, Kondo K, Van Schil P, Moran C. The Masaoka-Koga stage classification for thymic malignancies: clarification and definition of terms. *J Thorac Oncol.* (2011) 6:S1710–6. doi: 10.1097/JTO.0b013e31821e8c8f
25. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol.* (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128
26. Dong D, Tang L, Li ZY, Fang MJ, Gao JB, Shan XH, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol.* (2019) 30:431–8. doi: 10.1093/annonc/mdz001
27. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* (2015) 162:W1–73. doi: 10.7326/M14-0698
28. Dong D, Fang MJ, Tang L, Shan XH, Gao JB, Giganti F, et al. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. *Ann Oncol.* (2020) 31:912–20. doi: 10.1016/j.annonc.2020.04.003
29. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J.* (2019) 53:1800986. doi: 10.1183/13993003.00986-2018
30. Song J, Shi J, Dong D, Fang M, Zhong W, Wang K, et al. A new approach to predict progression-free survival in stage IV EGFR-mutant NSCLC patients with EGFR-TKI therapy. *Clin Cancer Res.* (2018) 24:3583–92. doi: 10.1158/1078-0432.CCR-17-2507
31. Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res.* (2017) 23:7253–62. doi: 10.1158/1078-0432.CCR-17-1038
32. Wang X, Zhao X, Li Q, Xia W, Peng Z, Zhang R, et al. Can peritumoral radiomics increase the efficiency of the prediction for lymph node metastasis in clinical stage T1 lung adenocarcinoma on CT? *Eur Radiol.* (2019) 29:6049–58. doi: 10.1007/s00330-019-06084-0
33. Han X, Gao W, Chen Y, Du L, Duan J, Yu H, et al. Relationship between computed tomography imaging features and clinical characteristics, Masaoka-Koga Stages, and World Health Organization histological classifications of thymoma. *Front Oncol.* (2019) 9:1041. doi: 10.3389/fonc.2019.01041
34. Coutant C, Olivier C, Lambaudie E, Fondrinier E, Marchal F, Guillemain F, et al. Comparison of models to predict nonsentinel lymph node status in breast cancer patients with metastatic sentinel lymph nodes: a prospective multicenter study. *J Clin Oncol.* (2009) 27:2800–8. doi: 10.1200/JCO.2008.19.7418
35. Pandey S, Jaipal U, Mannan N, Yadav R. Diagnostic accuracy of multidetector computed tomography scan in mediastinal masses assuming histopathological findings as gold standard. *Pol J Radiol.* (2018) 83:e234–42. doi: 10.5114/pjr.2018.76709
36. Carter BW, Benveniste MF, Truong MT, Marom EM. State of the art: MR imaging of thymoma. *Magn Reson Imaging Clin N Am.* (2015) 23:165–77. doi: 10.1016/j.mric.2015.01.005
37. Hwang Y, Park IK, Park S, Kim ER, Kang CH, Kim YT. Lymph node dissection in thymic malignancies: implication of the ITMIG lymph node map, TNM stage classification, and recommendations. *J Thorac Oncol.* (2016) 11:108–14. doi: 10.1016/j.jtho.2015.09.001

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Han, Luo, Liu, Chen, Gao, Wang, Liu, Yu, Zhang and Ma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Differentiation of Brain Abscess From Cystic Glioma Using Conventional MRI Based on Deep Transfer Learning Features and Hand-Crafted Radiomics Features

Linlin Bo<sup>1†</sup>, Zijian Zhang<sup>2†</sup>, Zekun Jiang<sup>1</sup>, Chao Yang<sup>1</sup>, Pu Huang<sup>1</sup>, Tingyin Chen<sup>3</sup>, Yifan Wang<sup>3</sup>, Gang Yu<sup>1</sup>, Xiao Tan<sup>2</sup>, Quan Cheng<sup>4,5\*</sup>, Dengwang Li<sup>1\*</sup> and Zhixiong Liu<sup>4,5\*</sup>

## OPEN ACCESS

### Edited by:

Yong Xu,  
Harbin Institute of Technology,  
Shenzhen, China

### Reviewed by:

Qi Zhu,  
Nanjing University of Aeronautics and  
Astronautics, China  
Chunshan Li,  
Harbin Institute of Technology, China

### \*Correspondence:

Zhixiong Liu  
zhixiongliu@csu.edu.cn  
Dengwang Li  
dengwang@sdnu.edu.cn  
Quan Cheng  
chengquan@csu.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

Received: 27 July 2021

Accepted: 06 October 2021

Published: 12 November 2021

### Citation:

Bo L, Zhang Z, Jiang Z, Yang C,  
Huang P, Chen T, Wang Y, Yu G,  
Tan X, Cheng Q, Li D and Liu Z (2021)  
Differentiation of Brain Abscess From  
Cystic Glioma Using Conventional MRI  
Based on Deep Transfer Learning  
Features and Hand-Crafted Radiomics  
Features. *Front. Med.* 8:748144.  
doi: 10.3389/fmed.2021.748144

<sup>1</sup> Shandong Key Laboratory of Medical Physics and Image Processing, Shandong Institute of Industrial Technology for Health Sciences and Precision Medicine, School of Physics and Electronics, Shandong Normal University, Jinan, China, <sup>2</sup> Department of Oncology, Xiangya Hospital, Central South University, Changsha, China, <sup>3</sup> Department of Network Information Center, Xiangya Hospital, Central South University, Changsha, China, <sup>4</sup> Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha, China, <sup>5</sup> National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

**Objectives:** To develop and validate the model for distinguishing brain abscess from cystic glioma by combining deep transfer learning (DTL) features and hand-crafted radiomics (HCR) features in conventional T1-weighted imaging (T1WI) and T2-weighted imaging (T2WI).

**Methods:** This single-center retrospective analysis involved 188 patients with pathologically proven brain abscess (102) or cystic glioma (86). One thousand DTL and 105 HCR features were extracted from the T1WI and T2WI of the patients. Three feature selection methods and four classifiers, such as k-nearest neighbors (KNN), random forest classifier (RFC), logistic regression (LR), and support vector machine (SVM), for distinguishing brain abscess from cystic glioma were compared. The best feature combination and classifier were chosen according to the quantitative metrics including area under the curve (AUC), Youden Index, and accuracy.

**Results:** In most cases, deep learning-based radiomics (DLR) features, i.e., DTL features combined with HCR features, contributed to a higher accuracy than HCR and DTL features alone for distinguishing brain abscesses from cystic gliomas. The AUC values of the model established, based on the DLR features in T2WI, were 0.86 (95% CI: 0.81, 0.91) in the training cohort and 0.85 (95% CI: 0.75, 0.95) in the test cohort, respectively.

**Conclusions:** The model established with the DLR features can distinguish brain abscess from cystic glioma efficiently, providing a useful, inexpensive, convenient, and non-invasive method for differential diagnosis. This is the first time that conventional MRI radiomics is applied to identify these diseases. Also, the combination of HCR and DTL features can lead to get impressive performance.

**Keywords:** brain abscess, deep transfer learning, radiomics, convolutional neural network, cystic glioma

## INTRODUCTION

Brain glioma is the most common intracranial brain tumor that is extremely difficult to treat. Currently, surgical resection is the standard treatment of resectable diseases, followed by postoperative radiotherapy and chemotherapy (1). The majority of gliomas are solid tumors, but some present cystic changes, such as cystic glioma, which has different clinicopathological features from other tumors. Brain abscess is an infectious disease that has high morbidity and mortality (2, 3). Though the treatment and prognosis of these two diseases are different, accurate and timely differential diagnosis is crucial. In many cases, CT and MR images lack specificity for cystic glioma and brain abscess, especially when the medical history and clinical manifestations of the diseases cannot provide a differential diagnosis for timely treatment measures. At present, the two diseases are mainly distinguished by pathological examination, with the caveat of invasive procedure and intra-operator variability. To accurately distinguish the two diseases, previous studies have proposed advanced MR images diagnosis techniques (2, 4), such as susceptibility-weighted imaging and apparent diffusion coefficients (ADC). However, these diagnosis techniques cannot obtain high accuracy, and they rely on the experience of radiologists (5). The use of the most rudimentary imaging modalities of T1-weighted imaging (T1WI) and T2-weighted imaging (T2WI) for a training model with a large sample size contributes to more universality and fewer errors.

As a method of machine learning, radiomics is used for quantitative image feature extraction from tumor regions of interest. It has great potential for oncology practice, including differential diagnosis, prediction of pathological classification, lymph node metastasis, and survival (6–10). Radiomics has been applied to brain tumor diseases (11–15), especially in differentiating brain tumors (16–21). For example, Qian et al. investigated the ability of radiomic analysis to distinguish between isolated brain metastases and glioblastoma (16); Dong et al. used the radiomic features derived from the areas of peripheral enhancing edema to differentiate glioblastoma from supratentorial single brain metastasis (17); Zhang et al. investigated the feasibility of contrast-enhanced T1WI radiomics features extracted by machine-learning algorithms to distinguish between low-grade oligodendroglioma and atypical anaplastic oligodendroglioma (18); Chen et al. applied radiomics analysis to distinguish between metastatic brain tumors and glioblastomas based on contrast-enhanced T1WI, and they validated the discriminative performance of this method (19); Artzi et al. used radiomics-based machine learning to differentiate between brain metastasis subtypes and glioblastoma based on conventional postcontrast T1WI (20). However, the radiomics

features are mainly the texture, size, volume, shape, and intensity characteristics of the tumor, limiting the potential of this method. Therefore, extracting more complex features and fusing them with radiomic features may improve the prediction and generalization capabilities of the model (21–23).

In recent years, deep convolutional neural networks (CNNs) (24) with complex network structures have achieved remarkable results in the field of computer vision, such as tumor grade prediction, patient prognosis, pathology classification, and organ segmentation (25, 26). The successful application of deep learning requires a large number of training cohort sets. Since the available medical data sets have a limited size, a pretrained CNN known as “transfer learning” can be employed to avoid overfitting and replace deep learning in many practical applications (21, 27, 28).

It is not clear whether T1WI and T2WI, as conventional routine images in hospitals, also have diagnostic values for distinguishing brain abscess from cystic glioma. In this study, we hypothesized that conventional T1WI and T2WI would also be valuable in distinguishing between these two diseases. To this end, the DLR features extracted from patients with brain abscesses or cystic gliomas were used to validate the diagnostic capability of T1WI and T2WI.

## MATERIALS AND METHODS

### Patients

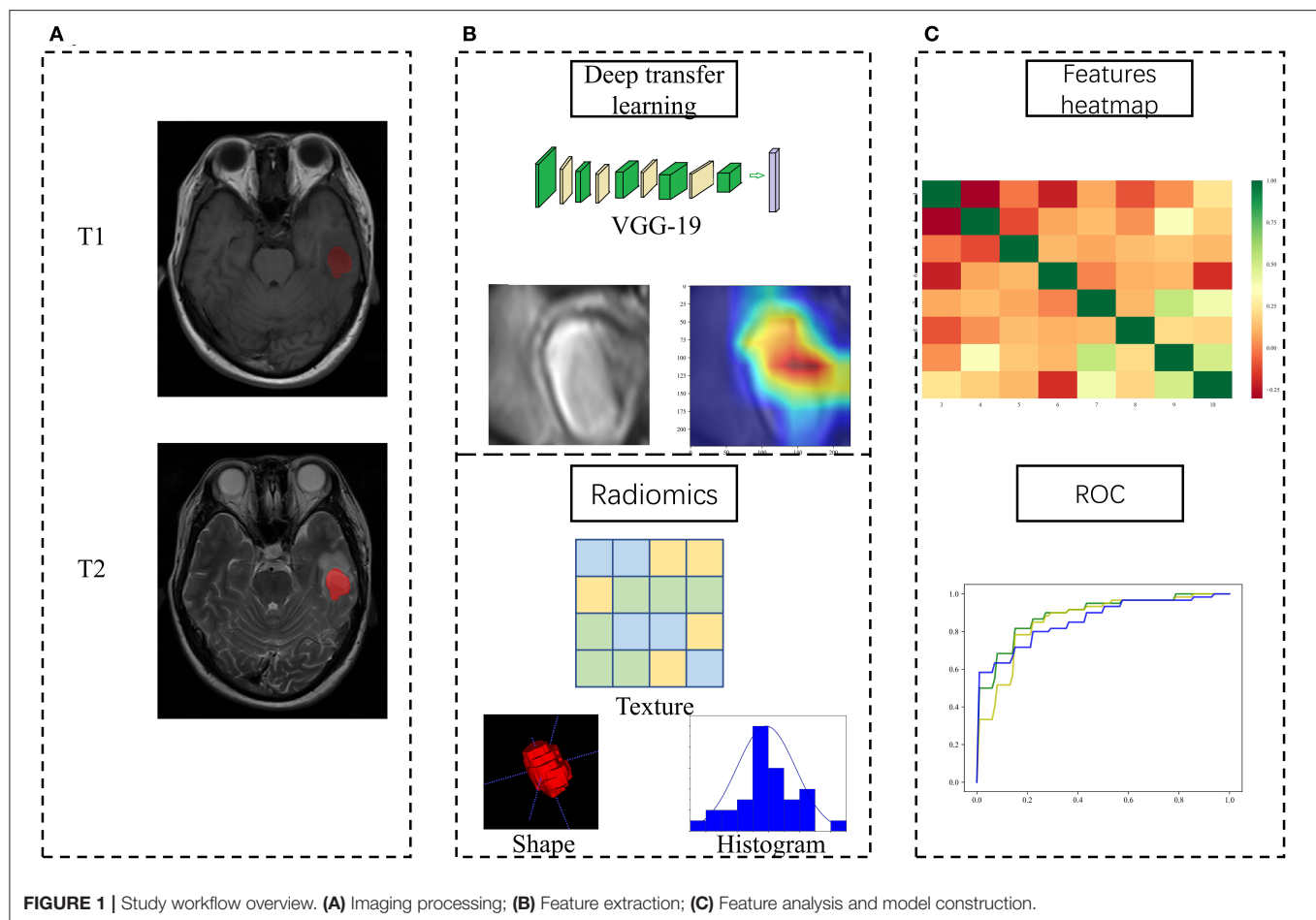
This study was reviewed and approved by the Institutional Review Board of Xiangya Hospital and informed consent was provided by the patient participating in this study. From January 2017 to October 2020, 216 patients who met requirements and underwent T1WI and T2WI MRI were included in the cohort after an initial case screening. Twenty-eight patients were excluded due to poor MRI quality caused by technical operations or inspection processes. Finally, 188 patients were enrolled in this study, among which, 102 patients were diagnosed with brain abscesses (age [mean  $\pm$  SD],  $47.8 \pm 17.6$  years; 33 males and 69 females) and 86 were diagnosed with cystic gliomas (age [mean  $\pm$  SD],  $46.2 \pm 15.1$  years; 27 males and 59 females). The training cohort and test cohort were divided by stratified sampling with a ratio of 7:3, and the distribution of the two diseases was almost the same as that of the overall data set. Then, a nested 5-fold cross-validation was performed on the training cohort.

**Figure 1** showed the flowchart of our study, consisting of image preprocessing, feature extraction, feature analysis, and model construction.

### Image Acquisition

All MRI examinations were conducted in the radiology department of Xiangya hospital with a 3.0T MR Scanner. High-quality MR images were obtained under the following configurations: ①axial T1WI: layer thickness = 5 mm, layer spacing = 1.5 mm, matrix =  $512 \times 416$ , field of view =  $24 \times 24$  cm. ②axial T2WI: layer thickness = 5 mm, layer spacing = 1.5 mm, matrix =  $416 \times 512$ , field of view =  $24 \times 24$  cm. All MR images were retrieved from the picture archiving and communication system for further image feature extraction.

**Abbreviations:** DTL, deep transfer learning; HCR, hand-crafted radiomics; T1WI, T1-weighted image; T2WI, T2-weighted image; KNN, k-nearest neighbors; RFC, random forest classifier; LR, logistic regression; SVM, support vector machine; AUC, Area Under the Curve; DLR, deep learning-based radiomics; ADC, apparent diffusion coefficients; CNN, convolutional neural networks; LASSO, least absolute shrinkage and selection operator; RFE, recursive feature elimination; comb-HCR, combined HCR; comb-DTL, combined DTL; comb-DLR, combined DLR; ROC, receiver operating characteristic; RQS, radiomics quality score.



## Image Preprocessing and Tumor Segmentation

To convert T1WI images to the space of T2WI images, automatic rigid registration was performed with the ITK-SNAP software (Version 3.8.0, <http://www.itksnap.org/>) to segment the structures in the 3D medical image. Meanwhile, manual segmentation of the lesions of all subjects was performed on registered T2WI and T1WI images by a neuroradiologist with 10 years of experience. Then, a radiologist (with 10-years experience) segmented 50 cases, consisting of 25 pathologically proven brain abscess cases and 25 cystic glioma cases randomly selected from all samples. In this way, the consistency of the extracted HCR and DTL features of the neuroradiologist and radiologist was evaluated, and the impact of inter-operator variation on model stability and generalizability was reduced. Besides, the intra-class correlation coefficient of each feature of these 50 cases is calculated.

Image preprocessing was performed as follows. Before DTL features were extracted, an image of the largest cross-sectional area and its upper and lower layers were chosen as a three-channel image. Then, a rectangular region of interest around the tumor contour was used to crop the MR image. Next, the size of the tumor patch was resized to  $224 \times 224$  to meet the input

size requirement of the pretrained CNN model. Before the HCR features were extracted, B-spline interpolation was adopted to resample all images to the same voxel size of  $1 \times 1 \times 1 \text{ mm}^3$ . To avoid the influence of different MR image machine scanners on feature extraction, all images were normalized. Moreover, it seems that deep learning is less affected by different MR machine types than radiomics.

## Feature Extraction

In line with the Imaging Biomarker Standardization Initiative, two kinds of features were extracted, i.e., DTL feature and HCR feature. As for the extraction of the DTL feature, ResNet-50 (29) and VGG-19 (30) pretrained on the natural image dataset ImageNet (<http://www.image-net.org/>) were taken as our base models. (Visual Geometry Group) VGG-19 contains 19 hidden layers (16 convolution layers and 3 full connection layers). It uses  $3 \times 3$  convolutional kernels in all layers to deepen the number of layers and avoid excessive parameters. As for ResNet, it integrates residual learning to avoid gradient dispersion and accuracy reduction in deep networks, improving the network efficiency, accuracy, and execution speed. The internal deep learning features in the image are also visualized while the convolutional layer receives the input features and generates

the output feature mapping. As for the extraction of the HCR feature, 105 original HCR features were extracted from each of the axial T1WI and T2WI images using PyRadiomics (Version 2.1.0, <https://pyradiomics.readthedocs.io/>).

## Feature Selection

To prevent overfitting, the multistep feature selection method was adopted to select the best features for distinguishing brain abscesses from cystic gliomas. First, all the HCR features were analyzed in order by the Spearman rank correlation test and mutual information method. The Spearman rank correlation test was used to investigate the internal linear correlation between individual features. The higher the absolute value of the correlation coefficient, the stronger the correlation. As for the non-redundant features (with a linear correlation coefficient  $< 0.95$ ), the mutual information method was used to capture arbitrary relationships (both linear and non-linear) between each feature and object variable. Then, according to the imaging modality and feature category, the remaining radiomics features and all DLR features were divided into four feature groups, i.e., T1WI-HCR, T2WI-HCR, T1WI-DTL, and T2WI-DTL group. Finally, the least absolute shrinkage and selection operator (LASSO) and recursive feature elimination (RFE) method based on LR and SVM were adopted to repeatedly create the model and select the best feature subset in each feature group.

## Feature Fusion

Feature fusion indicates that two feature groups are put together. The T1WI-HCR and T2WI-HCR feature groups were fused to combined HCR (comb-HCR), and 10 optimal MR image features were selected; similarly, T1WI-DTL and T2WI-DTL were fused to combined DTL (comb-DTL); the T1WI-HCR and T1WI-DTL feature groups were fused to T1WI-DLR, and 10 optimal MR image features were selected; the T2WI-HCR and T2WI-DTL feature groups were fused to T2WI-DLR, and 10 optimal MR image features were selected; the comb-HCR and comb-DTL feature groups were fused to combined DLR (comb-DLR), and 10 optimal MR image features were selected. Refer **Figure 2** for details of feature selection and fusion flow chart.

## Feature Analysis/Model Construction and Validation

After feature fusion and selection, we used each feature group separately to build machine learning classification models, including LR, RFC, KNN, and SVM, implemented by Python Scikit-learn ([https://scikit-learn.org/stable/user\\_guide.html](https://scikit-learn.org/stable/user_guide.html)). The performance of different classifiers was compared. To prevent overfitting, we performed 1,000 iterations of nested 5-fold cross-validation to select the best parameters for the classifier in the training cohorts. The discriminative power of the model was assessed by AUC values, Youden Index, and receiver operating characteristic (ROC) curves. Accuracy, precision, recall, specificity, and F1-score were also used as quantitative metrics. The AUC values for comparative disease identification were carried out using DeLong test.

## Clinical ADC Maps vs. Our Model

In this study, two experienced radiologists (with more than 10 years of experience in brain tumor MRI) were assigned to jointly perform image ADC diagnosis, but they were not involved in the quantitative image analysis described above. All clinicopathological information were removed, and the radiologists distinguished brain abscesses from cystic gliomas based on ADC images only. The proportion of ADC maps in our study cohort was counted. Besides, the diagnostic performance of the two radiologists on the same dataset following the current clinical practice (including the use of ADC maps) was compared to that of the established classifier.

## Statistical Analysis

The comparison of categorical variables was performed through chi-square tests or Fisher tests, and the comparison between quantitative variables was performed through *t*-tests or Mann-Whitney *U*-test. Meanwhile, the Spearman rank correlation test was adopted to evaluate the correlation and executed in Python. A  $p < 0.05$  (two-sided) indicates a significant difference in distinguishing cystic gliomas from brain abscesses. Statistical analysis was performed with IBM SPSS Statistics (version 25; IBM Corporation, Armonk, NY, USA), R (<https://cran.r-project.org/>), and Python (version 3.6.6, <https://www.python.org>). A pretrained CNN model was run using Keras with a Tensorflow backend (<https://keras.io/applications/#Resnet-50> and <https://keras.io/applications/#VGG-19>).

## RESULTS

### Patient Characteristics

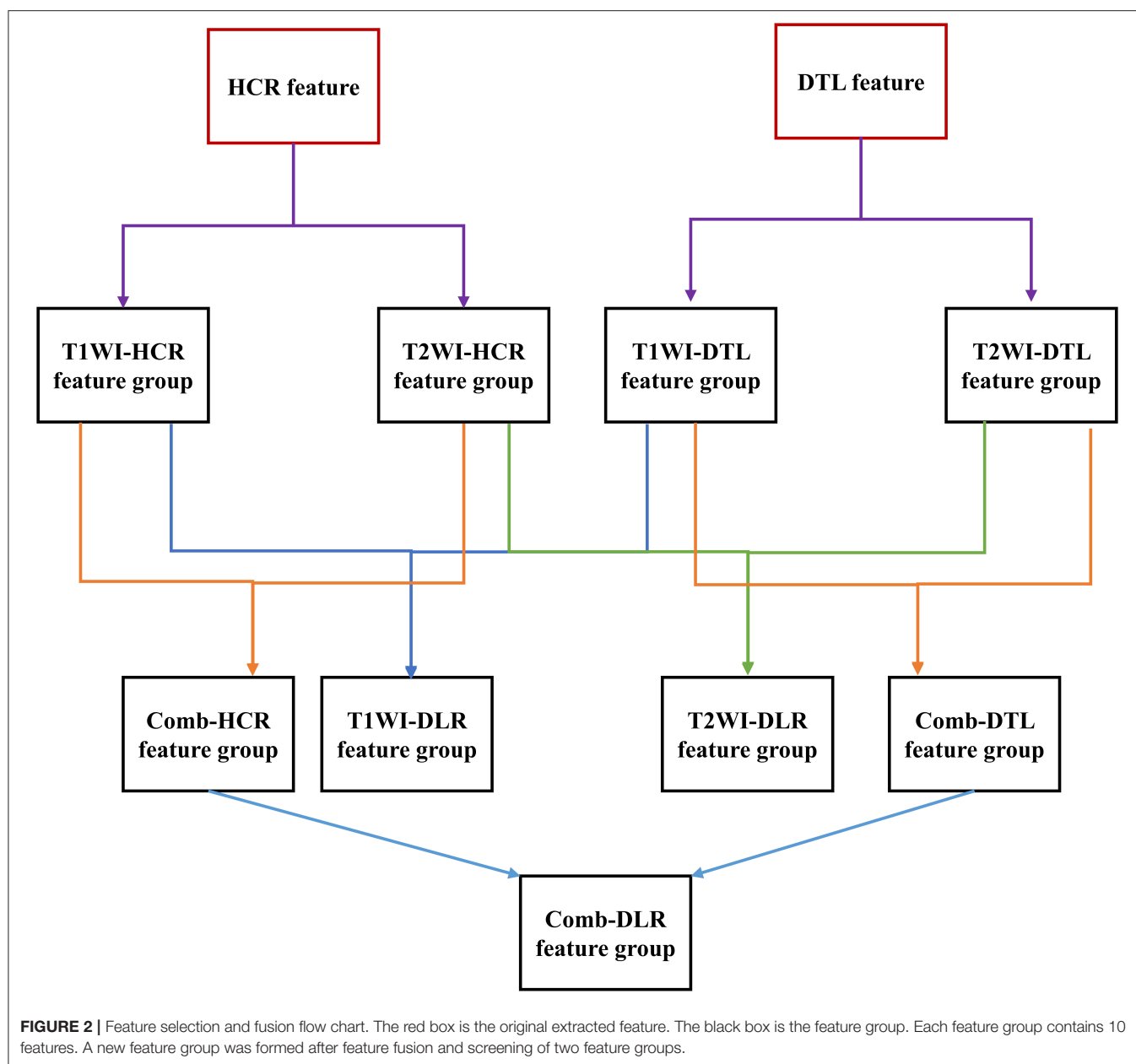
A total of 131 and 57 patients were involved in the training and test cohort of this study, respectively. Specifically, the training cohort involved 71 patients with brain abscesses and 60 patients with gliomas, while the test cohort involved 31 patients with brain abscesses and 26 patients with gliomas. The patient characteristics are provided in **Supplementary Table 1**. The gold standard for distinguishing between brain abscess and cystic glioma was confirmed by pathologists through pathological examination.

### Results of the Feature Extraction

To extract the DTL features, the tumor patch images were input to the pretrained CNN to extract 1,000 features from each MR image modality, and the extracted features were outputs from the last fully connected layer of VGG-19 and ResNet-50. The extracted HCR features included First Order Statistics (18 features), Shape-based (3D) (14 features), Gray Level Cooccurrence Matrix (22 features), Gray Level Run Length Matrix (16 features), Gray Level Size Zone Matrix (16 features), Neighboring Gray Tone Difference Matrix (5 features), and Gray Level Dependence Matrix (14 features).

### Results of the Feature Selection and Fusion

After feature selection and fusion, only the features with intra-class correlation coefficients  $> 0.95$  were retained, indicating that



these features are not affected by multiple tumor segmentation operators and present high reproducibility. For single image modality analysis, a feature selection method was used to preserve 10 optimal features in each group. In multimodality analysis, the two groups of each modality were combined, and 10 features were filtered out. Refer **Supplementary Table 2** for detailed feature selection results.

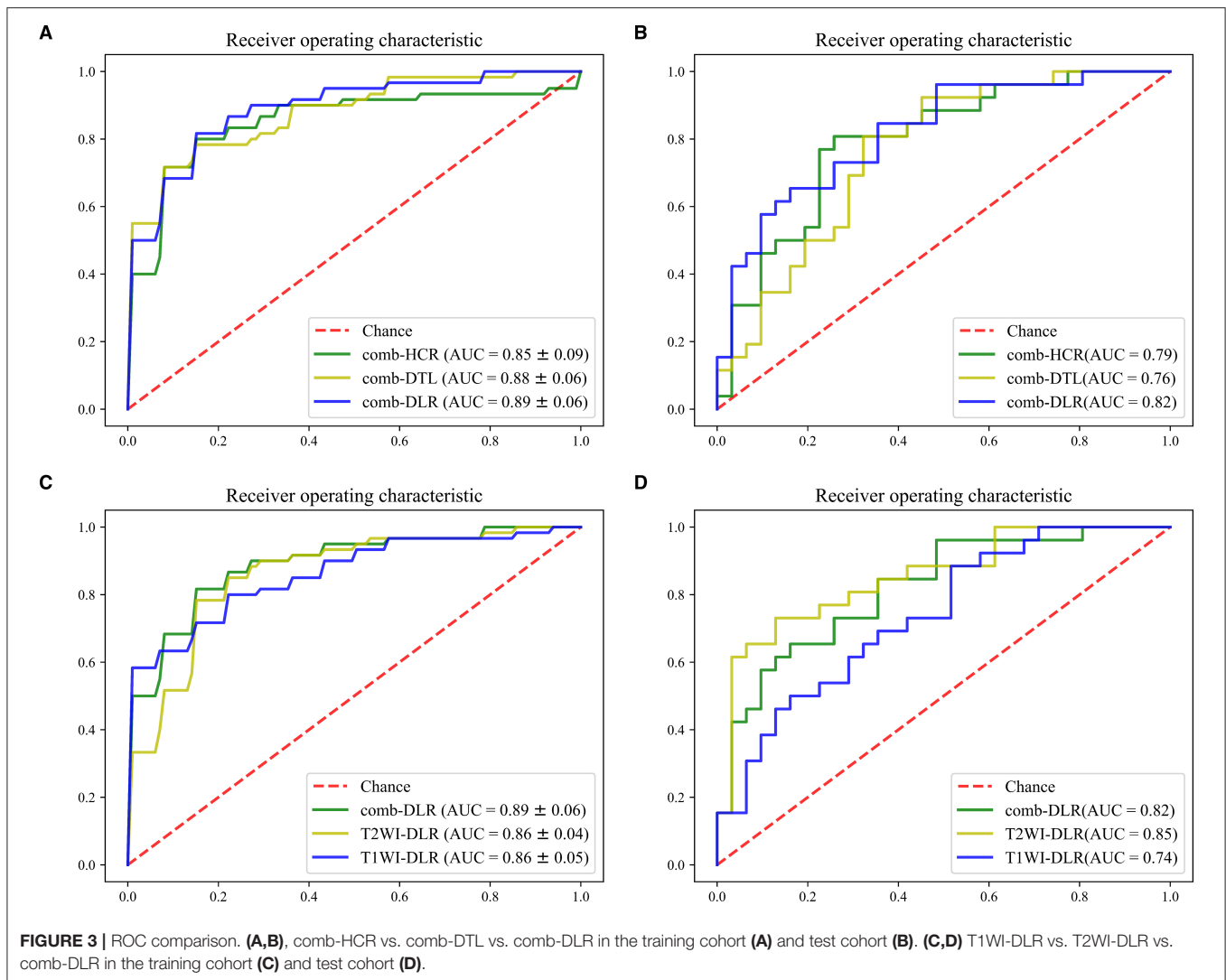
### DLR vs. DTL vs. HCR

The modeling effects of the combined modality feature with different category were compared. On the test cohort, the feature obtained from comb-DLR presented higher diagnostic accuracy than those from comb-HCR and comb-DTL for distinguishing brain abscess from cystic glioma. The AUC values of the models

established with the comb-DLR, comb-HCR, and comb-DTL models features were 0.82, 0.79, and 0.76, respectively; the AUC values of the models established with the T2WI-DLR, T2WI-HCR, and T2WI-DTL features were 0.85, 0.80, and 0.71, respectively; the AUC values of the models established with the T1WI-DLR, T1WI-HCR, and T1WI-DTL features were 0.74, 0.77, and 0.80, respectively. Refer **Figure 3**, **Table 1** for details of model comparison results.

### Multimodality vs. Single Modality

Since the AUC value of the model established with the DLR features was statistically higher than that of the models established with the HCR and DTL features in most cases, the DLR features were used in the multimodal experiments. On



the test cohort, the AUC value of the model established with the T2WI-DLR features was statistically higher than that of the model established with the T1WI-DLR and com-DLR features. Model comparison results are provided in **Figure 3**, **Table 1**.

## Construction and Validation of the Final Model

As can be seen from **Table 1**, the optimal model was obtained by using the T2WI-DLR features combined with an SVM-based RFE feature selection method, and an SVM classifier. The AUC value of the model on the training and test cohort reached 0.86 (95% CI: 0.81, 0.91) and 0.85 (95% CI: 0.75, 0.95) for distinguishing brain abscess from cystic glioma, respectively. **Figure 4** presents the performance of the optimal model. It can be seen from the figure that on the training set, the AUC values of the nested 5-fold were 0.86, 0.86, 0.92, 0.88, and 0.80, respectively. Besides, the standard deviation of the mean AUC value was 0.04, indicating that our model has good stability and robustness and reduces overfitting. The optimal cutoff value of the model was determined by Youden

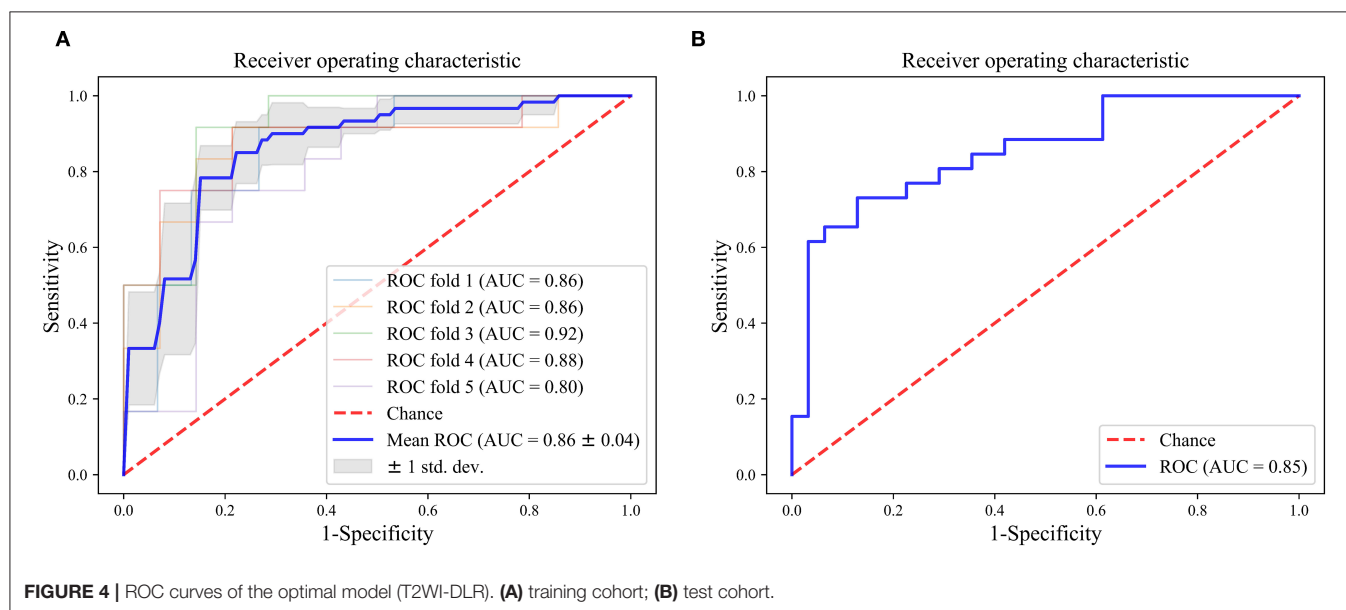
Index. On the test cohort, the sensitivity and specificity of the model were 73.1 and 87.1%, respectively, with an optimal critical value of 0.512 and a Jorden index of 0.601. The details of HCR and DTL feature selection and model construction are listed in **Tables 2, 3**, respectively. The selection of the hyperparameters for each model is listed in **Supplementary Table 3**. Moreover, the ROC curves for each model were compared, and the Delong test results are detailed in **Supplementary Table 4**. It can be seen from the table, that T2WI-DLR features (AUC, 0.85) are superior to T2WI-DTL (AUC, 0.71;  $P = 0.0058$ ; DeLong test).

## DTL Feature Visualization

As shown in **Figure 5**, the feature maps output by the last convolutional layer in the VGG-19 and Resnet-50 model are visualized. The feature map of the visually perceptible tumor region captures most of the details in the image. To a certain extent, it confirms the reliability of transfer learning for feature extraction. By visualizing the learning of the features in the image by these two networks, more insight into the working of

**TABLE 1** | Model representation.

Imaging modality and feature category		Features selection+ Classifier	Training cohort(Cross validation)	Test cohort					
			Mean AUC (95% CI)	AUC (95% CI)	Accuracy	Precision	Recall	F1-score	Specificity
HCR	T1 WI-HCR	RFE(LR)+LR	0.75 (95% CI: 0.61, 0.89)	0.79 (0.67–0.91)	0.67	0.65	0.77	0.61	0.74
	T2WI-HCR	RFE(LR)+SVM	0.84 (95% CI: 0.72, 0.96)	0.75 (0.62–0.88)	0.75	0.73	0.73	0.73	0.77
	comb-HCR	RFE(LR)+LR	0.85 (95% CI: 0.74, 0.96)	0.79 (0.68–0.91)	0.77	0.77	0.74	0.75	0.77
DTL	T1 WI-DTL	RFE(SVM)+SVM	0.89(95% CI: 0.84, 0.94)	0.81 (0.69–0.93)	0.74	0.76	0.62	0.68	0.84
	T2WI-DTL	RFE(SVM)+SVM	0.9(95% CI: 0.85, 0.95)	0.71 (0.58–0.85)	0.68	0.72	0.5	0.59	0.84
	comb-DTL	RFE(LR)+LR	0.88(95% CI: 0.81, 0.95)	0.77 (0.65–0.90)	0.65	0.69	0.42	0.52	0.84
DLR	T1 WI-DLR	RFE(LR)+LR	0.86(95% CI: 0.80, 0.92)	0.75 (0.63–0.88)	0.67	0.65	0.63	0.64	0.68
	T2 WI-DLR	RFE(SVM)+SVM	0.86(95% CI: 0.81, 0.91)	0.85 (0.75–0.95)	0.77	0.73	0.76	0.75	0.81
	comb-DLR	RFE(SVM)+SVM	0.89(95% CI: 0.82, 0.96)	0.83 (0.73–0.94)	0.72	0.65	0.71	0.68	0.77



the networks can be obtained, and the reasons why the disease may be correctly identified by transfer learning models can be understood.

### Clinical ADC Maps vs. Our Model

The cases containing ADC maps in the cohort were counted. There were 33 cases in total, including 21 cases of brain abscess and 12 cases of cystic glioma. A detailed comparison of the distinguishment performance between the clinical ADC maps diagnosis and our model is shown in **Table 4**. For each patient,

the diagnosis of the radiologist and the model prediction are listed in **Supplementary Table 5**. It can be found that our model has the same accuracy as the clinical ADC assessment, so it has great potential for distinguishing between the two diseases.

### DISCUSSIONS

Accurate identification of brain abscesses and cystic gliomas is essential to planning appropriate treatment, assessing outcome, and future prognosis. However, due to the similarity in the

**TABLE 2 |** HCR model construction.

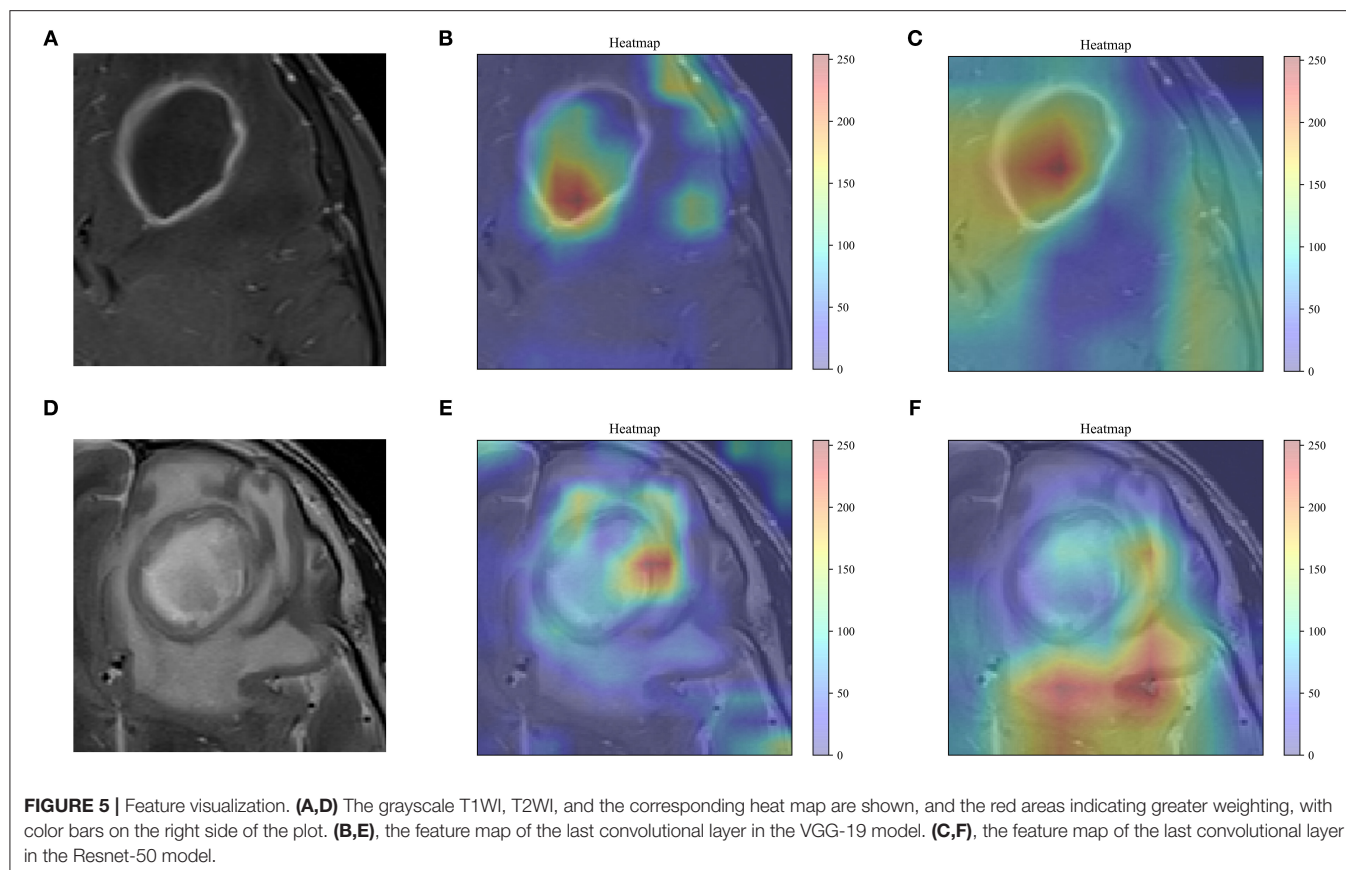
Classifier	Feature selection method	T1			T2		
		Optimal feature number	Mean AUC (training cohort)	AUC (test cohort)	Optimal feature number	Mean AUC (training cohort)	AUC (test cohort)
LR	RFE(LR)	10	0.75	0.77	10	0.84	0.81
	RFE(SVM)	10	0.74	0.76	10	0.83	0.81
	LASSO	5	0.72	0.75	11	0.81	0.80
SVM	RFE(LR)	10	0.72	0.75	10	0.84	0.80
	RFE(SVM)	10	0.75	0.71	10	0.84	0.80
	LASSO	5	0.72	0.75	11	0.82	0.79
KNN	RFE(LR)	10	0.70	0.75	10	0.75	0.88
	RFE(SVM)	10	0.71	0.78	10	0.74	0.81
	LASSO	5	0.71	0.75	11	0.79	0.82
RFC	RFE(LR)	10	0.69	0.75	10	0.76	0.83
	RFE(SVM)	10	0.73	0.72	10	0.75	0.80
	LASSO	5	0.71	0.72	11	0.74	0.86

**TABLE 3 |** DTL model construction.

Classifier	Feature selection method	T1			T2		
		Optimal feature number	Mean AUC (training cohort)	AUC (test cohort)	Optimal feature number	Mean AUC (training cohort)	AUC (test cohort)
LR	RFE(LR)	10	0.84	0.72	10	0.87	0.73
	RFE(SVM)	10	0.89	0.80	10	0.87	0.68
	LASSO	0	–	–	9	0.81	0.75
SVM	RFE(LR)	10	0.86	0.70	10	0.86	0.85
	RFE(SVM)	10	0.90	0.71	10	0.83	0.69
	LASSO	0	–	–	9	0.81	0.75
KNN	RFE(LR)	10	0.72	0.54	10	0.82	0.75
	RFE(SVM)	10	0.73	0.60	10	0.81	0.61
	LASSO	0	–	–	9	0.80	0.63
RFC	RFE(LR)	10	0.69	0.71	10	0.79	0.69
	RFE(SVM)	10	0.72	0.69	10	0.82	0.74
	LASSO	0	–	–	9	0.72	0.71

conventional MR images, i.e., ring enhancement, it is difficult to distinguish between the two diseases. In this study, a deep learning-based statistical analysis method based on multistep feature selection and fusion was demonstrated and verified. The experimental results indicate that the method can be used to distinguish between brain abscess and cystic glioma in conventional T1WI and T2WI. The previous literature on disease prediction prognosis and classification differential diagnosis for quantitative image analysis has shown that deep learning contributes to better performance of radiomics analysis (21, 27, 31, 32). Our study demonstrates that by extracting DTL features with VGG-19, a model with excellent feature learning and feature representation abilities can be obtained. Besides, as shown in **Figure 5**, VGG-19 can better focus on the details of the tumor region than Resnet-50.

According to the feature selection results of the optimal model, two “good” HCR features were selected for statistical significance analysis. It can be seen from the box plots in **Figure 6** that there is not much difference in the distribution of the features between brain abscess and cystic glioma. All corresponding *p*-values of the statistical tests for distinguishing the two diseases are presented in the figure. These results indicate that these two features have a good identification ability in this work, showing the reproducibility and usefulness of feature engineering. Besides, the two “good” HCR features are all texture features, which reflect the homogeneous phenomenon in the image, which once again demonstrates the superiority of texture features in distinguishing brain abscess from cystic glioma. Previous studies have also shown that texture features are highly predictive in many tasks, which is consistent with the results obtained in this study (33, 34).



**TABLE 4 |** Clinical ADC maps vs. our model.

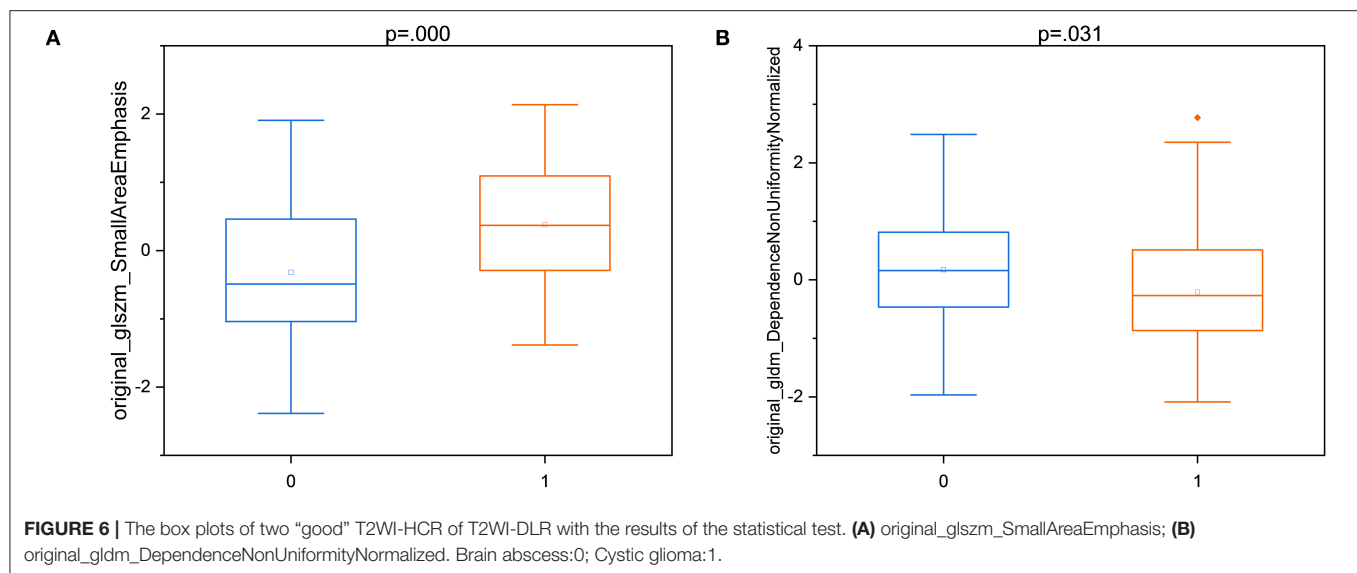
	Accuracy	Precision	Recall	F1-score	Specificity
ADC maps	0.848	0.818	0.75	0.783	0.905
T2WI-DLR	0.848	0.706	1	0.889	0.762

Compared with using T1WI and T2WI alone in DLR, the model based on combined modality does not achieve improved performance. This indicates that the single modality of T2WI is also a good predictor for distinguishing between brain abscesses and gliomas, and this is consistent with the fact that the T2 modality is more commonly used in imaging diagnosis of brain diseases in clinical practice. When the performance of the two models is stable and the results are complementary, model fusion can lead to better performance (35).

Previous studies (5, 36–38) have demonstrated that the advanced MRI techniques, such as magnetic resonance spectroscopy, susceptibility-weighted imaging, ADC, and dynamic susceptibility contrast-enhanced, can distinguish brain abscesses from gliomas, but these techniques have some limitations. Refer **Supplementary Table 6** for model performance comparison. First, the sample size of these techniques is small, and only a few cases of pyogenic abscess and glioblastoma are included, hindering the direct application of the results to daily clinical practice. Second, the model

based on the combination of intralesional susceptibility signal and ADC achieved a good AUC value (38). However, the combination does not lead to obvious improvement of differential diagnosis, because only a small number of patients with abscess/glioblastoma show atypical high/low ADC. Finally, none of the techniques were based on conventional MR images, increasing the image acquisition difficulty and cost. In this study, some measures were taken to overcome these limitations. The relatively larger sample size contributes to a better performance of conventional MR images for distinguishing brain abscesses and gliomas. The comparison of dataset size is listed in **Supplementary Table 6**. Also, the HCR and DTL features were extracted from conventional MR images. To our knowledge, there is no report on integrating HCR and DTL features for distinguishing brain abscesses from cystic gliomas. Besides, our research is based on some ordinary image data and does not require special training, so it has significant potential. In addition, DTL feature extraction uses a fixed-size bounding box for the tumor region, which not only provides information about intertumoral heterogeneity but also provides tumor microenvironment information to a certain extent.

To promote the development of radiomics as imaging biomarkers, a plethora of studies have used radiomics quality score (RQS) to evaluate and standardize radiomics (8, 39, 40). The RQS of our study was satisfactory at 15 points (41.7% of the ideal quality score), and the detailed result is listed in



**Supplementary Table 7.** The RQs of the relevant work (16–20) were analyzed in our study, but only our study is open to science and data, only one study conducted a multivariable analysis with non-radiomics features (20), and only one study based on multicenter validation (16). Besides, no research has conducted a phantom study, collected images of individuals at additional time points, discussed biological correlates, conducted a prospective study, or reported the cost-effectiveness of the clinical application.

The limitations of our work are as follows. First, due to the difficulty of obtaining external validation data, the patients in our study were single-center. The effects of clinical ADC maps diagnosis were compared with our proposed model, but this sequence was not added to our model due to the insufficient sample size of ADC maps. Multicenter validation, multi-MRI sequences, and prospective studies will be involved in our future work. Meanwhile, additional features such as proteomics, transcriptomics, pathomics, and genomic features were not considered in our study. Multi-omics joint analysis that integrates complex structural systems with multiple layers, levels, and functions may enhance the performance to identify brain abscess and cystic glioma and overcome the limitations of a single theoretical model. Besides, this study only used the image of the largest cross-section area with the upper and lower layers as the input to the VGG-19 or ResNet-50 model. The use of the 3D volume of the tumor/region of interest should be investigated in future research. Finally, the application of our study to identify other brain tumors and the enhancement of the algorithm will be explored.

## CONCLUSIONS

This paper first reports a model combining DTL features and HCR features from conventional MRI for distinguishing brain abscesses from cystic glioma. The study results provide an effective, inexpensive, convenient, and non-invasive method for differential diagnosis.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, the main code are available from <https://github.com/better0123/radiomics-analysis.git>. Further inquiries can be directed to the corresponding author/s.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board (IRB) of Xiangya Hospital. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

LB, ZZ, and ZJ: study design. ZL, QC, ZZ, XT, YW, and TC: data collection. LB, PH, ZZ, and CY: data analysis. DL, GY, and ZL: supervision. LB, QC, ZZ, and ZJ: manuscript writing. All authors contributed to the article and approved the submitted version.

## FUNDING

This study was supported by the National Nature Science Foundation of China (NO. 82073893, NO. 81873635, NO. 81703622 and NO. 81472693); the China Postdoctoral Science Foundation (NO. 2018M63302); the Natural Science Foundation of Hunan Province (NO. 2018JJ3838 and NO. 2018SK2101); the Hunan Provincial Health and Health Committee Foundation of China (C2019186); Xiangya Hospital Central South University postdoctoral foundation; the National Natural Science Foundation of China (61971271); the Taishan Scholars Project of Shandong Province (Tsqn20161023) and the Primary Research; Development Plan of Shandong Province (No. 2018GGX101018 and No. 2019QYTPY020).

## ACKNOWLEDGMENTS

The authors thank the Network and Information Center and Department of Radiology of Xiangya Hospital that have given support in data extraction and analysis.

## REFERENCES

- Ganau L, Paris M, Ligarotti GK, Ganau M. Management of gliomas: overview of the latest technological advancements and related behavioral drawbacks. *Behav Neurol.* (2015) 2015:862634. doi: 10.1155/2015/862634
- Afridi M, Asadollahi S, Nath K, Chawla SJJNPN. Differentiation of brain abscesses from necrotic high-grade gliomas using advanced MR imaging techniques: a mini review. *J Neuroimaging Psychiatry Neurol.* (2019) 4:17–22. doi: 10.17756/jnpn.2019-030
- Alvis Miranda H, Castellar-Leones SM, Elzain MA, Moscote-Salazar LR. Brain abscess: current management. *J Neurosci Rural Pract.* (2013) 4(Suppl 1):S67–81. doi: 10.4103/0976-3147.116472
- Toh CH, Wei KC, Ng SH, Wan YL, Lin CP, Castillo M. Differentiation of brain abscesses from necrotic glioblastomas and cystic metastatic brain tumors with diffusion tensor imaging. *Am J Neuroradiol.* (2011) 32:1646. doi: 10.3174/ajnr.A2581
- Toh CH, Wei K-C, Chang C-N, Ng S-H, Wong H-F, Lin C.-P. Differentiation of brain abscesses from glioblastomas and metastatic brain tumors: comparisons of diagnostic performance of dynamic susceptibility contrast-enhanced perfusion MR imaging before and after mathematic contrast leakage correction. *PLoS ONE.* (2014) 9:e109172. doi: 10.1371/journal.pone.0109172
- Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to radiomics. *J Nucl Med.* (2020) 61:488–95. doi: 10.2967/jnumed.118.222893
- Liu Z, Wang S, Di Dong JW, Fang C, Zhou X, Sun K, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics.* (2019) 9:1303. doi: 10.7150/thno.30309
- Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
- Giraud P, Giraud P, Gasnier A, El Ayachy R, Kreps S, Foy J-P, et al. Radiomics and machine learning for radiotherapy in head and neck cancers. *Front Oncol.* (2019) 9:174. doi: 10.3389/fonc.2019.00174
- Song J, Yin Y, Wang H, Chang Z, Liu Z, Cui L. A review of original articles published in the emerging field of radiomics. *Eur J Radiol.* (2020) 127:108991. doi: 10.1016/j.ejrad.2020.108991
- Xie H, Ma S, Wang X, Zhang XJ. Noncontrast computer tomography-based radiomics model for predicting intracerebral hemorrhage expansion: preliminary findings and comparison with conventional radiological model. *Eur Radiol.* (2020) 30:87–98. doi: 10.1007/s00330-019-06378-3
- Qiu Q, Duan J, Deng H, Han Z, Gu J, Yue NJ, et al. Development and validation of a radiomics nomogram model for predicting postoperative recurrence in patients with esophageal squamous cell cancer who achieved pCR after neoadjuvant chemoradiotherapy followed by surgery. *Front Oncol.* (2020) 10:1398. doi: 10.3389/fonc.2020.01398
- Dong F, Li Q, Xu D, Xiu W, Zeng Q, Zhu X, et al. Differentiation between pilocytic astrocytoma and glioblastoma: a decision tree model using contrast-enhanced magnetic resonance imaging-derived quantitative radiomic features. *Eur Radiol.* (2019) 29:3968–75. doi: 10.1007/s00330-018-5706-6
- Bae S, An C, Ahn SS, Kim H, Han K, Kim SW, et al. Robust performance of deep learning for distinguishing glioblastoma from single brain metastasis using radiomic features: model development and validation. *Sci Rep.* (2020) 10:12110. doi: 10.1038/s41598-020-68980-6
- Zhang Z, Yang J, Ho A, Jiang W, Logan J, Wang X, et al. A predictive model for distinguishing radiation necrosis from tumour progression after gamma knife radiosurgery based on radiomic features from MR images. *Eur Radiol.* (2018) 28:2255–63. doi: 10.1007/s00330-017-5154-8
- Qian Z, Li Y, Wang Y, Li L, Li R, Wang K, et al. Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classifiers. *Cancer Lett.* (2019) 451:128–35. doi: 10.1016/j.canlet.2019.02.054
- Dong F, Li Q, Jiang B, Zhu X, Zeng Q, Huang P, et al. Differentiation of supratentorial single brain metastasis and glioblastoma by using perienhancing oedema region-derived radiomic features and multiple classifiers. *Eur Radiol.* (2020) 30:3015–22. doi: 10.1007/s00330-019-06460-w
- Zhang Y, Chen C, Cheng Y, Teng Y, Guo W, Xu H, et al. Ability of radiomics in differentiation of anaplastic oligodendroglioma from atypical low-grade oligodendroglioma using machine-learning approach. *Front Oncol.* (2019) 9:1371. doi: 10.3389/fonc.2019.01371
- Chen C, Ou X, Wang J, Guo W, Ma XJF. Radiomics-based machine learning in differentiation between glioblastoma and metastatic brain tumors. *Front Oncol.* (2019) 9:806. doi: 10.3389/fonc.2019.00806
- Artzi M, Bressler I, Ben Bashat D. Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis. *J Magn Reson Imaging.* (2019) 50:519–28. doi: 10.1002/jmri.26643
- Han W, Qin L, Bay C, Chen X, Yu KH, Miskin N, et al. Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas. *Am J Neuroradiol.* (2020) 41:40. doi: 10.3174/ajnr.A6365
- Lao J, Chen Y, Li Z-C, Li Q, Zhang J, Liu J, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci Rep.* (2017) 7:10353. doi: 10.1038/s41598-017-10649-8
- Wu X, Li Y, Chen X, Huang Y, He L, Zhao K, et al. Deep learning features improve the performance of a radiomics signature for predicting kras status in patients with colorectal cancer. *Acad Radiol.* (2020) 27:e254–62. doi: 10.1016/j.acra.2019.12.007
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005
- Zhang W, Yang G, Huang H, Yang W, Xu X, Liu Y, et al. ME-Net: multi-encoder net framework for brain tumor segmentation. *Int J Imaging Syst Technol.* (2021) 1–15. doi: 10.1002/ima.22571
- Jin Y, Yang G, Fang Y, Li R, Xu X, Liu Y, et al. 3D PBV-Net: an automated prostate MRI data segmentation method. *Comput Biol Med.* (2021) 128:104160. doi: 10.1016/j.compbimed.2020.104160
- Xue L-Y, Jiang Z-Y, Fu T-T, Wang Q-M, Zhu Y-L, Dai M, et al. Transfer learning radiomics based on multimodal ultrasound imaging for staging liver fibrosis. *Eur Radiol.* (2020) 30:2973–83. doi: 10.1007/s00330-019-06595-w
- Hu Y, Xie C, Yang H, Ho JWK, Wen J, Han L, et al. Computed tomography-based deep-learning prediction of neoadjuvant chemoradiotherapy treatment response in esophageal squamous cell carcinoma. *Radiother Oncol.* (2021) 154:6–13. doi: 10.1016/j.radonc.2020.09.014
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition[C]. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2016) 770–8.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv [Preprint] arXiv:1409.1556.* (2014).
- Yu J, Deng Y, Liu T, Zhou J, Jia X, Xiao T, et al. Lymph node metastasis prediction of papillary thyroid carcinoma based on transfer learning radiomics. *Nat Commun.* (2020) 11:4807. doi: 10.1038/s41467-020-18497-3
- Zhang Y, Lobo-Mueller EM, Karanicolas P, Gallinger S, Haider MA, Khalvati F. Improving prognostic performance in resectable pancreatic ductal adenocarcinoma using radiomics and deep learning features fusion in CT images. *Sci Rep.* (2021) 11:1378. doi: 10.1038/s41598-021-80998-y

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2021.748144/full#supplementary-material>

33. Tian Q, Yan LF, Zhang X, Zhang X, Hu YC, Han Y, et al. Radiomics strategy for glioma grading using texture features from multiparametric MRI. *J Magn Reson Imaging*. (2018) 48:1518–28. doi: 10.1002/jmri.26010
34. Zhang X, Xu X, Tian Q, Li B, Wu Y, Yang Z, et al. Radiomics assessment of bladder cancer grade using texture features from diffusion-weighted imaging. *J Magn Reson Imaging*. (2017) 46:1281–8. doi: 10.1002/jmri.25669
35. Wu X, Dong D, Zhang L, Fang M, Zhu Y, He B, et al. Exploring the predictive value of additional peritumoral regions based on deep learning and radiomics: a multicenter study. *Med Phys*. (2021) 48:2374–85. doi: 10.1002/mp.14767
36. Fu J-H, Chuang T-C, Chung H-W, Chang H-C, Lin H-S, Hsu S-S, et al. Discriminating pyogenic brain abscesses, necrotic glioblastomas, and necrotic metastatic brain tumors by means of susceptibility-weighted imaging. *Eur Radiol*. (2015) 25:1413–20. doi: 10.1007/s00330-014-3518-x
37. De Simone M, Brogna B, Sessa G, Oliva G, Guida B, Magliulo M. Valuable contribution of magnetic resonance spectroscopy in differentiation of brain abscess from glioma. *Infect Dis*. (2017) 49:871–3. doi: 10.1080/23744235.2017.1331464
38. Lai P-H, Chung H-W, Chang H-C, Fu J-H, Wang P-C, Hsu S-H, et al. Susceptibility-weighted imaging provides complementary value to diffusion-weighted imaging in the differentiation between pyogenic brain abscesses, necrotic glioblastomas, and necrotic metastatic brain tumors. *Eur J Radiol*. (2019) 117:56–61. doi: 10.1016/j.ejrad.2019.05.021
39. Park JE, Kim D, Kim HS, Park SY, Kim JY, Cho SJ, et al. Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement. *Eur Radiol*. (2020) 30:523–36. doi: 10.1007/s00330-019-06360-z
40. Park JE, Kim HS, Kim D, Park SY, Kim JY, Cho SJ, et al. A systematic review reporting quality of radiomics research in neuro-oncology: toward clinical utility and quality improvement using high-dimensional imaging features. *BMC Cancer*. (2020) 20:29. doi: 10.1186/s12885-019-6504-5

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Bo, Zhang, Jiang, Yang, Huang, Chen, Wang, Yu, Tan, Cheng, Li and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Pyramid-Net: Intra-layer Pyramid-Scale Feature Aggregation Network for Retinal Vessel Segmentation

Jiawei Zhang<sup>1,2,3,4†</sup>, Yanchun Zhang<sup>4,5,6\*</sup>, Hailong Qiu<sup>1†</sup>, Wen Xie<sup>1†</sup>, Zeyang Yao<sup>1†</sup>, Haiyun Yuan<sup>1</sup>, Qianjun Jia<sup>1</sup>, Tianchen Wang<sup>3</sup>, Yiyu Shi<sup>3</sup>, Meiping Huang<sup>1\*</sup>, Jian Zhuang<sup>1\*</sup> and Xiaowei Xu<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Jun Feng,  
Northwest University, China

### Reviewed by:

Juanying Xie,  
Shaanxi Normal University, China  
Márton Szemenyei,  
Budapest University of Technology  
and Economics, Hungary  
Erlei Zhang,  
Northwest A&F University, China

### \*Correspondence:

Yanchun Zhang  
yanchun.zhang@vu.edu.au  
Meiping Huang  
huangmeiping@126.com  
Jian Zhuang  
Zhuangjian5413@163.com  
Xiaowei Xu  
xiao.wei.xu@foxmail.com

†These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

Received: 19 August 2021

Accepted: 05 November 2021

Published: 07 December 2021

### Citation:

Zhang J, Zhang Y, Qiu H, Xie W,  
Yao Z, Yuan H, Jia Q, Wang T, Shi Y,  
Huang M, Zhuang J and Xu X (2021)  
Pyramid-Net: Intra-layer  
Pyramid-Scale Feature Aggregation  
Network for Retinal Vessel  
Segmentation. *Front. Med.* 8:761050.  
doi: 10.3389/fmed.2021.761050

<sup>1</sup> Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Provincial People's Hospital, Guangdong Cardiovascular Institute, Guangdong Academy of Medical Sciences, Guangzhou, China, <sup>2</sup> Shanghai key Laboratory of Data Science, School of Computer Science, Fudan University, Shanghai, China, <sup>3</sup> Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, United States, <sup>4</sup> Oujiang Laboratory (Zhejiang Lab for Regenerative Medicine, Vision and Brain Health), Wenzhou, China, <sup>5</sup> Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou, China, <sup>6</sup> College of Engineering and Science, Victoria University, Melbourne, VIC, Australia

Retinal vessel segmentation plays an important role in the diagnosis of eye-related diseases and biomarkers discovery. Existing works perform multi-scale feature aggregation in an inter-layer manner, namely **inter-layer feature aggregation**. However, such an approach only fuses features at either a lower scale or a higher scale, which may result in a limited segmentation performance, especially on thin vessels. This discovery motivates us to fuse multi-scale features in each layer, **intra-layer feature aggregation**, to mitigate the problem. Therefore, in this paper, we propose Pyramid-Net for accurate retinal vessel segmentation, which features intra-layer pyramid-scale aggregation blocks (IPABs). At each layer, IPABs generate two associated branches at a higher scale and a lower scale, respectively, and the two with the main branch at the current scale operate in a **pyramid-scale** manner. Three further enhancements including pyramid inputs enhancement, deep pyramid supervision, and pyramid skip connections are proposed to boost the performance. We have evaluated Pyramid-Net on three public retinal fundus photography datasets (DRIVE, STARE, and CHASE-DB1). The experimental results show that Pyramid-Net can effectively improve the segmentation performance especially on thin vessels, and outperforms the current state-of-the-art methods on all the adopted three datasets. In addition, our method is more efficient than existing methods with a large reduction in computational cost. We have released the source code at <https://github.com/JerRuy/Pyramid-Net>.

**Keywords:** deep learning, neural network, feature aggregation, pyramid scale, retinal vessel segmentation

## 1. INTRODUCTION

The subtle changes in the retinal vascular, including vessel width, tortuosity, and branching features, indicate mass eye-related diseases, such as diabetic retinopathy (1), glaucoma (2), and macular degeneration (3). Meanwhile, those characteristics are important biomarkers for numerous systemic diseases, including hypertension (4) and cardiovascular diseases (5). Retinal

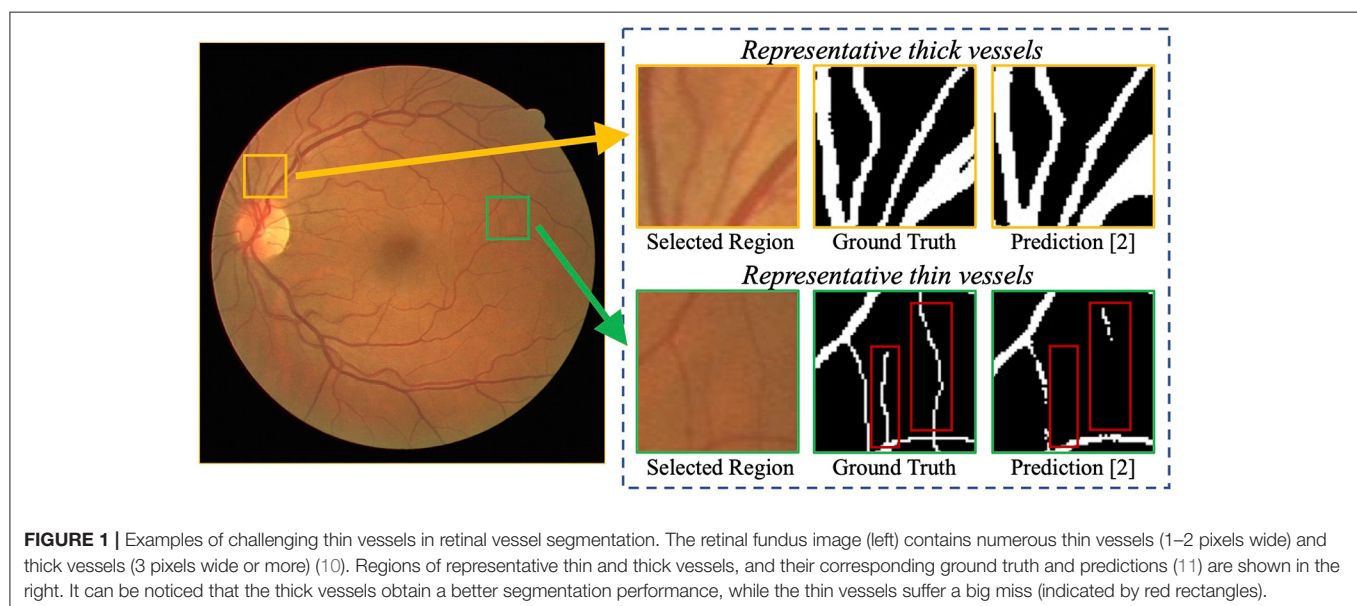
vessel segmentation is one of the cornerstones to access those characteristics, particularly for automatic retinal image analysis (6, 7). For example, hypertensive retinopathy is a retinal disease, which is caused by hypertension. Increased vascular curvature or stenosis can be found in patients with hypertension (8). Conventionally, manual segmentation is laborious and time-consuming, and suffers subjectivity among experts. To improve efficiency and reliability and reduce the workload of doctors, the clinical practice puts forward high requirements for automatic segmentation (9).

Recently, deep neural networks have boosted the segmentation performance of retinal vessel segmentation (10, 12) by a large margin compared with traditional methods (13, 14). However, thin vessels cannot be segmented accurately. For example, **Figure 1** demonstrates a commonly-seen fundus image containing numerous thin vessels and thick vessels, and corresponding segmentation (11) and ground truth. We can easily notice that the thick vessels enjoy a promising performance, but the thin vessels suffer a big miss. A potential reason is that the continuous pooling operations in most neural networks are used to encode the features, which leads to a mass loss of appearance information and harms the segmentation accuracy, especially on thin vessels. Note that in practice, it is also difficult to segment these thin vessels for experts due to low contrast and ambiguousness. Currently, some works have been proposed to tackle the above problems, e.g., a particular processing branch for thin vessels (12), a new loss function to emphasize thin vessels (10). However, the segmentation performance is still limited considering the clinical requirement of retinal image analysis.

Meanwhile, **multi-scale feature aggregation** to fuse coarse-to-fine context information has been popular to segment thin/small objects (15–19). There are mainly two approaches: input-output level category and intra-network level category. In the input-output level category, connections exist between inputs

at various scales and corresponding intermediate layers (15), or between the intermediate layers and the final predictions with corresponding scales (18). In the intra-network level category, features from previous layers are adjusted in channel numbers and spatial dimension and then aggregated with the ones in the later layer (16). However, current multi-scale feature aggregation works in an inter-layer manner, **inter-layer feature aggregation**, which can only fuse features at either a lower scale or a higher scale. For example, in the encoder, feature maps at the lower scale cannot be fused by that at the current scale because of the processing order of the layers. A possible solution is to fuse multi-scale features in each layer, **intra-layer feature aggregation**, to consider features at both the high scale and the low scale.

Motivated by the above discoveries, in this paper, we propose Pyramid-Net for accurate retinal vessel segmentation. In each layer of Pyramid-Net, intra-layer pyramid-scale aggregation blocks (IPABs) are employed in both the encoder and the decoder to aggregate features at pyramid scales (the higher scale, the lower scale, and the current scale). In this way, two associated branches at the higher scale and the lower scale are generated to assist the main branch at the current scale. Therefore, coarse-to-fine context information is shared and aggregated in each layer, thus improving the segmentation accuracy of capillaries. To further improve the performance, three optimizations, including pyramid inputs enhancement, deep pyramid supervision, and pyramid skip connections, are applied to IPABs. We have conducted comprehensive experiments on three retinal vessel image segmentation datasets, including DRIVE (20), STARE (21), and CHASE-DB1 (22) with various segmentation networks. The experimental results show that our method can significantly improve the segmentation performance, especially on thin vessels, and achieves state-of-the-art performance on the three public datasets. In addition, our method is more efficient than the existing method with a large reduction in computational cost.



Overall, this work makes the following contributions:

- 1) We discovered that thin vessels suffer a big miss in the segmentation results of existing methods;
- 2) We proposed Pyramid-Net for retinal vessel segmentation in which intra-layer pyramid-scale aggregation blocks (IPABs) aggregate features at the higher, current, and lower scales to fuse coarse-to-fine context information in each layer;
- 3) We further propose three enhancements: pyramid input enhancement, deep pyramid supervision, and pyramid skip connections to boost the performance;
- 4) We conducted comprehensive experiments on three public vessel image datasets (DRIVE, STARE, and CHASE-DB1), and our method achieves the state-of-the-art performance on three datasets.

The remainder of this paper is organized as follows. Section 2 introduces related works and the motivation of the proposed method. Section 3 details the overall framework of the proposed Pyramid-Net, including IPABs and three optimizations (pyramid inputs enhancement, deep pyramid supervision, and pyramid skip connections). Section 4 first introduces datasets, implementation, and evaluation. Second, quantitative evaluations on three vessel image datasets, comparisons with the state-of-the-art algorithms, and several visual retinal segmentation results are presented. Third, several ablation studies that included evaluating the thin vessel, ablation analysis, and cross-training evaluation are discussed. Section 5 concludes the paper.

## 2. RELATED WORK AND MOTIVATION

### 2.1. Vessel Image Segmentation

With the emergence of numerous public-available retinal image datasets (20–22), the supervised vessel segmentation methods became popular in the community. Commonly-seen supervised methods consist of two steps: feature extraction and classification. Some methods extracted the color intensity (24) and principle components (25) from the images, while some methods utilized wavelet (26) and edge responses (27). In terms of classification, various classic classifiers, including Support Vector Machine (SVM) (28), perceptron (29), random decision forests (30), and Gaussian model (26) are commonly seen and widely used in traditional supervised vessel image segmentation. Recently, in the light of fully convolutional networks (FCNs) (31) and U-Net (23), data-driven deep learning-based methods have demonstrated promising results and dominated the area of vessel image segmentation. Yan et al. (10) pointed out that the training loss tends to ignore the loss of thin vessels and is dominated by the thick vessels, which may be caused by the imbalance between thin vessels and thick vessels. Furthermore, Yan et al. (12) explored a three-stage network separating the segmentation of thick vessels, thin vessels, and the vessel fusion into different stages to make full use of the difference between thick and thin vessels to improve the overall segmentation performance. Considering that the consecutive pooling may lead to accuracy loss, CE-Net (32) encodes the high-dimension information and preserves spatial information to improve the

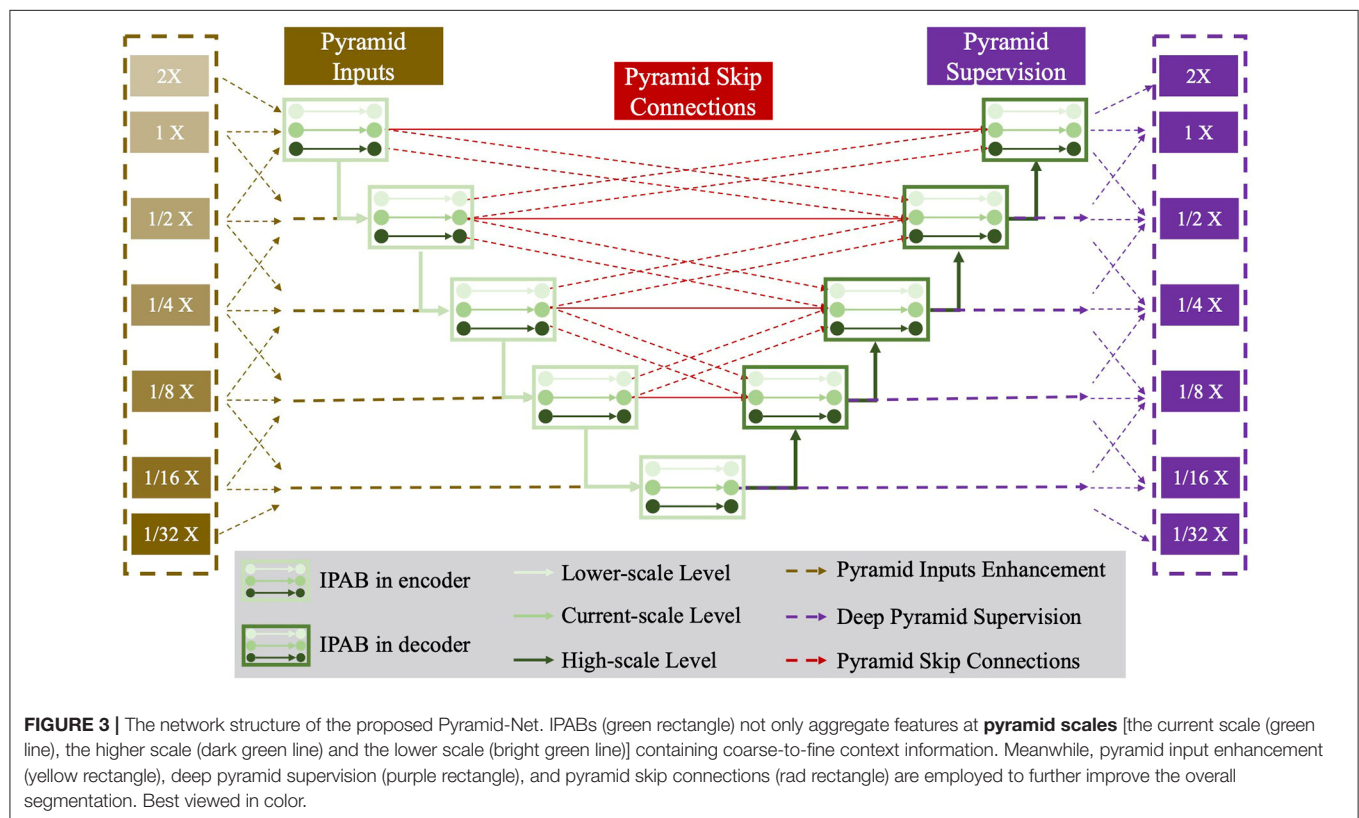
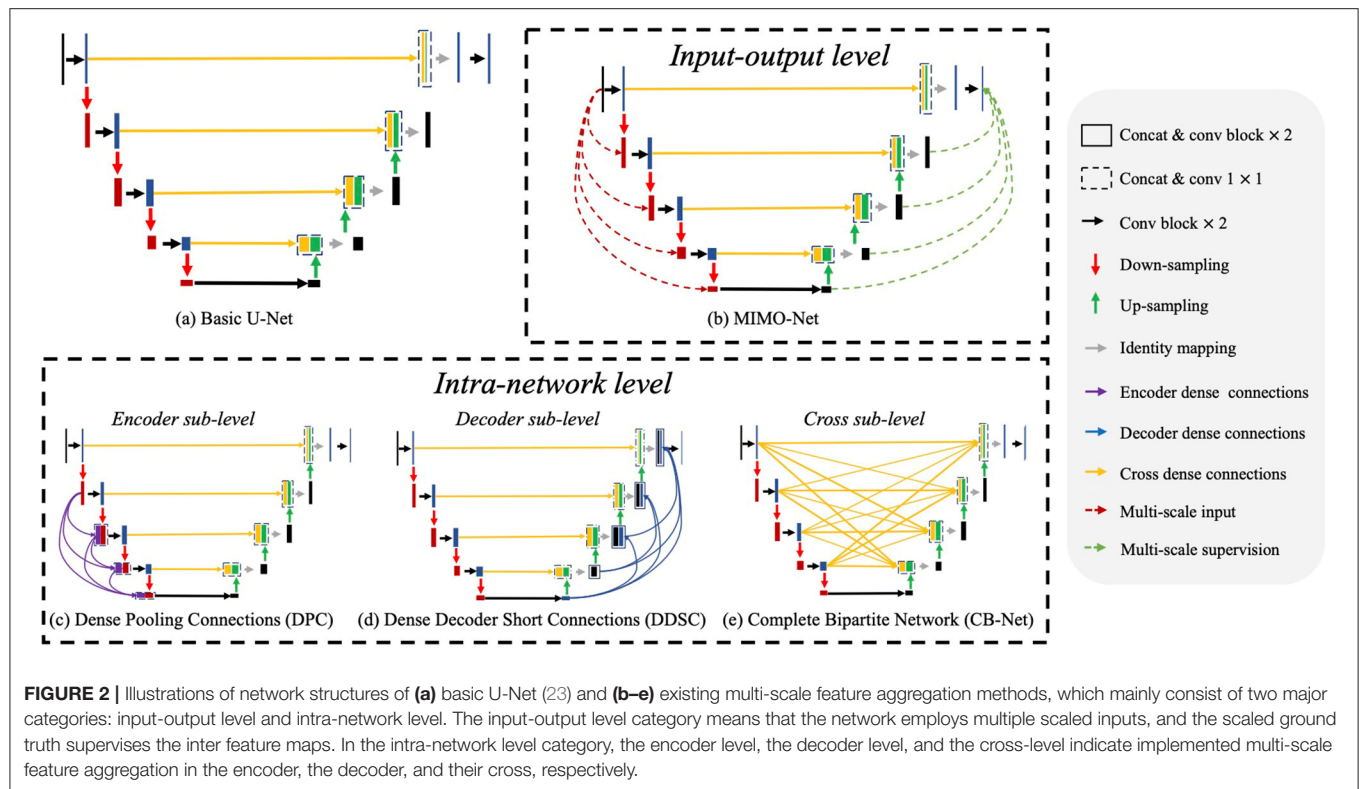
overall segmentation. HA-Net (33) dynamically assigns the regions in the image hard regions or simple regions, and then introduces attention modules to help the network concentrate on the hard region for accurate vessel image segmentation. Meanwhile, some works introduce the spatial attention (34) and the channel attention (34) to the vessel segmentation domain and achieve promising results. The proposed method extends considerably to our previous work (35), which only supply some simplified evaluation on two public available vessel segmentation datasets. In this work, we have added a new module named “pyramid skip connections,” which furthers boost the performance. Meanwhile, we have added another widely-used dataset (STARE) to demonstrate the generalization of our proposed Pyramid-Net. Moreover, in terms of the analysis, we have supplied in-depth analyses of our method including evaluation on thin vessel segmentation, ablation analysis, and cross-training evaluation.

### 2.2. Motivation

Multi-scale feature aggregation is widely used in medical image segmentation, which fuses the previous feature maps with different scales to improve the network performance. As shown in **Figure 2**, recent works (36–39) introduced multi-scale feature aggregation to strengthen feature propagation, alleviate the vanishing gradient problem, and improve the overall segmentation. We divide those methods into two major categories: input-output level and intra-network level.

**Input-output level category:** The connections exist between inputs at various scales and corresponding intermediate layers, or between the intermediate layers and the final predictions with corresponding scales. For example, Wu et al. (40) generated multi-scale feature maps by max-pooling and up-sampling layer and employed two sub-models to extract and aggregate features at multiple scales. MIMO-Net (41) fused scaled input images with multiple resolutions into the intermediate layers of the network in the encoder, and optimized the features in the decoder to improve the overall segmentation performance. MILD-Net (42) fused scaled original images with multiple resolutions to alleviate the potential accuracy decline caused by max-pooling.

**Intra-network level category:** In this approach, features from previous layers are adjusted in channel numbers and spatial dimension and then aggregated with the ones in the later layer. For ease of discussion, we discuss the network structures of related works based on the U-Net as shown in **Figure 2**. Note that U-Net is the most widely-used network in medical image segmentation. These works contain three main approaches: dense connections in the encoder (encoder sub-level), dense connections in the decoder (decoder sub-level) and dense connections in the cross of the encoder and the decoder (cross sub-level): (1) Encoder sub-level: (15) aggregated the scale inputs into the intermediate layers in the encoder to alleviate the accuracy decline caused by pooling; (2) Decoder sub-level: Dense decoder short connections (18) made full use of the feature maps in the decoder by fusing them with the feature maps in later layers; (3) Cross sub-level: Complete bipartite networks (16) inspired by the structure of complete bipartite graphs connected every layer in the encoder and the decoder.



Though multi-scale feature aggregation can significantly improve segmentation performance, we discover that they usually work in an inter-layer manner, **inter-layer feature aggregation**. In such a manner, features at either a lower scale or a higher scale are fused by the current layer. For example, in the encoder, feature maps at the lower scale cannot be fused by that at the current scale because of the processing order of the layers. The same phenomenon also exists in the decoder. Note that a successful segmentation needs to consider both feature maps at high scales for global localization information and low scales for detailed appearance information. Thus, we may mitigate the above problem by performing multi-scale feature aggregation in each layer of the network, **intra-layer feature aggregation**. How to obtain the multi-scale features in each layer becomes another problem. We may use pooling and upsampling to obtain two associated branches operating on a higher scale and a low scale, respectively. In this way, there exist three branches at three different scales (namely **pyramid scales**) in each layer, which is like a ResNet block (43). In this way, we may aggregate coarse-to-fine context information from pyramid-scale feature maps in each layer to further improve the segmentation performance.

### 3. METHODS

In this section, we first introduce IPABs and then describe three optimizations, including pyramid input enhancement, deep pyramid supervision, and pyramid skip connections. **Figure 3** presents the structure details of Pyramid-Net.

#### 3.1. Intra-layer Pyramid-Scale Aggregation Block

Intra-layer pyramid-scale aggregation block are based on the ResNet block (43), which is widely adopted in deep learning. **Figure 4** illustrates the structure of the ResNet block (43), which is formulated as

$$X_{l+1} = f(X_l) + X_l, \quad (1)$$

where  $X_l$  and  $X_{l+1}$  are the input and the output of the current layer, while  $f(\cdot)$  represents the main branch of the current layer. ResNet learns the additive residual function  $f(\cdot)$  with respect to the unit input through a shortcut connection between them. Meanwhile, the multi-scale feature aggregation inspires us to propose associated branches to learn coarse-to-fine features in each residual branch. **Figure 4** illustrates the detailed structures of traditional ResNet blocks and our IPABs. Different from ResNet blocks, in each layer, IPABs generate two associated branches to aggregate coarse-to-fine feature maps to assist the main branch at the current scale. In each branch, the processing steps are almost the same as those in traditional ResNet blocks. Some extra steps such as up-sampling and down-sampling are adopted at the higher and the lower scales to adjust scales. In order to reduce the potential increase of computational cost, the number of channels of the inputs  $X_l$  in the main branch has been reduced to half, while the number of channels of resized

inputs  $X_l^p$  and  $X_l^d$  in the associated branches is reduced to one-fourth. The feature maps with channel adjustment are fed to the processing steps at three scales and are processed in parallel. The three outputs at pyramid scales are then concatenated. The whole process is formulated as follows,

$$\tilde{X}_{l+1} = H(f(\hat{X}_l^p), f(\hat{X}_l), f(\hat{X}_l^d)) + X_l, \quad (2)$$

where  $X_l^p$  and  $X_l^d$  are the up-sampled and the down-sampled results of the current input  $X_l$  with channel adjustment, respectively.  $\hat{X}_l^p$ ,  $\hat{X}_l$  and  $\hat{X}_l^d$  are the enhanced results using pyramid input enhancement, which only exists in the encoder and is detailed in section 3.2. Meanwhile,  $\hat{X}_l^p$ ,  $\hat{X}_l$ , and  $\hat{X}_l^d$  are replaced by  $\hat{X}_l^p$ ,  $\hat{X}_l$ , and  $\hat{X}_l^d$  in the decoder, which represents the enhancement results by pyramid skip connections and are detailed in section 3.4.  $H(\cdot)$  represents the aggregation process, which performs re-scaling and feature concatenation.  $\tilde{X}_{l+1}$  is the strengthened results of  $X_{l+1}$  by IPAB.

The channel attention module selectively emphasizes interdependent channel maps by integrating associated features among all channel maps. To improve the efficiency of feature extraction, we also employ an attention mechanism (44, 45) in IPAB as follows,

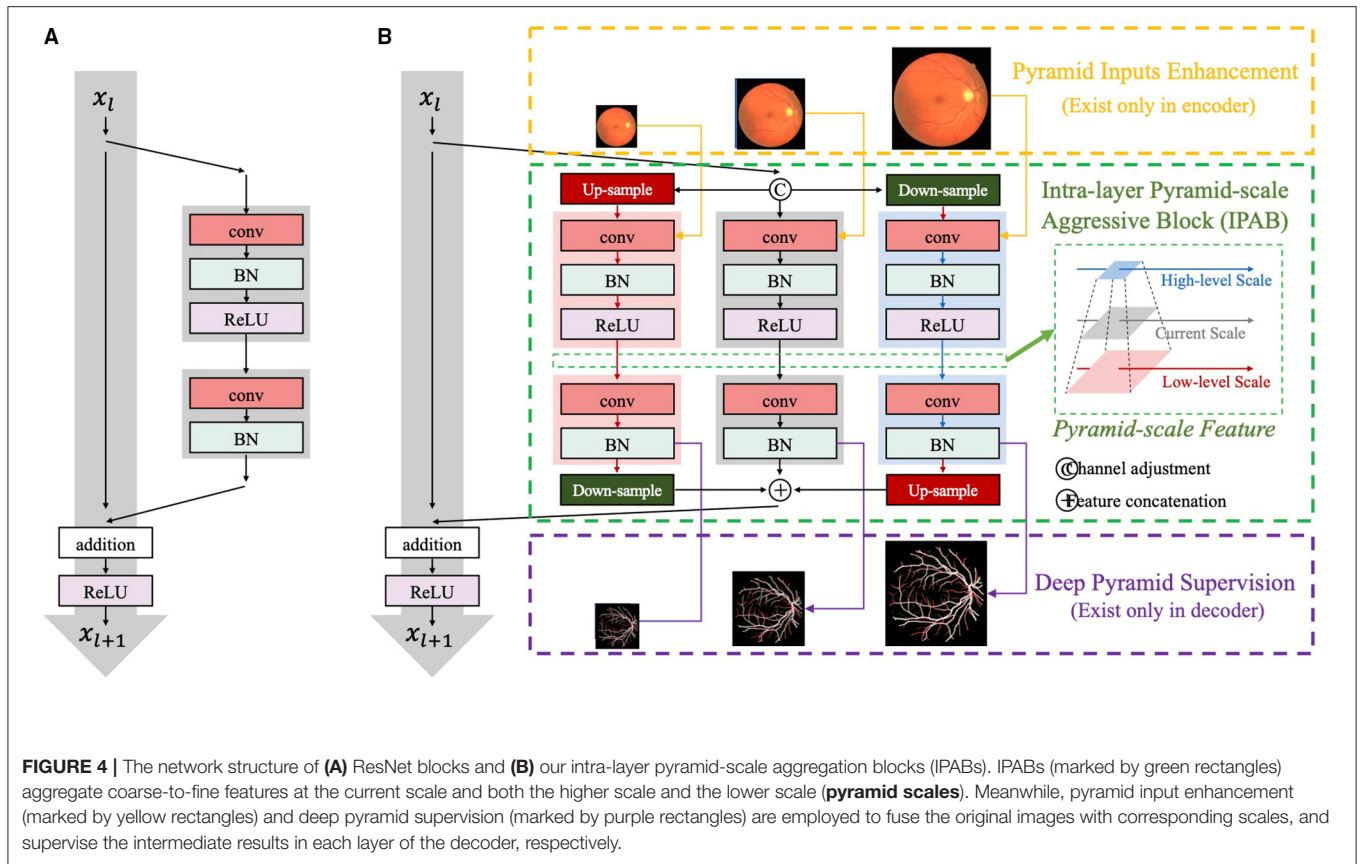
$$\Phi(\tilde{X}_{l+1}) = Q(\Phi_{Avg}(\tilde{X}_{l+1})) + Q(\Phi_{Max}(\tilde{X}_{l+1})). \quad (3)$$

$$\Psi(\tilde{X}_{l+1}) = \sigma(\Phi(\tilde{X}_{l+1})) \otimes \tilde{X}_{l+1}. \quad (4)$$

where  $\Psi(\cdot)$  is the operation of attention process,  $Q$  is the conventional operation using  $1 \times 1$  kernels for channel adjustment, and  $\sigma$  is the activation function. Average-pooling  $\Phi_{Avg}(\cdot)$  and max-pooling  $\Phi_{Max}(\cdot)$  are adopted to aggregate channel information. By utilizing IPAB, each layer of the network aggregates the feature with pyramid scales, which helps fuse coarse-to-fine context information to improve the overall segmentation performance.

#### 3.2. Pyramid Input Enhancement

Pyramid input enhancement fuses the input image with multiple scales to IPABs to reduce the loss of information caused by re-scaling and enhance feature fusion. Pooling operations with various pooling sizes are used to guarantee spatial resolution consistency. Particularly, in each layer, the input image is scaled at higher, current, and lower scales, and fed to three parallel processing steps at multiple scales in the IPAB. Pooling operations over larger regions successively reinforce the scale and translation invariance while reducing noise sensitivity at the same time as more and more context information is added. The aggregation should facilitate discrimination between relevant features and local noises. The above three pyramid-scale images are concatenated with corresponding outputs of up-sampling, down-sampling, and channel adjustment, respectively. Suppose that  $X_l$  is denoted as the input of the current layer, and  $X_l^p$ , and  $X_l^d$  are results at the higher scale and the lower scale, respectively. Meanwhile,  $I_{l-1}$ ,  $I_l$  and  $I_{l+1}$  are the scaled inputs of  $X_l^p$ ,  $X_l$ , and  $X_l^d$



with the same size, respectively. The fusion process of the current scale is formulated as follows,

$$\hat{X}_{l-1} = H(X_l^d, \mathbf{W}^d(I_{l-1})), \quad (5)$$

$$\hat{X}_l = H(X_l, \mathbf{W}(I_l)), \quad (6)$$

$$\hat{X}_{l+1} = H(X_l^p, \mathbf{W}^p(I_{l+1})), \quad (7)$$

where  $\mathbf{W}^p(\cdot)$ ,  $\mathbf{W}^d(\cdot)$ , and  $\mathbf{W}(\cdot)$  represents  $3 \times 3$  convolutional operations and is applied before concatenating to the pyramid-scale features, and  $H(\cdot)$  denotes channel adjustment.

### 3.3. Deep Pyramid Supervision

Deep pyramid supervision optimizes feature maps at multiple scales to improve the segmentation of multi-scale objects and fast the training process. Similar to pyramid input enhancement, deep pyramid supervision connects the intermediate layer to the final prediction thus fusing coarse-to-fine context information. Particularly, the feature maps at multiple scales from each IPAB in the decoder are fed into a plain  $3 \times 3$  convolutional layer followed by Sigmoid function. Deep pyramid supervision at the  $l$ th scale of the decoder can be defined as,

$$L_l = L(Y_l^p, M_{l-1}) + L(Y_l, M_l) + L(Y_l^d, M_{l+1}). \quad (8)$$

The ground truths  $M$  are scaled to the same size as the pyramid-scale feature maps for deep supervision, e.g.,  $Y_l^p$ ,  $Y_l$ , and  $Y_l^d$  are supervised by the corresponding ground truth  $M_{l-1}$ ,  $M_l$ , and  $M_{l+1}$ , respectively. Note that the feature maps in each layer can be directly fused with the final prediction and optimized without massive convolutional processing. Therefore, deep pyramid supervision can be adapted to different depths for different tasks in training, which supply adaptive model capacity, thereby facilitating the segmentation of objects with different scales.

### 3.4. Pyramid Skip Connections

Pyramid skip connections perform feature reuse among the three scaled feature maps (the higher scale, the current scale, and the lower scale) in each IPAB module. Suppose that  $X_l$  is the input of the current layer in the decoder, and  $X_l^p$  and  $X_l^d$  are the results at the higher scale and the lower scale, respectively. Meanwhile,  $(\tilde{X}_l^p, \tilde{X}_{l+1}^d, \tilde{X}_{l+2}^d)$ ,  $(\tilde{X}_{l-1}^p, \tilde{X}_l, \tilde{X}_{l+1}^d)$ , and  $(\tilde{X}_{l-2}^p, \tilde{X}_{l-1}, \tilde{X}_l^d)$  are three groups of learned feature maps from the encoder, and feature maps in each group have the same spatial dimension with the corresponding scaled input  $\hat{X}_{l-1}$ ,  $\hat{X}_l$ , and  $\hat{X}_{l+1}$ , respectively. The fusion process of the current scale is formulated as follows,

$$\hat{X}_{l-1} = H(X_l^d, H(\tilde{X}_l^p, \tilde{X}_{l+1}^d, \tilde{X}_{l+2}^d)), \quad (9)$$

$$\hat{X}_l = H(X_l, H(\tilde{X}_{l-1}^p, \tilde{X}_l, \tilde{X}_{l+1}^d)), \quad (10)$$

$$\hat{X}_{l+1} = H(X_{l+1}^p, H(\tilde{X}_l^d, \tilde{X}_{l-1}, \tilde{X}_{l+2}^p)), \quad (11)$$

where  $H(\cdot)$  denotes channel adjustment. We can see that features at the current-scale  $l$  can reuse and aggregate feature maps at most five scales ( $l - 2, l - 1, l, l + 1$ , and  $l + 2$ ).

## 4. EXPERIMENTS

### 4.1. Datasets

We used three public available retinal vessel datasets, DRIVE (20), STARE (21), and CHASE-DB1 (22) for evaluation. The images in the three datasets are collected using digital retinal imaging, a standard method of documenting the appearance of the retina. More details of the datasets are as follows.

**DRIVE:** The DRIVE dataset (20) consists of 40 images with a resolution of  $565 \times 584$  pixels, which were acquired using a Canon CR5 non-mydiatic 3CCD camera with a 45-degree field of view (FOV). Two trained human observers labeled the vessels in all images, and the ones from the first observer were used for network training. The dataset has been divided into a training and a test set (20), both of which contain 20 images.

**CHASE-DB1:** The CHASE-DB1 dataset (22) contains vascular patch images with a resolution of  $999 \times 960$ , which were acquired from 28 eyes of 14 ten-year-old children. Since images were captured in subdued lighting and the operators adjusted illumination settings, the images contain more illumination variation in CHASE-DB1 compared with the DRIVE datasets. Following the configuration in Li et al. (46), the first 20 images and the remaining 8 images are employed as the training set and the test set, respectively.

**STARE:** The STARE dataset (21) consists of 20 equal-sized images with a resolution of  $700 \times 605$  pixels. Each image is with a  $35^\circ$  FOV, and half of the images of eyes are with ocular pathology. As the training set and the test set are not explicitly specified, the same leave-one-out cross-validation is adopted (33) for performance evaluation, where models are iteratively trained on 19 images and tested on the rest images. Liking other methods (10), manual annotations generated by the first observer are used for both training and test.

### 4.2. Implementations

All experiments were conducted on an Nvidia GeForce Titan X (pascal) containing 12 GB memory. Meanwhile, we employed CE-Net (32), one of the state-of-the-art methods in retinal vessel segmentation, as the backbone models to implement IPABs, pyramid input enhancement, deep pyramid supervision, and pyramid skip connections. Normalization of the training data has been implemented. In order to express the details of multi-scale feature fusion more clearly, we use U-Net as the basic network to explain, which is widely used in the medical image segmentation domain. In practice, we use the state-of-the-art method CE-Net to replace U-Net to obtain better performance. During training, we adopted Adaptive Moment Estimation (Adam) as the learning optimizer with a batch size of 4. Data augmentation operations including horizontal flip, vertical flip, and diagonal flip are used

**TABLE 1 |** Performance comparison of Pyramid-Net and the state-of-the-art methods on the DRIVE dataset.

Method	Sens (%)	Spec (%)	Acc (%)	AUC (%)
FCN (31)	74.89	96.21	94.13	95.67
U-Net (23)	75.31	96.45	94.45	96.01
DeepVessel (11)	76.12	97.68	95.23	97.52
(10)	76.53	98.18	95.42	97.52
(47)	77.92	98.13	95.56	97.84
(40)	78.44	98.07	95.67	98.19
CE-Net (32)	83.09	97.47	95.45	97.79
BTS-DSN (48)	78.91	98.04	95.61	98.06
(49)	79.16	98.11	95.70	98.10
(50)	79.40	98.16	95.67	97.72
Vessel-Net (51)	80.38	98.02	95.78	98.21
MResU-Net (52)	79.69	97.99	-	97.99
CTF-Net (53)	78.49	98.13	95.67	97.88
Hybrid-Net (6)	<b>83.53</b>	97.51	95.79	-
HA-Net (33)	79.91	98.13	95.81	98.23
<b>Pyramid-Net</b>	82.38	<b>98.19</b>	<b>96.26</b>	<b>98.32</b>

*Bold values mean the state-of-the-art performance.*

to enlarge the train samples. We use a threshold to obtain the final segmentation from pixel probability vectors. Particularly, the pixels with values smaller than the threshold are assigned to the background class, and the remaining pixels with values equal to or greater than the threshold are categorized as the vessel class. The final prediction is the ensemble of the segmentation output of the vessel images, its rotation ( $90^\circ$ ), and its flip (horizontal and vertical).

### 4.3. Evaluation Metrics

We introduce four evaluation metrics including Sensitivity (Sens), Specificity (Spec), Accuracy (Acc), and Area Under the ROC Curve (AUC) to validate our proposed Pyramid-Net. The metrics are calculated as follows:

$$\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN}), \quad (12)$$

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}), \quad (13)$$

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}). \quad (14)$$

True positive (TP) and true negative (TN) present that pixels are correctly classified to objects or backgrounds, respectively. Meanwhile, pixels will be labeled as false positive (FP) or false negative (FN), if they are misclassified to objects or backgrounds, respectively.

### 4.4. Quantitative Results

We compared our Pyramid-Net with existing state-of-the-art works on three vessel image segmentation datasets (DRIVE, CHASE-DB1, and STARE). **Tables 1–3** illustrate the comparison results of Pyramid-Net and the current state-of-the-art methods.

**TABLE 2 |** Performance comparison of Pyramid-Net and the state-of-the-art methods on the CHASE-DB1 dataset.

Method	Sens (%)	Spec (%)	Acc (%)	AUC (%)
(54)	76.15	95.75	94.67	96.23
(46)	75.07	97.93	95.81	97.16
(55)	81.94	97.39	96.30	-
(10)	76.33	98.09	96.10	97.81
(47)	77.56	98.20	96.34	98.15
FCN (31)	76.41	98.06	96.07	97.76
(56)	81.55	97.52	96.10	98.04
(48)	78.88	98.01	96.27	98.40
(50)	80.74	98.21	96.61	98.12
(51)	81.32	98.14	96.61	98.60
Three-stage (12)	76.41	98.06	96.07	97.76
CTF-Net (52)	79.48	<b>98.42</b>	96.48	98.47
Hybrid-Net (6)	81.76	97.76	96.32	-
HA-Net (33)	<b>82.39</b>	98.13	96.70	98.70
Pyramid-Net	81.17	98.26	<b>96.89</b>	<b>98.92</b>

*Bold values mean the state-of-the-art performance.*

**TABLE 3 |** Performance comparison of Pyramid-Net and the state-of-the-art methods on the STARE dataset.

Method	Sens (%)	Spec (%)	Acc (%)	AUC (%)
(54)	73.20	98.40	95.60	96.70
(57)	77.91	97.58	95.54	97.48
(58)	76.80	97.38	-	-
(10)	75.81	98.46	96.12	98.01
(56)	75.95	98.78	96.41	98.32
Three-stage (12)	77.35	98.57	96.38	98.33
MResU-Net (52)	81.01	97.95	-	98.16
Hybrid-Net (6)	79.46	98.21	96.26	-
HA-Net (33)	81.86	98.44	96.73	98.32
<b>Pyramid-Net</b>	<b>82.35</b>	<b>98.87</b>	<b>97.19</b>	<b>98.62</b>

*Bold values mean the state-of-the-art performance.*

For the DRIVE dataset, Pyramid-Net achieves a high score of 82.38, 98.19, 96.26, and 98.32% on Sens, Spec, Acc, and AUC, respectively, and outperforms state-of-the-art methods in three metrics including Spec, Acc, and AUC. In terms of Sens, CE-Net achieves the best performance of 83.09%, while our method achieves a comparable result, which is 0.71% lower. Overall, Pyramid-Net achieves higher overall performance than CE-Net. For the CHASE-DB1 dataset, compared with the state-of-the-art results, the proposed Pyramid-Net achieves high score of 81.17, 98.26, 96.89, and 98.92% for Sens, Spec, Acc, and AUC, respectively, which consistently enjoys a better performance than all the current state-of-the-art methods. For the STARE dataset, Pyramid-Net achieves a promising score of 82.35, 98.87, 97.19, and 98.62% for Sens, Spec, Acc, and AUC, respectively, which is also consistently better than all the current state-of-the-art methods. The consistent improvements in **Tables 1–3** indicate the effectiveness and robustness of our Pyramid-Net.

## 4.5. Qualitative Results

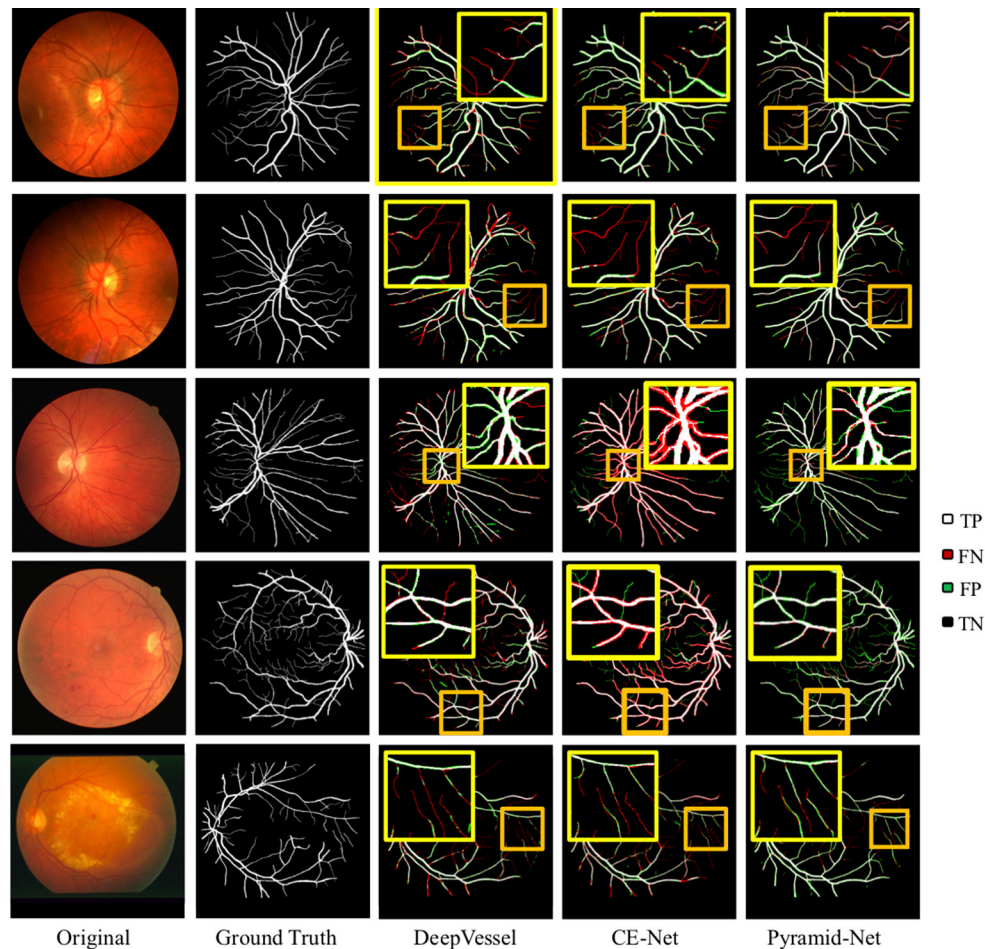
The visual comparisons between Pyramid-Net and the state-of-the-art methods, including DeepVessel and CE-Net on the DRIVE dataset and the CHASE-DB1 dataset are shown in **Figure 5**. White (TP) and black (TN) pixels are correct predictions of vessels and the background, respectively, while red (FP) and green (FN) pixels are incorrect predictions. In **Figure 5**, dark yellow rectangles contain the selected areas used for detail comparison, and the bright yellow rectangles contain the zoomed area in the dark yellow rectangle. We can notice that current methods enjoy a good performance on the segmentation of main retinal vessels, but the effect on some capillaries is poor. For example, Row 1 of **Figure 5** shows that the result of DeepVessel misses a large number of thin vessels on the DRIVE dataset, and that of CE-Net obtains a much better accuracy on thin vessels. However, in Row 2, there is no significant difference between the results of the two methods. In both Rows 1 and 2 of **Figure 5**, our method can achieve much higher accuracy, but we can still notice that our method cannot segment them correctly if the vessels are too thin. We can further observe that our method has much fewer false-negative pixels (indicated by green) than the other two. This may due to the fact that our proposed IPABs can consider more scales thus improving the segmentation accuracy. Overall, our proposed Pyramid-Net evidently improves the segmentation performance, especially for those narrow, low-contrast, and ambiguous retinal vessels.

## 4.6. Evaluation on Thin Vessels

In the previous subsection, the results in **Figure 5** indicate that though the main vessels enjoy a promising segmentation performance, the segmentation of thin vessels always suffers a big miss in the prediction. In practice, it is challenging to segment the thin vessels from the complex retina background, which are always low-contrast and extremely narrow (1–2 pixels). Thus, in this subsection, to evaluate the effectiveness of Pyramid-Net on thin vessels, we compared Pyramid-Net with the state-of-the-art methods on an additional dataset only containing thin vessel labels. Vessels with a width of 1 or 2 pixels are commonly regarded as the thin vessels in the DRIVE dataset. To avoid potential unfair in the evaluation on the manual addition label of the thin vessel, we distinguish thick vessels from thin vessels by an opening operation (10). The evaluation results are summarized in **Table 4**. It can be noticed that Pyramid-Net achieves a high ACC score of 96.26, 96.51, and 91.64% on all vessels, thick vessels, and thin vessels, respectively. Overall, our method outperforms the state-of-the-art methods on all metrics. As for the thin vessel segmentation, our methods achieve an improvement of 4.73% over backbone model CE-Net and outperforms the state-of-the-art method by about 3.86%. The experiment results indicate that our Pyramid-Net is particularly effective on thin vessels.

## 4.7. Ablation Analysis

To justify the effectiveness of IPABs, pyramid input enhancement, deep pyramid supervision, and pyramid skip connections in the proposed Pyramid-Net, we conduct ablation analysis using the DRIVE dataset as a vehicle. The ablation experimental results are summarized in **Table 5**. We use CE-Net



**FIGURE 5 |** Visual comparison of Pyramid-Net and the state-of-the-art methods including DeepVessel (11) and CE-Net (32) on DRIVE (Row 1–2), CHASE-DB1 (Row 3–4), and STARE (Row 5) datasets. White (TP) and black (TN) pixels indicate correct predictions of object and background, respectively, while red (FN) and green (FP) pixels indicate incorrect predictions. The dark yellow rectangle contains the area used to compare segmentation details, and the bright yellow rectangle contains the zoomed area in the dark yellow rectangle. Best viewed in color.

(32) as our backbone, which achieves a good score of 95.45 and 97.79% on Acc and on AUC, respectively. Firstly, we evaluate the effectiveness of IPABs on the backbone. Benefiting from aggregating coarse-to-fine context information from pyramid scale in each layer, the backbone model with IPABs achieves improvements of 0.62% on Acc and 0.30% on AUC. Second, we evaluate pyramid input enhancement and deep pyramid supervision to feed the original image at multiple scales into the network and supervise the immediate layers contains features at various scales. In **Table 5**, we can notice that the above two optimizations achieve improvements of more than 0.10 and 0.07% in AUC, respectively. Third, pyramid skip connections connect the encoder and the decoder and make full use of the features from multiple layers and scales in the encoder, which achieves an improvement of about 0.15% on AUC. Overall, integrating the pyramid-scale concept into the design of the basic unit and skip connections can obviously improve the network segmentation, and the other two optimizations also bring some improvement.

**TABLE 4 |** Performance comparison on thick and thin vessels of Pyramid-Net on the DRIVE dataset.

Method	All vessel (%)	Thick vessel (%)	Thin vessel (%)
(10)	95.42	95.78	87.78
CE-Net (32)	95.45	95.96	86.91
<b>Pyramid-Net</b>	<b>96.26</b>	<b>96.51</b>	<b>91.64</b>

*Bold values mean the state-of-the-art performance.*

#### 4.8. Cross-Training Evaluation

To evaluate the generalization of Pyramid-Net, we performed a cross-training evaluation on the DRIVE dataset and the STARE dataset. We directly implemented our models trained on the source dataset and tested on the target dataset for fair comparisons. The experimental results are summarized in **Table 6**. Overall, our method achieves the state-of-the-art transfer performance on both configurations. Particularly, for the configuration that models are trained on the STARE dataset

**TABLE 5 |** Ablation analysis of Pyramid-Net on the DRIVE dataset.

Method	Acc (%)	AUC (%)
Baseline	95.45	97.79
Baseline + IPABs	96.07	98.09
Baseline + IPABs + pyramid input	96.10	98.15
Baseline + IPABs + Pyramid supervision	96.15	98.12
Baseline + IPABs + pyramid skip connection	96.21	98.24
<b>Pyramid-Net</b>	<b>96.26</b>	<b>98.32</b>

*Bold values mean the state-of-the-art performance.*

**TABLE 6 |** Cross-training evaluation on the DRIVE dataset and the STARE dataset.

Method	Sens (%)	Spec (%)	Acc (%)	AUC (%)
<b>DRIVE (train) -&gt; STARE (test)</b>				
(12)	70.14	98.02	94.44	95.68
(56)	65.05	99.14	94.81	97.18
HA-Net (33)	71.40	98.79	95.30	97.58
Pyramid-Net	<b>75.71</b>	<b>98.86</b>	<b>95.57</b>	<b>97.78</b>
<b>STARE (train) -&gt; DRIVE (test)</b>				
(12)	73.19	98.40	95.80	96.78
(56)	70.00	97.59	94.74	97.18
HA-Net (33)	81.87	<b>98.79</b>	95.30	97.58
Pyramid-Net	<b>82.67</b>	98.76	<b>95.36</b>	<b>97.72</b>

*Bold values mean the state-of-the-art performance.*

and tested on the DRIVE dataset, it can be noticed that the transfer model can achieve competitive results on Spec and suffer a big loss of accuracy on Sens. The potential reason is the imbalance between thick vessels and thin vessels in the STARE dataset. Manual annotations of the STARE dataset contain more thick vessels than thin vessels, which led that the pre-trained model on the STARE dataset obtains a bad segmentation performance of thin vessels on the DRIVE dataset. When the conditions are reversed, the above situation is alleviated, and the corresponding scores on Sens, Spec, Acc, and AUC on the STARE dataset are comparable with the model trained on the STARE dataset.

#### 4.9. Comparison With Multi-Scale Aggregation Methods

To evaluate the effectiveness of the multi-scale information aggregated in the proposed Pyramid-Net, we compare existing multi-scale aggregation methods, including Dense Pooling Connections (15), Complete Bipartite Network (CB-Net) (16), Dense Decoder Short Connections (DDSC) (18), and U-Net++ (17) on the DRIVE dataset. For fair comparisons, we directly implement those different connection styles and our Pyramid-Net on U-Net (23). The comparison results and the  $p$ -values for the paired  $t$ -test are summarized in Table 7. Compared with existing methods, our method outperforms them by 0.65–0.99% and 0.67–1.50% on Acc and AUC, respectively. On the other hand, we also compare the computational cost of the proposed Pyramid-Net with existing methods. Obviously,

**TABLE 7 |** Comparison with existing multi-scale aggregation methods on the DRIVE Dataset.

Method	Acc (%)	AUC (%)	FLOPs	$p$ -values
U-Net (23)	94.45	96.01	334.95G	<0.01
DPC (15)	95.56	97.65	351.33G	<0.01
CB-Net (16)	95.61	97.52	441.62G	<0.01
DDSC (18)	95.42	97.48	381.07G	<0.01
U-Net ++ (17)	95.27	96.82	828.69G	<0.01
CE-Net (32)	95.45	97.79	-	<0.05
<b>Pyramid-Net</b>	<b>96.26</b>	<b>98.32</b>	188.15G	-

*Bold values mean the state-of-the-art performance.*

existing methods improve the network performance and increase the computational cost by 16.38–493.74G (104.9–247.4%) on FLOPs from the numerous feature reuse. Particularly, our proposed Pyramid-Net achieves state-of-the-art performance with a computational cost reduced by 216.8G (64.7%) on FLOPs. The reason for the above phenomenon is the channel reduction in each IPAB. The channels' main branch is reduced to half, while the number of channels at associated branches is half of that of the main branch. Overall, our method achieves the state-of-the-art performance of 96.26% on Acc and 98.32% on AUC with a 64.7% reduction on FLOPs.

## 5. CONCLUSION

In this paper, we introduced Pyramid-Net for accurate retinal vessel segmentation. In Pyramid-Net, the proposed IPABs are utilized to generalize two associated branches to aggregate coarse-to-fine feature maps at pyramid scales to improve the segmentation performance. Meanwhile, three optimizations including pyramid inputs enhancement, deep pyramid supervision, and pyramid skip connections are implemented with IPABs in the encoder, the decoder, and the cross of the two to further improve performance, respectively. Comprehensive experiments have been conducted on three retinal vessel segmentation datasets, including DRIVE (20), STARE (21), and CHASE-DB1 (22). Experimental results demonstrate that our IPABs can efficiently improve the segmentation performance, especially for thin vessels. In addition, our method is also much more efficient than existing methods with a large reduction in computational cost.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors/s.

## AUTHOR CONTRIBUTIONS

XX is the guarantor of the manuscript. JZh implemented the experiments and wrote the first draft of the manuscript. HQ, WX, and ZY managed the result analysis. All authors contributed to drawing up the manuscript.

## FUNDING

This work was supported by the National Key Research and Development Program of China (no. 2018YFC1002600), the Science and Technology Planning Project of

Guangdong Province, China (nos. 2017B090904034, 2017B030314109, 2018B090944002, and 2019B020230003), Guangdong Peak Project (no. DFJH201802), and the National Natural Science Foundation of China (no. 62006050).

## REFERENCES

- Winder RJ, Morrow PJ, McRitchie IN, Bailie J, Hart PM. Algorithms for digital image processing in diabetic retinopathy. *Comput Med Imaging Graph.* (2009) 33:608–22. doi: 10.1016/j.compmedimag.2009.06.003
- Mitchell P, Leung H, Wang JJ, Rochtchina E, Lee AJ, Wong TY, et al. Retinal vessel diameter and open-angle glaucoma: the blue mountains eye study. *Ophthalmology.* (2005) 112:245–50. doi: 10.1016/j.ophtha.2004.08.015
- Yannuzzi LA, Negr ao S, Tomohiro I, Carvalho C, Rodriguez-Coleman H, Slakter J, et al. Retinal angiomatous proliferation in age-related macular degeneration. *Retina.* (2012) 32:416–34. doi: 10.1097/IAE.0b013e31823f9b3b
- Ikram MK, Witteman JC, Vingerling JR, Breteler MM, Hofman A, de Jong PT. Retinal vessel diameters and risk of hypertension: the Rotterdam Study. *Hypertension.* (2006) 47:189–94. doi: 10.1161/01.HYP.0000199104.61945.33
- Gishti O, Jaddoe VW, Felix JE, Klaver CC, Hofman A, Wong TY, et al. Retinal microvasculature and cardiovascular health in childhood. *Pediatrics.* (2015) 135:678–85. doi: 10.1542/peds.2014-3341
- Yang L, Wang H, Zeng Q, Liu Y, Bian G. A hybrid deep segmentation network for fundus vessels via deep-learning framework. *Neurocomputing.* (2021) 448:168–78. doi: 10.1016/j.neucom.2021.03.085
- Guo S, Li T, Kang H, Li N, Zhang Y, Wang K. L-Seg: an end-to-end unified framework for multi-lesion segmentation of fundus images. *Neurocomputing.* (2019) 349:52–63. doi: 10.1016/j.neucom.2019.04.019
- Cheung CYL, Zheng Y, Hsu W, Lee ML, Lau QP, Mitchell P, et al. Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors. *Ophthalmology.* (2011) 118:812–8. doi: 10.1016/j.ophtha.2010.08.045
- Xu X, Lu Q, Yang L, Hu S, Chen D, Hu Y, et al. Quantization of fully convolutional networks for accurate biomedical image segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake city, UT: IEEE (2018). p. 8300–8.
- Yan Z, Yang X, Cheng KT. Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. *IEEE Trans Biomed Eng.* (2018) 65:1912–23. doi: 10.1109/TBME.2018.2828137
- Fu H, Xu Y, Lin S, Wong DWK, Liu J. Deepvessel: retinal vessel segmentation via deep learning and conditional random field. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Athens: Springer (2016). p. 132–9.
- Yan Z, Yang X, Cheng KT. A three-stage deep learning model for accurate retinal vessel segmentation. *Biomed Health Inf IEEE J.* (2019) 23:1427–36. doi: 10.1109/JBHI.2018.2872813
- Katz N, Goldbaum M, Nelson M, Chaudhuri S. An image processing system for automatic retina diagnosis. In: *Three-Dimensional Imaging and Remote Sensing Imaging*. Vol. 902. Los Angeles, CA: International Society for Optics and Photonics. (1988). p. 131–7.
- Spencer T, Olson JA, McHardy KC, Sharp PF, Forrester JV. An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. *Comput Biomed Res.* (1996) 29:284–302. doi: 10.1006/cbmr.1996.0021
- Playout C, Duval R, Cheriet F. A multitask learning architecture for simultaneous segmentation of bright and red lesions in fundus images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Granada: Springer (2018). p. 101–8.
- Chen J, Banerjee S, Grama A, Scheirer WJ, Chen DZ. Neuron segmentation using deep complete bipartite networks. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Quebec, QC: Springer (2017). p. 21–9. doi: 10.1007/978-3-319-66185-8\_3
- Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. U-net++: A nested u-net architecture for medical image segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Granada: Springer (2018). p. 3–11.
- Bilinski P, Priscacariu V. Dense decoder shortcut connections for single-pass semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Granada (2018) p. 6596–605.
- Ding H, Pan Z, Cen Q, Li Y, Chen S. Multi-scale fully convolutional network for gland segmentation using three-class classification. *Neurocomputing.* (2020) 380:150–61. doi: 10.1016/j.neucom.2019.10.097
- Staal J, Abràmoff MD, Niemeijer M, Viergever MA, Van Ginneken B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging.* (2004) 23:501–9. doi: 10.1109/TMI.2004.825627
- Hoover A, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans Med Imaging.* (2000) 19:203–10. doi: 10.1109/42.845178
- Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, et al. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans Biomed Eng.* (2012) 59:2538–48. doi: 10.1109/TBME.2012.2205687
- Ronneberger O, Fischer P, Brox T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Munich: Springer International Publishing (2015).
- Niemeijer M, Staal J, van Ginneken B, Loog M, Abramoff MD. Comparative study of retinal vessel segmentation methods on a new publicly available database. In: *Medical imaging 2004: image processing*. Vol. 5370. International Society for Optics and Photonics. San Diego, CA (2004). p. 648–56.
- Sinthanayothin C, Boyce JF, Cook HL, Williamson TH. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *Br J Ophthalmol.* (1999) 83:902–10. doi: 10.1136/bjo.83.8.902
- Soares JV, Leandro JJ, Cesar RM, Jelinek HF, Cree MJ. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Trans Med Imaging.* (2006) 25:1214–22. doi: 10.1109/TMI.2006.879967
- Rangayyan RM, Ayres FJ, Oloumi F, Oloumi F, Eshghzadeh-Zanjani P. Detection of blood vessels in the retina with multiscale Gabor filters. *J Electron Imaging.* (2008) 17:023018. doi: 10.1117/1.2907209
- Ricci E, Perfetti R. Retinal blood vessel segmentation using line operators and support vector classification. *IEEE Trans Med Imaging.* (2007) 26:1357–1365. doi: 10.1109/TMI.2007.898551
- Franklin SW, Rajan SE. Retinal vessel segmentation employing ANN technique by Gabor and moment invariants-based features. *Appl Soft Comput.* (2014) 22:94–100. doi: 10.1016/j.asoc.2014.04.024
- Zhang J, Chen Y, Bekkers E, Wang M, Dashtbozorg B, ter Haar Romeny BM. Retinal vessel delineation using a brain-inspired wavelet transform and random forest. *Pattern Recognit.* (2017) 69:107–23. doi: 10.1016/j.patcog.2017.04.008
- Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA: IEEE (2015). p. 3431–40.
- Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, et al. CE-Net: context encoder network for 2D medical image segmentation. *IEEE Trans Med Imaging.* (2019) 38:2281–2292. doi: 10.1109/TMI.2019.2903562
- Wang D, Haytham A, Pottenburgh J, Saedi OJ, Tao Y. Hard attention net for automatic retinal vessel segmentation. *IEEE J Biomed Health Inf.* (2020) 24:3384–96. doi: 10.1109/JBHI.2020.3002985
- Guo C, Szemenyei M, Yi Y, Zhou W, Bian H. Residual spatial attention network for retinal vessel segmentation. In: *International Conference on Neural Information Processing*. San Diego, CA: Springer (2020). p. 509–19.

35. Zhang J, Zhang Y, Xu X. Pyramid U-net for retinal vessel segmentation. In: *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Toronto, ON: IEEE (2021). p. 1125–9.
36. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The importance of skip connections in biomedical image segmentation. In: *Deep Learning and Data Labeling for Medical Applications*. Athens: Springer (2016). p. 179–87.
37. Zhang J, Jin Y, Xu J, Xu X, Zhang Y. Mdu-net: Multi-scale densely connected u-net for biomedical image segmentation. *arXiv preprint arXiv:181200352*. (2018).
38. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI: IEEE (2017). p. 6230–9.
39. Chen LC, Yang Y, Wang J, Xu W, Yuille AL. Attention to scale: Scale-aware semantic image segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* Las Vegas, NV: IEEE (2016). p. 3640–9.
40. Wu Y, Xia Y, Song Y, Zhang Y, Cai W. Multiscale network followed network model for retinal vessel segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Granada: Springer (2018). p. 119–26.
41. Raza SEA, Cheung L, Epstein D, Pelengaris S, Khan M, Rajpoot NM. MIMO-Net: a multi-input multi-output convolutional neural network for cell segmentation in fluorescence microscopy images. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* Melbourne, VIC: IEEE (2017). p. 337–40.
42. Graham S, Chen H, Gamper J, Dou Q, Heng PA, Snead D, et al. MILD-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images. *Med Image Anal.* (2019) 52:199–211. doi: 10.1016/j.media.2018.12.001
43. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* Las Vegas, NV: IEEE (2016). p. 770–8.
44. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* Salt Lake City, UT: IEEE (2018). p. 7132–41.
45. Li S, Zhang J, Ruan C, Zhang Y. Multi-stage attention-unet for wireless capsule endoscopy image bleeding area segmentation. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. San Diego, CA: IEEE (2019). p. 818–25.
46. Li Q, Feng B, Xie L, Liang P, Zhang H, Wang T. A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Trans Med Imaging.* (2015) 35:109–18. doi: 10.1109/TMI.2015.2457891
47. Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv[Preprint]*.arXiv:180206955. (2018) doi: 10.1109/NAECON.2018.8556686
48. Guo S, Wang K, Kang H, Zhang Y, Gao Y, Li T. BTS-DSN: deeply supervised neural network with short connections for retinal vessel segmentation. *Int J Med Inform.* (2019) 126:105–13. doi: 10.1016/j.ijmedinf.2019.03.015
49. Ma W, Yu S, Ma K, Wang J, Ding X, Zheng Y. Multi-task neural networks with spatial activation for retinal vessel segmentation and artery/vein classification. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Shenzhen: Springer (2019). p. 769–78.
50. Wang B, Qiu S, He H. Dual encoding u-net for retinal vessel segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Shenzhen: Springer (2019). p. 84–92.
51. Wu Y, Xia Y, Song Y, Zhang D, Liu D, Zhang C, et al. Vessel-Net: retinal vessel segmentation under multi-path supervision. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Shenzhen: Springer (2019). p. 264–72.
52. Li D, Dharmawan DA, Ng BP, Rahardja S. Residual u-net for retinal vessel segmentation. In: *2019 IEEE International Conference on Image Processing (ICIP)*. Taipei: IEEE (2019). p. 1425–9.
53. Wang K, Zhang X, Huang S, Wang Q, Chen F. Ctf-net: retinal vessel segmentation via deep coarse-to-fine supervision network. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. Iowa City, IA: IEEE (2020). p. 1237–41.
54. Roychowdhury S, Koozekanani DD, Parhi KK. Iterative vessel segmentation of fundus images. *IEEE Trans Biomed Eng.* (2015) 62:1738–49. doi: 10.1109/TBME.2015.2403295
55. Kassim YM, Palaniappan K. Extracting retinal vascular networks using deep learning architecture. In: *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Kansas City, MO: IEEE (2017). p. 1170–4.
56. Jin Q, Meng Z, Pham TD, Chen Q, Wei L, Su R. DUNet: A deformable network for retinal vessel segmentation. *Knowl Based Syst.* (2019) 178:149–62. doi: 10.1016/j.knosys.2019.04.025
57. Zhang J, Dashtbozorg B, Bekkers E, Pluim JP, Duits R, ter Haar Romeny BM. Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. *IEEE Trans Med Imaging.* (2016) 35:2631–2644. doi: 10.1109/TMI.2016.2587062
58. Orlando JI, Prokofyeva E, Blaschko MB. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE Trans Biomed Eng.* (2016) 64:16–27. doi: 10.1109/TBME.2016.2535311

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhang, Zhang, Qiu, Xie, Yao, Yuan, Jia, Wang, Shi, Huang, Zhuang and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Red Blood Cell Classification Based on Attention Residual Feature Pyramid Network

Weiying Song<sup>1</sup>, Pu Huang<sup>1\*</sup>, Jing Wang<sup>2</sup>, Yajuan Shen<sup>2</sup>, Jian Zhang<sup>2</sup>, Zhiming Lu<sup>2\*</sup>, Dengwang Li<sup>1\*</sup> and Danhua Liu<sup>1\*</sup>

<sup>1</sup> Shandong Key Laboratory of Medical Physics and Image Processing, Shandong Institute of Industrial Technology for Health Sciences and Precision Medicine, School of Physics and Electronics, Shandong Normal University, Jinan, China,

<sup>2</sup> Department of Clinical Laboratory, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China

## OPEN ACCESS

### Edited by:

Jun Feng,  
Northwest University, China

### Reviewed by:

Nurlan Dauletbayev,  
McGill University, Canada  
Karim A. Mohamed Al-Jashamy,  
SEGi University, Malaysia

### \*Correspondence:

Danhua Liu  
liudanhua@sdsu.edu.cn  
Dengwang Li  
dengwang@sdsu.edu.cn  
Zhiming Lu  
luzhiming@sdsu.edu.cn  
Pu Huang  
pu.wong@139.com

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

**Received:** 14 July 2021

**Accepted:** 25 November 2021

**Published:** 14 December 2021

### Citation:

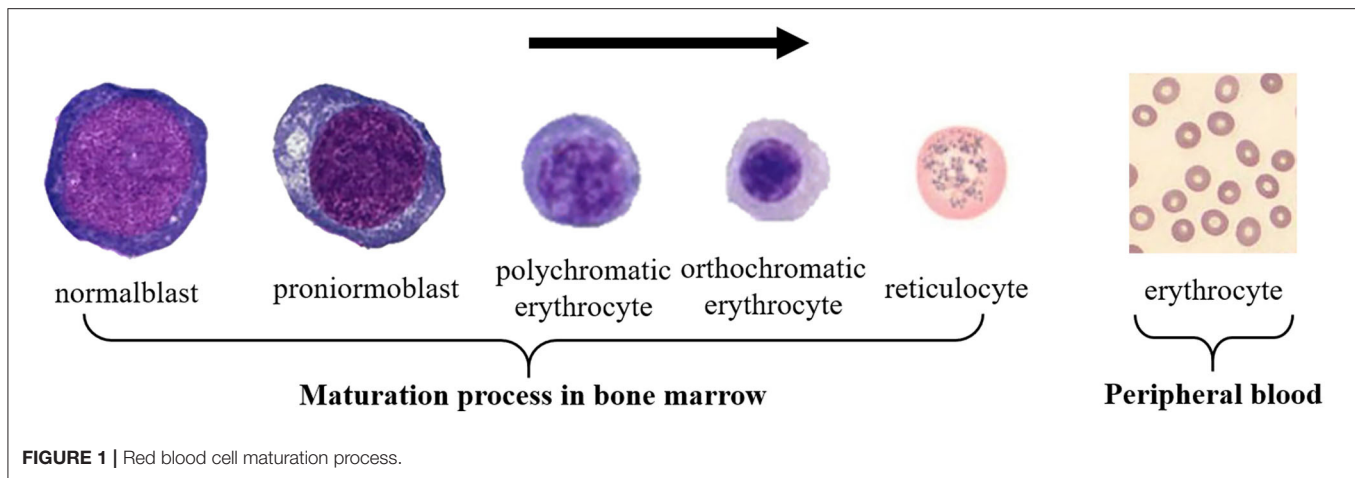
Song W, Huang P, Wang J, Shen Y,  
Zhang J, Lu Z, Li D and Liu D (2021)  
Red Blood Cell Classification Based  
on Attention Residual Feature Pyramid  
Network. *Front. Med.* 8:741407.  
doi: 10.3389/fmed.2021.741407

Clinically, red blood cell abnormalities are closely related to tumor diseases, red blood cell diseases, internal medicine, and other diseases. Red blood cell classification is the key to detecting red blood cell abnormalities. Traditional red blood cell classification is done manually by doctors, which requires a lot of manpower produces subjective results. This paper proposes an Attention-based Residual Feature Pyramid Network (ARFPN) to classify 14 types of red blood cells to assist the diagnosis of related diseases. The model performs classification directly on the entire red blood cell image. Meanwhile, a spatial attention mechanism and channel attention mechanism are combined with residual units to improve the expression of category-related features and achieve accurate extraction of features. Besides, the RoI align method is used to reduce the loss of spatial symmetry and improve classification accuracy. Five hundred and eighty eight red blood cell images are used to train and verify the effectiveness of the proposed method. The Channel Attention Residual Feature Pyramid Network (C-ARFPN) model achieves an mAP of 86%; the Channel and Spatial Attention Residual Feature Pyramid Network (CS-ARFPN) model achieves an mAP of 86.9%. The experimental results indicate that our method can classify more red blood cell types and better adapt to the needs of doctors, thus reducing the doctor's time and improving the diagnosis efficiency.

**Keywords:** attention mechanism, feature pyramid network, red blood cells, classification, microscopic image

## INTRODUCTION

As a connective tissue, blood has the following four forms, namely white blood cells (WBCs), red blood cells (RBCs), platelets, and plasma. Plasma can be regarded as an intercellular substance. The other three types of cells can be distinguished according to their shape, size, presence or absence of nucleus, color, and texture (1). RBCs are the majority component of blood cells, which transport oxygen to various parts of the human body and discharge the carbon dioxide produced by the human body (2, 3). The morphology of RBCs is non-nucleated, with biconvex and concave round pie-shaped cells. Its average diameter and thickness of this type of cell are about 7 and 2.5  $\mu\text{m}$ , respectively. RBCs are produced in the bone marrow, and the development of primitive RBCs into mature RBCs consists of four stages: basophilic normoblast, polychromatic normoblast, orthochromatic normoblast, and reticulocytes. After mature, RBCs enter the peripheral blood, as shown in **Figure 1**. The average life span of RBCs is about 120 days, and abnormal RBCs may live



**FIGURE 1 |** Red blood cell maturation process.

longer or shorter. Common RBC abnormalities include polycythemia, erythropenia, decreasing or increasing in size and hemoglobin, and changes in RBC morphology.

Diseases associated with RBCs include anemia, malaria, kidney tumors, malnutrition, and hemolytic disease, and anemia is the most common disease (4). These diseases cause many abnormal RBCs to appear in the peripheral blood. Mainly manifested as a change in the shape, size, and hemoglobin content of RBCs (5). Since abnormal RBCs may be a signal of certain diseases (6, 7), the detection and classification of RBCs are of great significance for the timely detection of diseases.

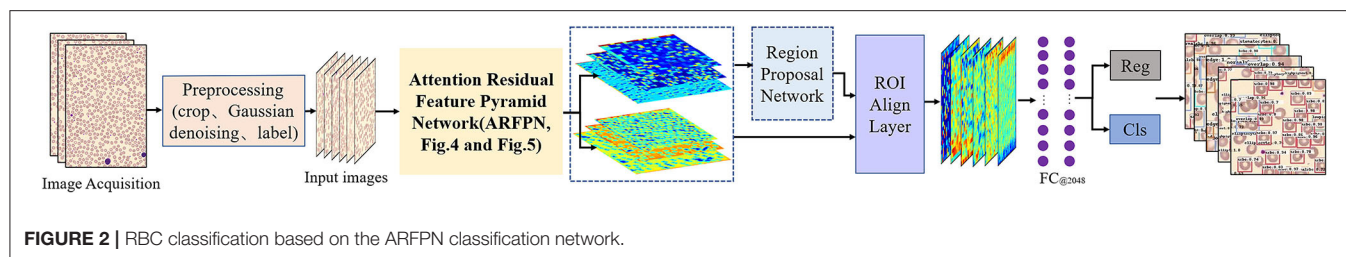
Clinically, doctors need to use a microscope to check whether there are abnormal RBCs or immature cells in the peripheral blood (8). In this case, there are usually hundreds of RBCs in the field of view, and a large number of images are obtained by microscopic image capturing equipment. This requires a lot of manpower. Meanwhile, the operation relies on the subjective judgment of the doctor, and different operators may produce different results (9), which will affect the accuracy of the test results.

In recent years, with the development of image processing technology, medical image analysis has become an indispensable tool in medical research, clinical disease diagnosis, and treatment (8). This technique has been used to analyze various types of medical images and extract more useful medical information from images to help clinical diagnosis. An automatic and effective cell classification method can be used to assist doctors in improving treatment plans and predicting treatment results. At present, the microscopic images generally have the following shortcomings: (1) The image capture process is affected by many factors such as light, color changes, blurring, etc.; (2) There may be interferences such as noise. In recent years, deep learning has developed into a research hotspot in medical image analysis. It can extract the hidden diagnosis features from medical images and solves the problems in medical image processing, such as object tracking (10), multi-label classification (11), pedestrian detection (12), and multi-class classification (13). Aiming at the challenges in RBC images, our study attempts to use the deep learning method to greatly improve the efficiency

of doctors and ensure the accuracy and objectivity of the detection results.

A lot of research works have been done on the detection and classification of RBCs. Yi et al. (14) proposed a method to analyze the equality of the covariance matrix in the Gabor filtered holographic image to automatically select the linear or non-linear classifier for RBC classification. This method used a single RBC image to classify three types of RBCs. Maji et al. (15) proposed to use mathematical morphology to automatically characterize RBCs. Mahmood et al. (16) used geometric features and Hough transform to detect the center of RBCs. Morphology was used to identify and extract RBCs from the background or other cells, Hough transform is used to identify the shape of RBCs. Besides, K-means clustering (17), boundary descriptors (18), and geometric features (19) were used to extract features. Sen et al. (20) used machine learning to divide RBCs into three categories. The method first divides RBCs into individual cells and then extracts features and classifications, which achieves an accuracy of 92%.

Lee et al. (21) proposed a hybrid neural network structure that combines parallel and cascading topologies for RBC classification. The authors used a single RBC image to extract shape features and clustering features. Then, the extracted features were input into a feedforward neural network with a three-layer structure for classification. Jambhekar et al. (22) studied the use of artificial neural networks to classify blood cells. The three-layer network achieves an accuracy of 81% for classifying sickle RBCs, WBCs, and overlapping cells. Elsalamony et al. (23) proposed to use a three-layer neural network to classify sickle cells and elliptical cells using the shape features of RBC. Xu et al. (24) used deep convolutional neural networks to classify eight types of RBCs, and the proposed method achieves an accuracy of 87.5%. Alzubaidi et al. (25) proposed a convolutional neural network using the ECOC model as a classifier. The method divides RBCs into normal cells, sickle cells, and other three categories, which achieves an accuracy of 88.11%. Kihm et al. (26) used a regression-based convolutional neural network to classify two types of RBCs ("slipper" and "croissant") in a flowing state. Parab et al. (27) used a convolutional neural network to extract



and classify individual RBCs after segmentation. They divided RBCs into nine categories and achieved an accuracy of 98.5%. Lin et al. (28) used FPN-ResNet-101 and Mask RCNN to classify two types of RBCs (hRBCs and tRBCs) in quantitative phase images, with an accuracy of 97%.

Most of the current works to segment red blood cell images into individual RBCs and then perform feature extraction and classification; very few works perform direct classification of the entire red blood cell image, and the number of RBCs in each image is small (about dozens). After fully understanding the needs of doctors and summarizing the methods in the research field, this paper proposed an Attention Residual Feature Pyramid Network (ARFPN). In this method, dense red blood cell images (each image contains about 230 red blood cells) are used for direct classification. Meanwhile, the feature pyramid network (29) is combined with spatial and channel attention mechanisms to focus on the multi-scale features related to categories, thus improving the expression of related features and suppressing background features. Besides, an anchor intensive strategy is adopted to better cover RBCs in the proposal stage. Moreover, the RoI align method is used to improve the extraction accuracy of RoI and locate the object more accurately. The contributions of this paper are summarized as follows: (i) the method can detect and classify 14 red blood cells; (ii) there is no single red blood cell segmentation, which simplifies the implementation steps and improves the efficiency; (iii) the method provides convenience for doctors, which can better adapt to the needs of doctors, and has better clinical applicability.

The rest of this paper is organized as follows. Section Materials and Methods introduces the used data set, data preprocessing methods, and the feature extraction and classification methods based on the channel and spatial attention feature pyramid network; section Results analyzes and introduced the experimental results; the results are discussed in section Discussion. Finally, conclusions are put forward in section Conclusions.

## MATERIALS AND METHODS

As shown in **Figure 2**, the workflow of our proposed method for RBC classification includes the image processing stage, feature extraction stage, post-processing stage, and cell classification stage. Each stage is described in detail in the following.

### Data Acquisition

The dataset was collected from the Department of Clinical Laboratory of Shandong Provincial Hospital, affiliated with Shandong First Medical University. The RBC images in the

dataset were collected by CellaVision DM96 (CellaVision AB, Lund, Sweden). The blood sample was put into a blood smear and then detected by the device to capture the image. The finished blood smear is shown in **Figure 3**.

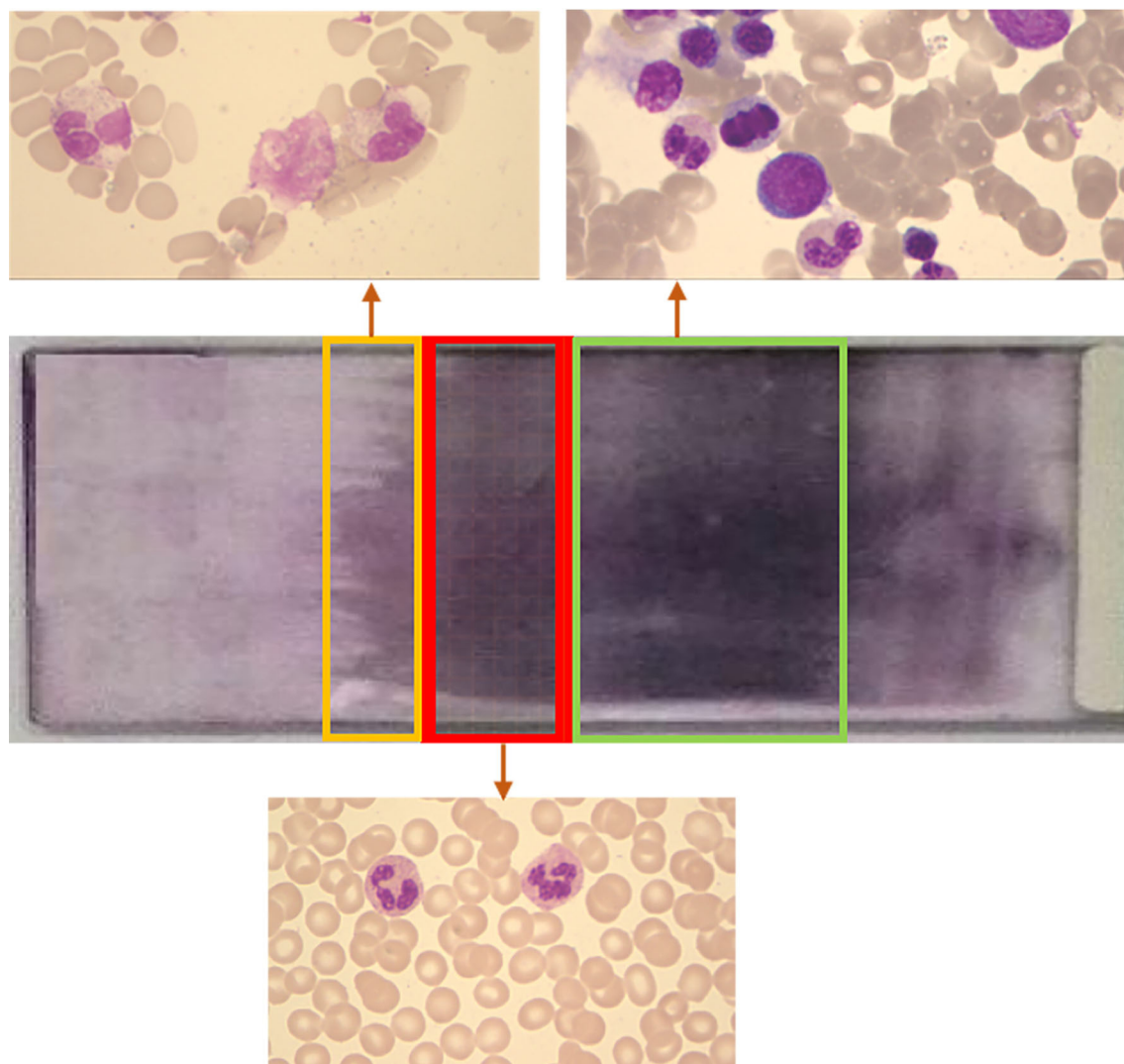
The resolution of each collected images is about  $1,418 \times 1,420$ . There are approximately 1,300 RBCs in each image (not including edge cells). All collected microscope images are in BMP format and contain RGB channels. Three types of cells are included in the images, i.e., RBC, white blood cells (WBC), and platelets. The obtained data set was verified by experienced doctors to avoid the interference of external factors, such as light.

### Pre-processing

The obtained images were preprocessed to make them more suitable for our study. First, the image in BMP format was converted to JPG format, and the noise was eliminated by a Gaussian filter. Then, the image was cropped according to the Pascal VOC dataset format. The size of the cropped image is  $375 \times 500$ , and the number of RBCs in the image is usually more than 200. The repeat parameter was set to 30% during cropping to expand the data. In the cropping process, the image containing many rare RBCs was horizontally flipped to expand the data and increase the sample size. Finally, LabelImg was adapted to label the RBCs in the image. Labeling and inspection were conducted by two experienced doctors. All the RBCs were divided into 14 categories (schistocyte, spherocyte, stomatocytes, target cells, hypochromic, elliptocytes, normal RBCs, overlapping RBCs, hyperchromic, microcyte, macrocyte, teardrop cells, basophilicstippling and the cells at the edge of the image). In RBC image, the resolution size of normal RBCs is  $21 \times 21$ , those with a resolution  $>24 \times 24$  are macrocyte, and those with a resolution  $<18 \times 18$  are microcyte. The schistocytes are broken red blood cells that resemble “fragments” in shape. Hyperchromic and hypochromic are related to the content of hemoglobin, and elliptocytes are shaped like ellipses. The target cell is shaped like a “shooting target,” and stomatocytes is shaped like a “mouth.” The corresponding quantity of each RBC category is listed in **Table 1**. The obtained dataset was used to evaluate our method and compare the results. After preprocessing, there are 588 images in total, each of which is a  $350 \times 500 \times 3$  RGB image. Four hundred and seventy images were used as the training set, and the remaining 118 images were used as the test set. As shown in **Table 2**.

### Feature Extraction of Shape, Size, and Hemoglobin Content

The size of normal RBCs is about  $7 \sim 8 \mu\text{m}$ , which is reflected in the image with a resolution of  $21 \times 21$ . The size of abnormal



**FIGURE 3 |** Schematic diagram of blood smear. A large number of red blood cells overlap in the green box area. The number of red blood cells in the yellow frame area is small. The number of cells in the red frame area is appropriate and evenly distributed, which is suitable for observation.

**TABLE 1 |** Various types of RBCs.

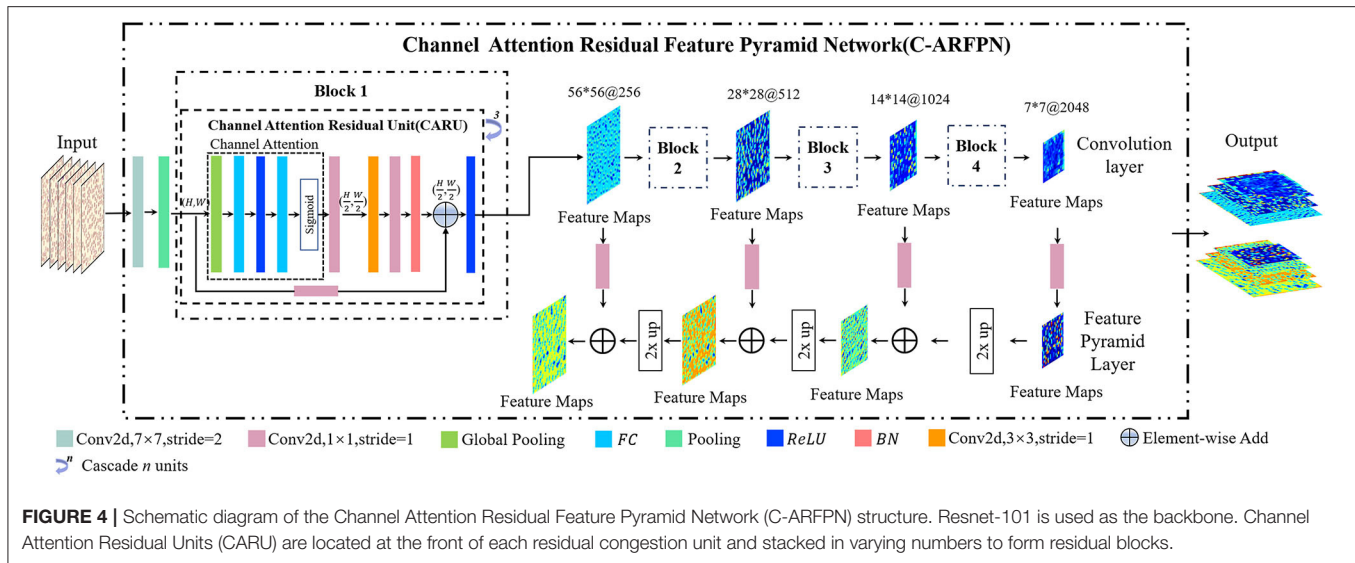
Name	Number	Name	Number
Schistocyte	772	Overlap	6,064
Spherocyte	4,021	Hyperchromic	17,693
Stomatocytes	615	Microcyte	17,060
Target cells	537	Macrocyte	9,084
Hypochromic	5,414	Teardrop cells	1,287
Elliptocytes	15,439	Basophilicstippling	451
Normal RBC	18,126	Edge cells	20,661

RBCs in the image varies widely, and each has its specific shape. The characteristics of the RBC images can be summarized as: (a) Large changes in cell size; (b) RBCs are small objects; (c) Cells are densely distributed; (d) The contrast between the RBC and

**TABLE 2 |** Allocation of training set and test set (number of cells and number of images).

	Cell number	Image number
Train	9,3885	470
Test	2,3366	118

the background is low. In deep learning object detection, the detection of small objects has always been a difficult problem due to low resolution, blurry pictures, less information, and weak feature expression. This study used feature pyramid network (FPN) to overcome the above problems because it can better deal with the multi-scale changes in object detection. The FPN makes reasonable use of the features of each layer in the convolutional network and merges the features of different layers. Specifically,



it constructs a top-down, horizontally connected hierarchical structure that combines low-resolution and strong semantic features with high-resolution and weak semantic features.

In recent years, attention network models have achieved good performance in classification tasks. In this research, channel attention mechanism (30) and spatial attention mechanism (31) were integrated into the feature extraction network to achieve accurate classification of RBCs. In the feature extraction stage, the attention mechanism (32) can highlight the features related to categories while focusing on the key features of red blood cells and generating more discriminative feature representations. The integration of these two mechanisms contributes to a great performance improvement when the number of growth parameters is small.

### Channel Attention Residual Feature Pyramid Network

The structure of Channel Attention Residual Feature Pyramid Network (C-ARFPN) is shown in **Figure 4**. ResNet-101 and ResNet-50 are used as the backbone of our network. Each bottleneck of the residual network is replaced with channel attention residual units (CARUs) that are located behind the residual unit. CARU first averages and pools the input features, so that the features can respond to the global distribution, thus expanding the global receptive field and reducing the calculation amount. The following two full connection layers map the channel feature representation to the sample label space, and the output represents the weight of each feature channel. The feature channel is weighted with the input feature to recalibration the input feature on the channel dimension.

### Channel and Spatial Attention Residual Feature Pyramid Network

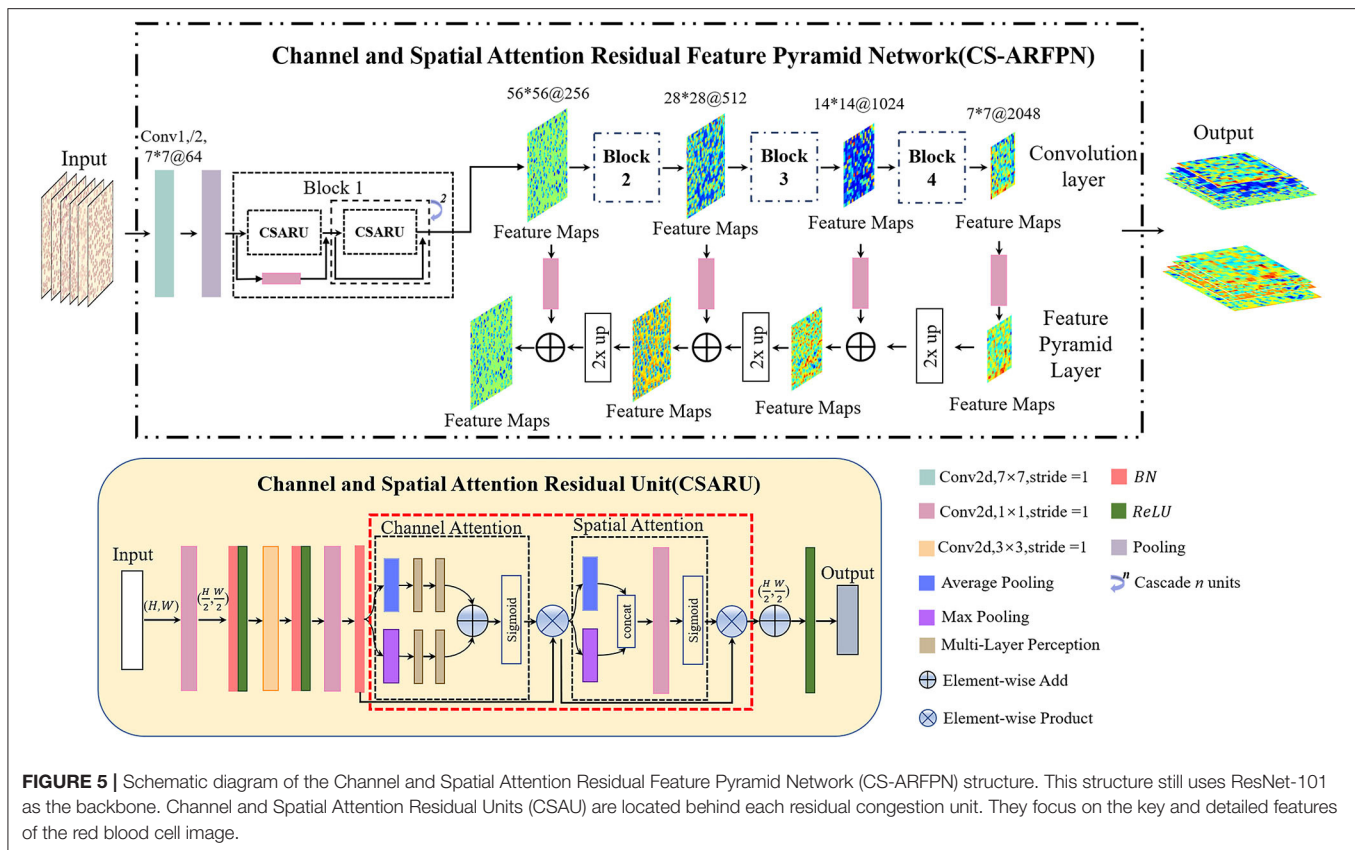
The structure of Channel and Spatial Attention Residuals Feature Pyramid Net-work (CS-ARFPN) is shown in **Figure 5**. The upper part of the figure illustrates the overall flow chart of feature extraction. Similarly, ResNet-50 and ResNet-101 are used as the

backbone. The lower part of the figure shows the structure of the Channel and Spatial Attention Residual Unit (CSARU), which is nested in each residual unit of the residual network and located behind the three convolution cores. The input feature is first compressed in the spatial dimension, and average pooling and maximum pooling are used to aggregate the spatial information of the feature map. The obtained feature space information is sent to the multi-layer perceptron for element-by-element summation, and the compressed space information is multiplied by the original feature points to obtain the channel attention feature. Then, the channel attention feature is input to the spatial attention unit, and average pooling and max pooling are used to compress channels and extract the maximum value. After dimensionality reduction through convolution operation, the attention feature is obtained by dot product with the original channel attention feature.

Both C-ARFPN and CS-ARFPN use ResNet-101 and ResNet-50 as the backbone. Each attention module is distributed in a residual unit according to its position. Since RBCs are small objects, the feature pyramid network combined with the attention mechanism can merge deep and shallow features and focus on category-related features. This improvement makes the characteristics of RBCs more accurate and richer, thus improving the model's detection and classification ability of small objects, and improving the performance of the model.

### Post-processing and Classification

After feature extraction, the features are input to the subsequent network for post-processing and classification. First, the feature map is input into the RPN to filter out the anchors containing the foreground. Then, the high-quality object candidate box is selected and input into the ROI pooling layer. In the pooling operation, RoI align (33) instead of RoI pooling operation is used. Compared with RoI pooling, RoI align removes the quantization rounding operation, so it can overcome the bounding box offset problem (34) and extract more accurate RoI. After the RoI



align operation is performed on the feature map, the candidate recognition regions of different sizes are normalized into a fixed-size object recognition region.

The features after RPN and RoI pooling are sent to the subsequent network for classification and regression. In this process, 14 types of RBCs are classified including schistocyte, spherocyte, stomatocytes, target cells, hypochromic, elliptocytes, normal RBCs, overlapping RBCs, hyperchromic, microcyte, macrocyte, teardrop cells, basophilic stippling, and the cells at the edge of the image. The schematic diagram is shown in **Figure 6**. Normal RBCs and macrocytes are displayed in one image to make their difference obvious. It can be seen from the figure that each cell has its characteristics. During the training process, the weight of the network is adjusted according to the input data to minimize the error between the input and the target. Then Fast RCNN (35) is used to perform cell classification, and the output of the classification prediction is converted into a probability distribution through softmax.

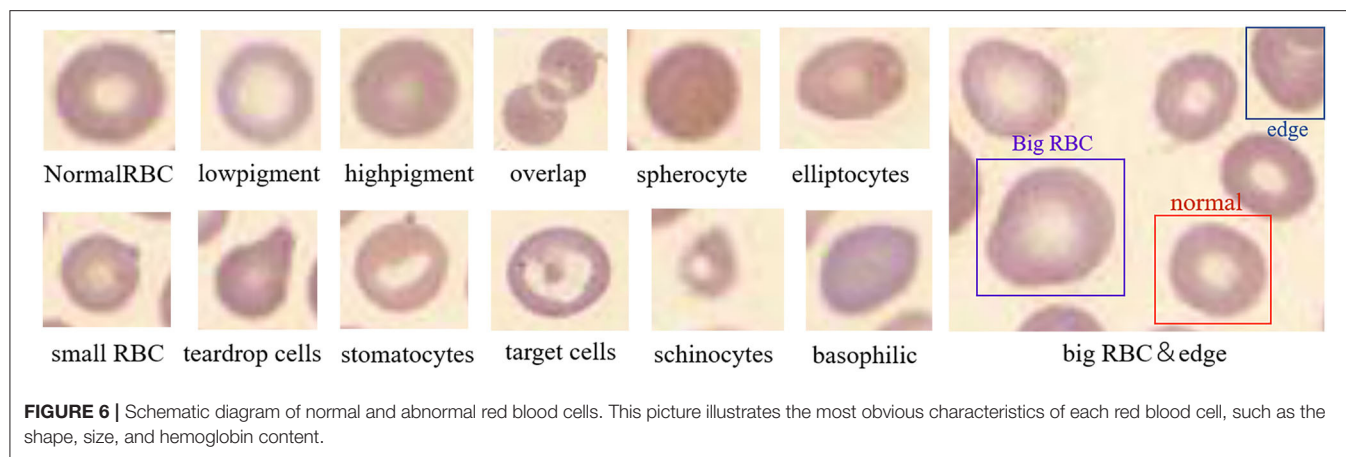
## Ablation Study

Our proposed method uses the attention module to learn the features related to categories. First, the effectiveness of the attention module was verified, and the performance of two different attention modules was compared. Meanwhile, the performance of RoI align and RoI pooling methods was compared. Besides, the impact of Adam optimizer, momentum optimizer, and various training parameters on the model

performance was investigated. All comparisons and analyses were performed under the same parameter settings. Moreover, the effectiveness of the proposed model on different was verified public datasets.

## Training Implementation

Our proposed method was implemented on a computer equipped with Intel® Core™ i7-8700k CPU@3.70GHz with 32GB memory, and the computationally intensive calculations were offloaded to an Nvidia Tesla P100 GPU with 16 GB HBM2 memory and 3,584 computer unified device architecture (CUDA) cores. To visually present the obtained model parameters, all experiments were conducted using Python programming language under the TensorFlow framework (36). In the training process, momentum and Adam optimizer were used in the parameter configuration to minimize the loss. The batch size was set to 1, and the number of iterations was set to 110,000. It takes 35 h to complete the optimization. In the early stage of training, a large learning rate was used to make the model easy to obtain the optimal solution; in the later stage of training, a small learning rate was used to ensure that the model will not fluctuate too much. The learning rate was divided by 10 after 60,000 and 80,000 iterations, and the minimum learning rate was set to  $10^{-6}$ . Besides, the momentum of the model was set to 0.9, and the weight decay was set to  $10^{-4}$ . During training, random initialization was used to initialize the weights, and used the cross-entropy loss function was adopted to evaluate the error



between the predicted value and the true value of our model. The calculation formula of the cross-entropy loss function is shown in Equation (1).

$$C = \sum_{i=1}^k y_i \log(p_i) \quad (1)$$

where  $k$  represents the number of classes;  $y_i$  represents the label of category  $i$ ;  $p_i$  represents the output probability of class  $i$ , and this value was calculated by Softmax.

## Evaluation Metrics

In our experiments, the metrics of precision, recall, and F1-score were taken to evaluate the performance of our proposed method. The calculation formulas of the evaluation metrics are expressed in Equations (2–4).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 \text{ score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Where, TP (True Positive) indicates the number of positive samples that are also judged by the model as positive; TN (True Negative) indicates the number of negative samples that are also judged by the model as negative; FN (False Negative) indicates number of positive samples that are judged by the model as negative. FP (False Positive) indicates the number of negative samples that are judged by the model as positive. Based on this, precision is the ratio of the number of correctly predicted positive examples to the number of samples predicted as positive; recall is the ratio of correctly predicted positive examples to the number of real positive samples. F1 score is the harmonic average of precision and recall, so it can comprehensively reflect the performance. In general, the higher the F1 score, the better the performance of the model.

**TABLE 3 |** The evaluation metrics of the model with/without the attention module. The best results are shown for each model.

Models	Recall	Precision	F1	mAP
FPN	0.756	0.759	0.759	0.798
Our proposed (S-ARFPN)	0.8	0.838	0.816	0.86
Our proposed (CS-ARFPN)	0.811	0.831	0.819	0.869

## RESULTS

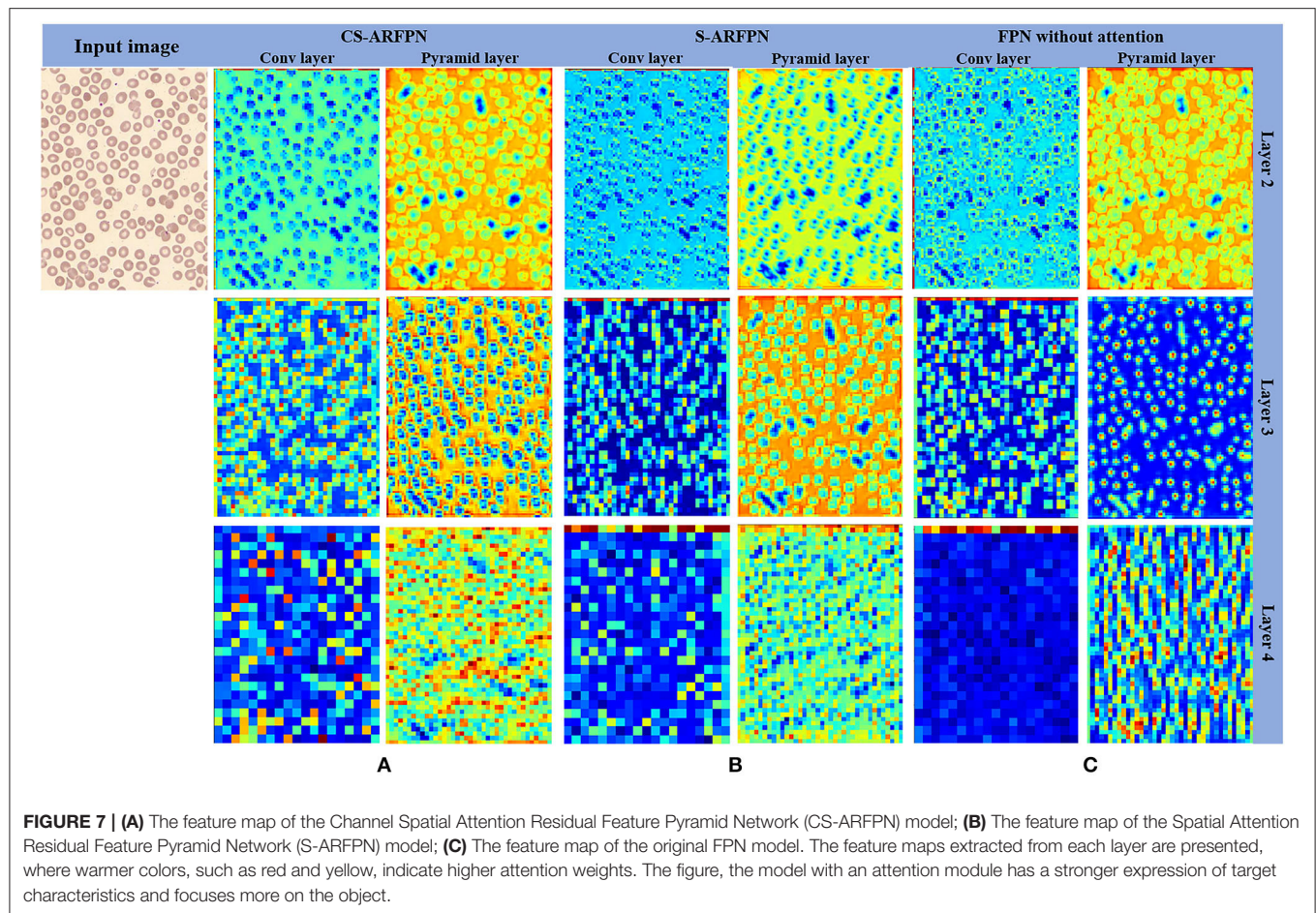
### Ablation Study: Comparison of FPN With or Without Attention Module, ROI Pooling or ROI Align, and Others

In the ablation study, ResNet-101 and ResNet-50 were used as the backbone in the training, and different learning rates were set. Since the RBC object is small and densely distributed, a small anchor size was used.

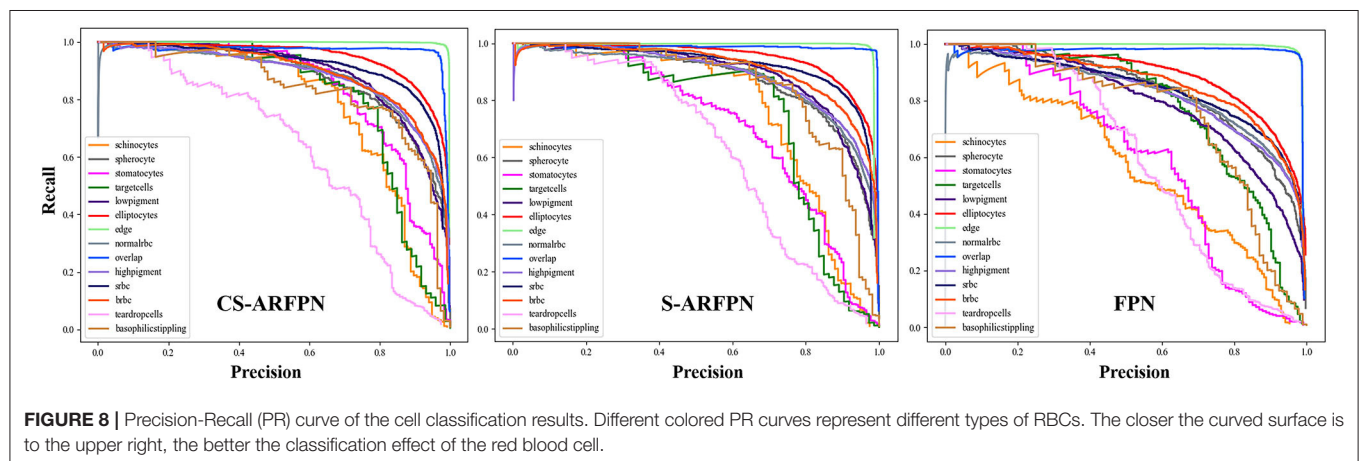
In **Table 3**, FPN is the original model without attention module; C-ARFPN is the feature pyramid network with channel attention residual unit; CS-ARFPN is the feature pyramid network with channel and spatial attention residual unit. It can be seen that the CS-ARFPN model achieves better performance. Compared with FPN, the accuracy, recall, F1 score, and mAP of CS-ARFPN and S-ARFPN are improved by {4.4, 7.9, 5.7, 6.1%} and {5.5, 7.4, 6.0, 7.2%}, respectively. Compared with the S-ARFPN model, the accuracy, recall, F1 score, and mAP of the CS-ARFPN model is improved by 1.1, −0.7, 0.3, and 0.9%.

**Figure 7** presents the feature map of the two models with different attention residual units and the original FPN model. The leftmost column shows the input image, and the next six columns show the feature maps of the three models including CS-ARFPN, S-ARFPN, and FPN (each model contains two columns of feature maps). They are the feature maps extracted by the convolutional layer {C2, C3, C4} and the pyramid layer {P2, P3, P4}. It can be seen that the feature maps extracted by the CS-ARFPN model pay more attention to the object to be recognized, so the model achieves better performance.

**Figure 8** shows the precision-recall curve (PR) of the three models of FPN, C-ARFPN, and CS-ARFPN. The closer the curve



**FIGURE 7 | (A)** The feature map of the Channel Spatial Attention Residual Feature Pyramid Network (CS-ARFPN) model; **(B)** The feature map of the Spatial Attention Residual Feature Pyramid Network (S-ARFPN) model; **(C)** The feature map of the original FPN model. The feature maps extracted from each layer are presented, where warmer colors, such as red and yellow, indicate higher attention weights. The figure, the model with an attention module has a stronger expression of target characteristics and focuses more on the object.



**FIGURE 8 |** Precision-Recall (PR) curve of the cell classification results. Different colored PR curves represent different types of RBCs. The closer the curved surface is to the upper right, the better the classification effect of the red blood cell.

to the upper right, the larger the area under the line and the better the performance of the model. The PR area under the curve (PR-AUC) of the three models is 0.798, 0.86, 0.869, respectively. Thus, the CS-ARFPN model achieves the best performance, followed by C-ARFPN.

Table 4 lists the recall, precision, F1-score, and AP of the CS-ARFPN model for classifying the 14 types of cells.

Table 5 shows the AP of the three models with two different optimization strategies. In the FPN, C-ARFPN, and CS-ARFPN models, the Momentum optimizer leads to 1.9, 18.7, and 9.7% higher performance than the Adam optimizer, respectively.

The performance of using two different RoI processing methods, i.e., Roi pooling and Roi align is shown in Table 6. The

**TABLE 4 |** Results of different red blood cells by the CS-ARFPN model.

Class	Recall	Precision	F1	AP
Schistocytes	0.681	0.841	0.753	0.787
Spherocyte	0.822	0.782	0.801	0.879
Stomatocytes	0.697	0.853	0.768	0.840
Target cells	0.744	0.821	0.786	0.815
Hypochromic	0.772	0.855	0.811	0.890
Elliptocytes	0.890	0.859	0.874	0.946
Edge	0.987	0.987	0.987	0.997
Normal RBC	0.804	0.810	0.807	0.880
Overlap	0.966	0.964	0.965	0.973
Hyperchromic	0.838	0.791	0.814	0.883
Microcyte	0.876	0.842	0.859	0.913
Macrocyte	0.857	0.784	0.819	0.893
Teardrop cells	0.570	0.686	0.622	0.631
Basophilic	0.837	0.761	0.797	0.843

**TABLE 5 |** Performance comparison of models using Adam and Momentum.

	FPN	S-ARFPN	CS-ARFPN
Adam	0.779	0.673	0.774
Momentum	0.798	0.860	0.869

**TABLE 6 |** Comparison of the precision of ROI pooling and ROI align methods.

	FPN	S-ARFPN	CS-ARFPN
ROI Pooling	0.786	0.823	0.858
ROI Align	0.798	0.860	0.869

model using the RoI align method achieves better performance than that using the RoI pooling method.

## Comparison With the State-of-the-Art Models and Comparison of Results Obtained on Other Data Sets

Our proposed method was compared with five classification methods based on deep learning, including Faster RCNN (32), RetinaNet (37), Cascade RCNN (38), R-FCN (39), and Cascade-FPN (40). All the models used were trained from scratch on the RBC dataset.

ResNet-50 and ResNet-101 were used as the backbone for model training; then, different parameters were set to finetune the models; finally, showed the best results for each model were obtained. **Table 7** lists the classification performance of different models. The highest performance values are bolded in the table. The mAP of the two proposed models is 0.86 and 0.869, respectively, and the two models achieve the best performance among all the models.

Besides, to verify the effectiveness of the proposed model, the performance of our proposed model on different datasets was compared, and the comparison results are listed in **Table 8**.

Among them, in the IDB data set, the accuracy of circular and elongated red blood cells are 99 and 94.4%, respectively. In the BCCD data set, the accuracy of WBC and Platelets are 97.43 and 92.89%, respectively. The proposed model achieves an mAP of 91% and 98.8% on the BCCD dataset and IDB dataset is 91.23%, respectively.

## DISCUSSION

To better assist doctors in diagnosing the diseases related to RBCs, this paper proposed an attention feature pyramid network model that can directly classify dense red blood cell images. Since RBCs are small objects, this paper combined the attention mechanism with the feature pyramid network to improve the detection of small objects. The experimental results show that the two proposed attention residual units can capture more key feature information of RBCs, which helps to classify RBCs more accurately.

In the training process, different backbones, learning strategies, and anchor settings were used, and the optimal parameter setting of the two models was obtained after a lot of training. The results show that different learning rates, anchor sizes, backbones, and attention modules led to performance differences. When ResNet-50 was used as the backbone, the CS-ARFPN model achieved the best performance under the learning rate of 0.001 and the anchor size of 32. When ResNet-50 was used as the backbone, the S-ARFPN model achieved the best performance under the learning rate of 0.002 and the anchor size of 4. The subsequent experiment and analysis were conducted based on the model and the above-mentioned optimal parameters.

In the experiment, the performance of FPN, C-ARFPN and CS-ARFPN with two different attention residual unit models was compared to verify the effectiveness of the attention mechanism. The effectiveness of our method was proved through evaluation metrics, feature maps, and PR curves.

It can be seen from **Table 3** that, compared with FPN, both CS-ARFPN and C-ARFPN achieved improved performance, which shows the effectiveness of the attention mechanism. Meanwhile, the CS-ARFPN model performed better than the C-ARFPN model, indicating that the CS-ARFPN model pays more attention to features, channel feature information and spatial feature information. To make the function of the attention module more intuitive, the feature maps of FPN, C-ARFPN, and CS-ARFPN are shown **Figure 7**. It can be seen that as the number of layers increases, the extracted features become more and more abstract and more difficult to understand. Compared with FPN, C-ARFPN and CS-ARFPN weaken the background characteristics and highlight the relevant components of the category, so the two models can more accurately capture the shape and size of RBCs. Besides, CS-ARFPN can better focus on the detailed features than C-ARFPN.

**Figure 8** shows the PR curves of the three models. The curves in the figure show that for each type of RBC, different attention modules led to different classification performances. In general, the PR curve obtained by the CS-ARFPN model is located on the

**TABLE 7** | Comparison of our proposed method with other advanced methods.

Backbone	Models	Recall	Precision	F1	mAP	Param (MB)
ResNet-50	Cascade RCNN (33)	0.361	0.687	0.463	0.385	254.47
	Faster RCNN (32)	0.398	0.652	0.481	0.394	54.07
	R-FCN (34)	0.530	0.757	0.638	0.551	95
	RetinaNet (31)	0.695	0.751	0.736	0.684	61.92
	Cascade-FPN (35)	0.759	0.757	0.736	0.736	98.14
	FPN	0.753	0.758	0.759	0.796	79.05
ResNet-101	Cascade RCNN (33)	0.429	0.699	0.525	0.416	290.69
	Faster RCNN (28)	0.447	0.687	0.526	0.434	63.70
	R-FCN (34)	0.528	0.880	0.620	0.548	95.66
	RetinaNet (31)	0.686	0.738	0.709	0.683	133.58
	FPN	0.756	0.759	0.759	0.798	115.27
ResNet-50	Our proposed (S-ARFPN)	0.754	0.758	0.753	0.792	88.67
	Our proposed (CS-ARFPN)	<b>0.811</b>	<b>0.831</b>	<b>0.819</b>	<b>0.869</b>	88.68
ResNet-101	Our proposed (S-ARFPN)	<b>0.800</b>	<b>0.838</b>	<b>0.816</b>	<b>0.860</b>	133.44
	Our proposed (CS-ARFPN)	0.791	0.789	0.788	0.833	133.43

Each model was trained under different parameter settings, and the best results of each model are shown. The classification effect of Cascade-FPN is not ideal when the backbone is Resnet101, so it is not displayed in the table. The best results are highlighted in bold.

**TABLE 8** | Comparison of the accuracy of the proposed method on different datasets.

Dataset	Images	Class	AP (%)
BCCD (41)	364	WBC	97.43
		RBC	83.05
		Platelets	92.89
IDB (42)	626	circular	99
		elongated	94.4
		other	80.3

upper right, which indicates that the area under the line is the largest, and the performance is the best.

The performance of the Adam optimizer and momentum optimizer was compared and analyzed. The results in **Table 5** show that the Momentum optimizer performs better for RBCs classification. During the training process, the Momentum optimizer had a slower convergence speed than the Adam optimizer, but it obtained better results and the generalization performance in our work.

As shown in **Table 6**, the RoI align method achieved higher AP than the RoI pooling method. The rounding operation in the RoI pooling method has little impact on the classification of large objects, but it will have a huge impact on the classification of small objects such as RBCs. The RoI align method removes the rounding operation, so it can accurately extract RoI and achieve better performance.

**Table 4** shows the performance metrics of classifying the 14 types of RBCs obtained by the CS-ARFPN model. Among them, the classification results of teardrop cells and schistocytes are not as accurate as other types. Although the attention mechanism focuses on the features related to categories,

inaccurate classifications are caused by certain features. This is because different types of feature extraction have different difficulties, and certain red blood cell types have a specific definition standard. In this case, the model fails to learn the abnormal RBC, thus resulting in misclassifications. Secondly, some types of red blood cell samples are small. Although the RBCs were expanded during the cutting process, the sample imbalance problem still existed in the study, which makes the model fail to learn the characteristics of red blood cells with few samples.

The comparison between our method and other advanced methods is shown in **Table 7**. Our method achieves better performance than other models. Meanwhile, our method was compared with other red blood cell classification methods, including Kihm et al. (26), Parab et al. (27), Lin et al. (28), and others. These methods all use a single red blood cell image for feature extraction and achieve good accuracy. The classification of the entire red blood cell image can be regarded as the classification of dense small objects with weak feature expression and diverse target changes, so feature extraction is more difficult. Due to this, our method obtains a slightly lower accuracy than the comparison methods. To better compare with other methods and verify the effectiveness and generalization of the proposed method, our method was evaluated on two public data sets, i.e., namely the BCCD dataset and the IDB dataset. As shown in **Table 8**, the classification results of WBCs and Platelets in the BCCD dataset in **Table 8** are better. The reason for the low accuracy of RBC classification is that the dataset is mainly provided for WBC classification, and most of the RBCs in the image are overlapping cells. In the IDB dataset, the classification results of circular and elongated are good. The reason for the low accuracy of the RBCs of the other category is that the category contains many small and medium categories, which poses a challenge to the

classification. The results indicate that our method is effective and generalizable, and the classification of the entire image can be further improved.

At present, due to the limited dataset and the imbalance of different types of RBC samples, it is difficult to improve the classification performance. After communicating with the doctor, we will collect RBC images under a microscope so that clear images with obvious RBC characteristics can be obtained. In future work, we will collect more data, especially the rare type of RBC. Meanwhile, we will investigate the use of a fully connected layer and loss function of the model to reduce the impact of sample imbalance to further improve the classification performance.

## CONCLUSIONS

Abnormal red blood cells can cause changes in shape, size, and amount of hemoglobin, which are closely related to the diagnosis of many diseases. This paper proposed a classification method that can directly classify 14 types of red blood cells on the entire red blood cell image. The feature pyramid network extracts the multi-scale features of RBCs, and the attention mechanism is used to improve the learning and representation of RBC features. Besides, the ROI alignment layer with good performance is used to unify the size of the candidate area. This method proposed in this study can achieve accurate red blood cell classification, which provides a clinically feasible, universal and convenient method for the diagnosis of red blood cell diseases.

## REFERENCES

- Chadha GK, Srivastava A, Singh A, Gupta R, Singla D. An automated method for counting red blood cells using image processing. *Proc Comput Sci.* (2020) 167:769–78. doi: 10.1016/j.procs.2020.03.408
- Yang M, Bourbakis N, editors. An overview of lossless digital image compression techniques. In: *48th Midwest Symposium on Circuits and Systems*, 2005. Covington, KY: IEEE (2005).
- Mazalan SM, Mahmood NH, Razak MAA, editors. Automated red blood cells counting in peripheral blood smear image using circular Hough transform. In: *2013 1st International Conference on Artificial Intelligence, Modelling and Simulation*. Kota Kinabalu: IEEE (2013). doi: 10.1109/AIMS.2013.59
- Jansen V. Diagnosis of anemia—a synoptic overview and practical approach. *Transfus Apher Sci.* (2019) 58:375–85. doi: 10.1016/j.transci.2019.06.012
- George LE. Comparative study using weka for red blood cells classification. *World Acad Sci Eng Technol Int J Med Health Pharm Biomed Eng.* (2015) 9:19–23. Available online at: [https://www.researchgate.net/publication/281208619\\_Comparative\\_Study\\_Using\\_Weka\\_for\\_Red\\_Blood\\_Cells\\_Classification](https://www.researchgate.net/publication/281208619_Comparative_Study_Using_Weka_for_Red_Blood_Cells_Classification)
- Chy TS, Rahaman MA, editors. Automatic sickle cell anemia detection using image processing technique. In: *2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*. Gazipur: IEEE (2018). doi: 10.1109/ICAEEE.2018.8642984
- Sun Q, Yang S, Sun C, Yang W, editors. An automatic method for red blood cells detection in urine sediment micrograph. In: *2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. Nanjing: IEEE (2018). doi: 10.1109/YAC.2018.8406379
- Di Ruberto C, Loddo A, Putzu L. Detection of red and white blood cells from microscopic blood images using a region proposal approach. *Comput Biol Med.* (2020) 116:103530. doi: 10.1016/j.compbiomed.2019.103530
- Venkatalakshmi B, Thilagavathi K, editors. Automatic red blood cell counting using hough transform. In: *2013 IEEE Conference*

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

WS and PH: conceptualization and methodology. JW, YS, and JZ: data production and curation. ZL: data curation, resources, and supervision. WS: validation, writing—original draft preparation, investigation, and visualization. DLI: supervision, project administration, and funding acquisition. DLIU: writing—review and editing and formal analysis. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was funded by the National Natural Science Foundation of China (61971271), the Taishan Scholars Project of Shandong Province (Tsqn20161023), and the Primary Research and Development Plan of Shandong Province (No. 2018GGX101018, No. 2019QYTPY02).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2021.741407/full#supplementary-material>

- on Information & Communication Technologies*. Thuckalay: IEEE (2013). doi: 10.1109/CICT.2013.6558103
- Araújo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, et al. Classification of breast cancer histology images using convolutional neural networks. *PLoS ONE.* (2017) 12:e0177544. doi: 10.1371/journal.pone.0177544
- Hamad A, Ersoy I, Bunyak F, editors. Improving nuclei classification performance in H&E stained tissue images using fully convolutional regression network and convolutional neural network. In: *2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. Washington, DC: IEEE (2018). doi: 10.1109/AIPR.2018.8707397
- Albehadili H, Alzubaidi L, Rashed J, Al-Imam M, Alwzawy HA, editors. Fast and accurate real time pedestrian detection using convolutional neural network. In: *The 1st International Conference on Information Technology (ICoIT'17)*. Irbil (2017). doi: 10.25212/ICoIT17.029
- Zeiler MD, Fergus R, editors. Visualizing and understanding convolutional networks. In: *European Conference on Computer Vision*. Cham: Springer (2014). doi: 10.1007/978-3-319-10590-1\_53
- Yi F, Moon I, Javidi B. Cell morphology-based classification of red blood cells using holographic imaging informatics. *Biomed Opt Expr.* (2016) 7:2385–99. doi: 10.1364/BOE.7.002385
- Maji P, Mandal A, Ganguly M, Saha S, editors. An automated method for counting and characterizing red blood cells using mathematical morphology. In: *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)*. Kolkata: IEEE (2015). doi: 10.1109/ICAPR.2015.7050674
- Mahmood NH, Mansor MA. Red blood cells estimation using hough transform technique. *Signal Image Process.* (2012) 3:53. doi: 10.5121/sipij.2012.3204
- Savkare S, Narote S, editors. Blood cell segmentation from microscopic blood images. In: *2015 International Conference on Information Processing (ICIP)*. Pune: IEEE (2015). doi: 10.1109/INFOP.2015.7489435

18. Lotfi M, Nazari B, Sadri S, Sichani NK, editors. The detection of dacrocyte, schistocyte and elliptocyte cells in iron deficiency anemia. In: *2015 2nd International Conference on Pattern Recognition and Image Analysis (IPRIA)*. Rasht: IEEE (2015). doi: 10.1109/IPRIA.2015.7161628
19. Dalvi PT, Vernekar N, editors. Computer aided detection of abnormal red blood cells. In: *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*. Bangalore: IEEE (2016). doi: 10.1109/RTEICT.2016.7808132
20. Sen B, Ganesh A, Bhan A, Dixit S, Goyal A. Machine learning based diagnosis and classification of sickle cell anemia in human RBC[C]. In: *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*. Tirunelveli: IEEE (2021). doi: 10.1109/ICICV50876.2021.9388610
21. Lee H, Chen Y-PP. Cell morphology based classification for red cells in blood smear images. *Pattern Recognit Lett.* (2014) 49:155–61. doi: 10.1016/j.patrec.2014.06.010
22. Jambhekar ND. Red blood cells classification using image processing. *Sci Res Rep.* (2011) 1:151–4. Available online at: [https://www.researchgate.net/publication/285841327\\_Red\\_blood\\_cells\\_classification\\_using\\_image\\_processing](https://www.researchgate.net/publication/285841327_Red_blood_cells_classification_using_image_processing)
23. Elsalamony HA. Healthy and unhealthy red blood cell detection in human blood smears using neural networks. *Micron.* (2016) 83:32–41. doi: 10.1016/j.micron.2016.01.008
24. Xu M, Papageorgiou DP, Abidi SZ, Dao M, Zhao H, Karniadakis GE. A deep convolutional neural network for classification of red blood cells in sickle cell anemia. *PLoS Comput Biol.* (2017) 13:e1005746. doi: 10.1371/journal.pcbi.1005746
25. Alzubaidi L, Al-Shamma O, Fadhel MA, Farhan L, Zhang J, editors. Classification of red blood cells in sickle cell anemia using deep convolutional neural network. In: *International Conference on Intelligent Systems Design and Applications*. Cham: Springer (2018). doi: 10.1007/978-3-030-16657-1\_51
26. Kihm A, Kaestner L, Wagner C, Quint S. Classification of red blood cell shapes in flow using outlier tolerant machine learning. *PLoS Comput Biol.* (2018) 14:e1006278. doi: 10.1371/journal.pcbi.1006278
27. Parab MA, Mehendale ND. Red blood cell classification using image processing CNN. *SN Comput Sci.* (2021) 2:70. doi: 10.1007/s42979-021-00458-2
28. Lin YH, Liao YK, Sung KB. Automatic detection and characterization of quantitative phase images of thalassemic red blood cells using a mask region-based convolutional neural network. *J Biomed Optics.* (2020) 25:116502. doi: 10.1117/1.JBO.25.11.116502
29. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S, editors. Feature pyramid networks for object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI: IEEE (2017). doi: 10.1109/CVPR.2017.106
30. Hu J, Shen L, Sun G, editors. Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT: IEEE (2018). doi: 10.1109/CVPR.2018.00745
31. Woo S, Park J, Lee J-Y, Kweon IS, editors. CBAM: Convolutional block attention module. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Cham: Springer (2018). doi: 10.1007/978-3-030-01234-2\_1
32. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *arXiv.* (2017) arXiv:1706.03762. Available online at: <https://arxiv.org/abs/1706.03762>
33. He K, Gkioxari G, Dollár P, Girshick R, editors. Mask R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision*. Venice: IEEE (2017). doi: 10.1109/ICCV.2017.322
34. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *arXiv.* (2015) arXiv:1506.01497. doi: 10.1109/TPAMI.2016.2577031
35. Girshick R, editor. Fast R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision*. Santiago: IEEE (2015). doi: 10.1109/ICCV.2015.169
36. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. editors. Tensorflow: a system for large-scale machine learning. In: *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*. Savannah, GA (2016).
37. Lin T-Y, Goyal P, Girshick R, He K, Dollár P, editors. Focal loss for dense object detection. In: *Proceedings of the IEEE International Conference on Computer Vision*. Venice: IEEE (2017). doi: 10.1109/ICCV.2017.324
38. Cai Z, Vasconcelos N, editors. Cascade R-CNN: delving into high quality object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT: IEEE (2018). doi: 10.1109/CVPR.2018.00644
39. Dai J, Li Y, He K, Sun J. R-fcn: object detection via region-based fully convolutional networks. *arXiv.* (2016) arXiv:1605.06409. Available online at: <https://arxiv.org/abs/1605.06409>
40. Wu Y, Chen Y, Yuan L, Liu Z, Wang L, Li H, et al., editors. Rethinking classification and localization for object detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA: IEEE (2020). doi: 10.1109/CVPR42600.2020.01020
41. Singh I, Pal Singh N, Singh H, Bawankar S, Ngom A. Blood cell types classification using CNN. In: *International Work-Conference on Bioinformatics and Biomedical Engineering*. Cham: Springer. (2020). p. 727–38. doi: 10.1007/978-3-030-45385-5\_65
42. Ferreira RL, Coelho Naldi M, Fernando Mari J. Morphological analysis and classification of erythrocytes in microscopy images. In: *Proceedings of the 2016 Workshop de Visão Computacional*. Campo Grande. (2016). p. 9–11.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Song, Huang, Wang, Shen, Zhang, Lu, Li and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Robust Training Method for Pathological Cellular Detector *via* Spatial Loss Calibration

Hansheng Li<sup>1†</sup>, Yuxin Kang<sup>1†</sup>, Wentao Yang<sup>2†</sup>, Zhuoyue Wu<sup>1</sup>, Xiaoshuang Shi<sup>3</sup>, Feihong Liu<sup>1</sup>, Jianye Liu<sup>1</sup>, Lingyu Hu<sup>4</sup>, Qian Ma<sup>4</sup>, Lei Cui<sup>1\*</sup>, Jun Feng<sup>1\*</sup> and Lin Yang<sup>1\*</sup>

<sup>1</sup> School of Information Science and Technology, Northwest University, Xi'an, China, <sup>2</sup> Fudan University Shanghai Cancer Center, Shanghai, China, <sup>3</sup> Department of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China, <sup>4</sup> AstraZeneca, Shanghai, China

## OPEN ACCESS

### Edited by:

Zhuotun Zhu,  
Johns Hopkins Medicine,  
United States

### Reviewed by:

Takaaki Sugino,  
Tokyo Medical and Dental University,  
Japan  
Yanning Zhou,  
The Chinese University of Hong Kong,  
China

### \*Correspondence:

Lei Cui  
leicui@nwu.edu.cn  
Jun Feng  
fengjun@nwu.edu.cn  
Lin Yang  
linyang@nwu.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

**Received:** 31 August 2021

**Accepted:** 15 November 2021

**Published:** 14 December 2021

### Citation:

Li H, Kang Y, Yang W, Wu Z, Shi X,  
Liu F, Liu J, Hu L, Ma Q, Cui L, Feng J  
and Yang L (2021) A Robust Training  
Method for Pathological Cellular  
Detector *via* Spatial Loss Calibration.  
Front. Med. 8:767625.  
doi: 10.3389/fmed.2021.767625

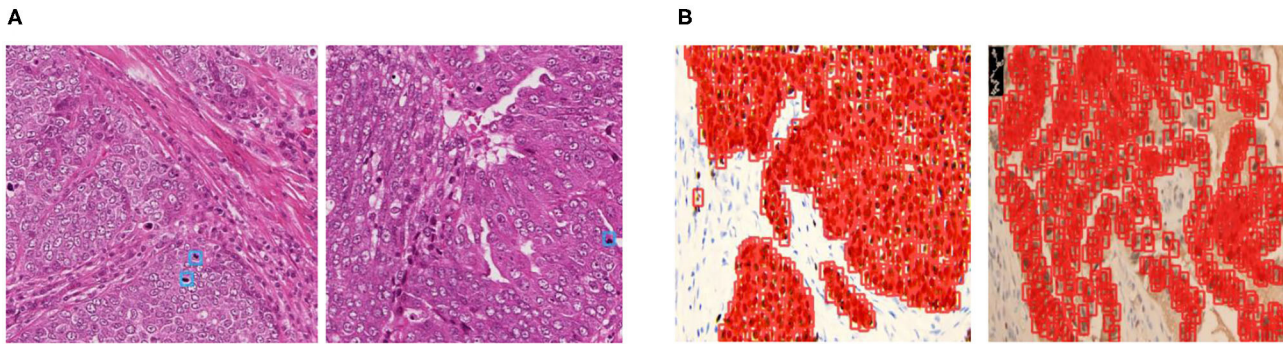
Computer-aided diagnosis of pathological images usually requires detecting and examining all positive cells for accurate diagnosis. However, cellular datasets tend to be sparsely annotated due to the challenge of annotating all the cells. However, training detectors on sparse annotations may be misled by miscalculated losses, limiting the detection performance. Thus, efficient and reliable methods for training cellular detectors on sparse annotations are in higher demand than ever. In this study, we propose a training method that utilizes regression boxes' spatial information to conduct loss calibration to reduce the miscalculated loss. Extensive experimental results show that our method can significantly boost detectors' performance trained on datasets with varying degrees of sparse annotations. Even if 90% of the annotations are missing, the performance of our method is barely affected. Furthermore, we find that the middle layers of the detector are closely related to the generalization performance. More generally, this study could elucidate the link between layers and generalization performance, provide enlightenment for future research, such as designing and applying constraint rules to specific layers according to gradient analysis to achieve "scalpel-level" model training.

**Keywords:** cellular detection, spatial loss calibration, sparsely annotated pathological datasets, convolutional neural network, object detection network

## 1. INTRODUCTION

Locating and counting cells in the pathological whole slide images (WSIs) is a direct way to find effective and important biomarkers, which is an essential and fundamental task of pathological image analysis (1–3). For instance, the spatial arrangement of tumor cells has been proved to be related to cancer grades (4, 5). Therefore, the qualitative and quantitative analysis of different types of tumors at cellular-level detection can help us better understand tumors and also explore various options for cancer treatment (6, 7).

Recently, object detection frameworks of Convolutional Neural Networks (obj-CNNs) have been proved powerful for locating instances in medical images [e.g., in CT images (8) and colonoscopy images (9)]. The big empirical success of obj-CNNs depends on the availability of a large corpus of fully annotated instances in training images (10). However, different from images of other modalities, we find two kinds of distributions of cells in pathological images, namely embedded and dense distribution, making full annotations of cellular-level instances difficult to



**FIGURE 1 |** The examples of annotations in the two kinds of datasets. **(A)** Cells of embedded distribution, which are sampled from the MITOS-ATYPIA-14 dataset. Obviously, mitoses that need to be annotated are often hidden among hundreds of other cells, tough to categorize and locate. **(B)** Cells of dense distribution in our Ki-67 dataset, usually more than hundreds of cells are required to be annotated in a small patch sampled from the whole slide image (WSI), which is an expensive and laborious task.

be guaranteed (refer to **Figure 1**). Specifically, the embedded distribution means that positive cells are hidden among hundreds of other cells, which are challenging for pathologists to categorize, locate, and then annotate. As for the dense distribution, a small patch sampled from the WSIs may contain hundreds of positive cells, making the annotation task expensive and laborious. Therefore, sparsely annotated datasets (SADs) are common in the field of the detection of cells.

In fact, when the training dataset contains a certain amount of sparse cellular annotations, the overfitting issue tends to easily occur, naturally leading to poor performance in generalization (11). In this study, we show the fundamental problems that decrease the generalization performance of the detector trained on SADs. First, deviation-loss, that is, numerous unannotated positive cells are mistaken for negative ones in the SADs, resulting in a serious miscalculated loss during training. Second, the deviation-loss dominates the early training process, and then drives the detector to learn only the features of the annotated cells, which yields the overfitting issue (Experimental testify can be seen in **Appendix A1**).

In this study, we point out that alleviating the deviation-loss during the training process can guide the detector to continuously learn the features of positive cells rather than only the annotated ones, and the SADs overfitting problem can be solved. In order to achieve that goal, the first cornerstone is how to identify those positive cells from negative ones when annotations are missing. We observe the more and more significant difference in densities between the predictions of the positive and negative cells during training (refer to **Figure 2**). Based on this observation, we propose a SADs training method named Boxes Density Energy (BDE), which utilizes densities' information to reduce the deviation-loss. Specifically, the more predictions for a cell, the more likely the cell is to be positive, and these predictions deserve smaller losses. In this way, deviation-loss disappears, and meanwhile, the overfitting problem is solved naturally.

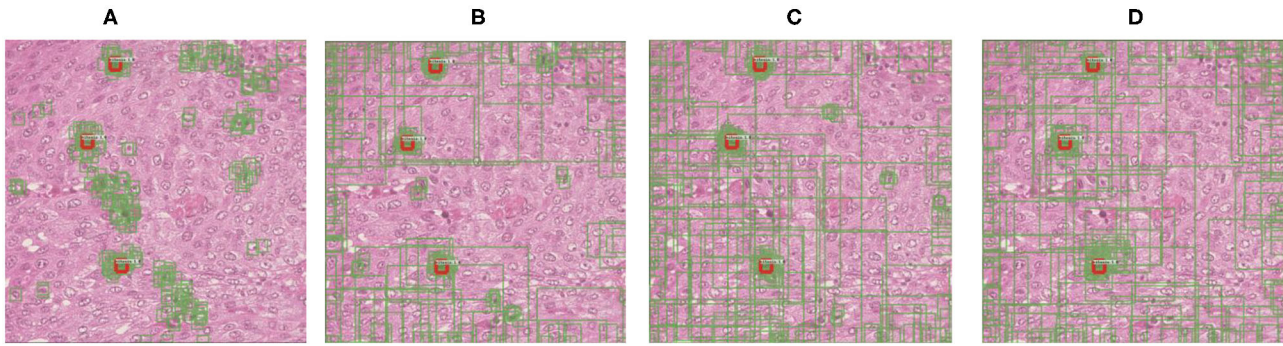
We have conducted experiments on two datasets, namely the MITOS-ATYPIA-14 dataset (embedded distribution)<sup>1</sup> and the Ki-67 dataset (dense distribution), which can be seen in **Figure 1**. Sufficient experimental results prove that our training method can significantly boost the performance of SADs. More importantly, we explore the gradient in the network and find that BDE brings a significant improvement on the middle layers (20–60 layers, 80 layers in total) of the network, indicating that the network's generalization performance seems to be closely related to the middle layers of the network. This may change the current training paradigm, such as applying constraint rules to specific layers according to gradient analysis to achieve the “scalpel-level” model training.

The organization of the study is as follows. The review of obj-CNNs and recent literature on SADs training methods is given in Section 2. Section 3 describes the proposed method in detail, and experimental results are presented in Section 4. Finally, we analyze the gradient of the trained network and conclude in Sections 5, 6, respectively.

A preliminary version of this study has been published in a conference study (12), which is only evaluated on the MITOS-ATYPIA-14 dataset. In this study, we have made significant extensions to generalize our methods on the Ki-67 dataset, aiming to provide a strong and comprehensive theory for relevant research. To be specific,

- We explore that some specific layers of CNN are strongly related to generalization performance, may provide theoretical guidance for future related research, e.g., one can improve the generalization of the network through more constraints on middle layers when training the network.
- In this study, we define the networks' training problems on SADs, from deviation-loss to the overfitting issue.

<sup>1</sup>MITOS-ATYPIA-14 dataset: <https://mitos-atypia-14.grand-challenge.org/dataset/>.



**FIGURE 2 |** The change of regression boxes' density during training on the MITOS-ATYPIA-14 dataset. Among them, the regression box is in green and manually annotated as red. This typical example shows that as the training process progresses, regression boxes increasingly surround the positive cells. **(A)** Regression boxes in 1 k training steps. **(B)** Regression boxes in 4 k training steps. **(C)** Regression boxes in 7 k training steps. **(D)** Regression boxes in 10 k training steps.

- This study formulated two cells' distribution in pathological images, namely embedded and dense distribution which may easily lead to SADs, and BDE can solve the SADs training problem on both embedded and dense distributions.

## 2. RELATED STUDY

### 2.1. Object Detection Networks

#### 2.1.1. The Framework

Object detection networks can be divided into two major categories, anchor-free and anchor-based frameworks. Among them, anchor-free frameworks (13, 14) are essentially making dense predictions, receiving higher recall rates but lower accuracy results (15), which do not meet the requirement of precisely pathological image analysis. On the other hand, anchor-based frameworks are more suitable for our tasks, and can be generally divided into one-stage methods (16, 17) and two-stage methods (18, 19). Both of them first tile a large number of preset anchors on the image, then predict the category and refine the coordinates of these anchors by one or several times, finally output these refined anchors as detection results. Because two-stage frameworks refine anchors several times more than one-stage frameworks (as shown in **Figure 3**), the former has greater accuracy. Hence, we choose the two-stage Feature Pyramid Network (FPN) (19) as the baseline in this paper.

#### 2.1.2. The Loss Function and Deviation Loss

In order to locate and recognize positive cells in the image, the object detection network has two parallel output layers to generate regression boxes ( $b$ ) with probability distribution ( $p$ ). The original loss ( $L$ ) consists of the classification loss  $L_{cls}$  and bounding-box regression loss  $L_{loc}$ :

$$L(p, u, b, v) = L_{cls}(p, u) + L_{loc}(b, v), \quad (1)$$

$$L_{cls}(p, u) = \sum_k -[u_k \cdot \log(p_k)], \quad (2)$$

$$L_{loc}(b, v) = \sum_k \text{smooth}_{L_1}(b_k - v_k), \quad (3)$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise.} \end{cases} \quad (4)$$

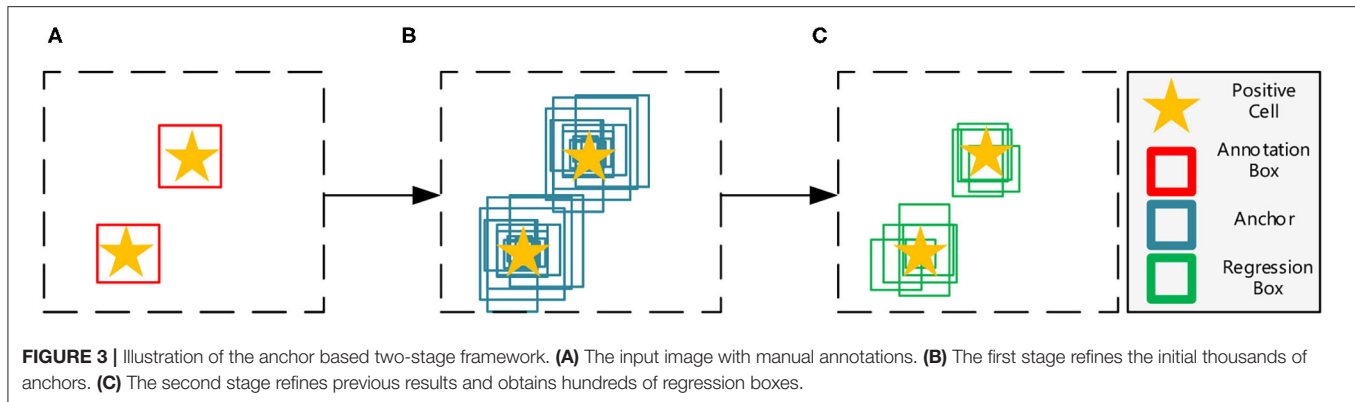
In Equation (2),  $u_k$  represents a one-hot label for a regression box indexed by  $k$ . When  $k$ -box's Intersection Over Union (IoU) with any instance annotation higher than a threshold, is assigned with a positive one-hot label ( $u_k \neq 0$ ), otherwise a negative ( $u_k = 0$ ). In Equation 3,  $v$  indicates the annotated bounding-boxes.

The loss function can accurately measure the margins between  $p$  and  $u$ ,  $b$ , and  $v$  on the fully annotated dataset. However, on the sparsely annotated cellular dataset, all unannotated positive cells are mistaken for negative, and  $u$  and  $v$  are translated into "untrustworthy" ground-truths. Thus,  $L_{cls}$  and  $L_{loc}$  may deviate seriously from the correct value, which we name deviation-loss. As a result, the deviation-loss confuses the training of networks, leading to limited performance.

### 2.2. Sparsely Annotated Datasets Training Methods

#### 2.2.1. Pseudo-Annotation Based Methods

In order to solve the SADs training problem, pseudo-annotation based methods have been proposed and achieved success on natural images (20, 21). They first train the detector using available instance-level annotations, then generate pseudo-annotations, and merge them with the original annotations to iteratively update the detector. For example, Niitani et al. (22) trained the detector to generate annotations using the Open Images Dataset V4 (OID). They then sampled the pseudo-annotations using assumptions such as "cars should contain tires." However, such a priori assumption in the field of cell detection is unknown. Other methods based on pseudo-annotations still need a certain number of fully annotated datasets, like Yan et al. (23) and Inoue et al. (24) employ a subset of fully annotated datasets to obtain a pre-trained detector, generating pseudo-annotations for the next training.



Obviously, such an iterative process brings uncontrollability into the training process, e.g., a bad pseudo-annotation generator may significantly influence the final results. In addition, there is not much consensus on how to utilize the pseudo-annotations until now, especially for object detection (22), e.g., determining the optimal number of iterations is tricky, therefore, it is urgent to solve the SADs training problem in a non-iterative way. Besides, considering that such methods are relatively difficult to replicate, with respect to, empirical and tricky parameter selection or special requirements of the forms of datasets, this study does not include such methods in the comparative experiment.

### 2.3. Loss-Calibration Based Methods

Compared with pseudo-annotation based methods, the loss-calibration methods for solving noise labels are more relevant to our study. The meaning of noise labels is wrong labels or missing labels (25, 26). These methods aim to reduce noise labels by establishing loss functions that are more noise-tolerant. For example, Müller et al. (27) softens the labels by adding a uniform distribution. Wang et al. (28) assumes that the network will become more and more reliable as the training continues and proposes reducing the loss gradually to reduce the influence of noise labels. However, these loss calibration methods also inevitably reduce the core contributions of correct labels for the training of the network. On the contrary, our BDE utilizes the regression boxes' density to encourage correct predictions and give relatively more significant losses to wrong predictions, whether the label is missing or not.

It is worth noting that in view of the class imbalance problem they, the have put forward many loss weighting schemes (17, 29). However, these methods may cause relatively large losses to correct predictions lacking corresponding annotations, which makes them ineffective on SADs.

## 3. BOXES DENSITY ENERGY

The overall process of our proposed BDE is shown in **Figure 4**. BDE is proposed to encourage the correct predictions of unannotated positive cells to ignore the adverse effect of the deviation-loss, which can be summarized into five core steps.

**Figure 4A** A sparsely annotated image is inputted for the training. At the second stage of the detector, each cell is surrounded by some regression boxes automatically that we regard as a group. **Figure 4B** Boxes Density: Calculate the average distance between each box and the others. **Figure 4C** Boxes Energy: Normalized operation by dividing the Box Density by the maximum distance between all boxes. **Figure 4D** Calculate the original total loss. **Figure 4E** BDE loss: Calibrate the original loss with Boxes Energy to guide the detector training in the right direction.

### 3.1. Boxes Density

The boxes density can be measured by the average distance between each box, so that denser boxes have smaller average distances than isolating ones. The density of a box indexed by  $i$  can be represented as:

$$\text{Density}(b_i) = \frac{1}{N} \sum_j^N D(b_i, b_j), \quad (5)$$

where  $N$  is the number of boxes per image,  $D$  is the distance function, we choose Manhattan distance (Equation 6) in this study considering the less computational cost.

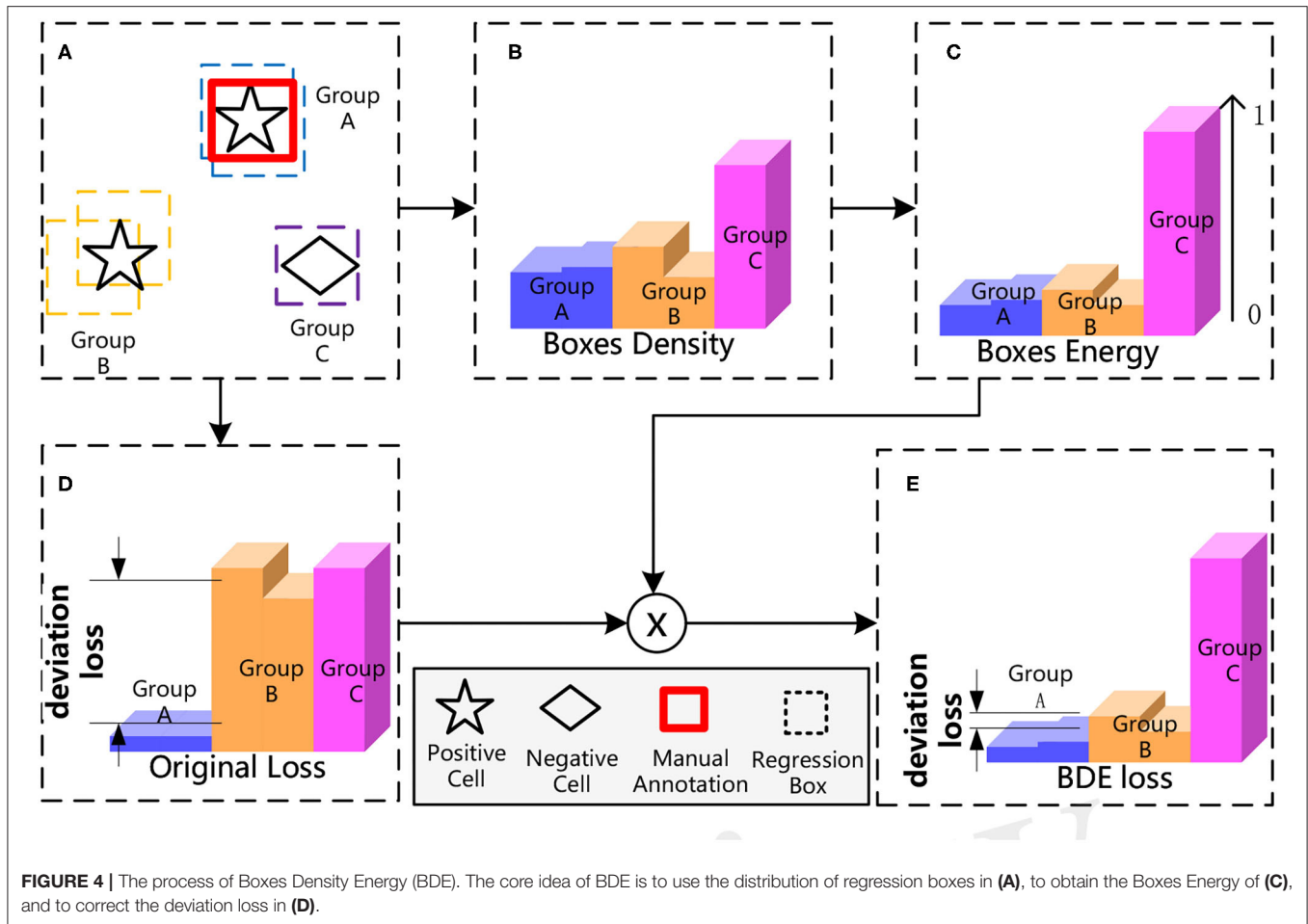
$$D(b_i, b_j) = |x_i - x_j| + |y_i - y_j|, \quad (6)$$

In which, the  $x_i$  and  $y_i$  represent the x-coordinate and y-coordinate of the center point of the box indexed by  $i$ .

We can prove that the average distance can measure the density effectively; if we treat regression boxes around a cell as a group, and assume that we have  $k$  groups  $\{G_1, \dots, G_j, \dots, G_k\}$ . Meanwhile, there are  $\{m_1, \dots, m_j, \dots, m_k\}$  boxes in the corresponding group.

For simplicity, we assume that the distances within a group are all close to 0, the distances between the groups are all  $d$ , and the total number of boxes is  $N$ , which means that  $N = \sum_{l=1}^k m_l$ . Thus, the average distance of each box in the  $j$ -group is Equation (7). This indicates that the box in a denser group (larger  $m_j$ ) of the  $j$ -group has a smaller density value.

$$\text{Density}(b_i) = \frac{0 \times m_j + (N - m_j) \times d}{N} = d \times (1 - \frac{m_j}{N}). \quad (7)$$



### 3.2. Boxes Energy and Loss Calibration

The main idea of our proposed method is that the more prediction boxes around a cell, the cell is more likely to be positive, and therefore, the predictions should have a smaller loss. The density of each box has been modeled, however, the range of density is not normalized. Therefore, we use Equation (8) to convert the Boxes Density to Boxes Energy which is normalized from 0 to 1. Afterward, Boxes Energy can be utilized as a weight of  $L_{cls}$  and  $L_{loc}$  (refer to Equations 9, 10). By that, the deviation loss is alleviated by calibrating the original loss.

$$\text{Energy}(b_i) = \frac{\text{Density}(b_i)}{\max(D(b))}. \quad (8)$$

$$L_{cls}^{BDE}(p, u) = \sum_k [1_{u_k=0} (\text{Energy}(b_k)) + 1_{u_k \neq 0}] \cdot [-u_k \cdot \log(p_k)], \quad (9)$$

$$L_{loc}^{BDE}(b, v, u) = \sum_k [1_{u_k=0} (\text{Energy}(b_k)) + 1_{u_k \neq 0}] \cdot [\text{smooth}_{L_1}(b_k - v_k)]. \quad (10)$$

In Equations 9, 10,  $u_k$  equals zero indicates the one-hot label of the box indexed by  $k$  is negative. With the loss-calibration of BDE, the detector can be trained along the right direction on the SADs. For example, if the box indexed by  $k$  is mistaken for negative ( $u_k$  is zero) due to SADs, but has a small  $\text{Energy}(b_k)$ , then, the original deviation-loss is calibrated by the term of  $\text{Energy}(b_k)$ . Finally, the total loss is improved from Equation (1) to:

$$L^{BDE}(p, u, b, v) = L_{cls}^{BDE}(p, u) + L_{loc}^{BDE}(b, v). \quad (11)$$

## 4. EXPERIMENTS

We utilize the FPN (19) with the backbone resnet50 (30) as the baseline. Our method is also compared with the representative loss-calibration methods, namely Label Smooth (LS) (27) and ProSelfLC (28). In Section 4.3, we conduct experiments to detect mitosis on the 2014 MITOS-ATYPIA Grand Challenge dataset and to detect tumor-cells on the Ki-67 dataset in Section 4.4. These two datasets can represent embedded and dense annotations. Experimental results demonstrate that BDE outperforms other methods on the SADs significantly, and BDE can address the training problem of SADs of both embedded and dense annotations.

## 4.1. Description and Implementation Details

The experiments for KI-67 and 2014 MITOS-ATYPIA datasets set the same hyperparameters. The inputted image is resized to the resolution of  $800 \times 800$  pixels. The number of training steps is 10 k. The learning rate is initially set to 0.001 and is divided by 10 at 5 k and 7.5 k steps. In order to objectively evaluate our method, we perform 4-fold cross-validation on the MITOS-ATYPIA-14 dataset and 3-fold cross-validation on the Ki-67 dataset. We implement our framework with the open source software library TensorFlow version 1.12.0 on a workstation equipped with two NVIDIA GeForce 2080 Ti GPUs.

## 4.2. Evaluation Metrics

The average precision (AP) and recall are used for performance evaluation. The recall is defined as the proportion of all positive examples ranked above a given rank. Precision is the proportion of all examples above that rank that are from the positive class. The AP summarizes the shape of the precision/recall curve and is defined as the mean precision at a set of eleven equally spaced recall levels  $[0, 0.1, \dots, 1]$ :

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, \dots, 1\}} p_{\text{interp}}(r). \quad (12)$$

The precision at each recall level  $r$  is interpolated by taking the maximum precision measured for a method for which the corresponding recall exceeds  $r$ :

$$p_{\text{interp}}(r) = \max_{\tilde{r} : \tilde{r} \geq r} p(\tilde{r}), \quad (13)$$

where  $p(\tilde{r})$  is the measured precision at recall  $\tilde{r}$  (31).

## 4.3. Experiments on the 2014 MITOS-ATYPIA Grand Challenge Dataset (embedded annotations)

### 4.3.1. Data Description

We have conducted experiments on the 2014 MITOS-ATYPIA Grand Challenge Dataset (MITOS-ATYPIA-14 dataset). The data samples were scanned by two slide scanners Aperio Scanscope XT and Hamamatsu Nanozoomer 2.0-HT, whole-slide histological images (WSIs) stained with standard hematoxylin and eosin (H&E) dyes. The centroids pixels of mitoses were manually annotated *via* two senior pathologists. In a situation of contradiction between the pathologists, the third one will provide the final say.

We choose the train-set of WSIs scanned from Hamamatsu Nanozoomer 2.0-HT, and we sample 393 patches that contain 743 mitoses with a sliding window of resolution of  $1,663 \times 1,485$  pixels. Annotations for training the FPN are generated by  $32 \times 32$  bounding boxes centered on all centroids pixels. For the MITOS-ATYPIA-14 dataset, we refer to the original data as a fully annotated dataset. Meanwhile, we randomly delete annotations until there is only one per training image and name it as an extremely sparse dataset. It is worth noting, we only conduct the sparse operations on the training dataset, and the testing dataset is intact.

### 4.3.2. Results of MITOS-ATYPIA-14 dataset

Boxes Density Energy can improve recall results on the fully annotated dataset. **Table 1** lists the recall and AP results on the fully annotated dataset. For the AP results, all methods have lower AP results than the baseline (FPN), which demonstrates that when loss-calibration methods are introduced to the training on fully annotated embedded annotations, interfering with the network's accuracy. On the other hand, for the recall results, BDE can improve the recall results significantly. FPN, LS, and ProSelfLC achieve 89.8, 85.5, and 88.7% average recall, respectively. While BDE achieves 94.6%, exceeding that of FPN by 4.8%.

Boxes Density Energy improves the network's performance in all aspects on the sparsely annotated dataset. As shown in **Table 2**, BDE outperforms other methods significantly on both AP and recall results. However, LS's overall performance is reduced compared with the baseline, which indicates that the assumption of annotation-distribution of LS is incompatible in the embedded annotations, whose positive and negative samples are extremely unbalanced.

## 4.4. Experiments on the Ki-67 Dataset (Dense Annotations)

### 4.4.1. Data Description

The Ki-67 dataset is used for training FPN to detect tumor-cells and count their number. We have 206 patches with a resolution of  $1,080 \times 1,920$  pixels sampled from WSIs, and the pathologists try their best to annotate all the tumor cells with key points in all patches. Finally, 21,025 tumor cells have been annotated. Then, we generate  $32 \times 32$  bounding boxes centered on all key points.

#### 4.4.1.1. The SAD of the Ki-67 Dataset

For the Ki-67 dataset, considering that there is an average of 102 annotated tumor cells in each patch, so we can retain different annotation rates to train the network to fully validate BDE, e.g., the retentive rate is 0.1 if 10% of annotations are retained. We have carried out experiments starting from the retentive rate of 0.1 and increasing it to 1 by 0.1. We believe that if the retentive rate is below 0.5, then the dataset we can define as a SADs because the number of unannotated instances is greater than the number of annotated instances in such a dataset. Experimental results have demonstrated the BDE can significantly boost the performance of networks trained on that SADs.

### 4.4.2. The Quantization Results

We evaluate the performance of our BDE which is trained on datasets with different retentive-rates, and observe that BDE is a robust training method, which is hardly affected by the quality of data annotations. For example, in **Table 3**, when the retentive-rate is dropped from 1.0 (original) to 0.1, BDE's AP result dropped from 49.02 to 46.45%, only reducing by 2.57%. On the other hand, FPN decreased by 23.88%, and LS decreased by 27.17%, and ProSelfLC decreased by 21.05%.

Similarly, **Table 4** lists the recall results of different methods trained on different retentive-rates. When the retentive-rate decreases from 1.0 to 0.1, BDE only reduces recall results by

**TABLE 1** | The recall and average precision (AP) results on the fully annotated MITOS dataset (original dataset).

Method	Fold1		Fold2		Fold3		Fold4		Avg. Recall	Avg. AP
	Recall	AP	Recall	AP	Recall	AP	Recall	AP		
FPN (Baseline)	80.2	41.8	89.4	46.9	95.8	44.6	93.6	60.7	89.8	<b>48.5</b>
LS (27)	75.6	36.7	84.6	47.7	91.6	41.7	90.4	64.2	85.5	47.6
ProSelfLC (28)	80.2	32.7	86.5	40.6	95.2	40.3	93.1	62.7	88.7	44.1
BDE (ours)	90.6	40.7	93.3	42.3	99.4	43.2	95.0	59.1	<b>94.6</b>	46.3

**TABLE 2** | The recall and AP results on the sparsely annotated MITOS dataset (retain one annotation in each image).

Method	Fold1		Fold2		Fold3		Fold4		Avg. Recall	Avg. AP
	Recall	AP	Recall	AP	Recall	AP	Recall	AP		
FPN (Baseline)	69.8	34.5	81.7	32.9	94.6	37.4	88.1	55.9	83.6	40.2
LS (27)	65.8	24.6	71.1	30.4	86.8	33.9	83.4	54.6	76.7	35.8
ProSelfLC (28)	80.2	28.8	84.6	28.4	95.8	30.1	85.7	50.1	86.5	34.3
BDE (ours)	88.5	41.8	89.4	37.1	95.8	40.2	91.3	60.1	<b>91.3</b>	<b>44.8</b>

**TABLE 3** | The AP results on different annotations-retentive rates on the Ki-67 dataset.

Retentive rate	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
FPN (Baseline)	26.22	33.57	39.92	41.94	43.88	45.15	46.17	47.28	48.22	50.1
LS (27)	24.30,	37.39	41.01	44.16	45.48	46.47	47.69	48.96	50.01	51.47
ProSelfLC (28)	30.67	38.85	43.07	45.37	46.57	<b>47.72</b>	<b>48.79</b>	<b>49.91</b>	<b>50.87</b>	<b>51.72</b>
BDE (ours)	<b>46.45</b>	<b>46.36</b>	<b>46.24</b>	<b>46.71</b>	<b>46.94</b>	47.52	47.24	48.05	48.60	49.02

**TABLE 4** | The recall results on different annotations-retentive rates on the Ki-67 dataset.

Retentive rate	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
FPN (Baseline)	38.84	44.81	48.34	49.84	51.01	51.73	52.39	53.14	53.64	54.22
LS (27)	31.45	45.23	48.01	50.82	51.93	52.84	<b>53.46</b>	<b>54.00</b>	<b>54.43</b>	<b>54.69</b>
ProSelfLC (28)	43.43	48.24	50.28	51.62	52.25	52.78	53.31	53.90	54.16	54.38
BDE (ours)	<b>52.70</b>	<b>53.07</b>	<b>53.12</b>	<b>53.32</b>	<b>53.25</b>	<b>53.68</b>	53.37	53.82	53.95	53.29

0.59%. While FPN, LS, and Proself LC decreased by 15.38, 23.24, and 10.95%, respectively. Furthermore, from **Figures 5, 6**, the robustness and stability of BDE can be demonstrated from the perspective of AP results and recall results' curves. Our method is almost unaffected by sparse annotations. In particular, when the retentive rate is in the range of 0.1–0.5, that is, sparse annotation, BDE achieves significant improvements.

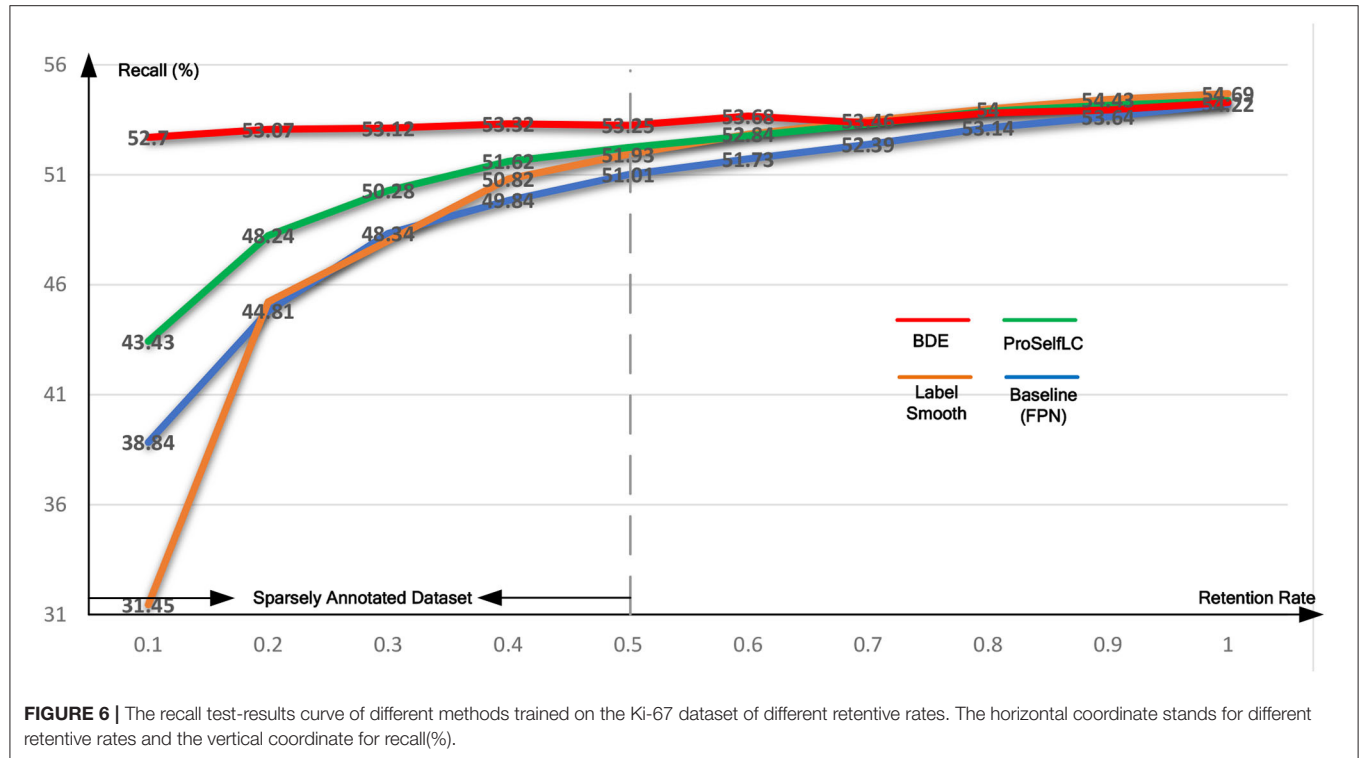
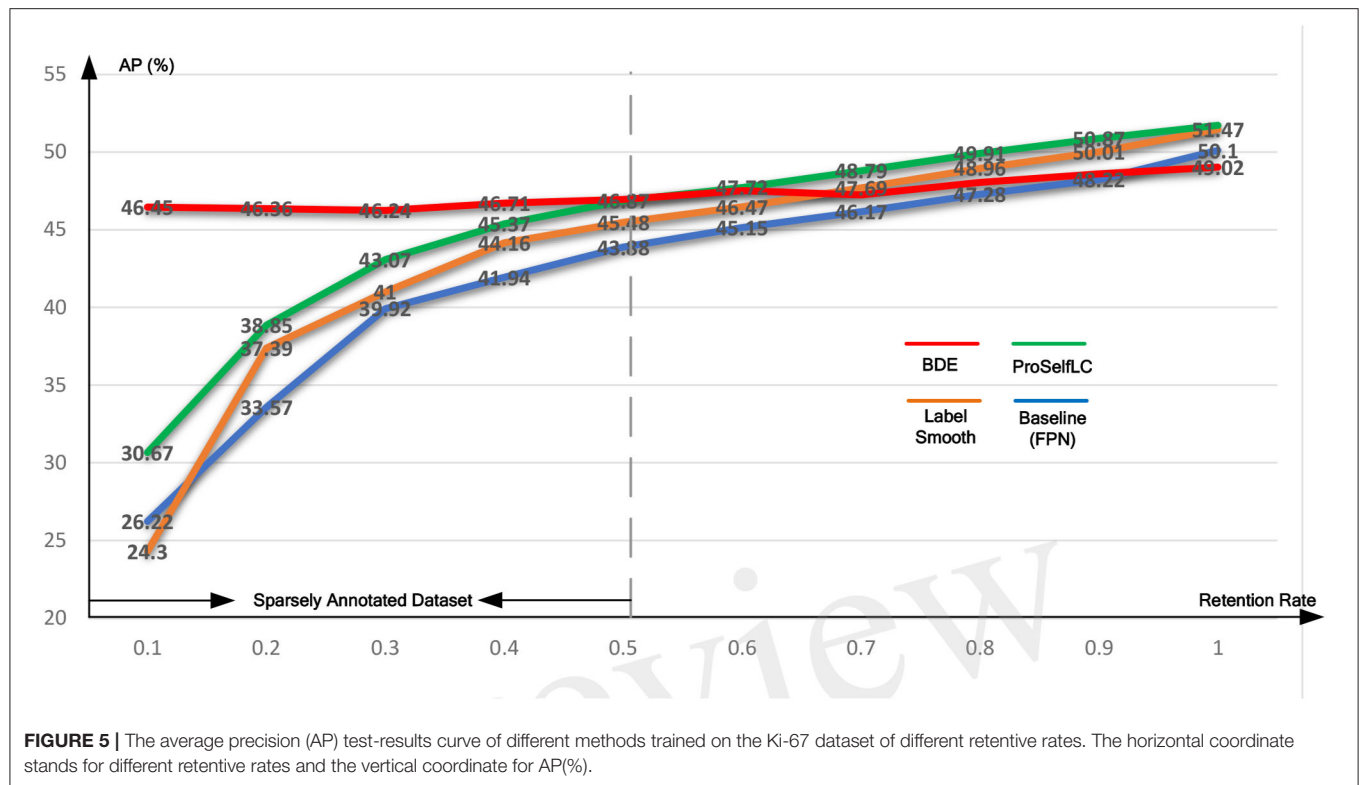
#### 4.4.3. The Qualitative Results

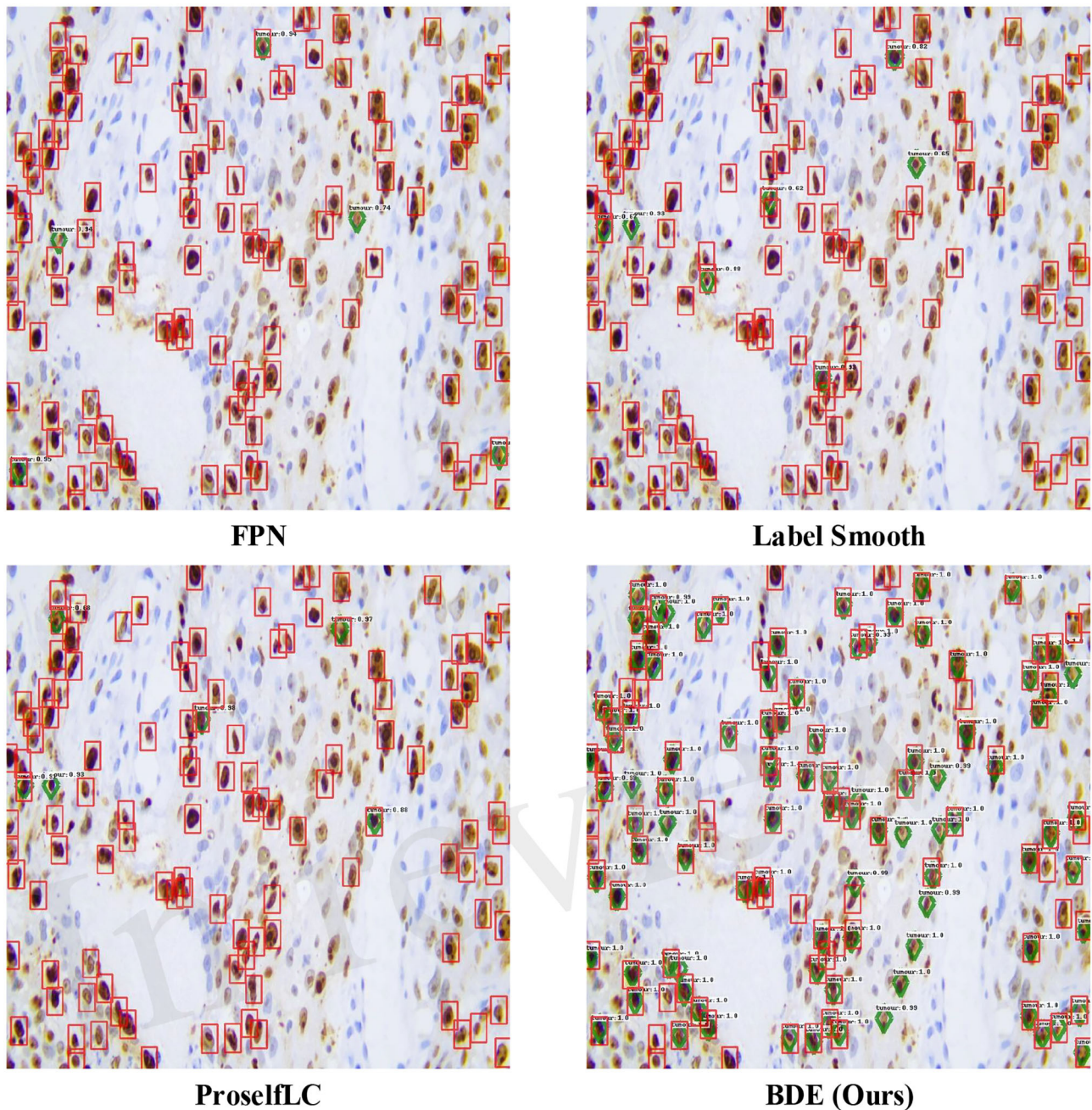
In **Figure 7**, we list some detection results produced by different methods. A score threshold of 0.6 is used for display. Obviously, other methods trained on the sparsely annotated dataset (the retentive rates is 0.1) tend to miss tumor cells, while our method largely avoids that mistake. Meanwhile, our BDE trained on the 0.1 retentive rate even achieve better performance than other methods trained on the 0.4 retentive rate.

## 5. LAYER-LEVEL GRADIENT ANALYSIS

### 5.1. Why Need Layer-Level Gradient Analysis

The gradient of a kernel is obtained by taking the chain derivative of the loss with respect to the weight, so that, the larger the weight of the kernel, not only its gradient is smaller but it also indicates that the kernel is more important. Thus, by comparing gradients of the same kernel but trained by different methods, we can know the advantages and disadvantages of training methods for this kernel. However, there are usually more than thousands of kernels in a single network, and it is not instructive to understand the superiority of kernel-level training. On the other hand, the same layer's kernels are responsible for similar feature extractions, e.g., kernels of a specific layer extract edges from different angles. Naturally, all kernels' average gradients in each





**FIGURE 7 |** Examples of detection results on the Ki-67 dataset. These results are trained on the dataset with the retentive rate of 0.1 and tested on the test-set. Predictions are drawn in green diamond, and manual annotations are in red boxes. A score threshold of 0.6 is used for display.

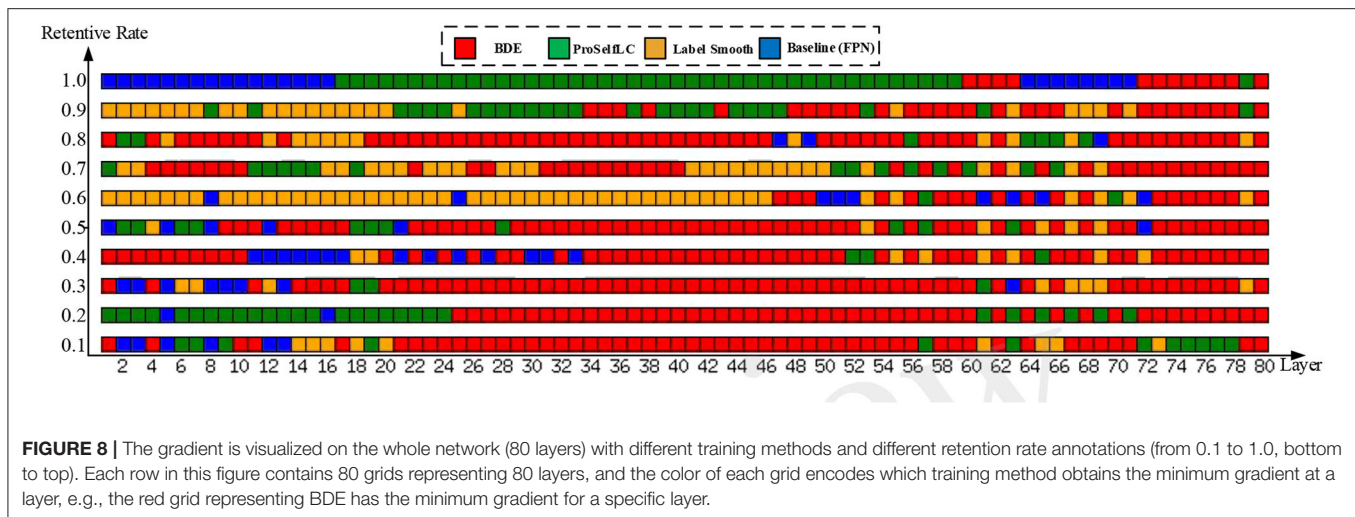
layer can be used as an objective evaluation standard for feature extraction ability. Therefore, we analyze the gradient of each layer to investigate why BDE can improve the performance.

## 5.2. How to Analyze the Gradient

We analyze the mean value of the gradients in each layer of the network by computing the back-propagation

via the testing loss. Specifically, for a layer indexed by  $l$ , whose mean gradient ( $\mu_l$ ) can be computed as follow:

$$\mu_l = \sum_{k=1}^{k=K} \frac{1}{K} \cdot A_{l,k}; \quad \mu_l \in \mathbb{R}^1, \quad (14)$$



in which,  $K$  is the number of convolution kernels in the layer indexed by  $l$ , and  $A_{l,k}$  can be obtained by Equation (15).

$$A_{l,k} = \frac{1}{d \times w \times h} \sum_i^d \sum_j^w \sum_m^h G_{l,k}^{i,j,m}, \quad (15)$$

where  $G_{l,k}$  is the gradient of  $k$ -th convolutional kernel in the  $l$ -th layer. Meanwhile,  $d$ ,  $w$ , and  $h$  are the depth, the width, and the height of this kernel.  $G_{l,k}$  can be computed by Equation (16).

$$G_{l,k} = \sum_i^N \frac{1}{N} \cdot \left| \frac{\partial L_{test}^i}{\partial W_{l,k}} \right|; \quad G_{l,k} \in R^{d \times w \times h}, \quad (16)$$

where  $L_{test}^i$  represents the loss computed on the  $i$ -th testing image, and there are  $N$  testing images, and  $W_{l,k}$  is the weights of the  $k$ -th convolutional kernel in the  $l$ -th layer. Further, the gradient represents the direction whether it is positive or negative, so that we perform an absolute operation on the calculated gradient.

### 5.3. Visualization and Discussion of the Gradient

As shown in **Figure 8**, we visualize the layer-level gradient of the networks (with 80 layers), which are trained on the Ki-67 dataset (retentive rates range from 0.1 to 1), and the gradient is obtained by the testing loss of the Ki-67 dataset. For each layer, we compare whose gradient is trained on different methods. Specifically, a grid with different colors indicates which method can obtain the minimum gradient, e.g., a red grid shows that our approach reduces the test gradient for a particular layer.

We can observe from **Figure 8** when the network is trained on a dataset whose retentive rate below 0.5, BDE improves most of the middle layers (roughly 20–60 layers), which does not seem to happen by accident. Therefore, we can further presume that the generalization performance improvement of the cell detection task is closely related to the middle layers of the network.

## 6. CONCLUSION

In this study, through theoretical analysis and experimental verification, we identify that the detector trained on sparsely annotated cellular datasets may fall into overfitting due to deviation-loss. In order to address the training limitation, we propose a novel training method, which is utilized to calibrate the deviation-loss based on the cues provided by the density of regression boxes. Extensive experiments demonstrated the strength of BDE to significantly improve the training performance of the cellular detector, even with 90% of annotations missing, the performance of our method is barely affected. Thus, our proposed BDE might enable better and faster development of accurate cellular detection. More importantly, through the visual analysis of the network gradient, we find that the improvement of generalization performance is closely related to the middle layer of the network, which is expected to provide a new theoretical direction for future research.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

HL: conception and design of study. FL, XS, and LH: drafting the manuscript. YK, ZW, and JL: analysis and/or interpretation of data. LY, WY, and QM: acquisition of data. LC and JF: funding acquisition. All authors contributed to the article and approved the submitted version.

## FUNDING

This study is supported by the National Natural Science Foundation of China (NSFC grant no. 6207326), and the Natural Science Foundation of Shaanxi Province of China (2021JQ-461, 2020JM-387).

## REFERENCES

- Li C, Wang X, Liu W, Latecki LJ. DeepMitosis: mitosis detection via deep detection, verification and segmentation networks. *Med Image Anal.* (2018) 45:121–33. doi: 10.1016/j.media.2017.12.002
- Xing F, Su H, Neltner J, Yang L. Automatic Ki-67 counting using robust cell detection and online dictionary learning. *IEEE Trans Biomed Eng.* (2014) 61:859. doi: 10.1109/TBME.2013.2291703
- Xing F, Cornish TC, Bennett T, Ghosh D, Yang L. Pixel-to-pixel learning with weak supervision for single-stage nucleus recognition in Ki67 images. *IEEE Trans Bio Med Eng.* (2019) 66:3088–97. doi: 10.1109/TBME.2019.2900378
- Sirinukunwattana K, Raza SEA, Tsang YW, Snead DR, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans Med Imaging.* (2016) 35:1196–206. doi: 10.1109/TMI.2016.2525803
- Lewis JS, Ali S, Luo J, Thorstad WL, Madabhushi A. A quantitative histomorphometric classifier (QuHbIC) identifies aggressive versus indolent p16-positive oropharyngeal squamous cell carcinoma. *Am J Surg Pathol.* (2014) 38:128. doi: 10.1097/PAS.0000000000000086
- Schmidt U, Weigert M, Broaddus C, Myers G. Cell detection with star-convex polygons. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Granada: Springer (2018). p. 265–73.
- Zhou Y, Chen H, Xu J, Dou Q, Heng PA. IRNet: instance relation network for overlapping cervical cell segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Shenzhen: Springer (2019). p. 640–8.
- Xie H, Yang D, Sun N, Chen Z, Zhang Y. Automated pulmonary nodule detection in CT images using deep convolutional neural networks. *Pattern Recognit.* (2019) 85:109–19. doi: 10.1016/j.patcog.2018.07.031
- Zhang R, Zheng Y, Poon CC, Shen D, Lau JY. Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker. *Pattern Recognit.* (2018) 83:209–19. doi: 10.1016/j.patcog.2018.05.026
- Xu M, Bai Y, Ghanem B, Liu B, Gao Y, Guo N, et al. Missing labels in object detection. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. Long Beach, CA (2019).
- Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:161103530*. (2016).
- Li H, Han X, Kang Y, Shi X, Yan M, Tong Z, et al. A novel loss calibration strategy for object detection networks training on sparsely annotated pathological datasets. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Lima: Springer (2020). p. 320–9.
- Tian Z, Shen C, Chen H, He T. FCOS: fully convolutional one-stage object detection. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul: IEEE (2020).
- Zhou X, Wang D, Krähenbühl P. Objects as Points. *arXiv.* (2019) *arXiv:1904.04467*.
- Zhang S, Chi C, Yao Y, Lei Z, Li SZ. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Virtual (2020) p. 9759–68.
- Huang R, Pedoem J, Chen C. YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers. In: *2018 IEEE International Conference on Big Data (Big Data)*. Seattle, WA: IEEE (2018). p. 2503–10.
- Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In: *Proceedings of the IEEE International Conference on Computer Vision*. Venice: IEEE (2017). p. 2980–8.
- Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. In: *Advances in Neural Information Processing Systems*. Montreal, QC (2015). p. 91–99.
- Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI (2017). p. 2117–25.
- Zhang X, Wei Y, Feng J, Yang Y, Huang TS. Adversarial complementary learning for weakly supervised object localization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT (2018). p. 1325–34.
- Zhang X, Wei Y, Kang G, Yang Y, Huang T. Self-produced guidance for weakly-supervised object localization. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Munich (2018). p. 597–613.
- Niitani Y, Akiba T, Kerola T, Ogawa T, Sano S, Suzuki S. Sampling techniques for large-scale object detection from sparsely annotated objects. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Long Beach, CA (2019). p. 6510–8.
- Yan Z, Liang J, Pan W, Li J, Zhang C. Weakly-and semi-supervised object detection with expectation-maximization algorithm. *arXiv preprint arXiv:170208740*. (2017).
- Inoue N, Furuta R, Yamasaki T, Aizawa K. Cross-domain weakly-supervised object detection through progressive domain adaptation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT: IEEE (2018) p. 5001–9.
- Sukhbaatar S, Bruna J, Paluri M, Bourdev L, Fergus R. Training convolutional networks with noisy labels. *arXiv preprint arXiv:14062080*. (2014).
- Patrini G, Rozza A, Krishna Menon A, Nock R, Qu L. Making deep neural networks robust to label noise: a loss correction approach. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI (2017). p. 1944–52.
- Müller R, Kornblith S, Hinton GE. When does label smoothing help? In: *Advances in Neural Information Processing Systems*. Vancouver, BC (2019). p. 4694–703.
- Wang X, Hua Y, Kodirov E, Clifton DA, Robertson NM. ProSelfLC: progressive self label correction for training robust deep neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Virtual (2021).
- Li B, Liu Y, Wang X. Gradient harmonized single-stage detector. In: *Computer Vision and Pattern Recognition*. Salt Lake City, UT (2018).
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV: IEEE (2016). p. 770–8.
- Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. *Int J Comput Vis.* (2010) 88:303–38. doi: 10.1007/s11263-009-0275-4

## ACKNOWLEDGMENTS

The authors would like to thank the medical team at AstraZeneca China and the technical team at DeepInformatics++ for their scientific comments on this study.

**Conflict of Interest:** WY, LH, and QM were employed by the company AstraZeneca, Shanghai, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

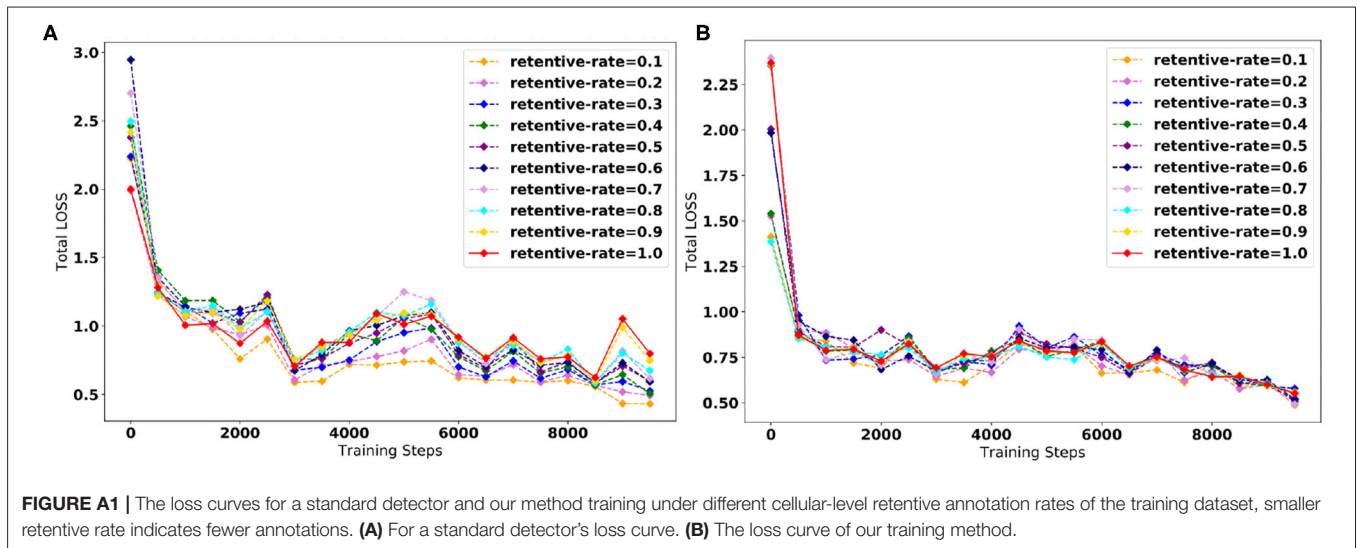
Copyright © 2021 Li, Kang, Yang, Wu, Shi, Liu, Liu, Hu, Ma, Cui, Feng and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## APPENDIX

### Overfitting Issues When Datasets Are Sparsely Annotated

**Figure A1A** in Appendix exhibits the loss curve of a standard object detector trained on the KI-67 dataset at different cellular-level retentive annotation rates. Before 3,000 steps, the detector trained on the datasets with

a lower retentive annotation rate leads to a larger loss, which indicates that the deviation-loss dominates the training process. After that, lower retentive annotation rates lead to smaller losses, which indicates that the detector tends to focus on the annotated instances and then drives the overfitting issue. As shown in **Figure A1B** In Appendix, our method can significantly solve the overfitting issue.





# Automatic Segmentation of Novel Coronavirus Pneumonia Lesions in CT Images Utilizing Deep-Supervised Ensemble Learning Network

Yuanyuan Peng<sup>1,2\*</sup>, Zixu Zhang<sup>1</sup>, Hongbin Tu<sup>1,3</sup> and Xiong Li<sup>4</sup>

<sup>1</sup> School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang, China, <sup>2</sup> School of Computer Science, Northwestern Polytechnical University, Xi'an, China, <sup>3</sup> Technique Center, Hunan Great Wall Technology Information Co. Ltd., Changsha, China, <sup>4</sup> School of Software, East China Jiaotong University, Nanchang, China

## OPEN ACCESS

### Edited by:

Jun Feng,  
Northwest University, China

### Reviewed by:

Wenjie Zhang,  
Changsha University of Science &  
Technology, China  
Cheng Liang,  
Shandong Normal University, China  
Hong Zhao,  
National University of Defense  
Technology, China  
Guanglei Zhang,  
Beihang University, China

### \*Correspondence:

Yuanyuan Peng  
3066@ecjtu.edu.cn

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

**Received:** 08 August 2021

**Accepted:** 29 November 2021

**Published:** 03 January 2022

### Citation:

Peng Y, Zhang Z, Tu H and Li X (2022)  
Automatic Segmentation of Novel  
Coronavirus Pneumonia Lesions in CT  
Images Utilizing Deep-Supervised  
Ensemble Learning Network.  
Front. Med. 8:755309.  
doi: 10.3389/fmed.2021.755309

**Background:** The novel coronavirus disease 2019 (COVID-19) has been spread widely in the world, causing a huge threat to the living environment of people.

**Objective:** Under CT imaging, the structure features of COVID-19 lesions are complicated and varied greatly in different cases. To accurately locate COVID-19 lesions and assist doctors to make the best diagnosis and treatment plan, a deep-supervised ensemble learning network is presented for COVID-19 lesion segmentation in CT images.

**Methods:** Since a large number of COVID-19 CT images and the corresponding lesion annotations are difficult to obtain, a transfer learning strategy is employed to make up for the shortcoming and alleviate the overfitting problem. Based on the reality that traditional single deep learning framework is difficult to extract complicated and varied COVID-19 lesion features effectively that may cause some lesions to be undetected. To overcome the problem, a deep-supervised ensemble learning network is presented to combine with local and global features for COVID-19 lesion segmentation.

**Results:** The performance of the proposed method was validated in experiments with a publicly available dataset. Compared with manual annotations, the proposed method acquired a high intersection over union (IoU) of 0.7279 and a low Hausdorff distance (H) of 92.4604.

**Conclusion:** A deep-supervised ensemble learning network was presented for coronavirus pneumonia lesion segmentation in CT images. The effectiveness of the proposed method was verified by visual inspection and quantitative evaluation. Experimental results indicated that the proposed method has a good performance in COVID-19 lesion segmentation.

**Keywords:** under CT imaging, deep learning, COVID-19 lesion segmentation, deep-supervised ensemble learning network, transfer learning, local and global features

## INTRODUCTION

Since the end of 2019, acute infectious pneumonia characterized by novel coronavirus disease 2019 (COVID-19) infection has been rapidly spread in the world, posing a huge threat to the lives of people (1). The outbreak of pneumonia caused by COVID-19 infection has been identified by the WHO as an emergency public health event of international concern, the number of patients with COVID-19 is rapidly growing in the world (2). So far, the cumulative confirmed cases of COVID-19 in the world exceeded 200 million, and the cumulative deaths reached 3.6 million. The early symptoms of pneumonia are not obvious but are strongly infective. It has caused huge economic losses to society and aroused wide concern in the world (3, 4).

The difficulty of prevention and treatment of COVID-19 has put forward urgent requirements to the research of rapid diagnosis methods. The prevention and control measures of early diagnosis, early isolation, and early treatment for patients with COVID-19 are one of the most effective strategies to solve the pneumonia epidemic (5). However, widely used nucleic acid testing and specific antibody detection technologies have several disadvantages, such as lagging, long time consuming, low detection efficiency, and serious risk of missed detection (6). As one of the most effective lung imaging modes, CT has the ability to identify the changes in lung lesions and pathological features in patients with COVID-19 (7, 8). Therefore, a large number of researchers have designed many different deep learning models to assist clinic doctors in the rapid diagnosis of COVID-19 in CT images (9).

Traditional methods achieve the purpose of different semantic segmentation tasks by extracting features of the target objects (10–12), but the segmentation performance was not good enough. To overcome the problem, various deep learning frameworks have been proposed to effectively segment the target objects (13). It could be divided into four classifications, deep-supervised learning, semi-supervised learning, weakly supervised learning, and unsupervised learning approaches in COVID-19 lesion segmentation. Compared with traditional methods, the segmentation performance has been largely improved by deep learning networks (14–16).

Although deep learning frameworks can achieve a good performance in COVID-19 lesion segmentation, they required a lot of data and the corresponding data annotations (17). To solve the problem, data augmentation is one of the most common operations, which can generate large amount of data through rotation, scaling, clipping, and transposing. However, using only data augmentation may cause some lesions to be undetected. Therefore, many advanced strategies, such as transfer learning, multi-task approach, and attention mechanism have been proposed to improve the performance of COVID-19 lesion segmentation. Based on this theory, a multi-task approach was designed by Yazdekhnasty et al. to reach the purpose (18), it had a good performance in lacking data and model generalization. Using a different strategy, Wang et al. integrated with transfer learning, UNet model, and multi-task learning to improve the segmentation performance of COVID-19 lesions (19). Recently, an attention mechanism (19, 20) was employed

to make up for the shortcoming of partial information missed caused by convolution operation in deep learning networks. To further improve the segmentation performance, semi-supervised learning strategies were proposed to train mounts of pseudo annotations. Based on this strategy, Zhao et al. presented a randomly selected propagation strategy to improve the segmentation performance of COVID-19 lesions (21). Similarly, Abdel-Basset et al. proposed an innovative semi-supervised few-shot segmentation method for COVID-19 lesion segmentation from a few amounts of annotated lung CT images (22). Existing supervised and semi-supervised methods require mounts of voxel-based annotations in the training stage (23). Unfortunately, it is difficult for clinicians to precisely annotate COVID-19 lesions due to the complex structural changes and blurred boundary information (24). To overcome this problem, mounts of weakly supervised methods have been proposed to segment COVID-19 lesions. The advantage of the weakly supervised approaches is that it can replace the complicated COVID-19 lesion labels with simple ones for training. Based on this strategy, Yang et al. presented a weakly supervised method based on a generative adversarial network (GAN) to improve the accuracy of COVID-19 lesion segmentation (25). A generator was adopted to remove lesions and generate healthy slices from input images, while a discriminator was used to force the generator to generate more accurate results with mounts of image-level annotations. The method was improved by Laradji et al. (26), where they utilized two encoder-decoder frameworks with shared weights. The first one encoded the original images, and the point annotations were treated as the corresponding supervised term. While the second one encoded the original image with geometric transformation, and the outputs of the first one with geometric transformation were regarded as the corresponding supervised term. Similarly, Wu et al. proposed a new 3D active learning framework called COVID-AL to segment COVID-19 lesions with volume-annotations (27). Recently, Wang et al. proposed a weakly supervised deep learning framework for COVID-19 lesion segmentation (28). First, a pre-trained UNet is applied to remove unrelated tissues for lung segmentation. Subsequently, the segmented lung is fed into the designed DeCoVNet to acquire the COVID-19 lesion feature map. Finally, a class activation mapping algorithm and a 3D connected component algorithm were combined for COVID-19 lesion localization. The fatal flaw of the deep supervised approach, semi-supervised approach, and weakly supervised approach is that mounts of data labels are required to supervise the training model in the training stage. Whereas data annotations need clinical experts to spend a lot of time to annotate it. Different from the above deep learning approaches, the unsupervised approach has a good performance in objects segmentation without any annotated labels. To alleviate the burden of data annotation, Yao et al. designed an unsupervised NormNet model to distinguish COVID-19 lesions from complex lung tissues (29). Based on the observation that parts of tracheae and vessels exhibit strong patterns, a three-stage (random shape, noise generation, and image filtering) strategy was used to generate lesion shape for subsequent segmentation. Taking the difference between COVID-19 lesions and other tissues into consideration,

**TABLE 1** | Coronavirus disease 2019 (COVID-19) lesion segmentation with different methods.

	Methods	Advantages	Disadvantages
Deep-supervised learning	Yazdekhnasty et al. (18) Wang et al. (19) Gao et al. (20)	Good performance in sufficient data and model generalization	Mounts of voxel-based Data annotation
Semi-supervised learning	Zhao et al. (21) Abdel-Basset et al. (22)	Good performance in lacking data	Parts of voxel-based Data annotation
Weakly supervised learning	Yang et al. (25) Laradji et al. (26) Wu et al. (27) Wang et al. (28)	Class annotation	Poor segmentation in small lesions
Unsupervised learning	Yao et al. (29)	No data annotations	Bad performance

a novel method named NormNet based on generative adversarial networks was presented for COVID-19 lesion segmentation. Unfortunately, the unsupervised NormNet model had a bad performance than some supervised methods. The advantages and disadvantages of different types of deep learning methods are summarized in **Table 1**.

Motivated by the fact that different deep learning methods have their own unique advantages, these advantages can be fused using a deep-supervised ensemble learning network to improve the COVID-19 lesion segmentation results in CT images. Unfortunately, the specific deep learning framework may take up more time and space.

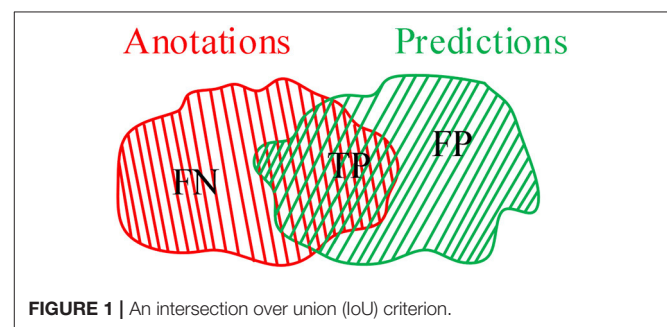
In this study, a deep-supervised ensemble learning network is proposed for COVID-19 lesion segmentation in CT images. To alleviate the overfitting problem on small datasets, a transfer learning strategy is used to acquire initialization parameters with better feature performance. Subsequently, an enumeration grid model is exploited to estimate the optimal weights for multiple deep learning model integration. In particular, we pay special attention to the COVID-19 lesion and its boundary segmentation, which can illustrate the effectiveness of the proposed method. Compared with several methods, the proposed model has a good performance in COVID-19 lesion segmentation in CT images.

## MATERIALS AND METHODS

In this study, we presented a deep-supervised ensemble learning network as an alternative model to segment COVID-19 lesions compared with several models in publicly available datasets. Visual inspection and quantitative evaluation were established in this study to verify the proposed ensemble learning network.

### Evaluation Criteria

To illustrate the validation of the proposed method, an IoU criterion and  $F_1$ -measure (10) are applied to verify the good performance. As shown in **Figure 1**, annotations A are divided into false negative (FN) and true positive (TP), whereas predictions B are divided into TP and false positive (FP). In which, TP is the common region between annotations A and



predictions B. Therefore, the mathematical description of IoU, precision (P), recall (R), and  $F_1$  can be defined as follows:

$$IoU = \frac{TP}{TP + FP + FN} \quad (1)$$

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (4)$$

In other words, the greater proportion of the common region between annotations and predictions, the greater IoU and  $F_1$  values will be. The smaller proportion of common region between annotations and predictions, the smaller IoU and  $F_1$  values will be. Especially in the case with small COVID-19 lesions, IoU has a strong ability to illustrate the effectiveness of the proposed method. Additionally, the  $F_1$ -measure criterion denotes the similarity between annotations and predictions.

As we all know that IoU and  $F_1$  pay more attention to the regional sensitivity of image segmentation, but the description of the segmentation boundary is also important. To further illustrate the effectiveness of the proposed method, Hausdorff distance is employed to evaluate the validation of the deep learning framework. The Hausdorff distance from annotations A to predictions B can be defined as

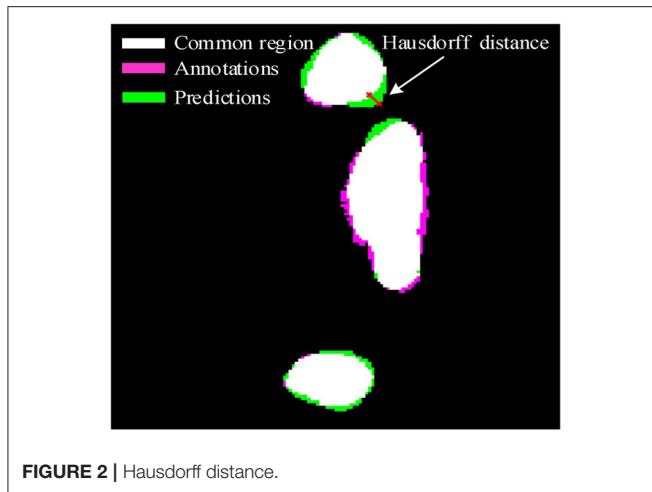


FIGURE 2 | Hausdorff distance.

$$h(A, B) = \max_{a \in A} \left\{ \min_{b \in B} \{d(a, b)\} \right\} \quad (5)$$

Where  $h(A, B)$  is the Hausdorff distance from A to B,  $d(a, b)$  is the distance between point a and point b. In which, point a and b are belonging to A and B, respectively. To illustrate the distance between A and B, a more general definition would be:

$$H(A, B) = \max \{h(A, B), h(B, A)\} \quad (6)$$

As shown in **Figure 2**, the pink area is the annotations A, the green area is the predictions B, and the white area is the common region covered by both predictions A and annotations B. Hausdorff distance measures the biggest distance from both predictions A and annotations B to common region. To avoid the influence of noise on Hausdorff distance, 95% Hausdorff distance (HD95) are treated as useful values sorted from small to large. To avoid Hausdorff distance being too large, a simple mathematical transformation is given as:

$$H = \sum_{i=1}^n \lg(HD95_i + 1) \quad (7)$$

Where  $i$  is the number of the test CT slices. In other words, the better detection of the COVID-19 lesion boundary, the smaller  $H$  value will be. Whereas the worse detection of the COVID-19 lesion boundary, the larger  $H$  value will be. The evaluation criteria IoU,  $F_1$ , and  $H$  are used to evaluate the validation of the proposed method.

## Data and Annotations

We employed a dataset from the China consortium of chest CT image investigation (30), which is a publicly available dataset. The dataset includes 150 CT scans. In which, 750 CT slices were selected from the dataset and annotated by four doctors with extensive clinical experience, 400 CT slices were treated as a training set, 200 CT slices were considered as a validation set, and 150 CT slices were used as a test set. In this work, we treated the corresponding annotations as the ground truth.

## Overview of the Proposed Method

In this article, we present a deep-supervised ensemble learning network for COVID-19 lesion segmentation in CT images in **Figure 3**. First, data augmentation is applied to increase the training data and improve the generalization ability of the model. Subsequently, a transfer learning strategy is employed to copy with small datasets and alleviate the overfitting problem. Finally, a deep-supervised ensemble learning network is presented to combine with local and global features for COVID-19 lesion segmentation in CT images.

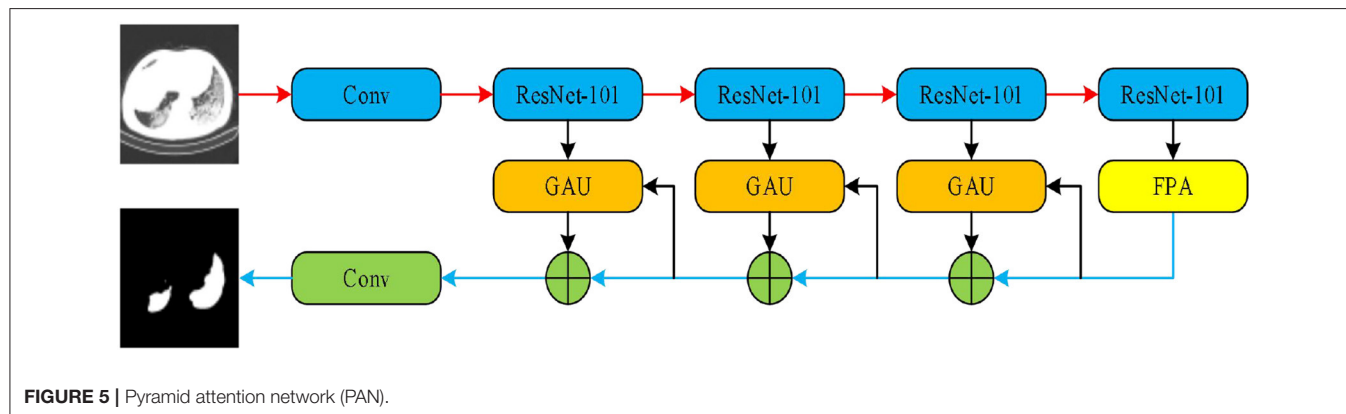
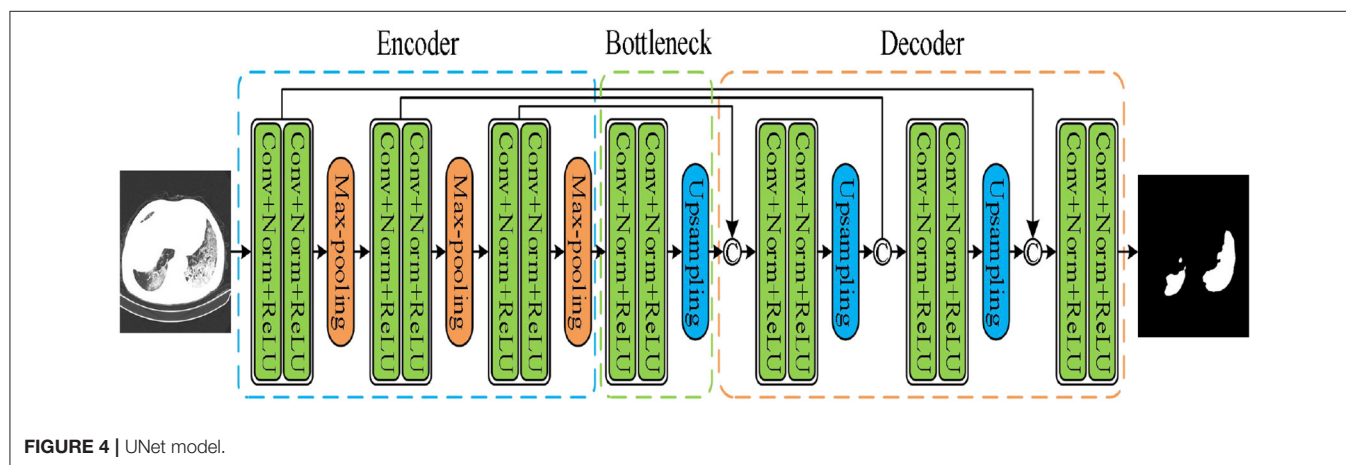
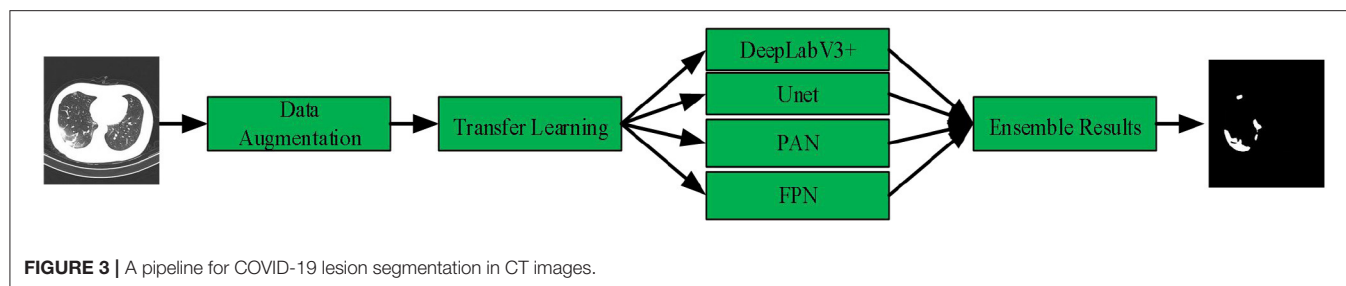
## Data Augmentation

Clinicians spend a great deal of time to annotate complex structures of COVID-19 lesions, which is too expensive. In general, data augmentation (31) is applied to process and increase the training data to make the data as diverse as possible and make the training model generalization ability stronger. To reach the purpose, CT image operations, such as horizontal flip, vertical flip, rotation, and scaling are adopted to increase training data and the corresponding COVID-19 lesion annotations. After data augmentation, the number of training data is varied from 400 to 4,000.

## Transfer Learning

Deep learning models have been widely applied in medical image processing (32). To acquire accurate COVID-19 lesion segmentation results, a large amount of CT images and manual annotations are required to adjust the parameters for the special deep learning model. However, it is too difficult to obtain mounts of CT images and the corresponding manual lesion annotations. In general, the designed deep learning models are trained with small datasets, which may lead to poor generalization ability and serious overfitting performance. To cope with the shortcomings, many technical means, such as data augmentation (31), multi-task learning (33), transfer learning (34), and attention mechanism (35) can be used to achieve the segmentation task. In this section, a transfer learning strategy is applied to solve the shortcomings.

It is a challenging task for doctors to manually annotate complex and variable COVID-19 lesions. To acquire the best parameters of the designed model with the limited COVID-19 lesion annotations, a transfer learning strategy is employed to accurately segment COVID-19 lesions in CT images. In this, the ImageNet dataset is treated as the pre-training dataset, which is one of the largest image datasets in the world (36). Whereas the EfficientNet model (37, 38) is considered as the pre-training model. First, the model is trained with the training data and the corresponding manual annotations to generate pseudo annotations, it was treated as the teacher in the “teacher-student” model. Whereas a powerful EfficientNet model is retrained by using manual and pseudo annotations, it was considered as the student in the “teacher-student” model. In the student learning stage, adding noise processing is used to make the generalization ability of students better than teachers (39). In this section, the estimated parameters are regarded as the initialized parameters for the deep-supervised ensemble learning network.

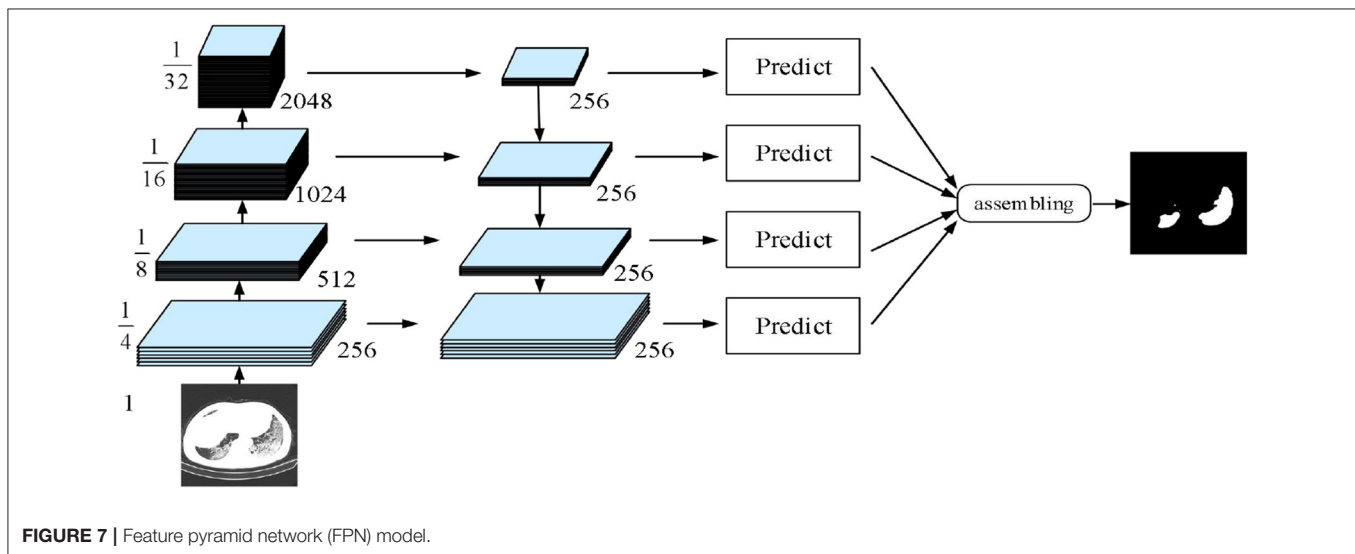
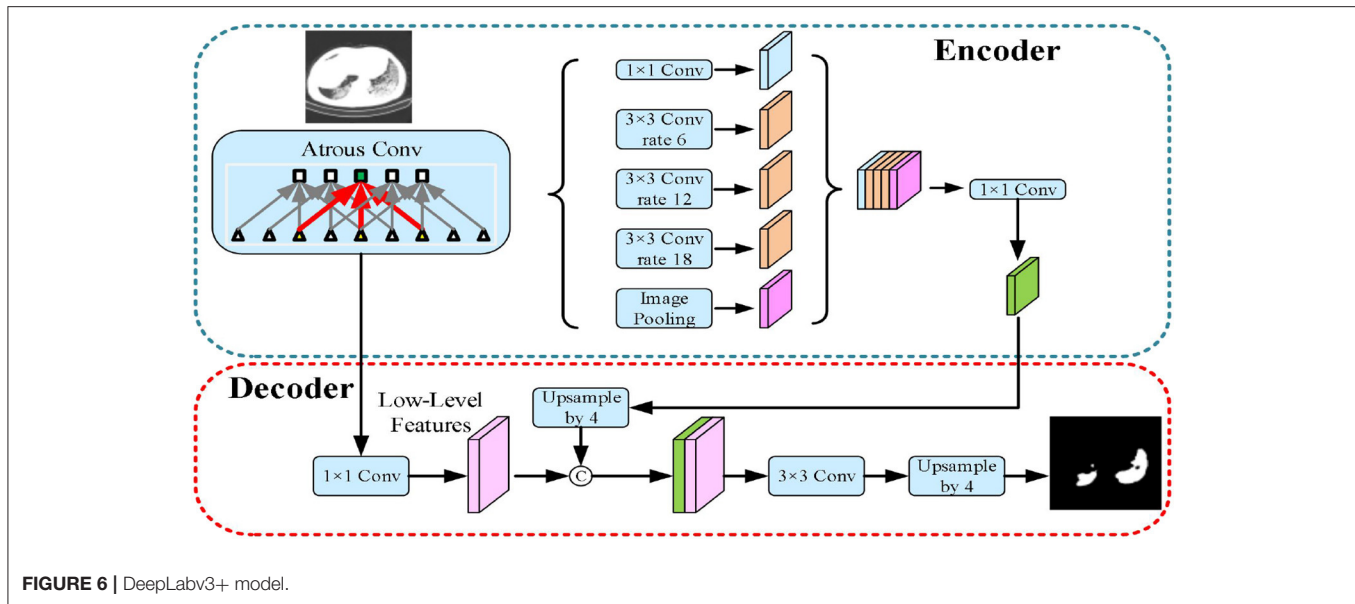


## Deep-Supervised Ensemble Learning Network

Different networks have different advantages and disadvantages, the advantages can be integrated together by effectively integrating various networks. To acquire the best COVID-19 lesions segmentation performance, a deep-supervised ensemble learning network is presented for COVID-19 segmentation in CT images.

The UNet model was applied according to Su et al. and Wang et al. (40, 41). As shown in **Figure 4**, the network had three encoding and decoding blocks, respectively. Encoding was designed using max-pooling, whereas the decoding was performed *via* a deconvolution. In addition, the encoder and the decoder were connected *via* skip connections.

As a second architecture, we implemented the pyramid attention network (PAN) by Li et al. (42). Unlike the traditional deep learning model, the basic principle of the pyramid attention network is to effectively extract local and global features of target objects by integrating the attention mechanism and spatial pyramid structures. As shown in **Figure 5**, an encoder-decoder scheme was adopted to locate target objects. In the encoder module, a feature pyramid attention (FPA) was introduced to the adopted spatial pyramid attention mechanism in the high-level output, and the global information was applied to learn stronger feature representation. In the decoder module, a global attention upsample (GAU) module was applied to extract the global information of the target objects to effectively segment COVID-19 lesions.



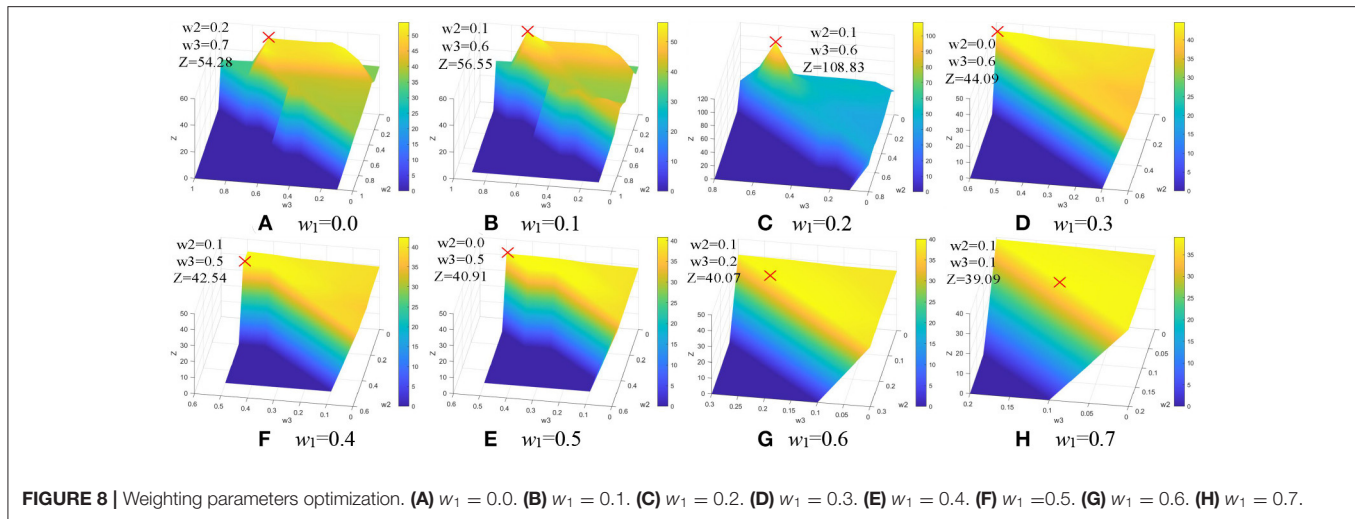
The third implemented architecture is the DeepLabv3+ by Chen et al. (43). As shown in **Figure 6**, it is the DeepLabv3 extended version by adding a decoder module to refine the segmentation results, especially along the novel coronavirus pneumonia lesion boundaries. Specifically, the Xception model was explored and the depthwise separable convolution was applied to both atrous spatial pyramid pooling and decoder modules, resulting in a faster and stronger encoder-decoder network.

The fourth implemented architecture is the multi-scale feature pyramid network (FPN) by Lin et al. (44). As shown in **Figure 7**, a top-down architecture with skip connections was designed to express high-level feature maps at all scales. Subsequently, the semantic feature maps with different scales were integrated to improve the segmentation performance of COVID-19.

In these deep learning networks, training was performed using the Adam optimizer with a learning rate of  $10^{-5}$ , the dice loss was considered as the loss function to express the relationship between the predicted probabilities and the corresponding lesion annotations. Inspired by the previous work of Golla et al. (45), we present an ensemble module to ensemble the probability feature maps of four networks. To give a mathematical expression:

$$E = w_1 \times E_1 + w_2 \times E_2 + w_3 \times E_3 + w_4 \times E_4 \quad (8)$$

Where  $w_1$ ,  $w_2$ ,  $w_3$ , and  $w_4$  are the weighting parameters,  $E_1$ ,  $E_2$ ,  $E_3$ , and  $E_4$  represent the predicted probabilities of PAN, FPN, Unet, and Deeplabv3+ networks. In which, the relationship among the weighting parameters  $w_1$ ,  $w_2$ ,  $w_3$ , and  $w_4$  can be represented as follows:

**TABLE 2 |** The maximum values of Z.

Maximum points				Maximum values
	$w_3$	$w_4$	$Z$	
0.0	0.2	0.7	0.1	54.28
0.1	0.1	0.6	0.2	56.55
0.2	0.1	0.6	0.1	<b>108.83</b>
0.3	0.0	0.6	0.1	44.09
0.4	0.1	0.5	0	42.54
0.5	0.0	0.5	0	40.91
0.6	0.1	0.2	0.1	40.07
0.7	0.1	0.1	0.1	39.09
0.8	0	0.1	0.1	37.40
0.9	0	0	0.1	35.81
1.0	0	0	0	35.08

$Z = 108.83$ , is the largest values. It illustrates that the proposed method has the best segmentation performance in this case.

**TABLE 3 |** Architecture parameters with different networks.

Parameters	Value
Input image size	$512 \times 512$
Output image size	$512 \times 512$
Learning rate	$10^{-5}$
Activation layers	Adam
Epochs	300
Batch size	4
Loss function	Dice

$$w_1 + w_2 + w_3 + w_4 = 1 \quad (9)$$

To acquire the most effective segmentation results of COVID-19 lesions in CT images, an enumeration grid model (46) is used to achieve the optimal weighting parameters.

Weighting parameters play an important role in COVID-19 lesion segmentation. To overcome the problem, a simple but

effective approach is designed to acquire the optimal weighting parameters with the enumeration grid model, which is a traversal method, it has the ability to enumerate all the parameters  $w_1$ ,  $w_2$ ,  $w_3$ , and  $w_4$ . First,  $w_1$  is set to be a fixed value varied from 0.0 to 1.0, and  $w_2$  and  $w_3$  are treated as variable values. As we all know that the sum of all the weighting parameters is 1.0, the weighting parameter  $w_4$  is  $1 - w_2 - w_3$ . In other words,  $w_4$  is defined by variable  $w_2$  and  $w_3$ . As a result, all the weighting parameters and the corresponding quantitative index IOU can be acquired with the enumeration grid model. Since the IoU value difference is very small, it is difficult to distinguish. To better distinguish the quantitative index IoU for a good presentation, a simple mathematical transformation is given as:

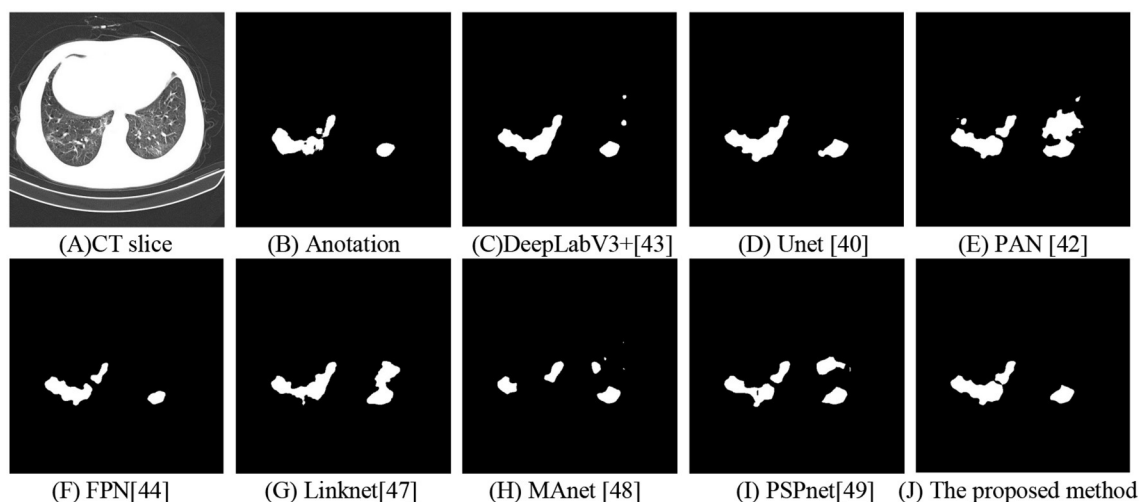
$$Z = \log_{9/10}^{|IoU - \max(IoU) - 0.0001|} \quad (10)$$

As shown in **Figure 8**, the maximum values of Z are marked with a red cross and the corresponding values are described in **Table 2**. The optimal weighting parameters are 0.2, 0.1, 0.6, and 0.1.

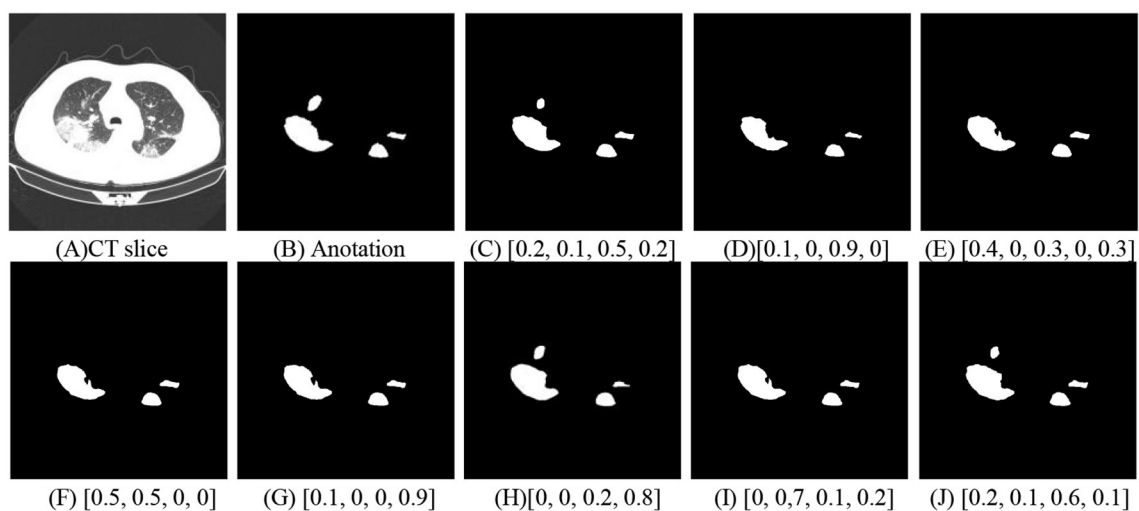
**Table 3** lists the architecture parameters with different networks. The dice loss function was regarded as the loss function. Adaptive moment estimation (Adam) was employed for the training process, which iteratively updates different network weights based on a publicly available novel coronavirus pneumonia dataset. The learning rate was initialized with  $10^{-5}$ . The above deep learning model was implemented in python using PyTorch with Lenovo Ren-9000 34IMZ, GPU GFX 2060, and CPU 32G.

## RESULTS

We present a deep-supervised ensemble learning network as an alternative model to segment COVID-19 lesions in CT images. The segmentation performance of the proposed method is validated in experiments with a publicly available dataset. The effectiveness of the proposed method was verified by visual inspection and quantitative evaluation.



**FIGURE 9 |** Segmentation of COVID-19 lesions with different deep learning methods. (A) CT slice. (B) Annotation. (C) DeepLabV3+. (D) Unet. (E) PAN. (F) FPN. (G) Linknet. (H) MAnet. (I) PSPnet. (J) The proposed method.



**FIGURE 10 |** COVID-19 lesion segmentation with different weighting parameters. (A) CT slice. (B) Annotation. (C) [0.2, 0.1, 0.5, 0.2]. (D) [0.1, 0, 0.9, 0]. (E) [0.4, 0, 0.3, 0, 0.3]. (F) [0.5, 0.5, 0, 0]. (G) [0.1, 0, 0, 0.9]. (H) [0, 0, 0.2, 0.8]. (I) [0, 0.7, 0.1, 0.2]. (J) [0.2, 0.1, 0.6, 0.1].

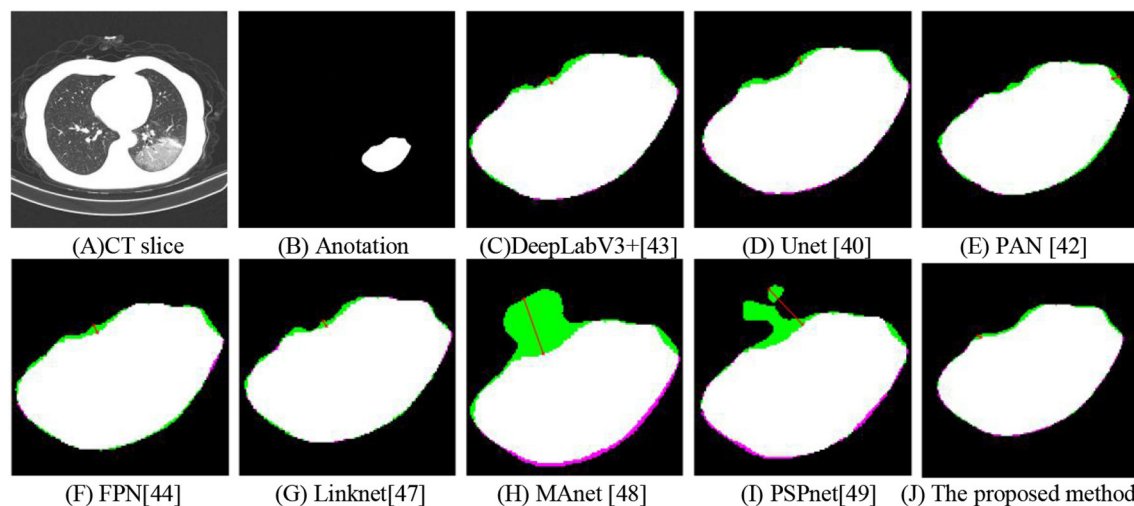
## Visual Inspection

For visual inspection, we selected a typical CT slice from a public dataset for demonstration. The CT slice, the corresponding annotation, DeepLabV3+ (43), UNet (40), PAN (42), FPN (44), Linknet (47), MAnet (48), PSPnet (49), and the proposed method are displayed in **Figure 9**. In this, many methods (40, 43, 48, 49) used the local features to segment COVID-19 lesions, the approaches may cause parts of small lesions to be undetected. While parts of methods (42, 44, 47) exploited the local and global features for COVID-19 lesion segmentation, the strategy may cause parts of clutters to be unremoved. On the contrary, a deep-supervised ensemble learning network is presented to combine with the advantages of different deep learning networks (40, 42–44) for COVID-19 lesion segmentation. As observed, the

proposed method has largely improved the segmentation results compared with seven deep learning networks (40, 42–44, 47–49). In other words, by using the deep-supervised ensemble learning network, the proposed method has a good performance in COVID-19 lesion segmentation.

To accurately segment novel coronavirus pneumonia lesions, the enumeration method is applied to estimate the best weighting parameters  $w_1$ ,  $w_2$ ,  $w_3$ , and  $w_4$ . In **Figure 10**, COVID-19 lesion segmentation results with different weighting parameters are displayed to evaluate the validation of the proposed method. It can be seen that the best weighting parameters are 0.2, 0.1, 0.6, and 0.1.

To investigate the effect of the proposed method in COVID-19 lesion segmentation, a CT slice is chosen in **Figure 11A** and



**FIGURE 11 |** Hausdorff distance with different methods. (A) CT slice. (B) Annotation. (C) DeepLabV3+. (D) Unet. (E) PAN. (F) FPN. (G) Linknet. (H) MAnet. (I) PSPnet. (J) The proposed method.

**TABLE 4 |** The IoU values with different methods.

Method	IoU	F <sub>1</sub>	H
DeepLabV3+ (43)	0.7058	0.7886	99.8686
Unet (40)	0.6927	0.7720	96.9404
PAN (42)	0.7081	0.7931	97.6020
FPN (44)	0.7031	0.7881	98.9161
Linknet (47)	0.6883	0.7618	101.9168
MAnet (48)	0.6067	0.7216	112.6191
PSPnet (49)	0.6696	0.7557	104.5176
The proposed method	<b>0.7279</b>	<b>0.8065</b>	<b>92.4604</b>

Compared with seven methods, the proposed method has the largest IoU value 0.7279 and F1 value 0.8065, which means that the proposed method has the best performance in regional sensitivity of COVID-19 lesion segmentation. What's more, the proposed method has the smallest H value 92.4604, which indicates that the computational model has the best performance in COVID-19 lesion boundary.

the corresponding annotation data are shown in **Figure 11B**. To better illustrate the effectiveness of the proposed method, the detected COVID-19 lesions with different methods were magnified. As shown in **Figure 11**, the pink region is the annotations, the green region is the predictions, and the white area is the common region covered by both predictions and annotations. In **Figures 11C–J**, the Hausdorff distance with different methods are 3.6056, 3.1623, 4.1231, 4.4721, 3.6056, 32.8938, 27.5862, and 2.2361. In other words, the proposed method has a lower Hausdorff distance than these typical methods (40, 42–44, 47–49).

## Quantitative Evaluation

We employed a dataset from the China consortium of chest CT image investigation, the dataset includes 750 CT images. As observed in **Table 4**, The IoU values corresponding to the DeepLabV3+ (43), Unet (40), PAN (42), FPN (44), Linknet

**TABLE 5 |** The relationship between the weighting parameters and the IoU index.

w <sub>1</sub>	w <sub>2</sub>	w <sub>3</sub>	w <sub>4</sub>	IoU
0.1	0.0	0.9	0.0	0.6929
0.4	0.3	0.0	0.3	0.7083
0.0	0.7	0.1	0.2	0.7101
0.4	0.3	0.1	0.2	0.7108
0.7	0.1	0.1	0.1	0.7116
0.1	0.1	0.2	0.6	0.7232
0.2	0.3	0.4	0.1	0.7219
0.3	0.0	0.1	0.6	0.7150
0.5	0.1	0.2	0.2	0.7142
0.6	0.0	0.2	0.2	0.7120
0.8	0.0	0.0	0.2	0.7075
0.0	0.1	0.8	0.1	0.6951
0.2	0.0	0.7	0.1	0.7252
0.0	0.1	0.6	0.3	0.7229
0.2	0.1	0.6	0.1	<b>0.7279</b>

Compared with different weighting parameters, the best weighting parameters are 0.2, 0.1, 0.6, and 0.1 and the maximum IoU value is 0.7279.

(47), MAnet (48), PSPnet (49), and the proposed method are 0.7058, 0.6927, 0.7031, 0.7081, 0.6883, 0.6067, 0.6696, and 0.7279. Whereas the F<sub>1</sub> values corresponding to different methods are 0.7886, 0.7720, 0.7881, 0.7931, 0.7618, 0.7216, 0.7557, and 0.8065, and the Hausdorff distance with different methods are 99.8686, 96.9404, 97.6020, 98.9161, 101.9168, 112.6191, 104.5176, and 92.4604. Both visual inspection and quantitative evaluation exhibited that our method can outperform these typical methods (40, 42–44, 47–49).

The relationship between the weighting parameters and the IoU index is shown in **Table 5**. Both visual inspection and quantitative evaluation exhibited that the best weighting parameters are 0.2, 0.1, 0.6, and 0.1.

## DISCUSSION

In this article, a deep-supervised ensemble learning network is presented for COVID-19 segmentation in CT images, the proposed method has many specific characteristics and advantages. Based on the fact that a large number of COVID-19 CT images and the corresponding lesion annotations are difficult to be obtained, a transfer learning strategy is employed to make up for the shortcoming and alleviate the overfitting problem. Second, a unique ensemble module is presented to improve the segmentation performance. While many studies use only one neural network to segment COVID-19 lesions, which cannot effectively discriminate COVID-19 lesions and other unrelated structures in CT images. Third, the proposed method is expected to preserve the completeness of COVID-19 lesions while maximally eliminating the unrelated structures. Last, the proposed method has a good performance in COVID-19 lesion segmentation in CT images.

The proposed method was validated in a publicly available dataset from the China consortium of chest CT image investigation. Both visual inspection and quantitative evaluation exhibited that the proposed approach could outperform these typical deep learning networks in COVID-19 lesion segmentation (40, 42–44, 47–49). Compared with manually defined annotations, our methods obtained a higher accuracy in COVID-19 lesion segmentation with an IoU index of 0.7279 and an  $F_1$  value of 0.8065 than these typical methods. In which, many typical methods (40, 43, 48, 49) used the local features to segment COVID-19 lesions, the approaches may cause parts of small lesions to be undetected. While parts of methods (42, 44, 47) exploited the local and global features for COVID-19 lesion segmentation, the strategy may cause parts of clutters to be unremoved. On the contrary, a deep-supervised ensemble learning network is presented to combine with the advantages of different deep learning networks (40, 42–44) for COVID-19 lesion segmentation. In other words, the proposed method uses weight parameters to measure the importance of local and global features for COVID-19 lesion segmentation. While the compared methods used only one neural network to segment COVID-19 lesions, it may cause parts of COVID-19 lesions to be undetected.

Compared with these conventional neural networks (40, 42–44, 47–49), the proposed method appears more efficient on COVID-19 lesion segmentation. This is ascribed to a well-designed fusion of transfer learning strategy, data augmentation, and multiple neural networks-ensemble approaches. In other words, the proposed method outperforms the conventional methods in that the merits of local and global features are efficiently combined. In addition, the segmentation of COVID-19 lesions has important clinical research significance. It can help doctors to diagnosis COVID-19 and develop the best treatment plan.

However, the designed deep-supervised ensemble learning network may take up more time and space than traditional

conventional neural networks (40, 42–44, 47–49). Additionally, based on the fact that a large number of COVID-19 CT images and the corresponding lesion annotations are difficult to be obtained, accurate segmentation of small coronavirus pneumonia lesions is still a long way off.

In conclusion, a deep-supervised ensemble learning network is presented for coronavirus pneumonia lesion segmentation in CT images. Based on the reality that mounts of COVID-19 CT images and the corresponding lesion annotations are difficult to acquire, a transfer learning strategy is used to alleviate the overfitting problem in a small dataset. Another contribution of the proposed method concerns the deep-supervised ensemble learning network. Using a single deep learning network, the accuracy of COVID-19 lesion segmentation results cannot reach a satisfactory performance. To overcome the problem, an ensemble strategy is presented to integrate multiple deep learning networks for COVID-19 lesion and its boundary segmentation in CT images. Experimental results indicated that our proposed deep-supervised ensemble learning model has a good performance in COVID-19 lesion and its boundary segmentation in CT images.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

YP was the major contributor in writing the manuscript. All authors analyzed the data, contributed to the article, and approved the submitted version.

## FUNDING

This research was funded by the Jiangxi Provincial Natural Science Foundation (Nos. 20212BAB202007, 20202BAB212004, 20212BAB211009, 20204BCJL23035, 20192ACB21004, and 20181BAB202017), by the Educational Science Research Project of China Institute of communications Education (No. JTYB20-33), by the Huxiang high-level talent gathering project (No. 2019RS1072), by the Scientific and Technological Research Project of Education Department in Jiangxi Province (No. GJJ190356), and the Science and Technology project of Changsha City (No. kq2001014).

## ACKNOWLEDGMENTS

We are grateful to the China consortium of chest CT image investigation for COVID-19 data sharing.

## REFERENCES

- Alon D, Paitan Y, Robinson E, Ganor N, Lipovetsky J, Yerushalmi R, et al. Downregulation of CD45 Signaling in COVID-19 patients is reversed by C24D, a novel CD45 targeting peptide. *Front Med.* (2021) 8:1251. doi: 10.3389/fmed.2021.675963
- Murphy A, Pinkerton LM, Bruckner E, Risser HJ. The impact of the novel coronavirus disease 2019 on therapy service delivery for children with disabilities. *J Pediatr.* (2021) 231:168–77. doi: 10.1016/j.jpeds.2020.12.060
- Comunale BA, Engineer L, Jiang Y, Andrews JC, Liu Q, Ji L, et al. Poliovirus vaccination induces a humoral immune response that cross reacts with SARS-CoV-2. *Front Med.* (2021) 8:1285. doi: 10.3389/fmed.2021.710010
- Li L, Cao Y, Fan J, Li T, Lang J, Zhang H, et al. Impact of COVID-19 pandemic on the clinical activities in obstetrics and gynecology: a national survey in China. *Front Med.* (2021) 8:1225. doi: 10.3389/fmed.2021.633477
- Meng Z, Wang T, Chen L, Chen X, Li L, Qin X, et al. The effect of recombinant human interferon alpha nasal drops to prevent COVID-19 pneumonia for medical staff in an epidemic area. *Curr Top Med Chem.* (2021) 21:920–7. doi: 10.2174/1568026621666210429083050
- Herpe G, Lederlin M, Naudin M, Obana M, Chaumoitre K, Gregory J, et al. Efficacy of chest CT for COVID-19 pneumonia diagnosis in France. *Radiology.* (2021) 298:81–7. doi: 10.1148/radiol.2020202568
- Zhao Y, Ji D, Li Y, Zhao X, Lv W, Xin X, et al. Three-dimensional visualization of microvasculature from few-projection data using a novel CT reconstruction algorithm for propagation-based X-ray phase-contrast imaging. *Biomed Opt Express.* (2020) 11:364–87. doi: 10.1364/BOE.380084
- Schalekarp S, Bleeker-Rovers CP, Beenen LFM, Quarles van Ufford HME, Gietema HA, Stoger JL, et al. Chest CT in the emergency department for diagnosis of COVID-19 pneumonia: dutch experience. *Radiology.* (2021) 298:98–106. doi: 10.1148/radiol.2020203465
- Bhargava A, Bansal A. Novel coronavirus (COVID-19) diagnosis using computer vision and artificial intelligence techniques: a review. *Multimed Tools Appl.* (2021) 80:19931–46. doi: 10.1007/s11042-021-10714-5
- Peng Y, Xiao C. An oriented derivative of stick filter and post-processing segmentation algorithms for pulmonary fissure detection in CT images. *Biomed Signal Proces.* (2018) 43:278–88. doi: 10.1016/j.bspc.2018.03.013
- Peng Y, Zhong H, Xu Z, Tu H, Li X, Peng L. Pulmonary lobe segmentation in CT images based on lung anatomy knowledge. *Math Probl Eng.* (2021) 2021:5588629. doi: 10.1155/2021/5588629
- Zhang Y, Hu Y, Zhao S, Cui C. The utility of PET/CT metabolic parameters measured based on fixed percentage threshold of SUVmax and adaptive iterative algorithm in the new revised FIGO staging system for stage III cervical cancer. *Front Med.* (2021) 8:1189. doi: 10.3389/fmed.2021.680072
- Chen S, Chen J, Yang Y, Chien C, Wang M, Lin L. Use of radiographic features in COVID-19 diagnosis: challenges and perspectives. *J Chin Med Assoc.* (2020) 83:644–7. doi: 10.1097/JCMA.0000000000000336
- Dastidar TR, Ethirajan R. Whole slide imaging system using deep learning-based automated focusing. *Biomed Opt Express.* (2020) 11:480–91. doi: 10.1364/BOE.379780
- Ali MJ, Hanif M, Haider MA, Ahmed MU, Sundas F, Hirani A, et al. Treatment options for COVID-19: a review. *Front Med.* (2020) 7:480. doi: 10.3389/fmed.2020.00480
- Hammoudi K, Benhabiles H, Melkemi M, Dornaika F, Arganda-Carreras I, Collard D, et al. Deep learning on chest X-ray images to detect and evaluate pneumonia cases at the ear of COVID-19. *J Med Syst.* (2021) 45:75. doi: 10.1007/s10916-021-01745-4
- Erion G, Janizek JD, Sturmfels P, Lundberg SM, Lee S. Improving performance of deep learning models with axiomatic attribution priors and expected gradients. *Nat Mach Intell.* (2021) 37:1–12. doi: 10.1038/s42256-021-00343-w
- Yazdekhnasty P, Zindar A, Nabizadeh-ShahreBabak Z, Roshandel R, Khadivi P, Karimi N, et al. Bifurcated autoencoder for segmentation of COVID-19 infected regions in CT images. arXiv preprint 2011. (2020) p. 00631. doi: 10.1007/978-3-030-68790-8\_46
- Wang Y, Zhang Y, Liu Y, Tian J, Zhong C, Shi Z, et al. Does non-COVID-19 lung lesion help? Investigating transferability in COVID-19 CT image segmentation. *Comput Meth Prog Bio.* (2021) 202:106004. doi: 10.1016/j.cmpb.2021.106004
- Gao K, Su J, Jiang Z, Zeng L, Feng Z, Shen H, et al. Dual-branch combination network (DCN): Towards accurate diagnosis and lesion segmentation of COVID-19 using CT images. *Med Image Anal.* (2021) 67:101836. doi: 10.1016/j.media.2020.101836
- Zhao S, Li Z, Chen Y, Zhao W, Xie X, Liu J, et al. SCOAT-Net: A novel network for segmentation COVID-19 lung opacification from CT images. *Pattern Recogn.* (2021) 119:108109. doi: 10.1016/j.patcog.2021.108109
- Abdel-Basset M, Chang V, Hawash H, Chakraborty RK, Ryan M. FSS-2019-nCov: a deep learning architecture for semi-supervised few-shot segmentation of COVID-19 infection. *Knowl-Based Syst.* (2021) 212:106647. doi: 10.1016/j.knsys.2020.106647
- Yang D, Xu Z, Li W, Myronenko A, Roth HR, Harmon S, et al. Federated semi-supervised learning for COVID region segmentation in Chest CT using multi-national data from China, Italy, Japan. *Med Image Anal.* (2021) 70:101992. doi: 10.1016/j.media.2021.101992
- Piccolo V, Neri I, Filippeschi C, Oranges T, Argenziano G, Battarra VC, et al. Chilblain-like during COVID-19 epidemic: a preliminary study on 63 patients. *J Eur Acad Dermatol.* (2020) 34:e291–e345. doi: 10.1111/jdv.16526
- Yang Y, Chen J, Wang R, Ma T, Wang L, Chen J, et al. Towards unbiased COVID-19 lesion localisation and segmentation via weakly supervised learning. In: IEEE 18th International Symposium on Biomedical Imaging. (2021). p. 1966–70. doi: 10.1109/ISBI48211.2021.9433806
- Laradji I, Rodriguez P, Manas O, Lensink K, Law M, Kurzman L, et al. A weakly supervised consistency-based learning method for COVID-19 segmentation in CT images. In: IEEE Winter Conference on Applications of Computer Vision. (2021). p. 2453–62. doi: 10.1109/WACV48630.2021.00250
- Wu X, Chen C, Zhong M, Wang J, Shi J, COVID-AL. the diagnosis of COVID-19 with deep active learning. *Med Image Anal.* (2021) 68:101913. doi: 10.1016/j.media.2020.101913
- Wang X, Deng X, Fu Q, Zhou Q, Feng J, Ma H, et al. A weakly supervised framework for COVID-19 classification and lesion localization from chest CT. *IEEE Trans Med Imaging.* (2020) 39:2615–25. doi: 10.1109/TMI.2020.2995965
- Yao Q, Xiao L, Liu P, Zhou SK. Label-free segmentation of COVID-19 lesions in lung CT. *IEEE Trans Med Imaging.* (2021) 40:2808–19. doi: 10.1109/TMI.2021.3066161
- Zhang K, Liu X, Shen J, Li Z, Sang Ye, Wu X, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell.* (2020) 181:1423–33. doi: 10.1016/j.cell.2020.04.045
- Zhu Y, Yeung CH, Lam EY. Digital holographic imaging and classification of microplastics using deep transfer learning. *Appl Optics.* (2021) 60:38–47. doi: 10.1364/AO.403366
- Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X. Deep learning in medical image registration: a review. *Phys Med Biol.* (2020) 65:20TR01. doi: 10.1088/1361-6560/ab843e
- Du S, Du J, Tang Y, Ouyang H, Tao Z, Jiang T. Achieving efficient inverse design of low-dimensional heterostructures based on a vigorous scalable multi-task learning network. *Opt Express.* (2021) 29:19727–42. doi: 10.1364/OE.426968
- Christensen CN, Ward EN, Lu M, Lio P, Kaminski CF, ML-SIM. universal reconstruction of structured illumination microscopy images using transfer learning. *Biomed Opt Express.* (2021) 12:2720–33. doi: 10.1364/BOE.414680
- Wang SH, Fernandes S, Zhu Z, Zhang YD. AVNC: attention-based VGG-style network for COVID-19 diagnosis by CBAM. *IEEE Sens J.* (2021). doi: 10.1109/JSEN.2021.3062442
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun Acn.* (2017) 60:84–90. doi: 10.1145/3065386
- Marques G, Agarwal D, Diez IDLT. Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. *Appl Soft Comput.* (2020) 96:106691. doi: 10.1016/j.asoc.2020.106691
- Munien C, Viriri S. Classification of hematoxylin and eosin stained breast cancer histology microscopy images using transfer learning with EfficientNets. *Comput Intel Neurosc.* (2021) 2021:5580914. doi: 10.1155/2021/5580914
- Xie Q, Luong M, Hovy E, Le QV. Self-training with noisy student improves ImageNet classification. In: CVPR. (2020) doi: 10.1109/CVPR42600.2020.01070

40. Su R, Zhang D, Liu J, Cheng C. MSU-Net: Multi-scale U-Net for 2D medical image segmentation. *Front Genet.* (2021) 12:140. doi: 10.3389/fgene.2021.639930
41. Wang B, Yang J, Ai J, Luo N, An L, Feng H, et al. Accurate tumor segmentation via octave convolution neural network. *Frontiers in Medicine.* (2021) 8:653913. doi: 10.3389/fmed.2021.653913
42. Li H, Xiong P, An J, Wang L. Pyramid attention network for semantic segmentation. arXiv preprint (2018). p. 1805.
43. Chen L C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *ECCV*, (2018). p. 801–18. doi: 10.1007/978-3-030-01234-2\_49
44. Lin TY, Dollar P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: *CVPR*, (2017). p. 2117–25. doi: 10.1109/CVPR.2017.106
45. Golla AK, Bauer DF, Schmidt R, Russ T, Norenberg D, Chung K, et al. Convolutional neural network ensemble segmentation with ratio-based sampling for the arteries and veins in abdominal CT scans. *IEEE Trans Bio-med Eng.* (2021) 68:1518–26. doi: 10.1109/TBME.2020.3042640
46. Hartley MG, Ralph E, Norville IH, Prior JL, Atkins TP. Comparison of PCR and viable count as a method for enumeration of bacteria in an A/J mouse aerosol model of Q fever. *Front Microbiol.* (2019) 10:1552. doi: 10.3389/fmicb.2019.01552
47. Chaurasia A, Culurciello E. Linknet: exploiting encoder representations for efficient semantic segmentation. In: *VCIP*. (2017). p. 1–4. doi: 10.1109/VCIP.2017.8305148
48. Fan T, Wang G, Li Y, Wang H. MA-Net: A Multi-Scale attention network for liver and tumor segmentation. *IEEE Access.* (2020) 8:179656–65. doi: 10.1109/ACCESS.2020.3025372
49. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: *CVPR*. (2017). p. 2881–90. doi: 10.1109/CVPR.2017.660

**Conflict of Interest:** HT was employed by company Technique Center, Hunan Great Wall Technology Information Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Peng, Zhang, Tu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# MPMR: Multi-Scale Feature and Probability Map for Melanoma Recognition

Dong Zhang<sup>1,2†</sup>, Hongcheng Han<sup>1,3†</sup>, Shaoyi Du<sup>1†</sup>, Longfei Zhu<sup>4</sup>, Jing Yang<sup>3</sup>, Xijing Wang<sup>1</sup>, Lin Wang<sup>5\*</sup> and Meifeng Xu<sup>4\*</sup>

<sup>1</sup> Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China, <sup>2</sup> School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, China, <sup>3</sup> School of Software Engineering, Xi'an Jiaotong University, Xi'an, China, <sup>4</sup> Dermatology Department, Second Affiliated Hospital of Xi'an Jiaotong University (Xibei Hospital), Xi'an, China, <sup>5</sup> School of Information Science and Technology, Northwest University, Xi'an, China

## OPEN ACCESS

### Edited by:

Jun Feng,  
Northwest University, China

### Reviewed by:

Qiang Yan,  
Taiyuan University of Technology,  
China  
Jialin Peng,  
Huaqiao University, China

### \*Correspondence:

Lin Wang  
wanglin@nwu.edu.cn  
Meifeng Xu  
xumf96@163.com

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

Received: 14 September 2021

Accepted: 08 December 2021

Published: 05 January 2022

### Citation:

Zhang D, Han H, Du S, Zhu L, Yang J,  
Wang X, Wang L and Xu M (2022)  
MPMR: Multi-Scale Feature and  
Probability Map for Melanoma  
Recognition. *Front. Med.* 8:775587.  
doi: 10.3389/fmed.2021.775587

Malignant melanoma (MM) recognition in whole-slide images (WSIs) is challenging due to the huge image size of billions of pixels and complex visual characteristics. We propose a novel automatic melanoma recognition method based on the multi-scale features and probability map, named MPMR. First, we introduce the idea of breaking up the WSI into patches to overcome the difficult-to-calculate problem of WSIs with huge sizes. Second, to obtain and visualize the recognition result of MM tissues in WSIs, a probability mapping method is proposed to generate the mask based on predicted categories, confidence probabilities, and location information of patches. Third, considering that the pathological features related to melanoma are at different scales, such as tissue, cell, and nucleus, and to enhance the representation of multi-scale features is important for melanoma recognition, we construct a multi-scale feature fusion architecture by additional branch paths and shortcut connections, which extracts the enriched lesion features from low-level features containing more detail information and high-level features containing more semantic information. Fourth, to improve the extraction feature of the irregular-shaped lesion and focus on essential features, we reconstructed the residual blocks by a deformable convolution and channel attention mechanism, which further reduces information redundancy and noisy features. The experimental results demonstrate that the proposed method outperforms the compared algorithms, and it has a potential for practical applications in clinical diagnosis.

**Keywords:** malignant melanoma, whole slide image, multi-scale feature, probability map, neural networks

## 1. INTRODUCTION

Malignant melanoma (MM) is a highly aggressive form of skin cancer whose incidence continues to increase at a great rate worldwide (1). It is characterized by an extraordinary metastasis capacity and chemotherapy resistance, and the difficulty of effective treatment increases with its continually developing aggression. Therefore, early diagnosis is essential to improve the survival rate of MM patients. Pathological examination is the gold standard for the diagnosis of MM (2), which enables the most reliable diagnosis based on pathological features at the cell level compared to other methods. Tissue cut from the lesion on the skin is made into pathological slices and scanned by a Digital Pathology Microscope Slide Scanner to get a whole-slide image (WSI). Through the WSI,

the pathologist finds out the property of the tissue and marks the MM region, if it exists, to measure related pathological indicators, such as lesion size, invasion depth, etc., which provide an important reference for treatment planning and surgical prognosis (3).

Analyzing WSIs is a challenging task (4). Even an experienced pathologist spends an average of 10–20 min recognizing the region of MM in a WSI, of which identifying the MM region takes up much time. First, a WSI has billions of pixels, and the physician needs to perform a scanned screening of the pathology images in a zoomed-in window. Second, the complex visual characteristics of the skin lesions, such as irregular-shaped texture, fuzzy boundaries, etc., increase the difficulty of recognition. Some MM tissues are hard to distinguish from some benign tissues (5), which is a challenge for MM recognition. These problems aggravate the work burden of pathologists, affecting the efficiency of pathological examination. Third, the difficulty in training and scarcity of pathologists, as well as the uneven distribution of medical resources, make it difficult to obtain a timely and accurate diagnosis for every melanoma patient. Therefore, there is an urgent need for an effective method for automatic MM recognition in WSIs.

MM region screening in WSIs is an image recognition task that utilizes computer vision. Since convolutional neural networks (CNNs) have provided state-of-the-art image classification and segmentation performance, medical image analysis methods based on CNNs have been developed. The U-Net proposed by Ronneberger et al. (6) and its derivative improved networks (7–10) have achieved considerable success in medical image segmentation in recent years. However, pixel-wise image segmentation methods have limitations in MM region recognition in WSIs. The huge size of WSIs poses problems to the computation of the network. Some MM recognition methods based on deep learning are proposed. For example, Hekler et al. (11) trained a CNN based on ResNet-50 (12) to realize the classification of histopathological images of melanomas and nevi with an accuracy of 81%. The limited feature extraction capabilities of ResNet make it challenging to achieve higher accuracy. Wang et al. (13) used a deep CNN to establish a diagnosis model through the patch of eyelid melanoma histopathological slides and obtained good results. Yu et al. (14) proposed a method for melanoma recognition by leveraging very deep CNNs and constructed a fully convolutional residual network for accurate MM segmentation. However, it applies only to dermoscopy images analysis, which is easier to realize but not as reliable and detailed as pathological analysis.

However, these methods only work for the region of interest marked by pathologists. They cannot achieve good results in WSIs. The huge number of pixels makes network training difficult or impossible. Resizing images by down-sampling will lead to the loss of detailed information, which is unacceptable for MM diagnosis focusing on pathological features at the cellular level. Furthermore, due to the characteristics of WSIs and the limited feature extraction capability of related networks, the existing methods are difficult to adapt for WSIs-based MM recognition.

Based on the above considerations, we proposed a novel MM recognition method based on a multi-scale feature representation and probability map to recognize the MM tissue region in WSIs, as shown in **Figure 1**. The following contributions are made to our work.

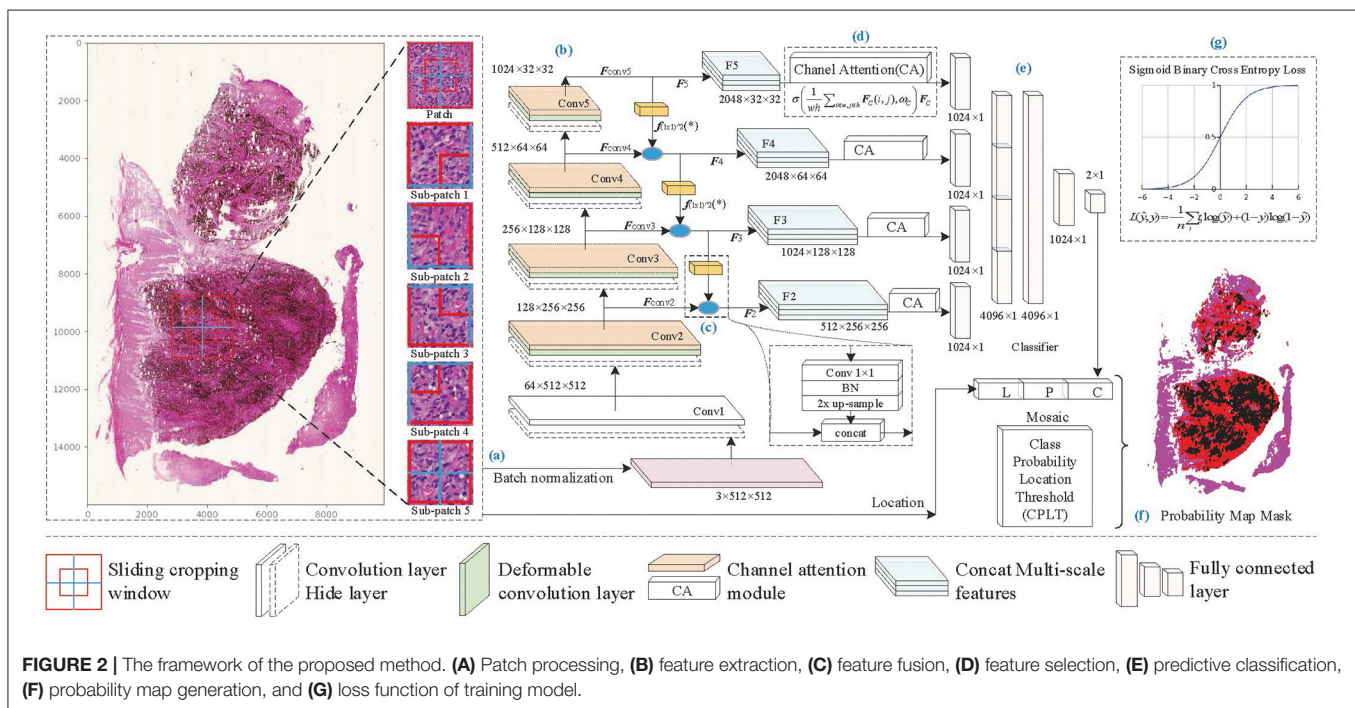
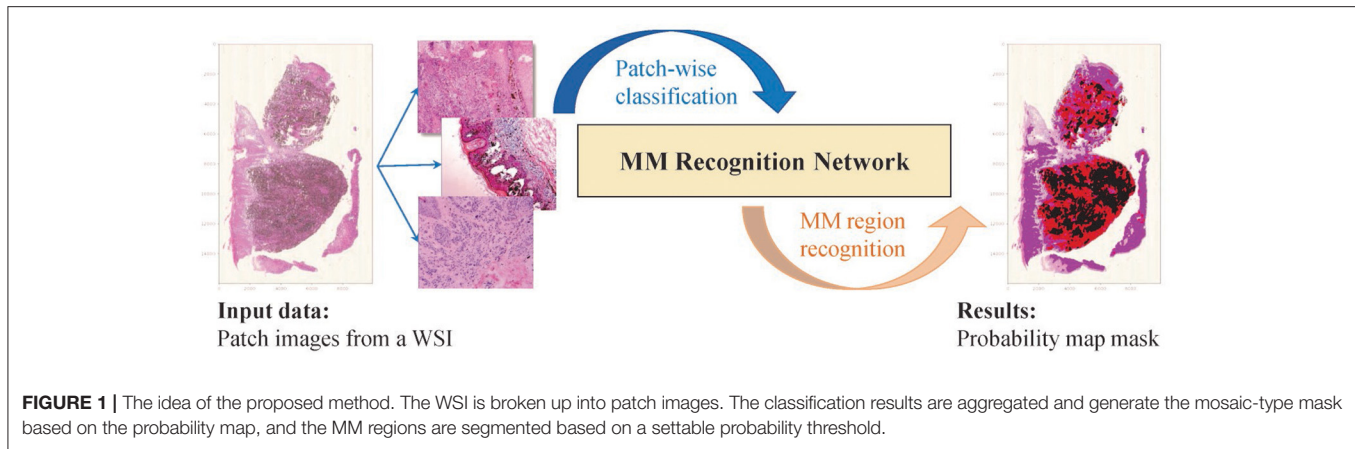
- Aiming at the difficult and inaccurate problems of recognizing the enormous size of WSIs, the breaking up the whole into parts idea is introduced to recognize melanoma. Furthermore, using predicted results and probabilities generates the mosaic-style mask and lesion region.
- To take both global and local features, we propose an efficient multi-scale network for improving melanoma recognition, combining high-level features with more semantic information and low-level features with more detail information. A multi-scale sliding cropping operation is used to obtain patch and sub-patch images.
- To enhance the feature representation of irregular-shaped lesions, highlight the critical features, and reduce the impact of information redundancy and data noise, we reconstruct the residual block by deformable convolution and channel attention.

The paper is arranged as follows. Section 2 details the proposed method, including the description of our method's framework, the realization principles, and the equations of each module. Section 3 shows the experimental results of our method, compares algorithms on an available WSI dataset, and further provides the ablation analysis to prove the effectiveness and rationality of the proposed method. Section 4 provides a further discussion on the feature representation capability of the proposed multi-scale network. And section 5 provides a brief summary and the conclusions of this work.

## 2. METHODOLOGY

### 2.1. Framework of Our Method

On super-large WSI images, patch-based recognition is necessary and feasible. Melanoma pathological analysis mainly focuses on cell-scale characteristics. We set patch size according to pathologists' professional advice, which ensures that cell morphology and local distribution are well represented in the patches. On the boundary of the patch, some cells may be torn, but the overlapping sampling method can effectively avoid the loss of information caused by incomplete splitting. For lesion areas without clear boundaries, mixed cell tissue limits feature extraction by conventional rectangular convolution. Therefore, the proposed method reconstructed the residual block by deformable convolution and channel attention to overcome the irregular-shaped lesion and focus on important features. Furthermore, to overcome the influence of cell-scale differences, we built multi-scale feature fusion layers to enhance feature information and improve identification accuracy. The framework of the proposed method shown in **Figure 2** consists of the following seven components: patch processing, feature extraction, feature fusion, feature selection, predictive classification, mask generation, and loss function in training.



- Patch processing:** A WSI is broken up into  $N$  patches through sliding cropping ( $N$  depends on the sliding window size and sliding stride), from which tissue-contained patches are picked out by a color analysis method. Each tissue-contained patch is broken up into sub-patches. And then, patch images and sub-patch images are normalized to a uniform size.
- Feature extraction:** The lesion features are extracted by backbone Conv1 to Conv5. Considering the irregular-shaped cells, and focusing on essential features, deformable convolution (DC) and channel attention (CA) operations are embedded in the Conv2 to Conv5 layers to enhance the feature extraction capability of the network. Then extracted features  $F_{conv i} (i = 2, 3, 4, 5)$  are produced separately from Conv2 to Conv5.
- Feature fusion:** As the network is gradually deepened, the resolution of the feature map decreases, and the semantic properties of the features are enhanced. The features of the next layer, which contains richer semantic information, are concatenated with those of the current layer, which contains richer detailed texture information, to enhance the lesion feature representation capability of the network.
- Feature selection:** After the fused features  $F_i (i = 2, 3, 4, 5)$ , the channel attention mechanism is separately used to select the critical features and to enhance the correlation between high-level semantic features and low-level detailed features.
- Predictive classification:** The output features from each branch are flattened into a vector, respectively, and then they are concatenated. Fully connected layers are constructed to obtain the predictive classification results of patch images.

- **Probability map generation:** The prediction results of patches (containing prediction labels and confidence probabilities) are combined with the location information to generate the probability map of malignant tissues. And a mosaic-type mask of MM regions is obtained through a confidence probability threshold.
- **Loss function:** Sigmoid binary cross-entropy loss function is used in training for parameter optimization.

## 2.2. Multi-Scale Features

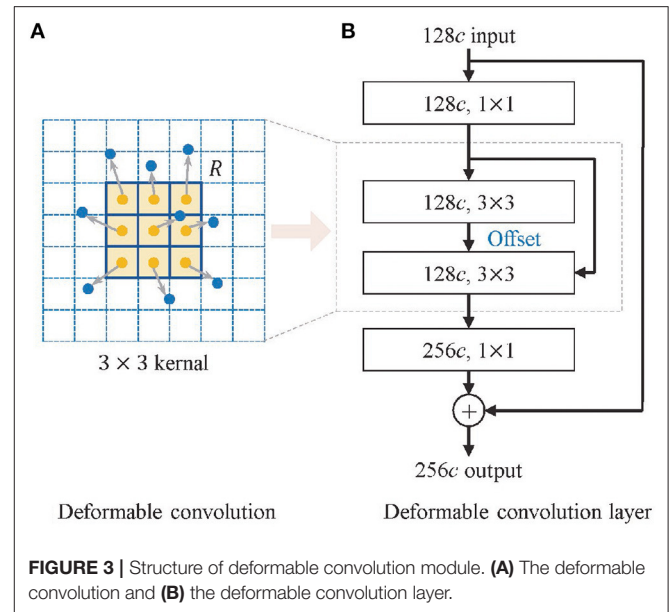
In the pathological examination, it is necessary to carry out comprehensive analysis according to various characteristics of lesions, such as tissue morphology and cell distribution, which are reflected on a large scale, and cell morphology and nuclear size, which are reflected on a smaller scale (15). Therefore, computational analysis of WSIs at different scales is beneficial to represent pathological features at different scales. A multi-scale sliding cropping method is embedded in the proposed algorithm. For each patch of a WSI, besides the whole patch image, cropped sub-patch images from the patch are normalized to a uniform size and input into the network together for the classification of the patch. The size and quantity of the sub-patch depend on the cropping method, for example, **Figure 2A** shows five sub-patches cropped from a patch image.

Furthermore, the idea of multi-scale is also reflected in the construction of the feature extraction network. As the network deepens, feature resolution decreases and channels increase, low-level detail information is being transformed into higher-level semantic information. However, factors such as data noise and chain derivative attenuate or lose the information in the forward and back propagation, which becomes more and more apparent with increasing network depth. The fusion of shallow features and deeper features, which are with different scales, to supplement the semantic information of high-level features is beneficial to improve the feature representation capability of the network. Based on the above considerations, a network with enhanced multi-scale feature extraction capability is constructed. As **Figure 2** shows, additional branches are added to the backbone network for feature fusion. In each branch, the feature of  $(i + 1)$ -th level  $F_{\text{Conv}i+1}$  ( $i = 2, 3, 4$ ) is concatenated with the feature of  $i$ -th level  $F_{\text{Conv}i}$  ( $i = 2, 3, 4$ ) after  $1 \times 1$  convolution and up-sampling, as shown in **Figure 2C**, and then  $F_i$  ( $i = 2, 3, 4$ ) is obtained, as shown in Equation (1).

$$F_i = f_{(1 \times 1) \times 2}(F_{i+1}) \oplus F_{\text{Conv}i} \quad (1)$$

where  $f_{(1 \times 1) \times 2}(\cdot)$  indicates that the  $F_i$  is obtained by the convolution of  $1 \times 1$  and double up-sampling of features  $F_{i+1}$  and has the same shape maps as the  $i$ -th conv output features  $F_{\text{Conv}i}$ .  $\oplus$  denotes the concatenation of the normalized features of the two groups.

The concatenated features  $F_2 \sim F_4$ , together with  $F_5$ , which is obtained from  $F_{\text{Conv}5}$ , are transferred to feature vectors and input into the fully connected layer via shortcut connections for classification.



## 2.3. Deformable Convolution

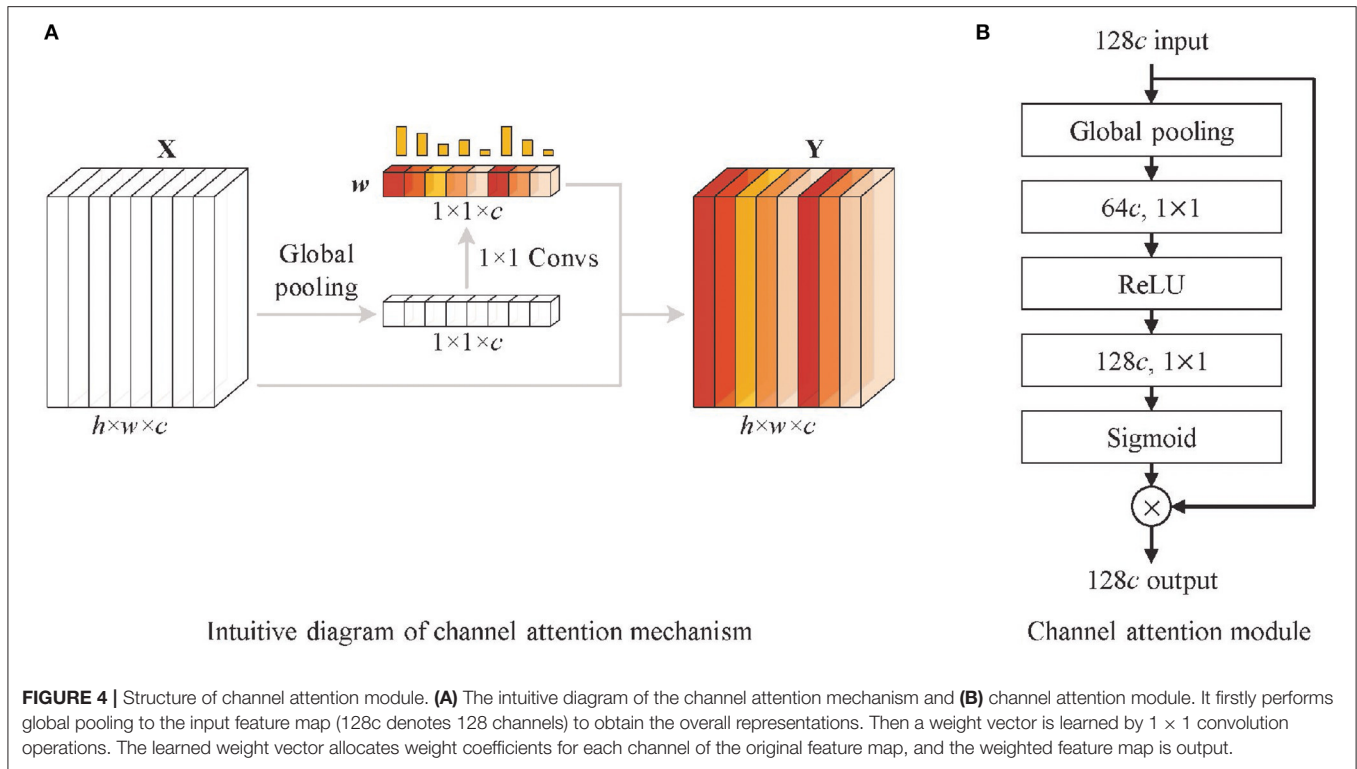
MM tissues in histopathological images mainly show as interstitial or heterogeneous tumor cells, which are mainly enlarged and darkly stained nuclei with varying shapes (16). This irregularity leads to the inadequate learning of melanoma feature information by traditional convolution for its fixed rectangular receptive field of the kernel. Inspired by Dai et al. (17) and Zhu et al. (18), we introduce offsets in the traditional convolution to make the geometry of the kernel more flexible, as shown in **Figure 3**, which improves the representation of irregular-shaped features. The deformable convolution format for each position  $p$  in the input feature map is shown in Equation (2).

$$y(p) = \sum_{p_k \in R} w(p_k) \cdot x(p + p_k + \Delta p_k) \quad (2)$$

where  $y(p)$  indicates the feature obtained by the convolution on one sampling point  $p$  of the feature map.  $R$  is the receptive field size of the regular kernel.  $p_k$  denotes the difference between the sampling points and  $y(p)$ ,  $k = 1, 2, 3 \dots N$ ,  $N = |R|$ ,  $\Delta p_k$  is the learned offset, and  $w$  is the kernel parameter. The offset of the deformable convolution has a dilated value, which determines the maximum distance for resampling and is set to 2.

## 2.4. Channel Attention

First, multi-scale feature fusion enriches the extracted feature information of the network, but while enhancing feature representation capability, it also brings some redundant features, which are unrelated to melanoma recognition, and interferes with model learning. It is particularly obvious in low-level features with higher resolution, and this effect becomes more prominent when low-level features are fused with high-level features through additional branch paths. Second, the deformable convolution helps in the feature extraction of irregular lesions and enhances lesion feature representations, but also generates



some noisy features influenced by non-lesion tissue. Therefore, to extract more valuable information and suppress the impact of redundant information and noises features, we need a mechanism that focuses on essential features and filters the irrelevant features.

Based on the above considerations, and inspired by the work of Hu et al. (19), a channel attention mechanism is used in the shortcut connection between features  $F_i (i = 2, 3, 4, 5)$  and fully connected layers, as shown in **Figure 2D**. It highlights high-value feature maps by a series of weights learned by the channel attention (CA) module. The filtering of channels is actually the weighting of different types of features. Although the convolution operation itself also correlates each channel of the feature map with each other, it is difficult to accurately assign appropriate weights to each channel due to the influence of the  $w$  and  $h$  dimensional feature distributions. To address this problem, the channel attention mechanism obtains a global representation of each channel by global pooling, and the weights of each channel are calculated by  $1 \times 1$  convolution based on the resulting feature vectors. In **Figure 4**, the CA module firstly performs global pooling to the input feature map to obtain the overall representation of it. Then a weight vector is learned by  $1 \times 1$  convolution. The learned weight vector allocates weight coefficients for each channel of the original feature map, and the weighted feature map is output. The mathematic description of the channel attention module is formatted as Equation (3).

$$Y = \sigma \left( W_{Conv2} \delta \left( W_{Conv1} \frac{1}{hw} \sum_{i \in h, j \in w} X(i, j) \right) \right) \otimes X \quad (3)$$

where  $X$  means the input feature map, and  $Y$  denotes the output feature map of the channel attention module,  $h$  and  $w$  are the height and width in the input feature maps.  $W_{Conv1}$  and  $W_{Conv2}$  indicate the parameters of two  $1 \times 1$  convolution operations, which are equivalent operations to fully connected layers.  $\delta$  is the ReLU activation.  $\sigma$  is the sigmoid function, and  $\otimes$  means the weighting calculation of the learned weight vector and the input feature map.

In addition to calculating the channel weights of feature weights, the channel attention mechanism also strengthens the correlation between channels through global pooling and  $1 \times 1$  convolution; making up for the defect of the weak correlation between channels in the convolution module is conducive to the enhancement of feature expression ability. Therefore, the channel attention modules are also embedded into the backbone network, as shown in **Figure 2B**.

### 3. EXPERIMENTS

#### 3.1. Experimental Setup

MM WSIs labeled by pathologists are rare and valuable data. The dataset is collected from the Second Affiliated Hospital of Xi'an Jiaotong University (Xibei Hospital), containing 30 WSIs labeled by experienced pathologists. Sliding window size is set to  $1,024 \times 1,024$ , sliding stride is set to 1024, 18,698 tissue-included patches are obtained, containing 7,369 malignant tissue patches and 11,329 benign tissue patches. They are divided into training, validation, and test datasets by a ratio of 6:2:2. Five sub-patches are cropped from each patch, as shown in **Figure 2A**, and

**TABLE 1** | The experimental results of the proposed algorithms and comparison algorithms, the higher the values of precision, recall, accuracy, and  $F1$ , the better the recognition performance.

Algorithms	Layer	Precision	Recall	Accuracy	$F1$
Inception V3 (2016)	50	0.8919	0.8876	0.8326	0.8897
ResNeXt (2017)	50	0.8974	0.9241	0.8618	0.9106
SENet (2018)	50	0.8915	0.9472	0.8721	0.9185
SENet (2018)	101	0.9120	0.9477	0.8812	0.9295
ResNeSt (2020)	50	0.9314	0.9327	0.8877	0.9321
ResNeSt (2020)	101	0.9526	0.9601	0.9355	0.9513
MPMR (ours)	50	0.9683	0.9709	0.9498	0.9696
<b>MPMR (ours)</b>	101	<b>0.9740</b>	<b>0.9861</b>	<b>0.9553</b>	<b>0.9749</b>

*Bold indicate maximum values.*

all images are resized to  $512 \times 512$  when input to the network. Considering that melanoma tissue features are non-directional and non-chiral, we introduce data augmentation operations by the mirror and random rotation in the range of  $(-90^\circ, +90^\circ)$ .

The proposed method is developed by Python 3.6 on Ubuntu18.04, and the hardware is RTX2080-12G with CUDA-10.1. The development libraries include MXNet-1.5, Gluoncv-0.5, Numpy-1.17, OpenCV-4.2, etc. The models iterate 30 epochs, and the batch size is 32. Gradient descent with momentum (20) is used for optimization. We set the momentum to 0.9. The learning rate is 0.001, and the decay rate is 0.99. Both recall ( $R$ ) and precision ( $P$ ) for MM recognition are considered in diagnosis, so the evaluation criterion  $F1$  score is used to comprehensively measure the performance of the proposed method, which is calculated as Equation (4).

$$F1 = \frac{2 \times P \times R}{P + R} \quad (4)$$

### 3.2. Results of Patch Classification

In order to verify the effectiveness and recognition performance of the algorithm, the proposed method is compared with some popular algorithms in recent years, including Inception V3 (21), ResNeXt (22), SENet (19), and ResNeSt (23). The experimental results are shown in **Table 1**, the higher the values of  $F1$ , recall, precision, and accuracy, the better the recognition performance. The  $F1$  values of all algorithms exceeded 90%, except Inception V3, and the scores of the proposed method also achieved the best results. SENet and ResNeSt, containing the channel attention module, outperform other comparison algorithms, indicating that the channel attention mechanism improves performance.

The proposed method outperforms all the comparison algorithms for the same number of layers, mainly benefiting from the deformable convolution, the channel attention, and the multi-scale feature fusion. In particular, the learning capability of multi-scale features in the proposed method effectively adapts the different scale samples. It sufficiently learns the feature information of melanoma in the training and validation datasets and has better robustness on the testing dataset. Therefore, the proposed algorithm outperforms other algorithms on the WSI test dataset.

### 3.3. Results of the Probability Map

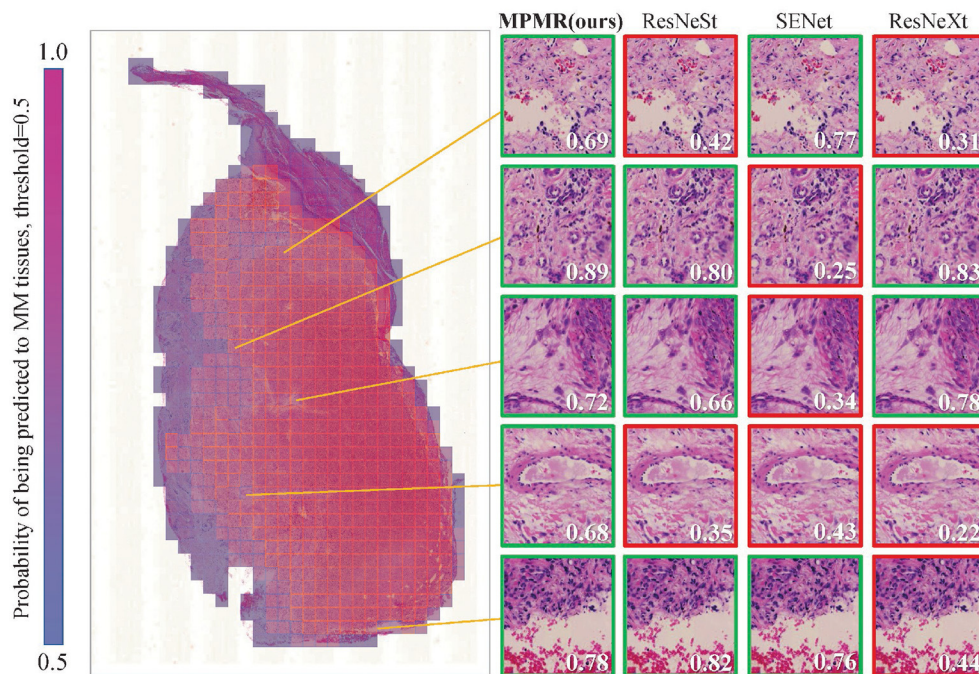
The prediction results of patch images containing prediction labels and confidence probabilities are combined with the location information to generate the probability map. The visualization results of a WSI containing malignant melanoma tissues are shown in **Figure 5**. The probability of being predicted as MM tissues is visualized as different colors, from 0 to 1. The threshold of malignant tissues and benign tissues is set to 0.5; red regions display the recognized MM tissues. The prediction results of some difficult samples of different algorithms are compared, and the proposed method provides the most correctly recognized patches, marked by green boxes, while other comparison algorithms provide some incorrect recognition results, marked by red boxes. The results indicate that the proposed method can obtain more accurate recognition results in WSIs.

### 3.4. Ablation Analyses

To analyze the contributions of multi-scale feature fusion, deformable convolution, and channel attention in the proposed method, ablation analyses are performed for these impacts. The results of ablation analyses are shown in **Tables 2–4**, the higher the values of precision, recall, accuracy, and  $F1$ , the better the recognition performance.

#### 3.4.1. Multi-Scale Features

The proposed method realizes multi-scale feature fusion by constructing additional branch paths and adopting shortcut connections between fused features and the fully connected layers. Low-level features containing more detail information are expected to supplement the semantic features of high-level features for enhancing the classification capability of the model. The experimental results of the networks with different numbers of branch paths are shown in **Table 2**. The more branch paths added, the higher the values of  $F1$ , recall, precision, and accuracy obtained. It indicates that the prediction method based on multi-scale features helps the proposed method to recognize melanoma. In the results of  $F5/F4/F3/F2$ , the accuracy of the proposed method decreases compared to  $F5/F4/F3$ , and the other evaluation indicators show weak increases. It indicates that multi-scale feature fusion should be carried out in an



**FIGURE 5 |** The mosaic-style mask result of a WSI in the test dataset is generated through the probability map obtained through patch image classification. The recognition results of some patches, which are difficult to recognize, through different algorithms are compared. A green box represents a correct prediction, and a red box represents a misclassification. The probabilities of being predicted as MM tissues are marked in each patch image, and the threshold is 0.5.

**TABLE 2 |** The experimental results of the ablation analysis of multi-scale features, the higher the values of precision, recall, accuracy, and  $F1$ , the better the recognition performance.

Multi-scale features	Layer	Precision	Recall	Accuracy	$F1$
F5	50	0.9400	0.9457	0.9367	0.9428
F5/F4	50	0.9448	0.9633	0.9493	0.9540
F5/F4/F3	50	0.9590	0.9675	<b>0.9567</b>	0.9632
F5/F4/F3/F2	50	<b>0.9683</b>	<b>0.9709</b>	0.9498	<b>0.9696</b>

*Bold indicate maximum values.*

**TABLE 3 |** The experimental results of the ablation analysis of deformable convolution, the higher the values of precision, recall, accuracy, and  $F1$ , the better the recognition performance.

Deformable convolution	Layer	Precision	Recall	Accuracy	$F1$
None	50	0.9380	0.9509	0.9387	0.9444
DConv5	50	0.9396	0.9522	0.9402	0.9458
DConv5/4	50	0.9464	0.9563	0.9462	0.9513
DConv5/4/3	50	0.9604	0.9619	<b>0.9569</b>	0.9611
DConv5/4/3/2	50	<b>0.9683</b>	<b>0.9709</b>	0.9498	<b>0.9696</b>

*Bold indicate maximum values.*

appropriate range, and an excess of fusions will cause information redundancy, which is not conducive to feature representation.

### 3.4.2. Deformable Convolution

The melanoma characteristics in the pathological images are mainly enlarged and darkly stained nuclei with varying shapes. This irregularity leads to the inadequate learning

of melanoma feature information by traditional convolution. Deformable convolution is embedded into the convolution layers of the proposed network to enhance irregular-shaped feature representation ability. The experimental results of the networks with different numbers of deformable convolution layers are shown in **Table 3**, indicating that the more deformable convolution layers embedded, the better the recognition

**TABLE 4 |** The experimental results of the ablation analysis of channel attention, the higher the values of precision, recall, accuracy, and *F1*, the better the recognition performance.

Channel attention	Layer	Precision	Recall	Accuracy	<i>F1</i>
None	50	0.9320	0.9216	0.9222	0.9242
Backbone (B)	50	0.9340	0.9321	0.9256	0.9331
Shortcut (S)	50	0.9472	0.9552	0.9460	0.9512
Both B and S	50	<b>0.9683</b>	<b>0.9709</b>	<b>0.9498</b>	<b>0.9696</b>

*Bold indicate maximum values.*

performance of the proposed method. Accuracy decrease occurs in the results of DConv5/4/3/2. It indicates that too many deformable convolution layers may amplify the impact of noise on learning and influence feature representation.

### 3.4.3. Channel Attention

Channel attention in the proposed method selectively enhances information-rich features, allowing subsequent processing of the networks to take full advantage of these features and suppress noisy features. The experimental results of channel attention with different numbers of layers are shown in **Table 4**, where the B-case indicates that channel attention modules are only embedded in the backbone network for feature extraction, and the S-case indicates that channel attention modules are used in the shortcut connection between fused features and fully connected layers. The recognition performance of the model is significantly improved after using channel attention modules. However, both the embedded B-case and S-case can obtain the best performance of the proposed method. This further demonstrates that the embedding of channel attention can facilitate positive network learning.

## 4. DISCUSSION

The pathological features related to melanoma are at different scales, such as tissue, cell, and nucleus, and enhancing the representation of multi-scale features is important for melanoma recognition. From the experimental results, it can be concluded that the residual block based on deformable convolution and multi-scale feature fusion brings considerable performance improvement in the patch-wise classification of WSIs.

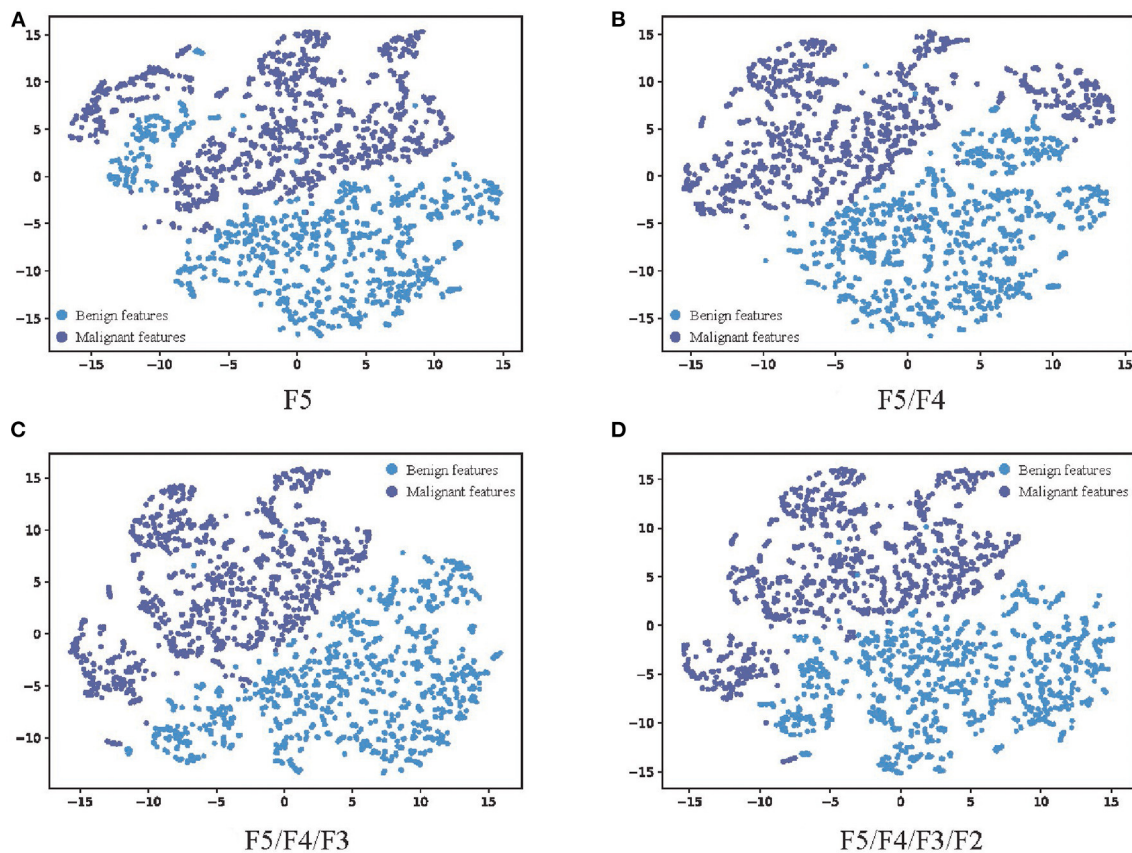
The visual features of malignant melanoma and benign nevi tissues are very similar, and the shape of the features is irregular and has uneven distribution, which further increases the difficulty of recognition. When learning the feature space of a histopathological skin image, the traditional convolutional network is limited by its fixed spatial geometric structure, as shown in **Figure 3**, which is not suitable for the irregular shape of lesions and the uneven distribution of melanoma cells. However, the deformable convolution layers effectively avoid the rectangular limitation of traditional convolution sampling. The experimental results in **Table 3** demonstrate that dynamic convolution can better extract the features of tissue images. The performance improvement of deformable convolution on the model grows as the number of layers increases, which introduces extra computational consumption but can be neglected for some tasks with low real-time requirements.

Shortcut connections, also known as skip connections, show considerable advantages in residual networks and U-shaped networks. The residual connections ensemble the feature at different layers through sum operation, and (24) put forward similar views. The connections between the encoder and the decoder in U-shaped networks, through deconvolution and concatenation, realize the fusion of features at different scales. In addition, extra information flows, brought by shortcut connections, provide shorter paths for the transmission of parameters in the forward- and back-propagation, reducing information attenuation. These ideas are embodied in the construction of the proposed networks. The additional branch paths in the proposed network realize multi-scale feature fusion through  $1 \times 1$  convolution,  $2 \times$  up-sampling, and concatenation. Another fusion of several fused features is performed through the shortcut connections between the fused features and the fully connected layers. The above operations are expected to enhance feature representation and make contributions to improve MM recognition precision.

In order to further analyze the influence of multi-scale fusion on the quality of features extracted from the network, t-distributed stochastic neighbor embedding (t-SNE) (25), a manifold learning dimensionality reduction method, is used to visualize the features extracted from the network with the different number of feature fusion branch paths. The feature vectors in the second-to-last layer of the full connection layers are transferred from 1,024 dimensions to 2 dimensions and visualized as shown in **Figure 6**. The higher the linear separability of benign features and malignant features after dimensionality reduction, the more beneficial the features extracted from the network are for classification. The dimension reduction results of F5, which represents the features extracted by the network without multi-scale feature fusion, are shown in **Figure 6A**, and some of the benign features are interspersed with the malignant features. The results of F5/F4 shown in **Figure 6B**, which represents the features extracted by the network with one branch path for multi-scale feature fusion, show considerable improvement. **Figures 6C,D** show more improvements, which indicates that the additional branch paths for multi-scale feature fusion improve the quality of the features extracted by the network, enhancing the feature representation for MM recognition, and finally provide more accurate MM recognition results. This is consistent with the experimental results in **Table 2**.

## 5. CONCLUSIONS

This work proposes a novel automatic MM recognition method in WSI based on multi-scale features and the probability map.



**FIGURE 6 |** The visualization results of the features extracted by networks with different numbers of feature fusion branch paths. **(A)** F5, **(B)** F5/F4, **(C)** F5/F4/F3, and **(D)** F5/F4/F3/F2. The feature vector in the second-to-last layer of the full connection layers are transferred from 1,024 dimensions to 2 dimensions through t-SNE. The blue scatter plots represent malignant features, and the purple scatter plots represent malignant features. The higher the linear separability of benign features and malignant features after dimensionality reduction, the more beneficial the features extracted from the network are for classification.

The idea that breaking up a WSI into patches and sub-patches through multi-scale sliding cropping solves the difficult-to-calculate problem of WSIs with huge sizes, and the probability map is generated based on the predicted class and confidence probabilities and location information of patch images to visualize the recognition result of MM tissues in WSIs. Additional branch paths and shortcut connections are established for multi-scale feature fusion, which realizes the information supplement of low-level features containing more detail information to deep features containing more semantic information. Deformable convolution operations are embedded into the backbone network to enhance the representation capability of irregular-shaped features in tissues. Channel attention modules are used in the shortcut connection between fused features and fully connected layers, and also the backbone network to highlight the high-value features and reduce the negative impacts of information redundancy caused by additional branch paths.

The results of comparison experiments indicate that the proposed method outperforms Inception V3, ResNeXt, SENet, and ResNeSt. The results of ablation analyses prove

the effectiveness of multi-scale feature fusion, deformable convolution, and channel attention modules. Through the proposed method, MM regions in WSIs can be recognized accurately and efficiently, which is a great help to pathological examination and the diagnosis of MM.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

DZ and HH were in charge of experiments and manuscript writing. LZ and MX were responsible for medical analysis and annotation of pathological data. SD and JY provided guidance for method formulation. LW and XW checked the experimental results. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Key Research and Development Program of China under (grant no. 2017YFA0700800), the National Natural Science

Foundation of China under (grant no. 61971343), the Shaanxi Provincial Social Science Fund under (grant no. 2021K014), and the Key Research and Development Program of Shaanxi Province of China under (grant no. 2020GXLH-Y-008).

## REFERENCES

- Palacios-Ferrer JL, García-Ortega MB, Gallardo-Gómez M, García MÁ, Díaz C, Boulaiz H, et al. Metabolomic profile of cancer stem cell-derived exosomes from patients with malignant melanoma. *Mol Oncol.* (2021) 15:407–28. doi: 10.1002/1878-0261.12823
- Shoo BA, Sagebiel RW, Kashani-Sabet M. Discordance in the histopathologic diagnosis of melanoma at a melanoma referral center. *J Amer Acad Dermatol.* (2010) 62:751–6. doi: 10.1016/j.jaad.2009.09.043
- Scolyer RA, Rawson RV, Gershenwald JE, Ferguson PM, Prieto VG. Melanoma pathology reporting and staging. *Mod Pathol.* (2020) 33:15–24. doi: 10.1038/s41379-019-0402-x
- Lu C, Mandal M. Automated analysis and diagnosis of skin melanoma on whole slide histopathological images. *Pattern Recognit.* (2015) 48:2738–50. doi: 10.1016/j.patcog.2015.02.023
- Khan MQ, Hussain A, Rehman SU, Khan U, Maqsood M, Mehmood K, et al. Classification of melanoma and nevus in digital images for diagnosis of skin cancer. *IEEE Access* (2019) 7:90132–44. doi: 10.1109/ACCESS.2019.2926837
- Ronneberger O, Fischer P, Brox T. “U-net: convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. (Springer) (2015). p. 234–41.
- Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. “Unet++: a nested u-net architecture for medical image segmentation,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer (2018). p. 3–11.
- Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*. (2018).
- Li C, Tan Y, Chen W, Luo X, Gao Y, Jia X, et al. “Attention Unet++: a nested attention-aware U-Net for liver CT image segmentation,” in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE (2020). p. 345–9.
- Huang H, Lin L, Tong R, Hu H, Zhang Q, Iwamoto Y, et al. “Unet 3+: a full-scale connected unet for medical image segmentation,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE (2020). p. 1055–9.
- Hekler A, Utikal JS, Enk AH, Berking C, Klode J, Schadendorf D, et al. Pathologist-level classification of histopathological melanoma images with deep neural networks. *Eur J Cancer* (2019) 115:79–83. doi: 10.1016/j.ejca.2019.04.021
- He K, Zhang X, Ren S, Sun J. “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2016). p. 770–8.
- Wang L, Ding L, Liu Z, Sun L, Chen L, Jia R, et al. Automated identification of malignancy in whole-slide pathological images: identification of eyelid malignant melanoma in gigapixel pathological slides using deep learning. *Brit J Ophthalmol.* (2020) 104:318–23. doi: 10.1136/bjophthalmol-2018-313706
- Yu Z, Jiang X, Zhou F, Qin J, Ni D, Chen S, et al. Melanoma recognition in dermoscopy images via aggregated deep convolutional features. *IEEE Trans Biomed Eng.* (2018) 66:1006–16. doi: 10.1109/TBME.2018.2866166
- Elmore JG, Barnhill RL, Elder DE, Longton GM, Pepe MS, Reisch LM, et al. Pathologists’ diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study. *BMJ* (2017) 357:j2813. doi: 10.1136/bmj.j2813
- Chopra A, Sharma R, Rao UN. Pathology of melanoma. *Surgical Clin.* (2020) 100:43–59. doi: 10.1016/j.suc.2019.09.004
- Dai J, Qi H, Xiong Y, Li Y, Zhang G, Hu H, et al. “Deformable convolutional networks,” in *Proceedings of the IEEE International Conference on Computer Vision*. (2017). p. 764–73.
- Zhu X, Hu H, Lin S, Dai J. “Deformable convnets v2: more deformable, better results,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (2019). p. 9308–16.
- Hu J, Shen L, Sun G. “Squeeze-and-excitation networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2018). p. 7132–41.
- Qian N. On the momentum term in gradient descent learning algorithms. *Neural Netw.* (1999) 1:145–51. doi: 10.1016/S0893-6080(98)00116-6
- Xia X, Xu C, Nan B. “Inception-v3 for flower classification,” in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*. (IEEE) (2017). p. 783–7.
- Xie S, Girshick R, Dollár P, Tu Z, He K. “Aggregated residual transformations for deep neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2017). p. 1492–500.
- Zhang H, Wu C, Zhang Z, Zhu Y, Lin H, Zhang Z, et al. Resnest: split-attention networks. *arXiv preprint arXiv:2004.08955*. (2020).
- Veit A, Wilber MJ, Belongie S. Residual networks behave like ensembles of relatively shallow networks. *Adv Neural Inf Process Syst.* (2016) 29:550–8.
- Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res.* (2008) 9:2579–605.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Han, Du, Zhu, Yang, Wang, Wang and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Three-Dimensional Liver Image Segmentation Using Generative Adversarial Networks Based on Feature Restoration

Runnan He<sup>1</sup>, Shiqi Xu<sup>2</sup>, Yashu Liu<sup>2</sup>, Qince Li<sup>1,2</sup>, Yang Liu<sup>2</sup>, Na Zhao<sup>3</sup>, Yongfeng Yuan<sup>2\*</sup> and Henggui Zhang<sup>1,4,5\*</sup>

<sup>1</sup> Peng Cheng Laboratory, Shenzhen, China, <sup>2</sup> School of Computer Science and Technology, Harbin Institute of Technology (HIT), Harbin, China, <sup>3</sup> School of Instrument Science and Engineering, Southeast University, Nanjing, China, <sup>4</sup> School of Physics and Astronomy, The University of Manchester, Manchester, United Kingdom, <sup>5</sup> Key Laboratory of Medical Electrophysiology of Ministry of Education and Medical Electrophysiological Key Laboratory of Sichuan Province, Institute of Cardiovascular Research, Southwest Medical University, Luzhou, China

## OPEN ACCESS

### Edited by:

Shuo Li,  
Western University, Canada

### Reviewed by:

Zhixin Li,  
Guangxi Normal University, China  
Hailan Shen,  
Central South University, China

### \*Correspondence:

Yongfeng Yuan  
yongfeng.yuan@hit.edu.cn  
Henggui Zhang  
henggui.zhang@gmail.com

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

**Received:** 14 October 2021

**Accepted:** 22 November 2021

**Published:** 07 January 2022

### Citation:

He R, Xu S, Liu Y, Li Q, Liu Y, Zhao N,  
Yuan Y and Zhang H (2022)  
Three-Dimensional Liver Image  
Segmentation Using Generative  
Adversarial Networks Based on  
Feature Restoration.  
Front. Med. 8:794969.  
doi: 10.3389/fmed.2021.794969

Medical imaging provides a powerful tool for medical diagnosis. In the process of computer-aided diagnosis and treatment of liver cancer based on medical imaging, accurate segmentation of liver region from abdominal CT images is an important step. However, due to defects of liver tissue and limitations of CT imaging procession, the gray level of liver region in CT image is heterogeneous, and the boundary between the liver and those of adjacent tissues and organs is blurred, which makes the liver segmentation an extremely difficult task. In this study, aiming at solving the problem of low segmentation accuracy of the original 3D U-Net network, an improved network based on the three-dimensional (3D) U-Net, is proposed. Moreover, in order to solve the problem of insufficient training data caused by the difficulty of acquiring labeled 3D data, an improved 3D U-Net network is embedded into the framework of generative adversarial networks (GAN), which establishes a semi-supervised 3D liver segmentation optimization algorithm. Finally, considering the problem of poor quality of 3D abdominal fake images generated by utilizing random noise as input, deep convolutional neural networks (DCNN) based on feature restoration method is designed to generate more realistic fake images. By testing the proposed algorithm on the LiTS-2017 and KiTS19 dataset, experimental results show that the proposed semi-supervised 3D liver segmentation method can greatly improve the segmentation performance of liver, with a Dice score of 0.9424 outperforming other methods.

**Keywords:** CT image, 3D segmentation of liver, semi-supervised, generative adversarial networks, feature restoration

## INTRODUCTION

Recent advances in deep convolutional neural networks (DCNN) have shown great promises in handling many computer vision tasks such as target detection, image classification, and semantic segmentation, which can usually reach human-level performance. However, one of the main limitations of DCNN is that they require a large amount of labeled data for training process.

This limitation is particularly prominent in dealing with medical image segmentation problems. At present, the acquisition of labeled three-dimensional (3D) medical images requires manual annotation, which is time-consuming and labor-intensive, limiting the further development of DCNN in medical image processing. Moreover, 3D image segmentation for medical applications needs great amount of computing resources, hurdling its practical application. Although the neural network has the characteristics of parameter sharing, it acquires a deeper network structure to improve the performance of the model. As the number of network layers increases, the parameter quantity is increased proportionally. Therefore, the deep neural network needs a large dataset to train the model for obtaining the model parameters. In the absence of sufficient training data, the neural network will have relatively low performance and poor generalization ability.

In addition, there are some problems with liver tissue's structure and CT imaging procession. Firstly, due to the differences in gender, age, and body type of patients, the shape and size of the liver of individual patients are different in appearance in their CT images. Moreover, there are many abdominal organs of compact structure, with tissue density similar to that of liver. Secondly, the area of the diseased area in patients' liver is not fixed in size and with random location that will cause interference to the network in the process of the liver recognition. Finally, there are problems such as sensitivity to noise, metal artifacts, and body motion during the imaging process of CT images, leading to variation of the gray value of liver area due to the influence of the imaging environment, resulting in uneven gray level of liver area, which affects the accuracy of liver segmentation.

Aiming at the difficulty of liver segmentation in abdominal CT images, this paper improves the contrast of the liver in the CT images by preprocessing abdominal CT images, which improves the recognition ability of the liver. Furthermore, semi-supervised learning algorithms reduce the need of large amount of labeled data. In recent years, generative adversarial networks (GAN) have shown great potentials for improving semantic segmentation in a semi-supervised manner (1). Thus, this study also employs GAN to generate fake images by combining labeled CT images to train the network in a semi-supervised manner, which can further improve the algorithm's performance of liver segmentation by expanding the dataset.

The rest of this paper is organized as follows. In section "Related Work", we give a brief overview of relevant work on liver segmentation. Section "Methods" then presents our 3D liver image segmentation approach based on GAN, which is evaluated and analyzed on the challenging task of liver segmentation in section "Experiments and Discussion". Finally, we conclude with a summary of our main contributions and results.

## RELATED WORK

Before deep learning was widely used, many methods have been proposed for the liver segmentation of abdominal CT images based on graphics, morphology, and traditional machine

learning. With the rapid development of deep learning and its blossom in the field of computer vision, the direction of research in the field of medical image segmentation has also begun to transform to deep learning. In the field of liver image segmentation, more and more methods based on deep learning have also appeared.

## Traditional Methods

Apollon proposed a hybrid liver segmentation algorithm based on pixel intensity threshold (2). It manually selects multiple initial seed points in the image and calculates the average pixel intensity value of nine adjacent pixels of the selected seed points to obtain the segmentation results of the liver image. Amir proposed a two-step liver segmentation method based on threshold and active contour by the contrast characteristics of liver CT image data set, liver shape diversity, and uneven texture (3). Seong proposed an abnormal liver segmentation method based on the adaptive threshold and angle line (4). Moreover, Farzaneh applied the Bayesian-based adaptive threshold to address the issue of liver segmentation (5). This algorithm adjusts the threshold through self-learning to obtain the initial segmentation result of the target area. Then, super pixels are used to constrain the boundary of the liver area for obtaining the final segmentation result. Chen proposed an improved slice-to-slice region growing method, which introduced centroid detection and intensity analysis, and applied morphological operations to extract the liver region (6). Gambino proposed a texture-based volume region growth algorithm, which effectively reduced the impact of artificially selected seed points (7). Lu proposed an improved region growing algorithm for liver segmentation (8). Firstly, the original image is preprocessed by the non-linear mapping. Then, the feature region of the liver is selected through human-computer interaction. Finally, it used the quasi-Monte Carlo method to generate seed points in the feature region, improving the region growth criterion. Rafiei proposed an innovative preprocessing and adaptive 3D region growth method, which uses the map intensity and position of the most probable voxel in the probability map as the region growth boundary to limit the region growth so as to realize the dynamic changes of region growth criterion during the training process (9).

The level set method was first proposed by Osher (10). It has become a classic image segmentation algorithm and has been successfully applied to medical image segmentation problems. Yang proposed a semi-automatic method based on level set and threshold, which includes two level set methods (11). Zhou proposed a liver tumor segmentation algorithm of unified level set by combining regional with boundary information, which is better than applying a single information-driven level set method (12). Alirri proposed a method for automatically segmenting the liver from CT dataset (13). This algorithm utilizes the local shape model and the estimated liver intensity range to establish the initial mask, and then the active contour algorithm is utilized to make the initial mask into the liver boundary. Kass first proposed the active contour model (Snake model) in 1988 (14). This method modeled the problem of image segmentation and transformed it into the problem of minimizing

the energy generalization function, which provided a new way of image segmentation. Chi proposed an automatic strategy-based active contour segmentation method for accurate and repeatable liver volume segmentation, which combines rotating template matching, K-means clustering, and local edge enhancement with gradient vector flow model (15). Bereciartua proposed a method for automatic 3D liver segmentation, which achieved liver segmentation by minimizing the fully variable dual (16).

Chen proposed a two-step liver segmentation method based on low-contrast images (17). In the first step, K-Means clustering algorithm and prior knowledge are applied to find and identify liver and non-liver pixels. In the second step, the liver is segmented from the low-contrast image based on graph cutting. Sangewar proposed a new variational model for segmentation of liver regions based on the idea of intensity probability distribution and regional appearance propagation, which overcomes the poor segmentation results caused by the low contrast and edge blur of liver CT images (18).

## Deep Learning Methods

Deep learning (DL), as a branch of machine learning, has shown potentials in medical image segmentation (19). When using deep convolutional neural networks for organ segmentation, thanks to its powerful feature extraction capabilities, it can accurately extract the complex and semantically rich feature information of organs, making the network have high segmentation capabilities (20). The advantages of deep learning are incomparable to traditional machine learning algorithms. Therefore, the current mainstream segmentation algorithms are based on deep learning, and the segmentation accuracy is generally better than traditional segmentation algorithms.

There are various types of segmentation methods. Some researchers applied two-dimensional (2D) convolutional neural networks (CNN) to deal with liver segmentation by learning 2D context of the image (21, 22). Others designed models with 3D contexts only in small voxels due to the high computation cost and memory consumption of 3D CNN (23–25). Furthermore, they used several 2D CNNs that are combined to enhance 2D contexts during the liver segmentation (26, 27). Finally, the 2D and 3D contexts were considered to fuse for training the network (28–31).

For the research of live segmentation, Ben-Cohen directly applied the full convolutional network on a relatively small liver dataset for liver and lesion segmentation. However, the segmentation results were not ideal (32). Christ proposed a way to automatically segment liver and lesions in CT abdominal images using cascaded fully convolutional networks and dense 3D conditional random fields for the joint segmentation of the liver and its lesions to achieve ideal effect (33). Yao proposed a cascade structure to realize automatic segmentation of liver CT images (34). A fully convolutional network is trained to roughly segment the liver, and then the conditional random field model is used as a post-processing refinement liver segmentation to improve the effect.

## METHODS

### Overview of the Framework

This paper proposed a semi-supervised 3D liver segmentation method based on deep convolutional GAN (DCGAN), which consists of the discriminator and generator. Among them, the improved 3D U-Net network is applied as a discriminator to identify real images and generated fake images and obtain the 3D segmentation results of the liver. Then, we design a DCNN based on feature restoration method to generate fake images by the feature map of the real images. The network structure of the optimization segmentation algorithm based on the GAN is shown in **Figure 1**.

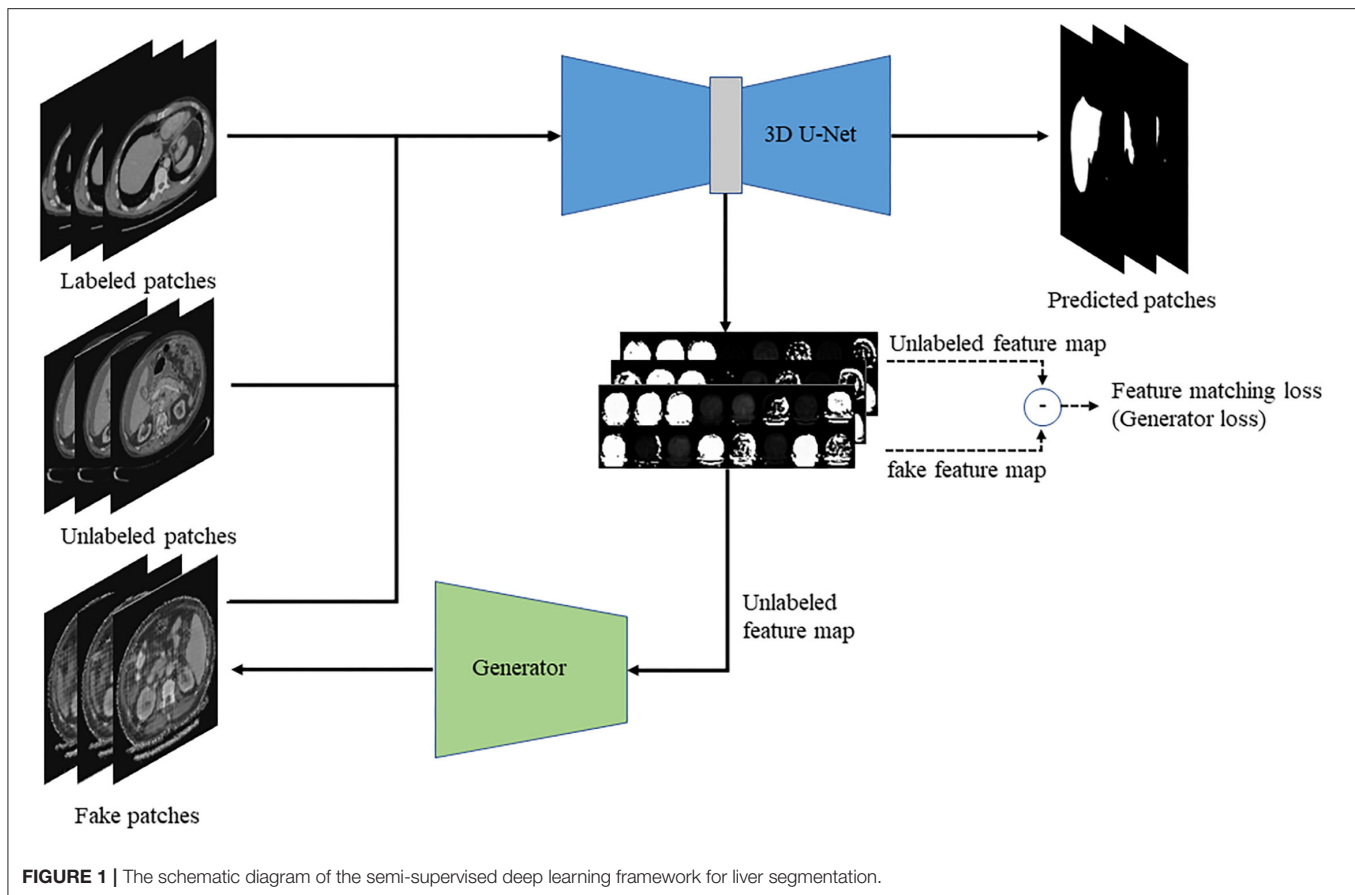
### Preprocessing

Data preprocessing is an important means to improve the effect of deep learning training, which can adjust the overall distribution of the sample to make it more suitable for training. Improving the quality of the sample can make the model easier to fit the feature distribution of the sample. In CT images, CT value is used to measure the density of human tissues or organs by the Hounsfield Unit (HU). The CT value range is generally  $[-1,000\text{HU}, +1,000\text{HU}]$ , in which air is  $-1,000\text{HU}$  and dense bone is  $+1,000\text{HU}$ . Therefore, the CT value in the image needs to be converted into gray value before the liver segmentation. The main steps are as follows:

- (1) CT value truncation: As the highest contrast range of the liver in the images is  $[-200\text{HU}, +200\text{HU}]$ , so we cut the CT value to a certain range, which the CT values smaller and larger than  $-200\text{HU}$  and  $+200\text{HU}$  are set to  $-200\text{HU}$  and  $+200\text{HU}$  to accomplish the CT value truncation. Clipping is used to improve the contrast between the liver and other tissues. This is a key step in preprocessing CT images. Without clipping, the segmentation performance of CT images is poor, and after the clipping processing, the network can also converge faster. In addition, this paper is mainly to segment the liver, which is not sensitive to bony structure information. By clipping the intensity range of CT images, the interference of bones and other tissues can be reduced.
- (2) CT value normalization: CT images in the dataset were obtained from several image acquisition sources with various scanning equipment and imaging environments, which led to different imaging effects and grayscales. Such difference in the gray level has a greater impact on the training process of the samples. Therefore, it is necessary to eliminate the influence of imaging differences as much as possible in the process of converting the CT value to the gray value. The normalized formula is shown in Formula (1):

$$H'(x, y, z) = \frac{H(x, y, z) - HU_{\min}}{HU_{\max} - HU_{\min}}, \quad (1)$$

where the values of  $HU_{\max}$  and  $HU_{\min}$  are  $+200$  and  $-200$ ,  $H(x, y, z)$  represents the CT value of the voxel with coordinates  $(x, y, z)$  in the CT image before normalization, and  $H'(x, y, z)$  represents the normalized value.



**FIGURE 1** | The schematic diagram of the semi-supervised deep learning framework for liver segmentation.

- (3) Gray value interval mapping: The normalization is set to facilitate the calculation and supervision of the training process. In the normalization stage, we multiply the normalized value by 255 according to the range of RGB value and convert it to an integer. Moreover, the normalized value below zero is invisible to the naked eye, therefore, it needs to be multiplied by a reasonable value in order to make the contrast of the image clearer. Thereby the CT value of  $[-200, +200]$  is mapped to the gray value interval of  $[0, 255]$ .

### Improved 3D U-Net Network Structure

The U-Net model was originally designed to solve the task of 2D medical image segmentation, which all network layers in the model are 2D. In order to realize the 3D segmentation of the images, a 3D version of the U-Net model needs to be applied. On the basis of not changing the original encode-decode structure of model, all the network layers in the model are replaced with a 3D type to obtain the 3D U-Net segmentation model (35). The 3D U-Net network structure is used to segment 3D images through the extension of the classic U-Net network in processing data dimensions. Compared with the classic U-Net network structure, in addition to the difference in the dimension of the convolution kernel, the 3D U-Net network only performs three down-sampling operations followed by one

batch normalization (BN) layer. In this paper, the liver 3D segmentation algorithm is to increase the performance of the network by adding some modules on the basis of the 3D U-Net network. The specific improvements and operations are described as follows.

### Squeeze and Excitation (SE) Module

The original 3D U-Net model only uses convolution to extract features. This paper has added SE structure to extract image features, which can weight each feature channel according to the value of the feature image to increase the weight of important features and reduce the weight of the irrelevant features, thereby improving the effect of feature extraction. The SE structure is an attention mechanism based on feature channel weighting (36). In this paper, the 3D SE structure and convolutional layer are combined as the basic convolution module, which is called the SE module. The SE module consists of two convolutional layers. The first convolutional layer adjusts the resolution and the number of channels of the input feature map to a specific size. In addition, it can compress the feature channel to reduce the amount of calculation. The second convolutional layer is utilized in conjunction with the SE structure. The SE structure first performs global pooling on the feature map and applies the bottleneck structure to finally obtain the weights

of each channel with a value range  $[0, 1]$  by the sigmoid activation function. The structure of the SE module is shown in **Figure 2**.

Among them,  $W \times H \times D \times C$  represents the size of the feature map,  $W$ ,  $H$ , and  $D$  represent the width, height, and depth of the feature map, respectively,  $C$  represents the number of channels of the feature map, and  $r$  is the multiple of the restoration of the number of channels in the Excitation operation. Scale is a weighted operation which is followed by the rectified linear unit (ReLU) activation function to output the result.

### Pyramid Pooling Module

The 3D U-Net model only uses three downsamplings to obtain the receptive field of the 3D image, and it elevates the role of shallow features by jump connections allowing the model to determine the importance of different scale receptive fields. In this study, we have introduced the pyramid pooling module to obtain a larger receptive field (37). The pyramid pooling module applies multiple scales of pooling operations to obtain and fuse feature information of multiple scales, which can improve the degree of freedom of the model for multi-scale receptive field selection. It also can add multi-scale information of features without affecting the original features. Furthermore, in order to splice the original feature image with the pooling results of different scales, it is necessary to make the pooling results of different scales to be the same size as the original feature map. This module enables each location to obtain the information of multiple ranges. Thus, the maximum range can directly reach the global size, and the module can quickly acquire a larger variety of information. This article will use the 3D structure of the module to obtain a larger range of receptive fields. Firstly, we applied the  $1 \times 1$  convolution operation to perform feature channel fusion for each scale pooling result, and it is up-sampled to the original feature map size and spliced. Then, the number of feature channels is also reduced through the  $1 \times 1$  convolution operation, and finally spliced with the original feature map as the output.

### Improved 3D U-Net Model

In order to improve the segmentation accuracy of the original 3D U-Net network, an improved network based on 3D U-Net is proposed to perform the 3D segmentation of the liver. Firstly, all the convolutional layers in the original 3D U-Net network are replaced with the SE module. The improved structure is a deep learning model based on the encoder-decoder structure, in which the encoder is composed of SE module and down-sampling, which is mainly responsible for extracting features and expanding the receptive field. The decoder consists of SE module and up-sampling, which has the main function of extracting features and expanding resolution. There is a skip connection between the encoder and the decoder. The skip connection splices the shallow features in the encoder with the deep features in the decoder, and the shallow features provide detailed information for the decoder. The purpose is to make the network pay more attention to the feature information related to the liver and increase its importance while reducing the role of irrelevant information such as background, enabling it to obtain

a finer segmentation boundary. In addition, the pyramid pooling module was introduced to make the network obtain multi-scale feature information and expand the receptive field of network. In the improved 3D U-Net network, the pyramid pooling module is added at the end of the encoding path that has the smallest resolution in the entire network. The modified network structure is shown in **Figure 3**.

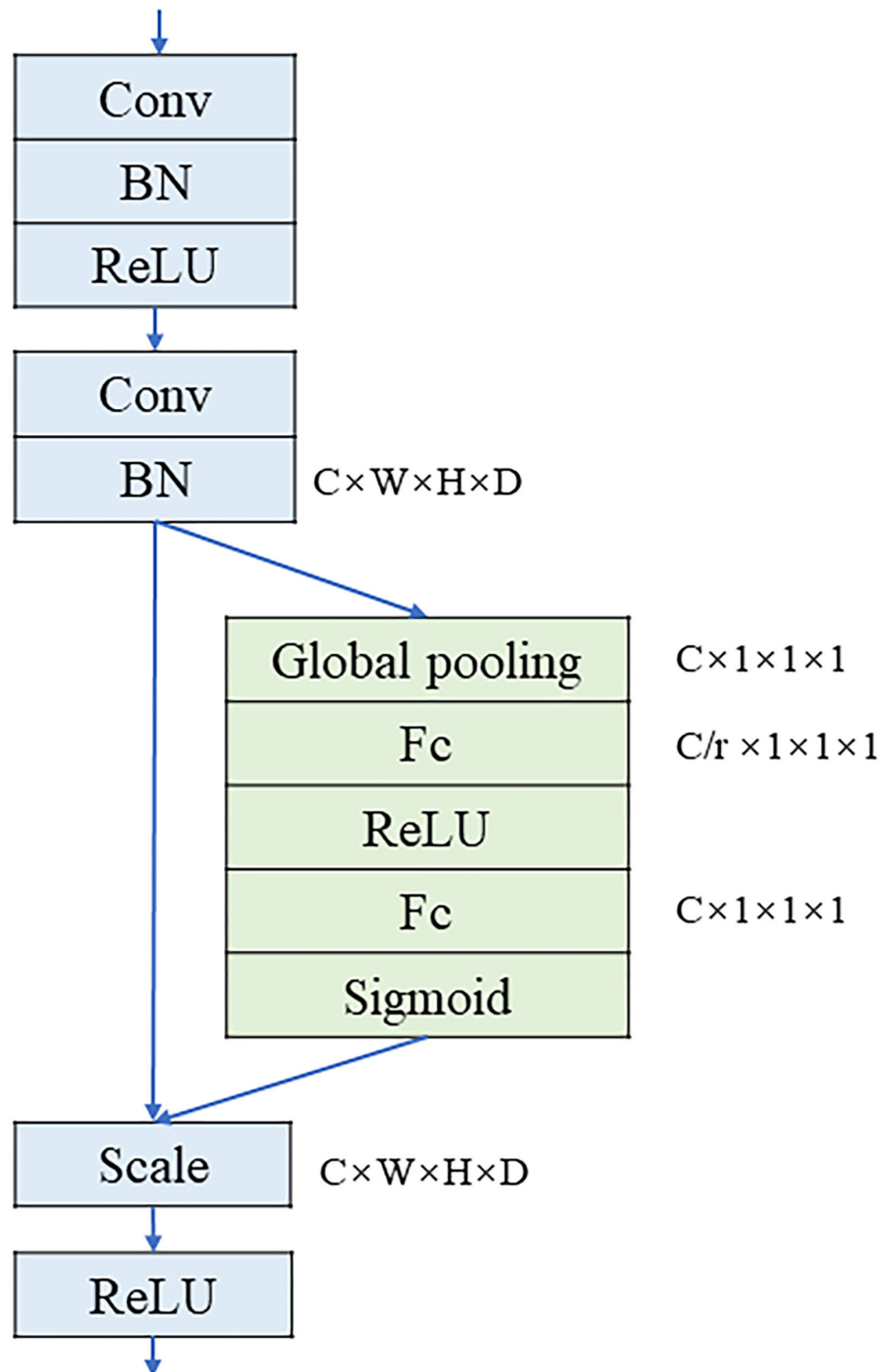
The encoding part is shown in the left half of the figure. In the encoding part, the model applies four SE module groups of which each module group consists of two SE modules. The first convolution step in the first SE module is 2 while the number of channels is increased. After the four SE module groups, we have added the pyramid pooling module that contains 3 parallel average pooling layers with sizes of 1, 2, and 5, respectively. The encoding part includes a total of 4 down-sampling, which are all performed by convolution with a step size of 2. The pyramid pooling module is added to the end of the encoding path that is the place with the smallest resolution in the entire network.

The decoding part is shown in the right half of the figure. The first part of decoding consists of three SE module groups—each of which contains two SE modules. The first SE module has two inputs: one is the output from the previous layer of the decoding part, and the other is the output from the corresponding position of the encoding part. The module first applies the transposed convolution with a step size of 2 to expand the output resolution of the previous layer to double the original resolution, reduce the number of channels to  $1/2$  of the original, and splice the output at the corresponding position of the encoding part. After that, it performs another SE convolution operation. After three SE module groups, the feature image is restored to the original image size using the transposed convolution. After splicing with the original image, it obtains the final segmentation result.

## Deep Convolutional Generative Adversarial Networks (DCGAN)

At present, convolutional neural networks have been widely applied in generative adversarial networks. However, generative adversarial networks lacked a general network architecture until the emergence of DCGAN, which is an unsupervised learning algorithm combining deep convolutional neural networks and generative adversarial networks (38). The design idea of DCGAN is to restrict the network structure based on the original network framework to achieve a more powerful generative model.

The generator of DCGAN generates fake images by convolution and up-sampling of random noise, which is widely used in the task of generating 2D images. To generate a 3D image, the effect of using random noise is very poor because of the difficulty to learn the 3D image distribution through a deep neural network. Learning high-dimensional image distribution is very slow by using noise as input, while using 3D U-Net network as a discriminator is very quick to converge during the training process. The contradiction will cause the problem of gradient dispersion. In addition, we have used random noise to generate CT images. However, the contours of the CT images are difficult to generate, let alone the internal distribution. In order to generate more realistic 3D images, it is necessary to add more

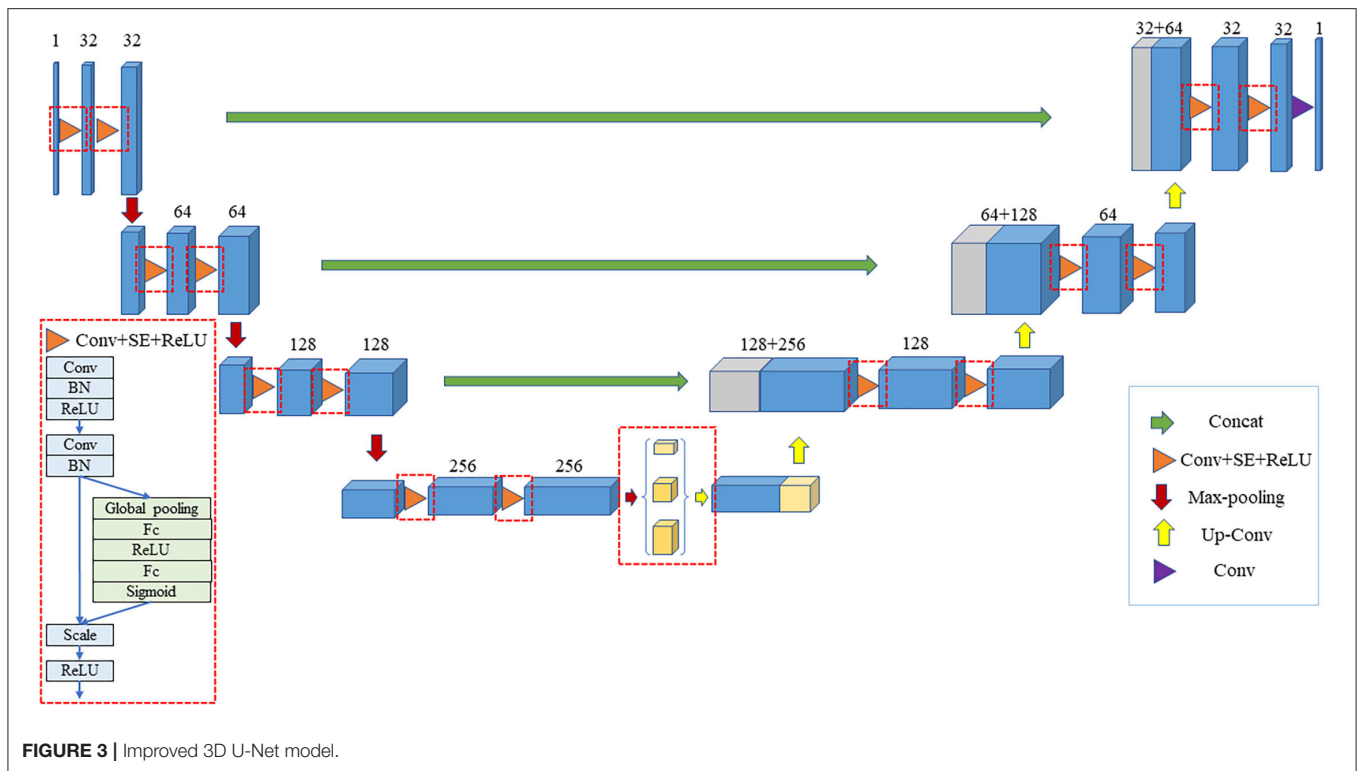


**FIGURE 2 |** The structure of SE module.

real image distribution information to replace random noise, which can speed up the learning rate of the generator.

In this article, we design a convolutional neural network based on feature restoration method. By extracting the feature map generated by the improved 3D U-Net network, a part of the

feature map is randomly selected as the input of neural network due to the following reasons. Recovering all the feature maps is a reverse process of feature extraction, which the distribution obtained is the same as the real image. If a part of the feature map is selected, the generator will learn the real image distribution



and complete the missing parts. Moreover, the image obtained is different from the real image which increases the variety of images and achieves the purpose of expanding the dataset. Because the feature map is randomly selected, the missing part is also random, and, therefore, the generator can be trained to restore the real image at any position. In addition, the reason why the feature map, rather than the partially missing real image, is applied is that the real image contains much useless information, which causes the generator to converge slowly. Through up-sampling and convolution operations, a fake image with the same size as the real image slice is obtained. After that, the feature map of fake and real images are, respectively, extracted through the improved 3D U-Net network, and the mean difference between the two feature maps is applied as the loss. Then, the network parameters of generator are updated through multiple iterations, making the generator better restore the feature map and, also, making a fake image closer to the real image. The generator structure is shown in **Figure 4**.

## The Definition of Label and Loss Function

The proposed segmentation optimization algorithm is semi-supervised. There are fully supervised learning for labeled data and unsupervised learning for unlabeled data and fake images. Therefore, the labels need to be redefined so that the discriminator can identify fake images. The original label defines the background and liver as 0 and 1. Now a new label category needs to be added to mark the fake images, in which the label is defined as 2. The output size of the improved 3D U-Net network is  $H \times W \times D \times 3$ , 3 represents the number of labels, the

output vector of each voxel is  $[l_{i,1}, l_{i,2}, l_{i,3}]$ , which represents the probability that the current voxel is false. In order to learn from unlabeled data, its output is forced to be a label of the real data that is achieved by maximizing the output vector.

The loss function of the discriminator is shown in Formula (2):

$$L_{discriminator} = L_{labeled} + L_{unlabeled} + L_{fake} \quad (2)$$

where  $L_{discriminator}$  is the loss of the discriminator,  $L_{labeled}$  is the loss of labeled data,  $L_{unlabeled}$  is the loss of unlabeled data, and  $L_{fake}$  is the loss of fake images generated by the generator.

For labeled data, we use cross entropy loss function to calculate as shown in Formula (3):

$$L_{labeled} = -E_{x,y \sim p_{data}(x,y)} \sum_{i=1}^{H \times W \times D} \log P_{model}(y_i | x, y_i < K+1) \quad (3)$$

where  $x$  represents the input image,  $K$  represents the number of classes labels,  $y$  represents labeled image,  $x, y \sim p_{data}(x, y)$  represents that the input image is labeled, and  $P_{model}(y_i | x, y_i < K+1)$  represents the probability of the voxel prediction category is  $y_i$  in the image.

For unlabeled data, the loss function is shown in Formula (4):

$$L_{unlabeled} = -E_{x \sim p_{data}(x,y)} \sum_{i=1}^{H \times W \times D} \log (1 - P_{model}(y_i | x, y_i < K+1)) \quad (4)$$

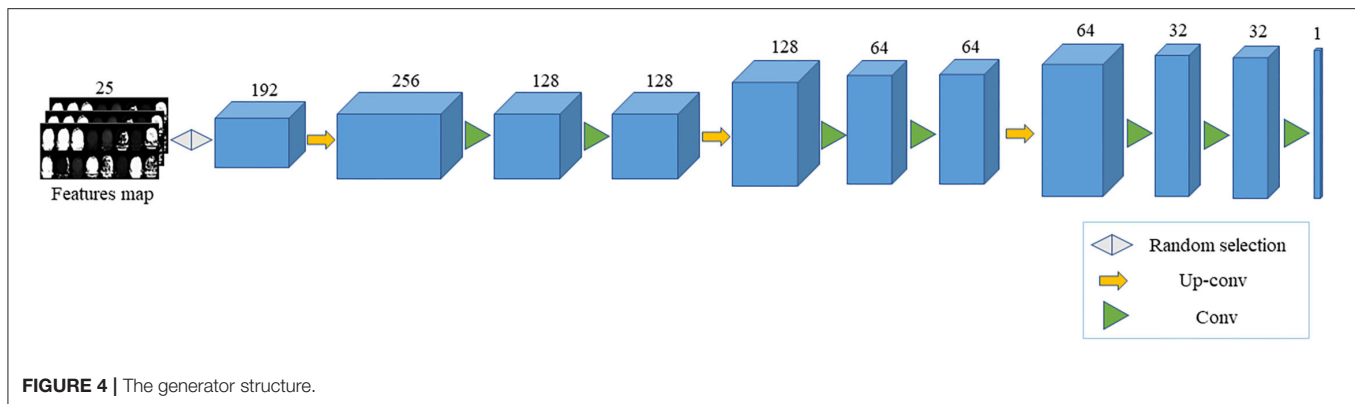


FIGURE 4 | The generator structure.

where  $x$  represents the input image,  $x \sim p_{data}(x, y)$  represents that the input image is unlabeled, and  $p_{model}(y_i|x, y_i < K+1)$  represents the probability of the voxel prediction category is fake images.

For the fake images, the loss function is shown in Formula (5):

$$L_{fake} = -E_{e \sim Encoder(x)} \sum_{i=1}^{H \times W \times D} \log P_{model}(y_i = K+1 | G_{\theta_G}(e)) \quad (5)$$

where  $x$  represents the input unlabeled image,  $e \sim Encoder(x)$  represents the fake images generated by the generator based on the unlabeled image distribution, and  $p_{model}(y_i = K+1 | G_{\theta_G}(e))$  represents the probability that the voxel prediction category is a fake image in the fake images.

The loss function of generator is shown in Formula (6):

$$L_{generator} = \|E_{e \sim Encoder(x)} f(x) - E_{e \sim Encoder(x)} f(G_{\theta_G}(e))\|_2^2 \quad (6)$$

where  $x$  represents the input unlabeled image,  $f(x)$  represents the extracted feature map from the unlabeled images through the 3D U-Net network decoder,  $e$  represents the fake image generated by the generator,  $f(G_{\theta_G}(e))$  represents the extracted feature map from the fake images through the 3D U-Net network decoder.

## EXPERIMENTS AND DISCUSSION

### Datasets

LiTS-2017 is a liver tumor segmentation challenge dataset launched in 2017 (39). The data set includes 100 3D abdominal CT scan images (nii format). In the experiment, 60 images were selected as the training set, and the validation and test set contain 20 images, respectively. The training set is divided into 50 labeled and 10 unlabeled images. The labeled and unlabeled images are applied to train the segmentation network and generate fake images. Firstly, we scale each image in the dataset and intercept the liver position to change the size of the image to  $256 \times 256 \times N$ . Then, the image is cut into patches with the size of  $256 \times 256 \times 16$  to obtain a total of 500 patches for training, which the number of labeled and unlabeled patches is 400 and 100. In the meantime, the validation and test set are processed to contain 200 patches, respectively. Moreover, in order to verify

the generalization of the proposed model, we applied another dataset, named KiTS19, which is a kidney tumor segmentation challenge dataset launched in 2019 (40). The dataset includes 200 3D abdominal CT scan images (nii format). For the dataset, we perform the same operations as the LiTS-2017. Finally, a total of 1,000 patches are generated for training, of which the number of labeled and unlabeled patches is 900 and 100. In the meantime, the validation and test set are processed to contain 210 patches, respectively.

### Evaluation Metrics

To evaluate the segmentation performance of the proposed network, we adopted the widely used segmentation evaluation metric: Dice coefficient (Dice) (41). Dice is a function of ensemble similarity measurement to calculate the similarity of two samples with the range [0, 1] at the pixel level. The real target (Ground truth) appears in a certain area  $A$ , and the target area of the model prediction result is  $B$ . Then the calculation of Dice is shown in formula (7):

$$Dice(A, B) = \frac{2|A \cap B|}{A + B} \quad (7)$$

where  $|A \cap B|$  represents the intersection between  $A$  and  $B$ ,  $|A|$  and  $|B|$  represent the total number of  $A$  and  $B$  pixels, respectively. Because there are overlapping elements between  $A$  and  $B$  in the denominator  $A + B$ , it adds a coefficient 2 to the numerator. In the problem of medical image segmentation,  $A$  and  $B$  represent the real label image and the segmented image predicted by the model.

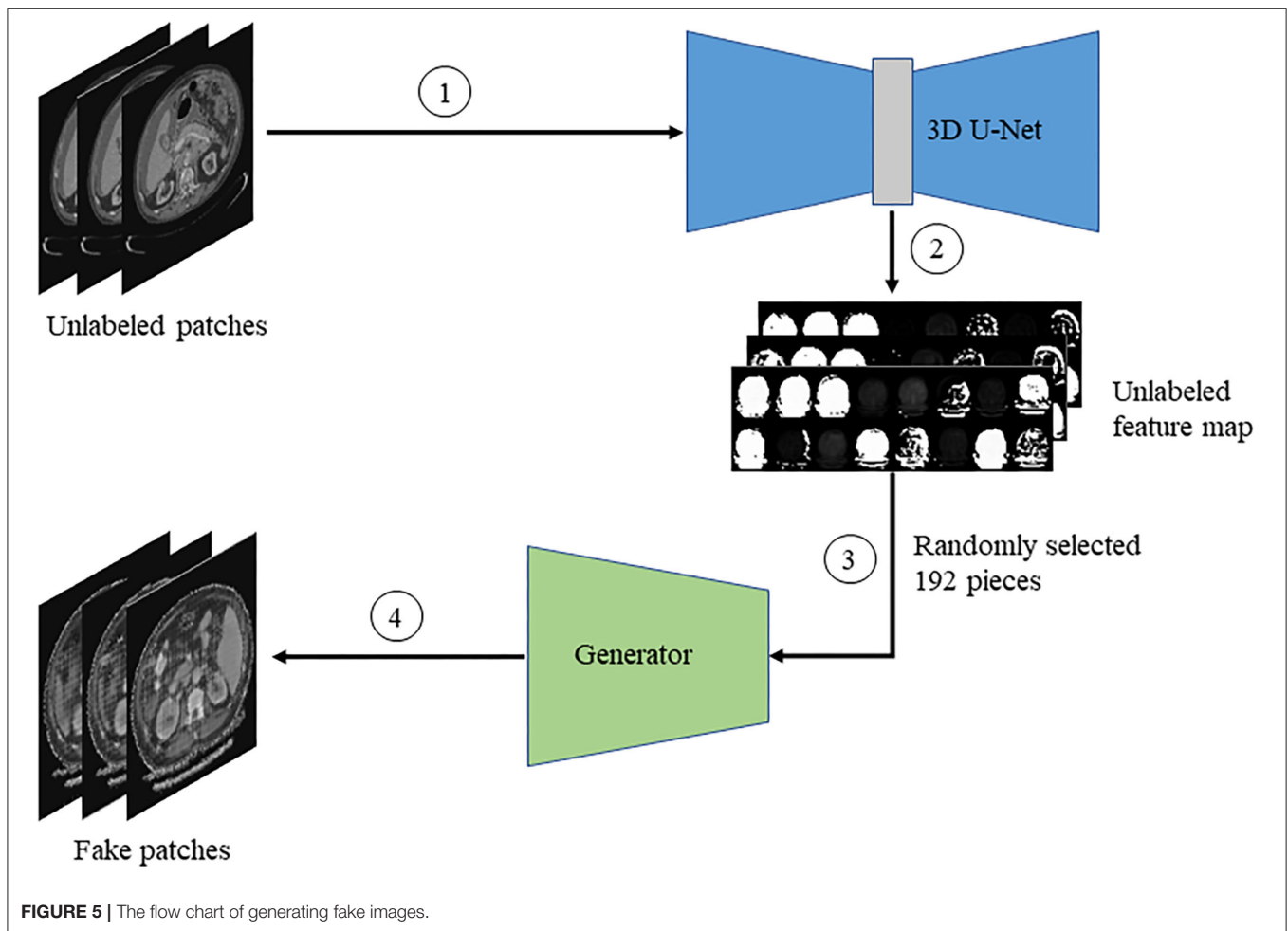
### Training Process

#### Generation of Fake Images

Firstly, the unlabeled image is input into the improved 3D U-Net network to obtain the feature map. Then, we randomly select a part of the feature map as the input of the generator. Finally, the generator generates fake images base on feature restoration method. The flowchart is shown in Figure 5.

#### Discriminator Training

The labeled, unlabeled, and fake images are, respectively, passed into the discriminator, and then the loss of the discriminator



is calculated to update the gradient of the discriminator. The flowchart is shown in **Figure 6**.

### Generator Training

The fake and unlabeled images are, respectively, passed to the improved 3D U-Net network after the gradient update to obtain the feature map, and the generator loss is calculated according to the similarity of the feature map to update the gradient of the generator. The flowchart is shown in **Figure 7**.

### The Results of Generated Images

The experiment proves that the fake image is very close to the real image by the feature restoration method, and the speed of the network learning distribution is faster than that of the network using random noise as the input, which avoids the problem of gradient dispersion. The fake images generated by the generator is shown in **Figure 8**.

## Experimental Results

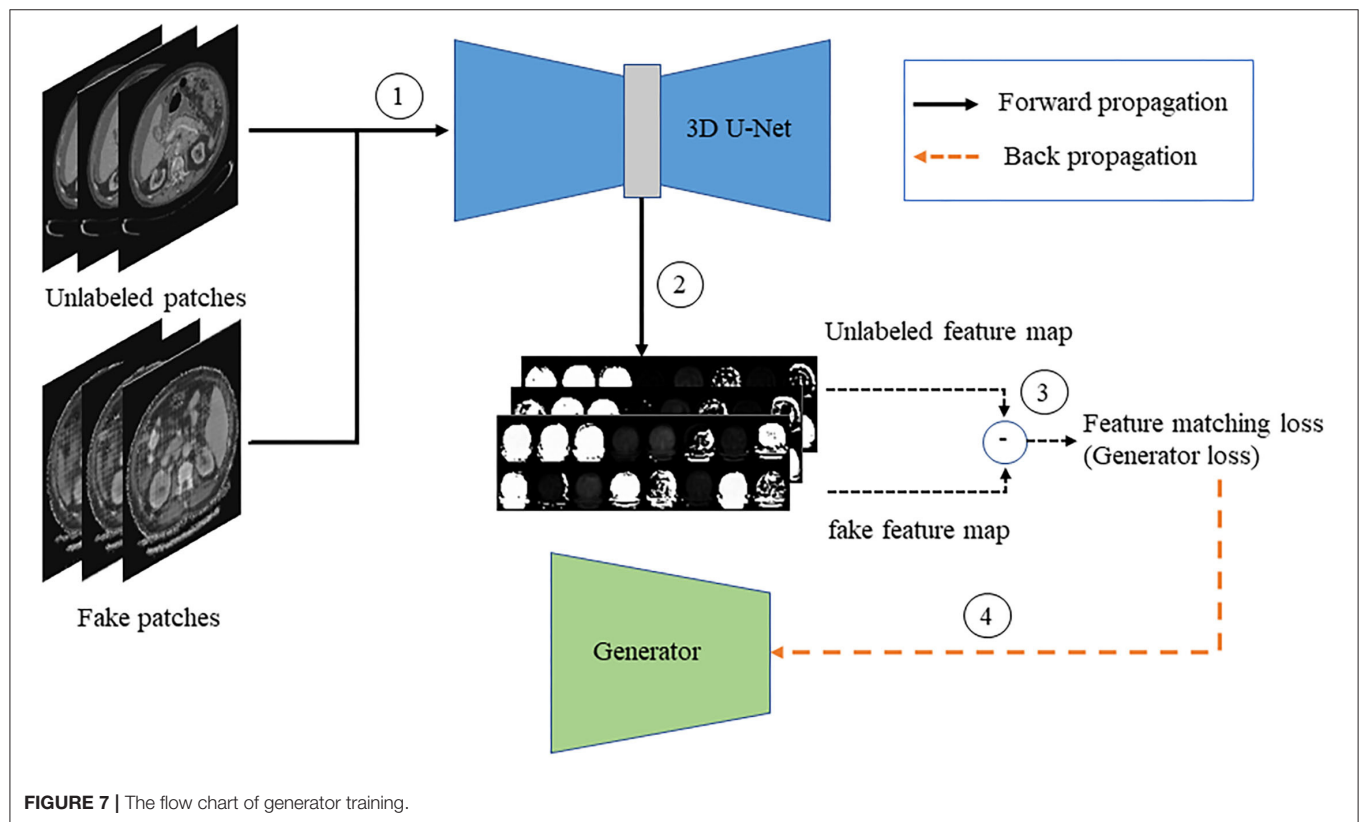
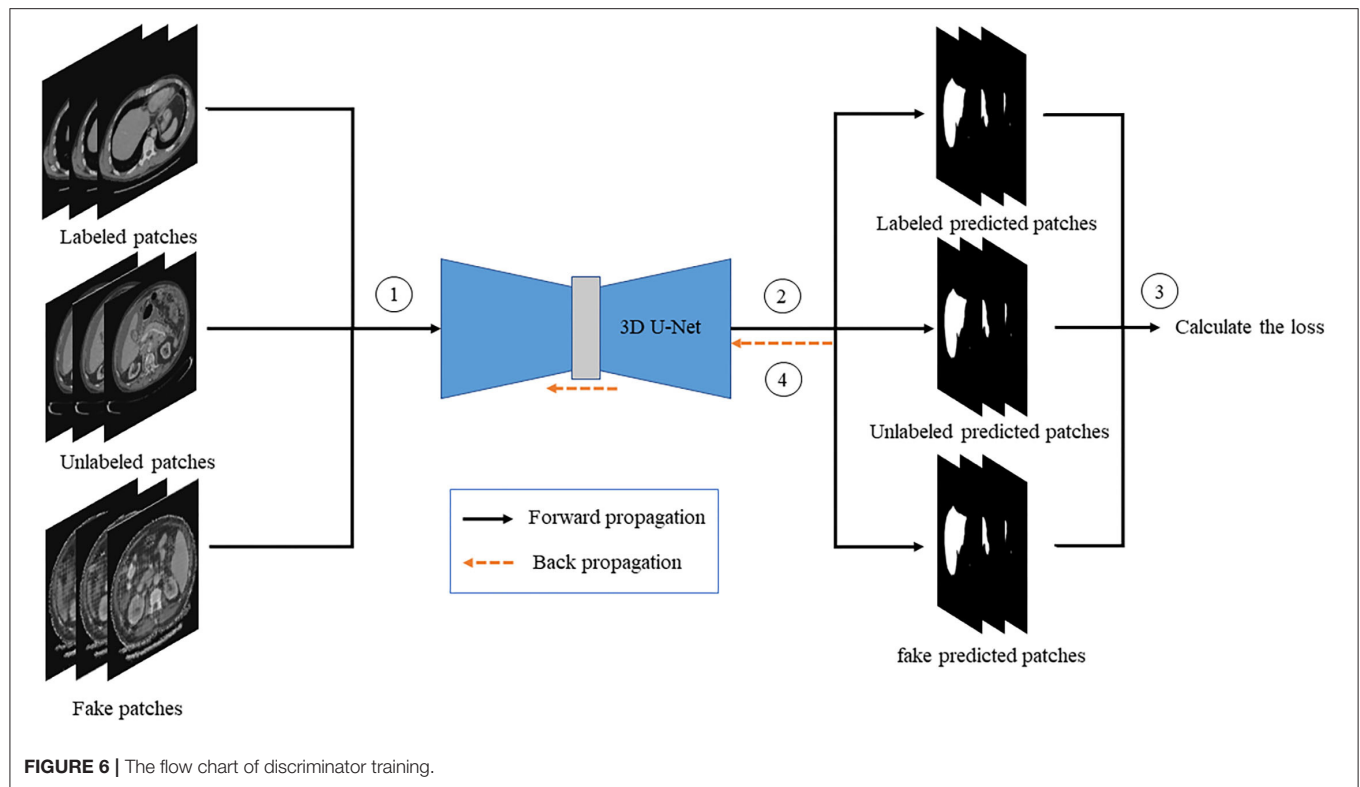
### Comparison of Experimental Results

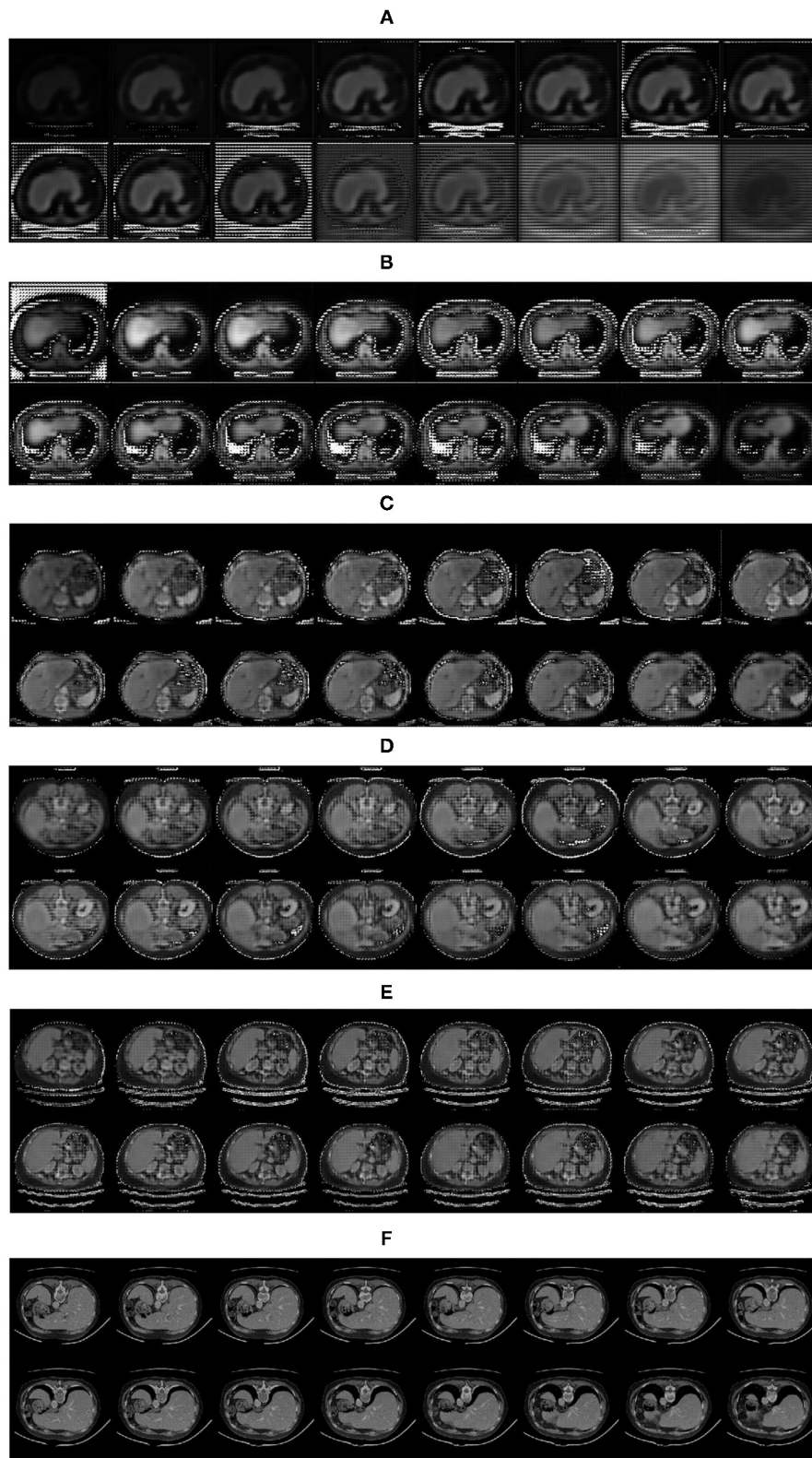
For each group of experiments, the segmentation results were verified on the same public dataset LiTS-2017. In the process of network training, the Adam optimization algorithm is applied,

the learning rate of discriminator and generator are set to  $10^{-4}$  and  $3 \times 10^{-4}$ , and the batch size is set to 1. We have trained the original 3D U-Net network, the improved 3D U-Net network, and the generative adversarial network based on the feature restoration method, respectively, and set up comparative experiments for the improvement of each part.

In this study, each ablation experiment uses the same training set, verification set and the test set, of which results are shown in **Table 1**. From the comparison experiment, it can be seen that the improved 3D U-Net network has improved compared with the original 3D U-Net network in the validation set. The improved 3D U-Net network also has a certain improvement over the original 3D U-Net in the test, especially the introduction of the GAN. The addition of the GAN not only improves the segmentation effect of the training set and validation set, but also improves the performance of the test set. For the improvement of segmentation performance, we have analyzed the contribution of each module, which are presented as follows.

Similar to the attention mechanism, the SE module integrated in the original 3D U-Net network extracts the relationship between the channels using the global information of each channel and weights each channel. The SE module can determine the importance of various features according to the value of each





**FIGURE 8 |** Fake images generated by the generator: **(A)** 1,000 iterations; **(B)** 5,000 iterations; **(C)** 10,000 iterations; **(D)** 15,000 iterations; **(E)** 20,000 iterations; **(F)** Real images.

**TABLE 1** | Comparison of experimental results on the LITS-2017 dataset.

Method	Validation set dice coefficient	Test set dice coefficient
3D U-Net	0.9160	0.881
3D U-Net+SE+Pyramid pooling	0.9304	0.905
3D U-Net+SE+Pyramid pooling+GAN	<b>0.9638</b>	<b>0.942</b>

The bold values represent the highest score.

**TABLE 2** | Comparison of experimental results with other methods.

Method	Dice coefficient
DenseNet (42)	0.923
3D DenseUNet-65 (43)	0.929
FCN+ACM (44)	0.943
GIU-Net (45)	<b>0.951</b>
3D U-Net+SE+Pyramid pooling+GAN	0.942

The bold values represent the highest score.

channel, which can effectively improve the performance of the model. Adding the pyramid pooling module, the application of multi-scale pooling layer allows the model to obtain different sizes of receptive fields, while further expanding the receptive field of model. The two operations, together, improve the feature extraction capability of the model. Furthermore, the GAN, based on the feature restoration method, can generate more diverse images containing real distribution information. On a limited dataset, it can improve the segmentation performance and generalization of the model, which is a promising way for the generation of 3D medical images.

Finally, we have compared the proposed algorithm with others. As shown in **Table 2**, our algorithm is better than the DenseNet and 3D DenseUNet-65 algorithms (42, 43). In addition, the performance of the proposed method is slightly lower than the FCN+ACM method (44). The GIU-Net algorithm (45) outperformed our proposed model for liver segmentation. A suitable explanation for that is that they utilized more data to train the model, while our model is trained on a limited dataset. In the future, we will further improve the performance and reduce the complexity of the model.

## The Segmentation Results

The partial segmentation result is selected from the test set as shown in **Figure 9**. The first column is the original image, the second column is the ground truth, and the third and last column are the segmentation results by the original 3D U-Net and proposed algorithm. It can be found that using the proposed algorithm, the 3D segmentation results of the liver are very close to the real labels, which the problem of poor segmentation accuracy of head and tail regions of the liver has been improved. In order to further analyze the segmentation results, we have also supplied the 3D surface plots with the color bar of the Hausdorff Distance (HD).

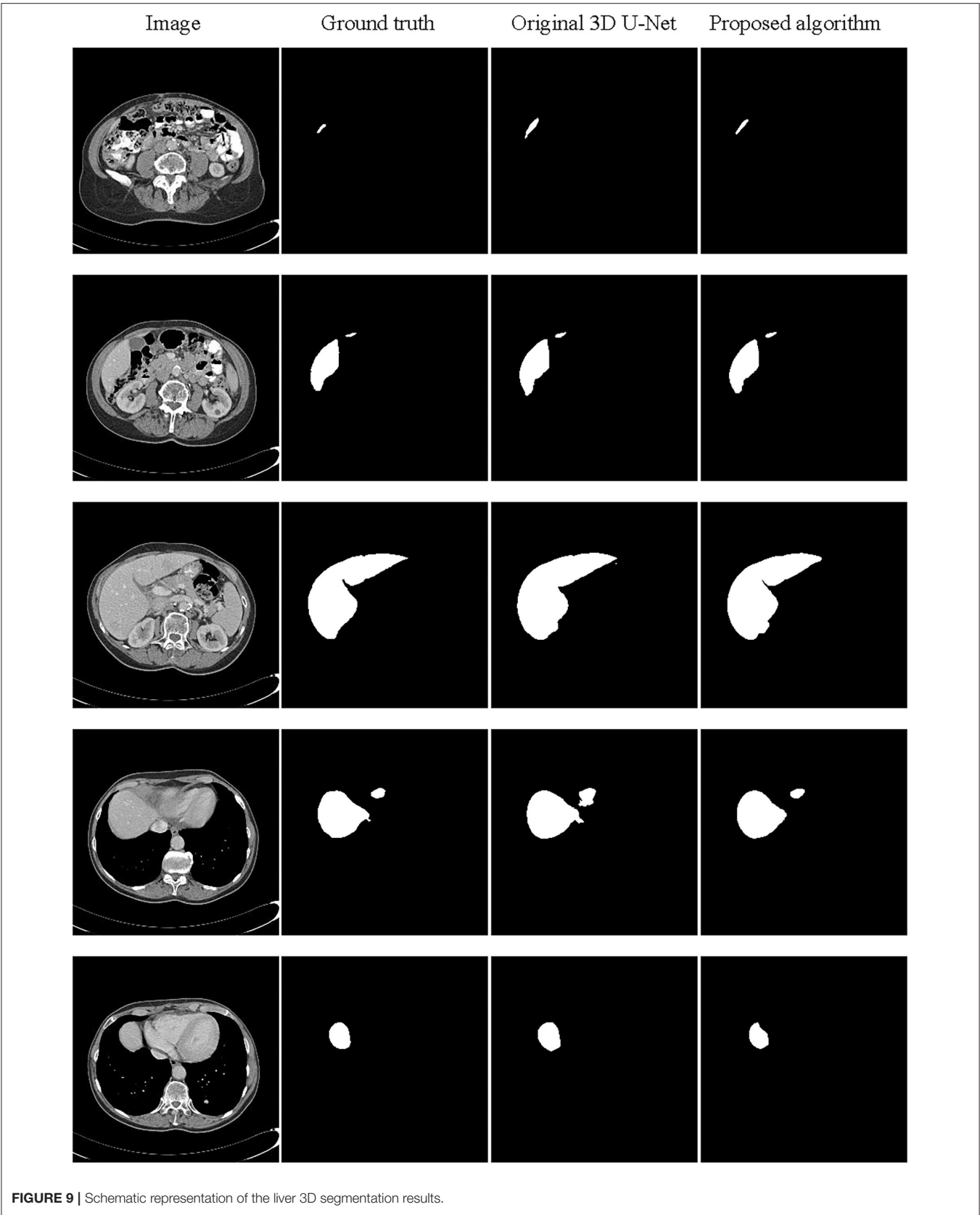
In **Figure 10**, the first row is the 3D surface plot generated by the original 3D U-Net network for two representative test data, while the second row is generated by the proposed algorithm. As we can see in the left column, the HD values of these two algorithms are relatively high. The 3D surface plot above, generated by the original 3D U-Net network, has a redundant prediction (the blue area) and obtains an HD value of 18.68 mm. The 3D surface plot below is generated by the proposed algorithm that has some outliers with the purple area, which has a greater impact on HD obtaining the HD value of 15.33 mm. In addition, the average symmetric surface distance (ASSD) values are 0.93 and 0.64 mm for the original 3D U-Net network and proposed algorithm, respectively. For this test data, the segmentation results of the two algorithms are not very good. However, the proposed algorithm still performs better than the original 3D U-Net network. For the right column, the segmentation effect of the two algorithms is better than the previous one. The 3D surface plot above was generated by the original 3D U-Net network whose boundary segmentation is not ideal. In this case, the HD value of 9.70 mm. The 3D surface plot below is generated by the proposed algorithm that is better than that the original 3D U-Net network and reduces the HD value (8.06 mm). Moreover, the average symmetric surface distance (ASSD) value of the original 3D U-Net network and proposed algorithm are 0.93 and 0.87 mm. For the scores of HD and ASSD, they are not as ideal as the Dice coefficient, because that we use the Dice coefficient to determine the end of training and the model is more inclined to calculate Dice to a certain degree. In the future, we will further improve the performance of the model.

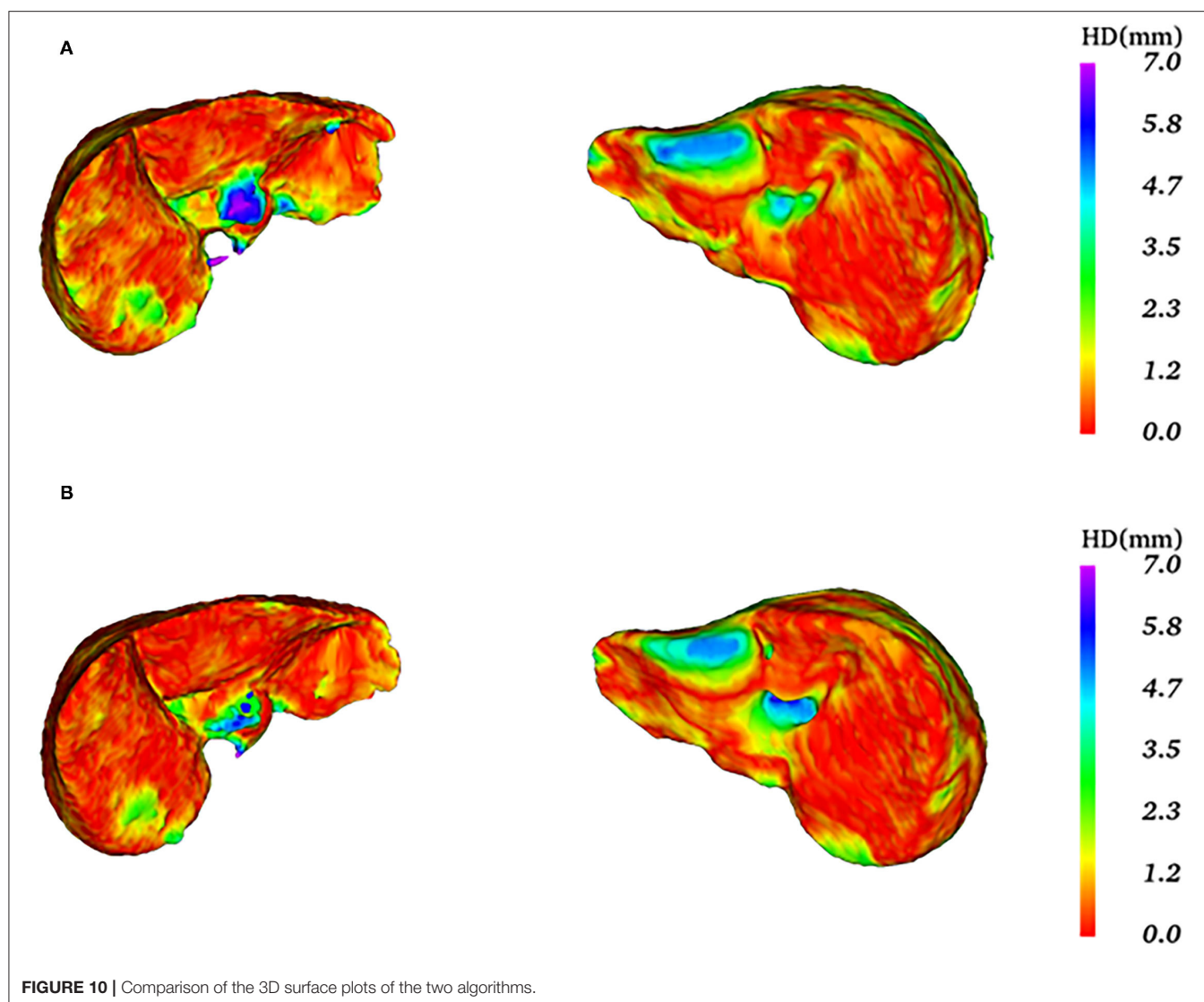
## Evaluation of Model Generalization

Generally speaking, a network model is proposed to solve and optimize a certain problem for obtaining satisfactory results. Therefore, most of models have poor generalization ability and can only be applied to a certain area or dataset. In order to establish a general network model, the accuracy may be sacrificed. However, thanks to the powerful generation ability of the generative adversarial network, the distribution of different organs can be learned through iteration to generate fake images. By this expansion of the dataset, the generalization ability of the model can be greatly improved. In this study, the proposed 3D liver segmentation model is applied to the KiTS19 kidney dataset to test the generalization ability of the model. The results are shown in **Table 3**.

It can be seen from the experimental results that the model still performs well for the 3D segmentation of the kidney. Comparing with the classic 3D U-Net network, the accuracy of kidney segmentation is significantly improved on the verification set. Furthermore, the segmentation ability of the model is still strong on the test set, even exceeding the performance of the verification set.

The partial segmentation result is selected from the test set as shown in **Figure 11**. The left column is the original image, the middle column is the annotated kidney image, and the right column is the segmentation result. It can be found that the 3D segmentation results of the kidney are also very close to the real





labels. Since it is only to verify the generalization of the image generation method based on feature restoration, it has not been further analyzed. In the future, we will further study and verify the general 3D image generation method.

## The Contribution and Future Work

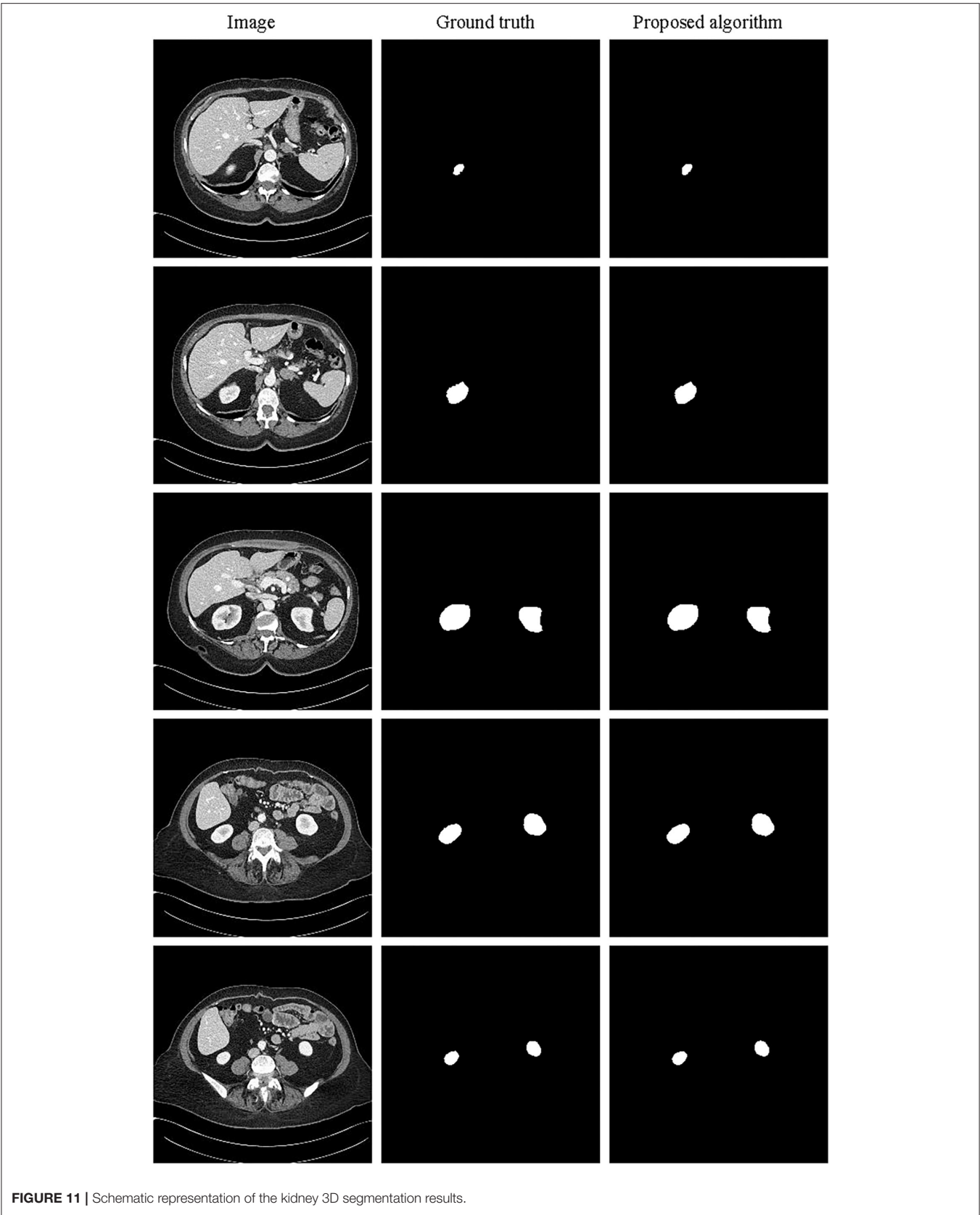
Aiming at the problem of low segmentation accuracy of the original 3D U-Net network, an improved network based on 3D U-Net is proposed to perform the 3D segmentation of the liver. In order to make the network pay more attention to the characteristic information of the liver and reduce the role of irrelevant information, such as background, we introduce the squeeze and excitation (SE) module to the network. Meanwhile, in order to allow the network to obtain feature information of multiple scales and expand the receptive field of the network, we also introduce the pyramid pooling module to the model. Through the combination of the two modules, we have improved the overall segmentation performance of the liver.

**TABLE 3 |** Comparison of experimental results on the KiTS19 dataset.

Method	Validation set dice coefficient	Test set dice coefficient
3D-U-Net	0.906	0.871
3D-U-Net+SE+Pyramid pooling+GAN	<b>0.959</b>	<b>0.959</b>

*The bold values represent the highest score.*

In view of the lack of labeled 3D data, we embed the improved 3D U-Net network in the GAN as the discriminator and propose a semi-supervised liver segmentation method. The limited labeled images and unlabeled images are used to train the learning model to generate fake images for expanding the dataset. Aiming at the poor quality of generating 3D abdominal fake images by using random noise as input, a DCNN based on feature restoration method is designed to



generate more realistic fake images using randomly selected feature maps, which is embedded in the GAN as the generator. Based on the feature restoration method, the generator can make better use of the real image distribution information to generate more realistic images, which increases the diversity of the images.

However, the network has a large number of parameters, which leads to a long training period when the computing resources are limited. Because the data is 3D volume of the liver, the encoder-decoder structure is applied to extract features and restore 3D images every time during the training. In addition, the generator needs to use the random feature map generated by the encoding part of the improved 3D U-Net model to generate fake images. Therefore, it takes up massive video memory during the training of the model. In this study, the training of the proposed model (100 epochs) required approximately 50 h on a single NVIDIA GTX 1080Ti with 11 GB, indicating that each epoch takes about 30 min. The total parameters of the proposed network are about 150 million, which is a relatively complex model.

Because the images take up too much memory, we can optimize the network structure and adjust the efficiency of the video memory to reduce the training time of the model. Moreover, in the field of medical image processing, there are many methods that use GAN to expand datasets. However, by now, there is no meaningful and universal quantitative way to judge the authenticity of the composite images generated by these methods. Therefore, the improvement and application of generative adversarial networks in the field of medical image processing is a direction worthy of in-depth research.

## CONCLUSION

In this study, we mainly conduct a lot of research on the 3D segmentation of liver CT images, which has mainly achieved the following research results. Firstly, in view of the poor effect of 3D U-Net network feature extraction and insufficient accuracy of liver segmentation results, the SE

module and pyramid pooling module are introduced into the 3D U-Net network to improve the accuracy of the segmentation results. Secondly, in view of the difficulty in obtaining labeled 3D CT images, the improved 3D U-Net network is embedded in generative adversarial network as the discriminator. In view of the poor quality of using random noise as input to generate 3D abdominal CT fake images, more real image distribution information is added to the input of the network, and a deep convolutional neural network is designed as the generator based on feature restoration method to generate more realistic fake images. Finally, the network model was applied in the 3D segmentation of kidney to test the generalization ability of the model, which have showed that the model can also obtain better segmentation results on the kidney dataset.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

RH and SX conceived the study. SX performed the design and implementation of the algorithm. YY and YashuL helped the algorithm design and implementation. RH, SX, and HZ wrote the manuscript. QL commented on and approved the manuscript. NZ and YangL contributed to the segmentation of liver. All authors have read and approved the final manuscript.

## FUNDING

This work was supported by the Collaborative Innovation Center for Prevention and Treatment of Cardiovascular Disease of Sichuan Province (CICPTCDSP) [Grant Number xtcx2019-01] (to HZ).

## REFERENCES

1. Mondal AK, Dolz J, Desrosiers C. Few-shot 3D multi-modal medical image segmentation using generative adversarial learning. arXiv:1810.12241 (2018).
2. Apollon Z, Dionissios K, Dimitrios K, Ioannis M, Vasileios M. A hybrid segmentation approach for rapid and reliable liver volumetric analysis in daily clinical practice. In: *IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE)*, Belgrade, Serbia. (2015). p. 1–6. doi: 10.1109/BIBE.2015.7367715
3. Foruzan AH, Chen YW, Zoroofi RA, Furukawa A, Sato Y, Hori M. Multi-mode narrow-band thresholding with application in liver segmentation from low-contrast CT images. In: *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. (2009). p. 1293–6. doi: 10.1109/IIH-MSP.2009.78
4. Seong W, Kim JH, Kim EJ, Park JW. Segmentation of abnormal liver using adaptive threshold in abdominal ct images. In: *IEEE Nuclear Science Symposium & Medical Imaging Conference*. (2010). p. 2372–5.
5. Farzaneh N, Habbo-Gavin S, Soroushmehr S, Patel H, Fessell DP, Ward KR, et al. Atlas based 3D liver segmentation using adaptive thresholding and superpixel approaches. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. (2017). p. 1093–7. doi: 10.1109/ICASSP.2017.7952325
6. Chen Y, Wang Z, Zhao W, Yang X. Liver segmentation from CT images based on region growing method. In: *2009 3rd International Conference on Bioinformatics and Biomedical Engineering*. (2009). p. 1–4. doi: 10.1109/ICBBE.2009.5163018
7. Gambino O, Vitabile S, Re GL, Tona GL, Librizzi S, Pirrone R, et al. Automatic Volumetric Liver Segmentation Using Texture Based Region Growing. In: *2010 International Conference on Complex, Intelligent and Software Intensive Systems*. (2010). p. 146–52. doi: 10.1109/CISIS.2010.118
8. Lu X, Wu J, Ren X, Zhang B, Li Y. The study and application of the improved region growing algorithm for liver segmentation. *Optik Int J Light Electron Opt*. (2014) 125:2142–7. doi: 10.1016/j.ijleo.2013.10.049
9. Rafiei S, Karimi N, Mirmahboub B, Najarian K, Soroushmehr S. Liver segmentation in abdominal CT images using probabilistic atlas and adaptive 3D region growing. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. (2019). p. 6310–3. doi: 10.1109/EMBC.2019.8857835

10. Osher S, Sethian JA. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. *J Comput Phys.* (1988) 79:12–49. doi: 10.1016/0021-9991(88)90002-2
11. Yang X, Yu HC, Choi Y, Lee W, Wang B, Yang J, et al. A hybrid semi-automatic method for liver segmentation based on level-set methods using multiple seed points. *Comput Meth Prog Biomed.* (2014) 113:69–79. doi: 10.1016/j.cmpb.2013.08.019
12. Zheng Z, Zhang X, Xu H, Liang W, Zheng S, Shi Y. A unified level set framework combining hybrid algorithms for liver and liver tumor segmentation in CT images. *BioMed Res Int.* (2018):1–26. doi: 10.1155/2018/3815346
13. Alirri OI, Rahni A. Automatic liver segmentation from ct scans using intensity analysis and level-set active contours. *J Eng Sci Technol.* (2018) 13:3821–39. doi: 10.1007/s11548-018-1801-z
14. Kass M. Active contour models. *Int J Comput.* (1988) 1:321–31. doi: 10.1007/BF00133570
15. Chi Y, Cashman P, Bello F, Kitney RI. A discussion on the evaluation of a new automatic liver volume segmentation method for specified CT image datasets. In: *MICCAI Workshop 3D Segmentation in the Clinic: A Grand Challenge, 2007* (2008).
16. Bereciartua A, Picon A, Galdran A, Iriondo P. Automatic 3D model-based method for liver segmentation in mri based on active contours and total variation minimization-sciencedirect. *Biomed Signal Process Control.* (2015) 20:71–7. doi: 10.1016/j.bspc.2015.04.005
17. Chen YW, Tsubokawa K, Foruzan AH. Liver segmentation from low contrast open MR scans using K-means clustering and graph-cuts. In: *Advances in Neural Networks-ISNN 2010, 7th International Symposium on Neural Networks, Shanghai, China.* (2010). doi: 10.1007/978-3-642-13318-3\_21
18. Sangewar S, Peshattiwari AA, Alagade V, Balpande R. Liver segmentation of CT scan images using K means algorithm. In: *International Conference on Advanced Electronic Systems (ICAES).* (2013). p. 6–9. doi: 10.1109/ICAES.2013.6659350
19. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436. doi: 10.1038/nature14539
20. Qi D, Yu L, Hao C, Jin Y, Heng PA. 3D deeply supervised network for automated segmentation of volumetric medical images. *Med Image Anal.* (2017) 41:40. doi: 10.1016/j.media.2017.05.001
21. Han Y, Ye JC. Framing U-Net via deep convolutional framelets: application to sparse-view CT. *IEEE Trans Med Imag.* (2018) 37:1418–29. doi: 10.1109/TMI.2018.2823768
22. Seo H, Huang C, Bassenne M, Xiao R, Xing L. Modified U-Net (MU-NET) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images. *IEEE Trans Med Imag.* (2019) 39:1316–25. doi: 10.1109/TMI.2019.2948320
23. Deng Z, Guo Q, Zhu Z. Dynamic regulation of level set parameters using 3D convolutional neural network for liver tumor segmentation. *J Healthc Eng.* (2019) 2019:1–17. doi: 10.1155/2019/4321645
24. Alalwan N, Abozeid A, ElHabshy AA, Alzahrani A. Efficient 3D deep learning model for medical image semantic segmentation. *Alexandria Eng J.* (2020) 60:1231–9. doi: 10.1016/j.aej.2020.10.046
25. Meng L, Tian Y, Bu S. Liver tumor segmentation based on 3D convolutional neural network with dual scale. *J Appl Clin Med Phys.* (2020) 21:144–57. doi: 10.1002/acm2.12784
26. Budak Ü, Guo Y, Tanyildizi E, Sengür A. Cascaded deep convolutional encoder-decoder neural networks for efficient liver tumor segmentation. *Med Hypotheses.* (2020) 134:1–8. doi: 10.1016/j.mehy.2019.109431
27. Albishri AA, Shah SJH, Lee Y. CU-Net: Cascaded U-Net model for automated liver and lesion segmentation and summarization. In: *IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* (2019). p. 1416–23. doi: 10.1109/BIBM47256.2019.8983266
28. Zhang J, Xie Y, Zhang P, Chen H, Xia Y, Shen C. Light-weight hybrid convolutional network for liver tumor segmentation. In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19).* (2019). p. 4271–7. doi: 10.24963/ijcai.2019/593
29. Jin Q, Meng Z, Sun C, Cui H, Su R. RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans. *Front Bioeng Biotechnol.* (2020) 8:1471. doi: 10.3389/fbioe.2020.605132
30. Tian J, Liu L, Shi Z, Xu F. Automatic couinaud segmentation from CT volumes on liver using GLC-UNet. *Machine Learning in Medical Imaging, MLMI 2019. Lecture Notes Comput Sci.* (2019) 11861:274–82. doi: 10.1007/978-3-030-32692-0\_32
31. Song L, Wang H, Wang ZJ. Bridging the gap between 2D and 3D contexts in CT volume for liver and tumor segmentation. *IEEE J Biomed Health Inform.* (2021) 25:3450–9. doi: 10.1109/JBHI.2021.3075752
32. Ben-Cohen A, Diamant I, Klang E, Amitai M, Greenspan H. *Fully Convolutional Network for Liver Segmentation and Lesions Detection.* New York City: Springer International Publishing. (2016). p. 77–85. doi: 10.1007/978-3-319-46976-8\_9
33. Christ PF, Elshaer M, Ettlinger F, Tatavarty S, Bickel M, Bilic P, et al. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* New York City: Springer, Cham (2016). p. 415–23. doi: 10.1007/978-3-319-46723-8\_48
34. Yao Z, He Z, Cheng Z, Yang Z, Shi Z. Fully convolutional neural network with post-processing methods for automatic liver segmentation from CT. *Chinese Automation Congress (CAC).* (2017). p. 3864–9.
35. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation.* Springer, Cham (2016).
36. Jie H, Li S, Gang S, Albanie S. Squeeze-and-excitation networks. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence.* (2017). p. 99.
37. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: *IEEE Computer Society.* (2016). doi: 10.1109/CVPR.2017.660
38. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. *Comput Sci.* arXiv:1511.06434 (2015).
39. Bilic P, Christ PF, Vorontsov E, Chlebus G, Wu J. The liver tumor segmentation benchmark (lits). arXiv:1901.04056 (2019).
40. Heller N, Sathianathan N, Kalapara A, Walczak E, Moore K, Kaluzniak H, et al. The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. *arXiv e-prints.* (2019).
41. Voiculescu I, Yeghiazaryan V. *An Overview of Current Evaluation Methods Used in Medical Image Segmentation.* (2015). p. 22.
42. Kaluva KC, Khened M, Kori A, Krishnamurthy G. 2D-Densely connected convolution neural networks for automatic liver and tumor segmentation. arXiv:1802.02182 (2018).
43. Guo X, Schwartz LH, Zhao B. Automatic liver segmentation by integrating fully convolutional networks into active contour models. *Med Phys.* (2019) 46:4455–69. doi: 10.1002/mp.13735
44. Li X, Chen H, Qi X, Dou Q, Fu C-W, Heng P-A. H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans Med Imaging.* (2018) 37:2663–74. doi: 10.1109/TMI.2018.2845918
45. Liu Z, Song Y-Q, Sheng VS, Wang L, Jiang R, Zhang X, et al. Liver CT sequence segmentation based with improved U-Net and graph cut. *Exp Syst Appl.* (2019) 126:54–63. doi: 10.1016/j.eswa.2019.01.055

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 He, Xu, Liu, Li, Liu, Zhao, Yuan and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Deep Learning in Prostate Cancer Diagnosis Using Multiparametric Magnetic Resonance Imaging With Whole-Mount Histopathology Referenced Delineations

Danyan Li<sup>1,2†</sup>, Xiaowei Han<sup>2†</sup>, Jie Gao<sup>3</sup>, Qing Zhang<sup>3</sup>, Haibo Yang<sup>4</sup>, Shu Liao<sup>4</sup>, Hongqian Guo<sup>3†</sup> and Bing Zhang<sup>1,2\*†</sup>

<sup>1</sup> Department of Radiology, Nanjing Drum Tower Hospital Clinical College of Nanjing Medical University, Nanjing, China,

<sup>2</sup> Department of Radiology, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, China, <sup>3</sup> Department of Urology, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, China, <sup>4</sup> Department of Research and Development, Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

## OPEN ACCESS

### Edited by:

Kuanquan Wang,  
Harbin Institute of Technology, China

### Reviewed by:

Xiaohu Li,  
First Affiliated Hospital of Anhui  
Medical University, China  
Marcelo Mamede,  
Federal University of Minas  
Gerais, Brazil

### \*Correspondence:

Bing Zhang  
zhangbing\_nanjing@njmu.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

Received: 08 November 2021

Accepted: 16 December 2021

Published: 13 January 2022

### Citation:

Li D, Han X, Gao J, Zhang Q, Yang H,  
Liao S, Guo H and Zhang B (2022)  
Deep Learning in Prostate Cancer  
Diagnosis Using Multiparametric  
Magnetic Resonance Imaging With  
Whole-Mount Histopathology  
Referenced Delineations.  
Front. Med. 8:810995.  
doi: 10.3389/fmed.2021.810995

**Background:** Multiparametric magnetic resonance imaging (mpMRI) plays an important role in the diagnosis of prostate cancer (PCa) in the current clinical setting. However, the performance of mpMRI usually varies based on the experience of the radiologists at different levels; thus, the demand for MRI interpretation warrants further analysis. In this study, we developed a deep learning (DL) model to improve PCa diagnostic ability using mpMRI and whole-mount histopathology data.

**Methods:** A total of 739 patients, including 466 with PCa and 273 without PCa, were enrolled from January 2017 to December 2019. The mpMRI (T2 weighted imaging, diffusion weighted imaging, and apparent diffusion coefficient sequences) data were randomly divided into training ( $n = 659$ ) and validation datasets ( $n = 80$ ). According to the whole-mount histopathology, a DL model, including independent segmentation and classification networks, was developed to extract the gland and PCa area for PCa diagnosis. The area under the curve (AUC) were used to evaluate the performance of the prostate classification networks. The proposed DL model was subsequently used in clinical practice (independent test dataset;  $n = 200$ ), and the PCa detective/diagnostic performance between the DL model and different level radiologists was evaluated based on the sensitivity, specificity, precision, and accuracy.

**Results:** The AUC of the prostate classification network was 0.871 in the validation dataset, and it reached 0.797 using the DL model in the test dataset. Furthermore, the sensitivity, specificity, precision, and accuracy of the DL model for diagnosing PCa in the test dataset were 0.710, 0.690, 0.696, and 0.700, respectively. For the junior radiologist without and with DL model assistance, these values were 0.590, 0.700, 0.663, and 0.645 versus 0.790, 0.720, 0.738, and 0.755, respectively. For the senior radiologist, the values were 0.690, 0.770, 0.750, and 0.730 vs. 0.810, 0.840, 0.835, and 0.825, respectively. The diagnosis made with DL model assistance for radiologists were significantly higher than those without assistance ( $P < 0.05$ ).

**Conclusion:** The diagnostic performance of DL model is higher than that of junior radiologists and can improve PCa diagnostic accuracy in both junior and senior radiologists.

**Keywords:** prostate cancer, deep learning, magnetic resonance imaging, segmentation, detection

## INTRODUCTION

Prostate cancer (PCa) is a major public health problem, representing the most common cancer type and the second highest cancer mortality among men in western countries (1). Multiparametric magnetic resonance imaging (mpMRI) plays an important role in diagnosis, targeted puncture guidance, and prognosis assessment of PCa in the current clinical setting (2). However, the performance of mpMRI usually varies based on the experience of radiologists at different levels (3), and the demand for MRI interpretation is ever-increasing. A convolutional neural network (CNN) approach, which can surpass human performance in natural image analysis, is anticipated to enhance computer-assisted diagnosis in prostate MRI (4, 5).

The CNN-based deep learning (DL) method revolutionizes and reshapes the existing work pattern. Diffusion weighted imaging (DWI), apparent diffusion coefficient (ADC), and T2-weighted imaging (T2WI) sequences are probably the most important and practical components of clinical prostate MRI examinations (6, 7). Several previous studies on DL involved a PCa diagnosis using only one or two of the above sequences and thus cannot be directly compared with clinical performance (8, 9).

Some studies focused on DL models with MRI data labeling based on biopsy locations that were determined by radiologists (10), which could result in inaccurate labeling. Whole-mount tissue sections, in which the entire cross-section of tissue from the gross section is mounted to the slide, provide pathologists with a good overview facilitating the identification of tumor foci (11–13). The use of prostate specimen whole-mount sectioning provides significantly superior anatomical registration for PCa than just mpMRI. Herein, we propose that the radiologists label PCa lesions on the MRI images using whole-mount histopathology images as reference to increase the accuracy of the labels.

In this study, a DL method was proposed to automatically conduct prostate gland segmentation, classification, and regional segmentation of PCa lesions, and subsequently compare its diagnostic efficiency with different level radiologists in clinical practice.

## MATERIALS AND METHODS

This retrospective study was approved by the Ethics Institution of Nanjing Drum Tower Hospital, and informed consent was waived since T2WI, DWI, and ADC sequences are part of the routine protocols for prostate MRI scans.

## Patients

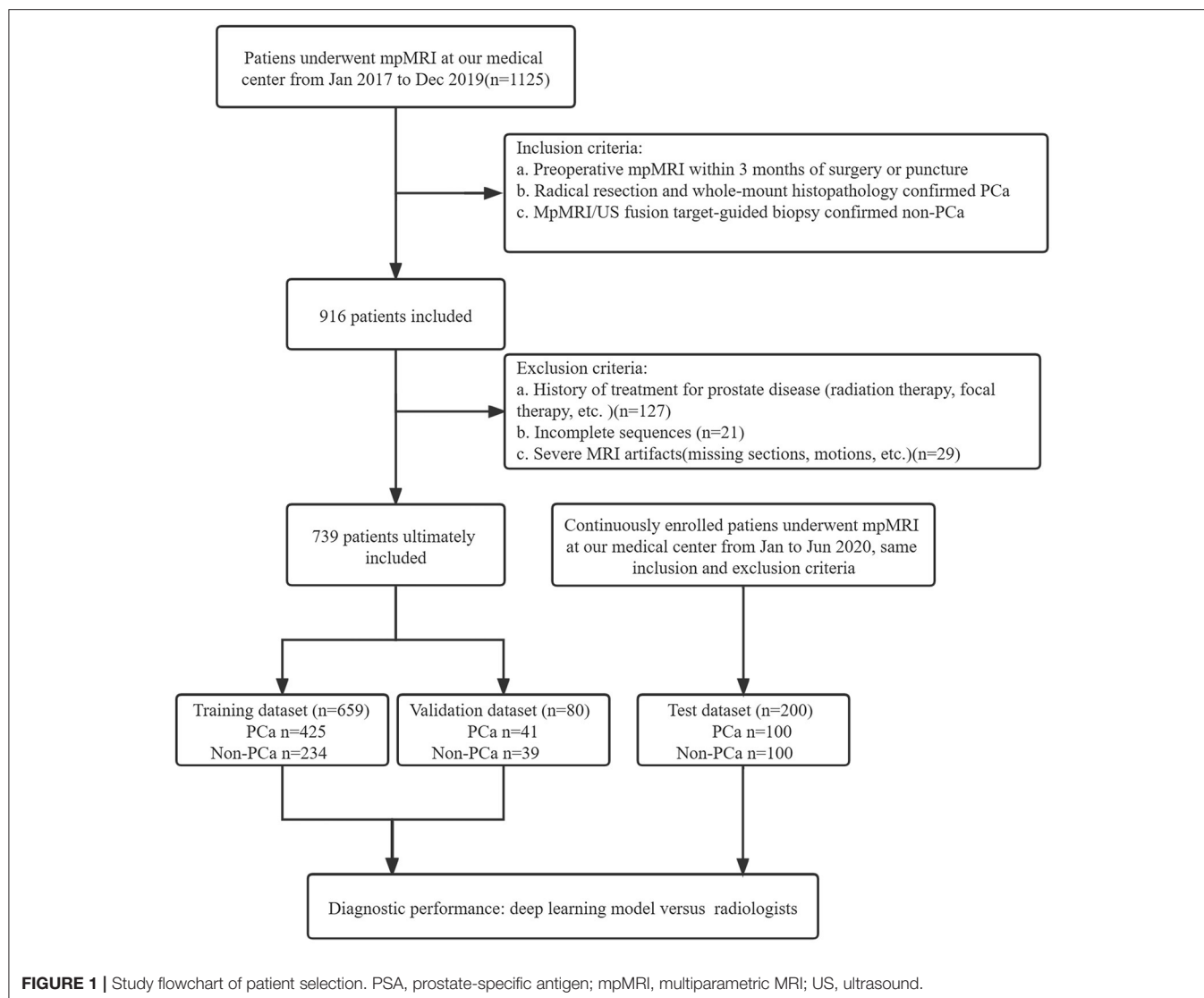
A total of 1125 patients who underwent prostate mpMRI between January 2017 and December 2019 were enrolled in the study. The inclusion criteria were as follows: (a) preoperative mpMRI within 3 months of surgery or puncture, (b) radical resection and whole-mount histopathology-confirmed PCa, and (c) mpMRI/ultrasonography (US) fusion target-guided biopsy or surgery confirmed non-PCa. Patients without PCa were defined as having negative biopsy or surgery. The exclusion criteria were (a) a history of treatment for prostate disease (radiation therapy, focal therapy, etc.), (b) incomplete imaging sequences, (c) severe MRI artifacts (missing sections, motions, etc.), and (d) unavailable whole-mount history. All MRI scans were reviewed in consensus by two radiologists, a 5-year junior and 10-year senior radiologist specializing in genitourinary imaging. A total of 739 patients, including 466 patients with PCa and 273 patients without PCa, were included for training and validation in the model. The independent dataset was consecutively collected from January 2020 to June 2020 with the same inclusion and exclusion criteria as mentioned above. A flowchart of the patient selection is shown in **Figure 1**.

## MR Imaging

Patients were scanned using two 3.0 T MRI scanners (uMR770; United Imaging, Shanghai, China and Ingenia; Philips Healthcare, Best, the Netherlands) with the same sequences and standard phased array surface coils according to the European Society of Urogenital Radiology guidelines. T<sub>1</sub>WI, T<sub>2</sub>WI, DWI and ADC sequences were acquired. Detailed parameters for transverse DWI (b-values of 50, 1,000, and 1,500 s/mm<sup>2</sup>) were as follows: repetition time (TR), 5,100 ms; echo time (TE), 80 ms; field of view, 26 × 22 cm; and thickness, 3 mm. Low b-value images were acquired at 50 s/mm<sup>2</sup> to avoid perfusion effects at a b-value of 0 s/mm<sup>2</sup>. ADC maps were calculated from the b-value (1,500 s/mm<sup>2</sup>) using the scanner software. T<sub>2</sub>WI, DWI (b-values of 1,500 s/mm<sup>2</sup>), and ADC (b-values of 1,500 s/mm<sup>2</sup>) sequences were used in this study.

## Histopathology

All the cases were confirmed by mpMRI/US fusion-guided targeted biopsy, and patients with PCa were further confirmed by radical resection and whole-mount histopathology. All the biopsies were conducted using the MRI-trans rectal ultrasound scan (TRUS) image registration system (Esaote® and RVS®). Whole-mount specimens were sliced from the apex to the base at 3-mm intervals following prostatectomy. All the specimens were examined by two independent urological pathologists.



## Prostate Gland and Cancer Region Labeling Referenced by Whole-Mount Histopathology Image in the Training and Validation Datasets

Based on the whole-mount histopathology images, the prostate gland and all the cancer regions on T2WI, DWI, and ADC sequences were labeled by two radiologists (with 5 and 10 years of expertise, respectively) under the supervision of a superior radiologist (with 15 years of expertise) using the open-source software ITK-SNAP (<http://www.simpleitk.org>, version 3.8.0). The workflow is illustrated in **Figure 2**.

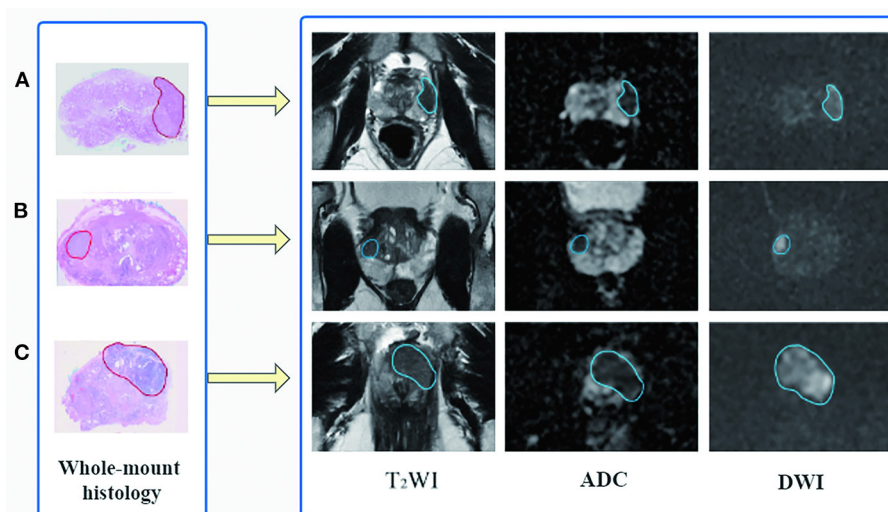
## The DL Network Structure

A CNN was constructed for prostate gland segmentation, classification, and cancer region segmentation/detection tasks. The model structure is illustrated in **Figure 3**.

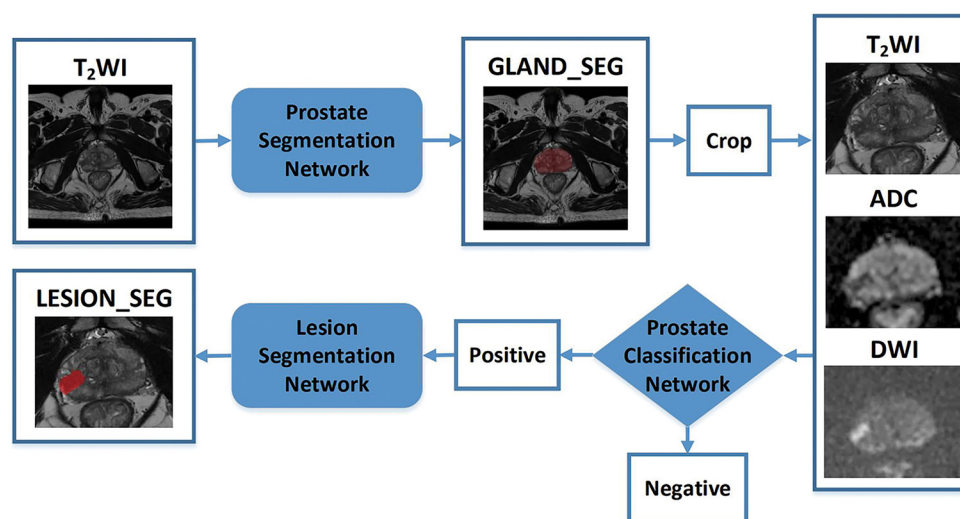
First, a prostate gland segmentation network based on the T2WI sequence was implemented to obtain a mask of the gland.

The mask was subsequently cropped to obtain three image patches including the gland on T2WI, DWI, and ADC sequences. Second, a prostate classification network based on the image patches from the T2WI, DWI, and ADC sequences was used to determine whether the gland had PCa lesion(s). If the gland was abnormal, a PCa segmentation network was used to obtain the lesion region. It is worth noting that the T2WI, DWI, and ADC patches were obtained based on the prostate gland segmentation results and were of fixed and similar sizes, including the gland.

The prostate gland segmentation network was based on V-Net (14), as shown in **Supplementary Figure 1**. The classification network was based on dense convolutional network (DenseNet) (15), which was used to determine whether the gland was normal. DenseNet connects each layer to every other layer in a feed-forward fashion. The feature maps of all the preceding layers were used as inputs for each layer, and their feature maps were used as inputs to all the subsequent layers. Prostate cancer lesion segmentation was also performed based on the image patches of T2WI, DWI, and ADC sequences. To obtain a more accurate



**FIGURE 2 |** Flowchart of region of interest delineation for prostate cancer lesion. All the prostate cancer lesions were manually labeled on the magnetic resonance images using whole-mount histopathology as a reference. Representative cases of prostate cancer in different zone distributions: **(A)** the lesion is in the left peripheral zone, **(B)** in the right peripheral and transition zone, and **(C)** in the transition zone and anterior fibromuscular stroma.



**FIGURE 3 |** Flowchart of the study. The blocks highlighted in blue (prostate gland segmentation network, prostate cancer classification network, prostate cancer segmentation/detection network) denote network models used in our study. “Crop” represents a fixed size region of interest (ROI) to crop the prostate gland according to the result of the prostate gland segmentation network. The cropped ROI of ADC and DWI would be registered to the cropped ROI of the T2-weighted imaging (T2WI) and then three cropped ROI would be fed into the prostate cancer classification network. “Positive” represents the positive output of the classification network; in that case, the cropped ROI would be fed into the prostate cancer segmentation network to obtain the lesion region. “Negative” represents the negative output of the classification network; in that case, the cropped ROI would not be fed into the prostate cancer segmentation network.

cancer region, the Up-Block in V-Net was changed to an Up SE-Block, which adds a squeeze-and-excitation operation following two convolutions, as shown in **Supplementary Figure 2**.

## Training and Optimization Details

In this study, all the networks were implemented using the PyTorch framework and Python 3.7. All the learning computations were performed on a Tesla V100 DGXS GPU with

32 GB of memory. The adaptive moment (Adam) algorithm was applied to optimize the parameters of the prostate segmentation network. The training dataset was randomly shuffled, and a batch size of four was selected. The stochastic gradient descent (SGD) algorithm was applied to optimize the parameters of the PCa network (15). The training dataset was randomly shuffled, and a batch size of 12 was selected. Finally, the Adam algorithm was applied again to optimize the parameters of the PCa

**TABLE 1** | Clinical and imaging characteristics of the included patients.

Characteristics	Training dataset <i>n</i> = 659	Validation dataset <i>n</i> = 80	Test dataset <i>n</i> = 200	<i>F</i> / $\chi^2$	<i>P</i>
Age, mean $\pm$ SD(y)	68.1 $\pm$ 7.8	67.5 $\pm$ 5.4	67.7 $\pm$ 7.4	0.53	0.58
Prostate cancer, <i>n</i> (%)	425 (64.5)	41 (51.3)	100 (50)	17.22	<0.01
Non prostate cancer, <i>n</i> (%)	234 (35.5)	39 (48.7)	100 (50)		
tPSA level (ng/ml)	17.8 $\pm$ 22.2	14.9 $\pm$ 15.7	15.9 $\pm$ 22.1	0.91	0.40
Prostate cancer	15.3 $\pm$ 21.3	22.0 $\pm$ 23.1	13.1 $\pm$ 17.7	0.04	0.96
Non prostate cancer	9.0 $\pm$ 5.3	7.35 $\pm$ 4.7	9.1 $\pm$ 6.4	1.18	0.31
Prostate cancer lesion numbers	500	59	127		
Prostate cancer zone distribution, <i>n</i> (%)					
PZ	315 (63.0)	42 (71.2)	78 (61.4)	5.37	0.49
TZ	146 (29.2)	10 (16.9)	38 (29.9)		
AFS	3 (0.6)	0	1 (0.8)		
Mixed	36 (7.2)	7 (11.9)	10 (7.9)		

The data are reported as the mean  $\pm$  standard deviations.

PSA, prostate-specific antigen; PZ, peripheral zone; TZ, transitional zone; AFS, anterior fibromuscular stroma.

region segmentation network. The training dataset was randomly shuffled, and a batch size of four was selected. During the training process for the prostate gland and cancer segmentation networks, the Dice loss was adopted, and the network weights were updated using the Adam optimizer with an initial learning rate of 0.0001. During the training process for the classification network, the cross-entropy loss was adopted, and the network weights were updated using SGD with an initial learning rate of 0.1.

### Image Analysis of the Junior and Senior Radiologists for the Test Dataset Without and With DL Assistance

The T2WI, DWI, and ADC images were imported from the DICOM format into ITK-SNAP (version 3.8.0). The MR images with and without DL delineations were independently reviewed by two radiologists, a 5-year junior and 10-year senior radiologist specializing in genitourinary imaging, who were blinded to the pathological results. PI-RADS v2.1 (16) recommendations were used by the junior and senior radiologists to evaluate the PCa likelihood of suspicious areas on mpMRI for each patient (17), and the results were divided into PCa (PI-RADS score 4-5 and partly PI-RADS score 3 cases) and non-PCa. Particularly, referring to PI-RADS score 3 cases, the final diagnosis would be further made by another 20-year radiologist specializing in genitourinary imaging.

### Statistical Analysis

Continuous variables are described using mean  $\pm$  standard deviation, while categorical variables are described using frequency and ratio. The chi-square test was used for the sample size and location distribution. The DL model was verified using the validation and test datasets. The Dice loss was used to evaluate the performance of prostate gland and PCa lesion

segmentation networks. The cross-entropy loss and AUC were used to evaluate the performance of the classification networks. Furthermore, the sensitivity, specificity, precision and accuracy were used to evaluate the diagnostic performance of the model in clinical application.

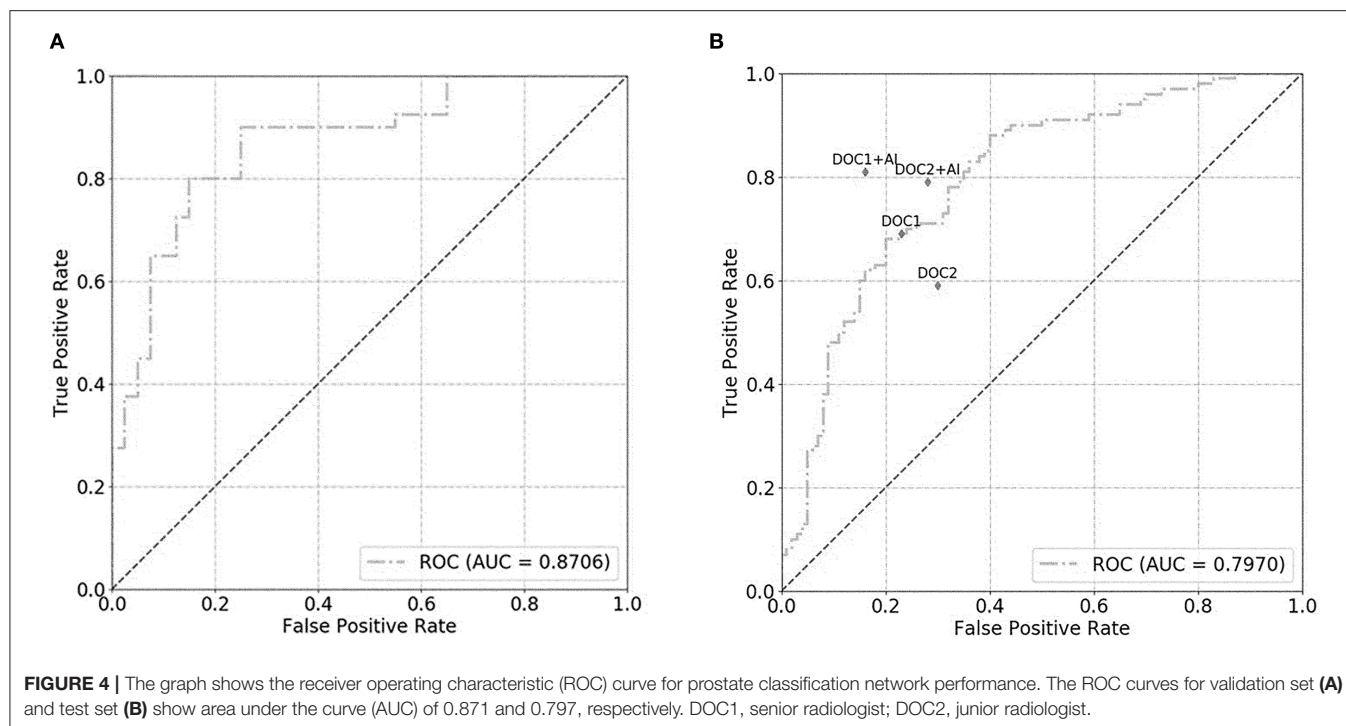
## RESULTS

### Study Sample Characteristics

Patient demographic data and characteristics of the training, validation, and test datasets are shown in **Table 1**. There were no significant differences in the patient age or total prostate-specific antigen (PSA) values among the three groups. In the training dataset, there were 500 pathologically proven cancer lesions, with 315 lesions in the peripheral zone (PZ), 146 in the transitional zone (TZ), 3 in the anterior fibromuscular stroma (AFS), and 36 in the mixed region. In the validation data set, there were 59 pathologically proven cancer lesions, with 42 lesions in the PZ, 10 in the TZ, 0 in the AFS, and 7 in the mixed region. In the test dataset, there were 127 pathologically proven cancer lesions, with 78 lesions in the PZ, 38 in the TZ, 1 in the AFS, and 10 in the mixed region.

### Performance of the DL Model in the Training Dataset

The training epoch was set as 700 for the prostate gland segmentation network, while the Dice loss values converged to 0.068; the convergence graph is shown in **Supplementary Figure 3**. A total of 330 epochs were set for training the prostate classification model, and the cross-entropy loss converged to 0.120; the convergence graph is shown in **Supplementary Figure 4**. For the PCa segmentation model, the network was trained for 240 epochs, when the value of the loss function converged to 0.167. A convergence graph is shown in **Supplementary Figure 5**.



**TABLE 2 |** Diagnostic performance of prostate cancer by different radiologists and DL model.

Group	Sensitivity	Specificity	Precision	Accuracy
Junior radiologist	0.590	0.700	0.663	0.645
Senior radiologist	0.690	0.770	0.750	0.730
DL model	0.710	0.690	0.696	0.700
DL+Junior	0.790	0.720	0.738	0.755
DL+Senior	0.810	0.840	0.835	0.825

Junior and senior radiologist, experienced in interpreting prostate MRI (5 and 10 years, respectively); DL, deep learning.

## Performance of the DL Model in the Validation Dataset

For the prostate gland automatic segmentation efficacy, the Dice loss values converged to 0.076, and the convergence graph is shown in **Supplementary Figure 3**. For the prostate automatic classification efficacy, the cross-entropy loss converged to 0.224, and the convergence graph is shown in **Supplementary Figure 4**. The AUC value for the prostate classification network was 0.871 (**Figure 4A**). For the prostate cancer automatic segmentation efficacy, the Dice loss values converged to 0.484, as shown in **Supplementary Figure 5**.

## Diagnostic Performance of Prostate Cancer by Different Radiologists and DL Model in the Test Dataset

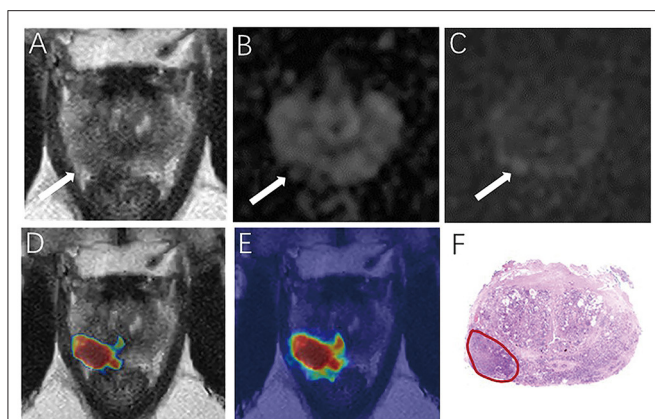
For the prostate automatic classification efficacy, the cross-entropy loss converged to 0.236. The AUC value for the prostate

classification network was 0.797 in the test dataset (**Figure 4B**). **Table 2** shows the evaluation of the model's diagnostic efficiency in practical applications based on the sensitivity, specificity, precision, and accuracy, with values of 0.710, 0.690, 0.696, and 0.700, respectively. For the junior radiologist without and with DL model assistance, these values were 0.590, 0.700, 0.663, and 0.645 vs. 0.790, 0.720, 0.738, and 0.755, respectively. For the senior radiologist, the values were 0.690, 0.770, 0.750, and 0.730 vs. 0.810, 0.840, 0.835, and 0.825, respectively. The values obtained with DL model assistance for radiologists were significantly higher than those without assistance ( $P < 0.05$ ). **Figure 5** shows a representative PCa example of radiologist-negative but DL model positive.

## DISCUSSION

We proposed a DL model for improving the diagnostic ability of PCa using mpMRI and whole-mount histopathology images referenced delineations. The DL model diagnostic ability was higher than that of a junior radiologist and can improve PCa diagnostic accuracy in both junior and senior radiologists in clinical practice.

MpMRI plays an important role in the diagnostic workflow of patients with suspected PCa (18). DWI, ADC, and T2WI are probably the most important sequences in the detection, identification, and staging of PCa (19–21), and the DCE sequence offers limited added value compared to T2+ADC+DWI (22). According to PI-RADS V2.1, the role of the DCE sequence is only helpful for score 3 lesions in the PZ (7). Some study also observed that for DWI score 3 lesions in the PZ of biopsy-negative patients, the DCE sequence had no significant increased value



**FIGURE 5 |** Demonstrate representative prostate cancer (PCa) example of radiologists negative (A–C) and deep learning (DL) model positive (D,E). Images show a case of DL model segmentation in a 60-year patient in a test set with prostate-specific antigen (PSA) of 5.59 ng/mL. Axial T2-weighted image (A) shows an ill-defined area of little low signal in the right peripheral zone (arrow), with slight restricted diffusion on apparent diffusion coefficient (ADC) maps (B). (C) Diffusion weighted imaging (DWI) (b-value 1,500 sec/mm<sup>2</sup>) shows slightly increased signal in this region, with an obvious conspicuity over background normal signal; this lesion would be PI-RADS score 3 for magnetic resonance imaging (MRI). (D,E) show overlapping areas between DL focused PCa region and genuine cancer location. The overlapped areas are colored in red. The software ITK-SNAP was used to open the probability map and MR images at the same time. Through the software function, the probability map is displayed as a jet type color map and overlapped on the T2 weighted imaging (T2WI) to obtain (E); The window width and window level of the probability map is adjusted to 0.5 and 0.75 respectively to display the probability map of the detected cancer area and overlapped on the image to obtain (D).

in improving the identification of PCa (13). So, we proposed a DL model based on DWI, ADC, and T2WI sequences without contrast medium injections. Furthermore, some previous studies on DL model using only one or two of the above sequences and thus cannot be directly compared with clinical performance (8, 9).

The use of whole-mount histopathological specimens is a strong reference standard. Moreover, Padhani et al. (23) suggested that training datasets with spatially well-correlated histopathologic validation should be used. Our previous studies confirmed that whole-mount sections can be used as a reference to obtain a highly accurate prostate lesion label on prostate mpMRI (13). We subsequently labeled the PCa lesions on MR images using whole-mount histopathology images as references. Furthermore, the DL model was developed based on the classic V-Net and DenseNet networks; SE-Block integrated variables controlling also helped in improving the model accuracy and performance (24). In our study, the DL model was used to extract the gland and PCa areas, and accurately identify PCa compared to the gold standard of histopathology. The AUC value was 0.797 for the prostate classification network in the test dataset, and the accuracy of PCa detection/diagnosis was 0.700, which is higher than that of several reports. For example, Ishioka et al. (25) proposed a CNN deep learning model with AUCs of 0.645 and 0.636 in two validation sets, respectively. Moreover, our

independent test dataset is imported without gland or lesion labeling in order to evaluate the model in real clinical work scenarios. The average PSA level of non-PCa group in the test dataset was  $9.1 \pm 6.4$  ng/mL. It was a little high because all the patients were confirmed by targeted biopsy or resection for prostatitis, hyperplasia, or other prostate benign diseases; thus, the differential diagnosis could be challenging (26).

Castillo et al. (27) systematically reviewed the performance of machine learning applications in PCa classification based on MRI, and found that only one paper (27 publications) compared the performance of radiologists with or without DL model assistance, and presented that evaluation should be performed in a real clinical setting since the ultimate goal of these models is to assist the radiologists in diagnosis. Seetharaman et al. (28) developed a SPCNet model accurately detected aggressive PCa. In our study, we evaluated the DL model in an independent test dataset to assess its clinical application value and to compare it with junior and senior radiologists. This DL model showed higher accuracy than junior radiologist in diagnosing PCa and slightly lower than the senior radiologist. Furthermore, the DL model improved PCa diagnostic accuracy for both junior and senior radiologists. This is similar to the findings of Cao et al. (29), who presented a DL algorithm (FocalNet) that achieved slightly but not significantly lower PCa diagnosis performance than genitourinary radiologists. Additionally, some studies demonstrated diagnostic accuracy for prostate cancer using PI-RADS was 71.0–83.5%, which was similar with our results, but PI-RADS usually varies based on the experience of radiologists at different levels (30, 31).

Currently, most DL models are not fully automated diagnosis systems; rather, they are adjunct tools that aid radiologists in reading prostate mpMRI results. Kotter et al. (32) determined that new DL technology would not threaten a radiologist's career but rather help strengthen his or her diagnostic ability. In summary, our proposed DL model can improve PCa diagnostic performance for both senior and junior radiologists, indicating that DL assistance can potentially improve the clinical workflow.

## Limitations and Outlook

There are several limitations to this study. First, all the patients were recruited from a single center. This may have negatively affected the performance of the model because larger and more diverse patient groups improve the generalizability of the classification algorithms. Second, the study was retrospective, the clinical data and traditional image parameters were not used in this study. Future studies should focus on multicenter data, biomarkers, and optimized algorithms to produce more reliable models for improving diagnosis, staging, and recurrence prediction of PCa. At last, all the patients were scanned using two 3.0 T MRI scanners in this study. The DL model may not perform so well using images provided by different machines or by a machine with a lower magnetic field.

## CONCLUSION

In this study, we proposed an automated DL model for the segmentation and detection of PCa based on mpMRI

and whole-mount histopathology referenced delineations. The diagnostic performance of DL model is higher than junior radiologist and could be capable of improving the diagnostic accuracy for both junior and senior radiologists and applying for young radiologist training.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and this retrospective study was approved by the Ethics Institution of Nanjing Drum Tower Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* (2020) 70:7–30. doi: 10.3322/caac.21590
2. Verma S, Choyke PL, Eberhardt SC, Oto A, Tempany CM, Turkbey B, et al. The current state of MR imaging-targeted biopsy techniques for detection of prostate cancer. *Radiology.* (2017) 285:343–56. doi: 10.1148/radiol.2017161684
3. Stabile A, Giganti F, Rosenkrantz AB, Taneja SS, Villeirs G, Gill IS, et al. Multiparametric MRI for prostate cancer diagnosis: current status and future directions. *Nat Rev Urol.* (2020) 17:41–61. doi: 10.1038/s41585-019-0212-4
4. Suarez-Ibarrola R, Hein S, Reis G, Gratzke C, Miernik A. Current and future applications of machine and deep learning in urology: a review of the literature on urolithiasis, renal cell carcinoma, and bladder and prostate cancer. *World J Urol.* (2020) 38:2329–47. doi: 10.1007/s00345-019-03000-5
5. Syer T, Mehta P, Antonelli M, Mallett S, Atkinson D, Ourselin S, et al. Artificial intelligence compared to radiologists for the initial diagnosis of prostate cancer on magnetic resonance imaging: a systematic review and recommendations for future studies. *Cancers (Basel).* (2021) 13:3318. doi: 10.3390/cancers13133318
6. Tamada T, Kido A, Yamamoto A, Takeuchi M, Miyaji Y, Moriya T, et al. Comparison of Biparametric and Multiparametric MRI for Clinically Significant Prostate Cancer Detection With PI-RADS Version 21. *J Magn Reson Imaging.* (2021) 53:283–91. doi: 10.1002/jmri.27283
7. Rudolph MM, Baur ADJ, Cash H, Haas M, Mahjoub S, Hartenstein A, et al. Diagnostic performance of PI-RADS version 21 compared to version 20 for detection of peripheral and transition zone prostate cancer. *Sci Rep.* (2020) 10:15982. doi: 10.1038/s41598-020-72544-z
8. Alkadi R, Taher F, El-Baz A, Werghi N. A deep learning-based approach for the detection and localization of prostate cancer in t2 magnetic resonance images. *J Digit Imaging.* (2019) 32:793–807. doi: 10.1007/s10278-018-0160-1
9. Clark T, Zhang J, Baig S, Wong A, Haider MA, Khalvati F. Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted MRI using convolutional neural networks. *J Med Imaging (Bellingham).* (2017) 4:041307. doi: 10.1117/1.JMI.4.4.041307
10. Yoo S, Gujrathi I, Haider MA, Khalvati F. Prostate cancer detection using deep convolutional neural networks. *Sci Rep.* (2019) 9:19518. doi: 10.1038/s41598-019-55972-4
11. Ito K, Furuta A, Kido A, Teramoto Y, Akamatsu S, Terada N, et al. Detectability of prostate cancer in different parts of the gland with 3-Tesla multiparametric magnetic resonance imaging: correlation with whole-mount histopathology. *Int J Clin Oncol.* (2020) 25:732–40. doi: 10.1007/s10147-019-01587-8
12. Gao J, Zhang C, Zhang Q, Fu Y, Zhao X, Chen M, et al. Diagnostic performance of 68Ga-PSMA PET/CT for identification of aggressive cribriform morphology in prostate cancer with whole-mount sections. *Eur J Nucl Med Mol Imaging.* (2019) 46:1531–41. doi: 10.1007/s00259-019-04320-9
13. Wang B, Gao J, Zhang Q, Zhang C, Liu G, Wei W, et al. Investigating the equivalent performance of biparametric compared to multiparametric MRI in detection of clinically significant prostate cancer. *Abdom Radiol (NY).* (2020) 45:547–55. doi: 10.1007/s00261-019-02281-z
14. Zhu Y, Wei R, Gao G, Ding L, Zhang X, Wang X, et al. Fully automatic segmentation on prostate MR images based on cascaded fully convolution network. *J Magn Reson Imaging.* (2019) 49:1149–56. doi: 10.1002/jmri.26337
15. Huang G, Liu Z, Pleiss G, Van Der Maaten L, Weinberger K. Convolutional networks with dense connectivity. *IEEE Trans Pattern Anal Mach Intell.* (2019). doi: 10.1109/TPAMI.2019.2918284
16. Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, et al. Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. *Eur Urol.* (2019) 76:340–51. doi: 10.1016/j.eururo.2019.02.033
17. Lo GC, Margolis DJA. Prostate MRI with PI-RADS v21: initial detection and active surveillance. *Abdom Radiol (NY).* (2020) 45:2133–42. doi: 10.1007/s00261-019-02346-z
18. Barrett T, Rajesh A, Rosenkrantz AB, Choyke PL, Turkbey B. PI-RADS version 21: one small step for prostate. *MRI Clin Radiol.* (2019) 74:841–52. doi: 10.1016/j.crad.2019.05.019
19. Xu L, Zhang G, Shi B, Liu Y, Zou T, Yan W, et al. Comparison of biparametric and multiparametric MRI in the diagnosis of prostate cancer. *Cancer Imaging.* (2019) 19:90. doi: 10.1186/s40644-019-0274-9
20. Zawaideh JB, Sala E, Shaida N, Koo B, Warren AY, Carmisciano, et al. Diagnostic accuracy of biparametric versus multiparametric prostate MRI: assessment of contrast benefit in clinical practice. *Eur Radiol.* (2020) 30:4039–49. doi: 10.1007/s00330-020-06782-0
21. Christophe C, Montagne S, Bourrelle S, Roupert M, Barret E, Rozet F, et al. Prostate cancer local staging using biparametric MRI: assessment and comparison with multiparametric MRI. *Eur J Radiol.* (2020) 132:109350. doi: 10.1016/j.ejrad.2020.109350
22. Vargas HA, Hötter AM, Goldman DA, Moskowicz CS, Gondo T, Matsumoto K, et al. Updated prostate imaging reporting and data system (PI-RADS v2) recommendations for the detection of clinically significant prostate cancer using multiparametric MRI: critical evaluation using whole-mount

## AUTHOR CONTRIBUTIONS

DL and XH acquired and analyzed data and drafted the manuscript. HY and SL analyzed and explained the imaging data. JG and QZ acquired the clinical information and revised the manuscript. HG and BZ designed the study and revised the manuscript. All authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

We sincerely thank Prof. Yao Fu and Department of Pathology in Nanjing Drum Tower Hospital for retrieval of pathological data of patients in this study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2021.810995/full#supplementary-material>

- pathology as standard of reference. *Eur Radiol.* (2016) 26:1606–12. doi: 10.1007/s00330-015-4015-6
23. Padhani AR, Turkbey B. Detecting prostate cancer with deep learning for MRI: a small step forward. *Radiology.* (2019) 293:618–9. doi: 10.1148/radiol.2019192012
  24. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell.* (2020) 42:2011–23. doi: 10.1109/TPAMI.2019.2913372
  25. Ishioka J, Matsuoka Y, Uehara S, Yasuda Y, Kijima T, Yoshida S, et al. Computer-aided diagnosis of prostate cancer on magnetic resonance imaging using a convolutional neural network algorithm. *BJU Int.* (2018) 122:411–7. doi: 10.1111/bju.14397
  26. Cuocolo R, Cipullo MB, Stanzione A, Ugga L, Romeo V, Radice L, et al. Machine learning applications in prostate cancer magnetic resonance imaging. *Eur Radiol Exp.* (2019) 3:35. doi: 10.1186/s41747-019-0109-2
  27. Castillo T JMC, Arif M, Niessen WJ, Schoots IG, Veenland JF. Automated classification of significant prostate cancer on MRI: a systematic review on the performance of machine learning applications. *Cancers (Basel).* (2020) 12:1606. doi: 10.3390/cancers12061606
  28. Seetharaman A, Bhattacharya I, Chen LC, Kunder CA, Shao W, Soerensen SJC, et al. Automated detection of aggressive and indolent prostate cancer on magnetic resonance imaging. *Med Phys.* (2021) 48:2960–72. doi: 10.1002/mp.14855
  29. Cao R, Mohammadian Bajgiran A, Afshari Mirak S, Shakeri S, Zhong X, Enzmann D, et al. Joint Prostate Cancer Detection and Gleason Score Prediction in mp-MRI via FocalNet. *IEEE Trans Med Imaging.* (2019) 38:2496–506. doi: 10.1109/TMI.2019.2901928
  30. Mussi TC, Yamauchi FI, Tridente CF, Tachibana A, Tonso VM, Recchimuzzi DR, et al. Interobserver agreement and positivity of PI-RADS version 2 among radiologists with different levels of experience. *Acad Radiol.* (2019) 26:1017–22. 2 doi: 10.1016/j.acra.2018.08.013
  31. Xu L, Zhang G, Zhang D, Zhang X, Bai X, Yan W. et al. Comparison of PI-RADS version 21 and PI-RADS version 2 regarding interreader variability and diagnostic accuracy for transition zone prostate cancer. *Abdom Radiol.* (2020) 45:4133–41. doi: 10.1007/s00261-020-02738-6
  32. Kotter E, Ranschaert E. Challenges and solutions for introducing artificial intelligence (AI) in daily clinical workflow. *Eur Radiol.* (2021) 31:5–7. doi: 10.1007/s00330-020-07148-2

**Conflict of Interest:** HY and SL were employed by company Shanghai United Imaging Intelligence Co.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li, Han, Gao, Zhang, Yang, Liao, Guo and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# U-shaped GAN for Semi-Supervised Learning and Unsupervised Domain Adaptation in High Resolution Chest Radiograph Segmentation

Hongyu Wang<sup>1</sup>, Hong Gu<sup>1</sup>, Pan Qin<sup>1\*</sup> and Jia Wang<sup>2\*</sup>

<sup>1</sup> Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China, <sup>2</sup> Department of Surgery, The Second Hospital of Dalian Medical University, Dalian, China

## OPEN ACCESS

### Edited by:

Kuanquan Wang,  
Harbin Institute of Technology, China

### Reviewed by:

Nurlan Dauletbayev,  
McGill University, Canada  
Zhixin Li,  
Guangxi Normal University, China

### \*Correspondence:

Pan Qin  
qp112cn@dlut.edu.cn  
Jia Wang  
wangjia77@hotmail.com

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

**Received:** 24 September 2021

**Accepted:** 14 December 2021

**Published:** 13 January 2022

### Citation:

Wang H, Gu H, Qin P and Wang J  
(2022) U-shaped GAN for  
Semi-Supervised Learning and  
Unsupervised Domain Adaptation in  
High Resolution Chest Radiograph  
Segmentation. *Front. Med.* 8:782664.  
doi: 10.3389/fmed.2021.782664

Deep learning has achieved considerable success in medical image segmentation. However, applying deep learning in clinical environments often involves two problems: (1) scarcity of annotated data as data annotation is time-consuming and (2) varying attributes of different datasets due to domain shift. To address these problems, we propose an improved generative adversarial network (GAN) segmentation model, called U-shaped GAN, for limited-annotated chest radiograph datasets. The semi-supervised learning approach and unsupervised domain adaptation (UDA) approach are modeled into a unified framework for effective segmentation. We improve GAN by replacing the traditional discriminator with a U-shaped net, which predicts each pixel a label. The proposed U-shaped net is designed with high resolution radiographs ( $1,024 \times 1,024$ ) for effective segmentation while taking computational burden into account. The pointwise convolution is applied to U-shaped GAN for dimensionality reduction, which decreases the number of feature maps while retaining their salient features. Moreover, we design the U-shaped net with a pretrained ResNet-50 as an encoder to reduce the computational burden of training the encoder from scratch. A semi-supervised learning approach is proposed learning from limited annotated data while exploiting additional unannotated data with a pixel-level loss. U-shaped GAN is extended to UDA by taking the source and target domain data as the annotated data and the unannotated data in the semi-supervised learning approach, respectively. Compared to the previous models dealing with the aforementioned problems separately, U-shaped GAN is compatible with varying data distributions of multiple medical centers, with efficient training and optimizing performance. U-shaped GAN can be generalized to chest radiograph segmentation for clinical deployment. We evaluate U-shaped GAN with two chest radiograph datasets. U-shaped GAN is shown to significantly outperform the state-of-the-art models.

**Keywords:** semi-supervised learning, unsupervised domain adaptation, generative adversarial network, medical image segmentation, chest radiograph

# 1. INTRODUCTION

Recently, deep learning models have gained increasing popularity in medical segmentation. However, deep learning models with supervision require substantial pixel-level annotated data to achieve sufficient accuracy and prevent over-fitting (1–4). Pixel-level annotation is expensive, especially with medical images, because it is time-consuming and requires highly skilled experts (3, 5). Therefore, medical image datasets are usually small, which cannot meet the requirement of deep learning, due to a lack of annotations (6, 7). Even if a model is well-trained on a certain medical dataset, its accuracy decreases when it is applied to unseen domains (8, 9). The deep learning models suffer an accuracy drop between two domains due to domain shift (8). These problems limit the application of deep learning models in clinical environments.

An alternative to supervised learning is semi-supervised learning, which involves using unannotated data to identify specific hidden features of the dataset to facilitate prediction. Semi-supervised deep learning using generative adversarial networks (GANs) (10) has been highly successful (1, 2), especially with nonmedical images (3). A semi-supervised GAN model was proposed to distinguish between predicted probability maps and the ground truth (2). In Souly et al. (1), a GAN was used to generate fake images close to real images as learned by the segmentation network. In medical segmentation, a few semi-supervised models have been developed using the model in Hung et al. (2) to improve the prediction accuracy with specially designed loss functions for particular image types, such as computed tomography of the liver (3), retinal optical coherence tomography (11), and pediatric MRI (12).

Domain adaptation (DA) suffers domain shift by transferring knowledge from the source domain to the target domain (13–16). A popular solution is transfer learning, which fine-tunes parts of a pre-trained model with annotated target domain data (13). However, transfer learning approaches rely on additional annotated data on the target domain, which is expensive or sometimes impractical. Instead, unsupervised domain adaptation (UDA) is more appealing to generalize models in clinical practice. UDA using GAN is becoming increasingly popular in the medical field (14–16). GAN-based UDA models use generators to transform the target data to the source domain, discriminate the source data from the target data, and improve segmentation accuracy with a specific training method (16), net structure (14), or training loss (15).

Deep learning algorithms require large amounts of data, which cannot be collected from a single medical center. Therefore, data from multiple collection centers, comprising large medical centers and small clinics, are required (17–19). The large medical centers provide partly annotated datasets for semi-supervised learning, while the clinics provide unannotated data. The annotated and unannotated data may come from either the same or different domains in the dataset collected from multiple centers. A single model that can deal with the semi-supervised and UDA approach at the same time is urgently needed.

To tackle the aforementioned problems, we propose an improved GAN model, called U-shaped GAN, for medical

image segmentation. U-shaped GAN is improved by replacing the traditional discriminator with a U-shaped net to assign each pixel a label. Training the segmentation model with images of high resolution is effective; however, it increases the computational burden (20, 21). U-shaped GAN is designed with high resolution radiographs for effective segmentation while considering computational burden. The pointwise convolution is applied to U-shaped GAN for dimensionality reduction, which decreases the number of feature maps while retaining their salient features. Moreover, the U-shaped net takes a pretrained ResNet-50 as an encoder to reduce the computational burden of training from scratch. A pixel-level semi-supervised loss is proposed to leverage the unannotated data to assist the annotated data for semi-supervised learning. U-shaped GAN is extended to UDA with minimal modification. The semi-supervised learning approach and UDA approach are merged into a single model to handle datasets from multiple medical centers conveniently and efficiently. We evaluate U-shaped GAN on lung segmentation for radiographs.

To conclude the introduction, we outline the major contributions of this work as follows:

- (1) U-shaped GAN is proposed for high resolution medical image segmentation while taking computational burden into account.
- (2) A semi-supervised learning approach is proposed to overcome the lack of annotated data. We employ a pixel-level semi-supervised loss that leverages the unannotated data to assist the annotated data for segmentation.
- (3) U-shaped GAN is extended to UDA with minimal modification to transfer knowledge among different domains without additional annotated data on the target domain.
- (4) In our framework, the semi-supervised learning approach and UDA approach are merged into a single model to handle datasets from multiple medical centers conveniently and efficiently.

# 2. METHODS

## 2.1. Background

In recent years, GAN has garnered considerable attention because of its superior performance in terms of generating images (2). GAN consists of a generator network  $G$  and discriminator net  $D$ .  $G$  generates fake images close to real data from a noise distribution deceiving the discriminator, while  $D$  distinguishes the real images from fake ones.  $G$  and  $D$  can be considered as two competitors in a min-max game with the following formulation:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x)_{real})] + E_{z \sim p_{noise}(z)} [\log(1 - D(G(z))_{real})], \quad (1)$$

where  $E$  is the expectation of a random variable,  $p_{data}(x)$  is the real data distribution, and  $p_{noise}(z)$  is a noise distribution.  $D(*)_{real}$  stands for the possibility that the sample is from the real data.  $G$  transforms the noise variable  $z$  from the distribution  $p_{noise}(z)$  into  $G(z)$ . The min-max game provides a useful feature representation for auxiliary supervised discrimination tasks (22).

## 2.2. Proposed Model

The goal of this study is to develop a unified framework for semi-supervised learning and UDA. Analyzing the influence factors in semi-supervised learning and UDA on chest radiographs, we propose a similar solution for the semi-supervised learning approach and UDA approach. A pixel-level semi-supervised loss is proposed to leverage the unannotated data to assist the annotated data for segmentation.

### 2.2.1. Semi-Supervised Learning

We propose a U-shaped GAN for semi-supervised lung segmentation from chest radiograph datasets. U-shaped GAN is based on the following hypothesis: the features in an ideal representation correspond to the underlying causes of the data. If label  $y$  is among the salient causes of data  $x$ , a suitable representation for the probability distribution  $p(x)$  may also be a suitable representation for computing the distribution of conditional probability  $p(y|x)$  (23). The marginal probability  $p(x)$  is related to the conditional probability  $p(y|x)$  through the Bayes rule:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \quad (2)$$

Under this hypothesis, we use the unannotated and annotated data to find a representation for the radiographs. A particular semi-supervised loss, which can be divided into a supervised loss and an unsupervised loss, is proposed. The supervised loss using the annotated data is employed for segmentation prediction, and the unsupervised loss using the unannotated data is utilized for a better representation of the whole dataset, as shown in **Figure 1A**. The unannotated data generalize the model as a regularizer. In U-shaped GAN, we employ a generator to generate realistic data segmented by a multiclass classifier (our discriminator) from the noise input, which in addition to classifying the pixels into lungs, determines whether a given pixel belongs to the real or generated data. The generator converges the realistic data to the real distribution  $p(x)$  of the partly annotated real data. This enables the discriminator to learn better features to represent the radiographs and to filter irrelevant individual features. Moreover, we employ the annotated data to find the relations among those features and the segmentation task. We modify a GAN by replacing the original discriminator with a U-shaped net (24) for assigning a label to each pixel. Training convolutional neural networks (CNNs) with a high resolution is effective for lung segmentation (20, 21). U-shaped GAN is designed for high-resolution chest radiographs. Unlike the segmentation GAN (1), we use the semantic classes and the background as an additional class for segmentation as the background contains several unannotated organs that reflect the imaging condition.

We improve GAN by replacing the traditional discriminator with a U-shaped net, which, instead of predicting each image a label, assigns to each pixel a label. The proposed end-to-end deep learning model is illustrated in **Figure 2**. The discriminator acts as a segmentation network to assign one of the following labels to each pixel: lung class, background, or fake data. The annotated

data is used to train the discriminator  $D$  to minimize the loss function  $L_I$ :

$$L_I = -E_{x \sim p_{data_i}(x,y)} [\log(D(y|x))], \quad (3)$$

where  $p_{data_i}(x, y)$  is the joint distribution of the pixel-level labels  $y$  and pixel values  $x$  of the annotated data; the discriminator  $D$  predicts the possibility  $D(y|x)$  of pixel  $x$  belonging to label  $y$ .

In semi-supervised learning, where the labels are partly available among the training images, it is convenient to leverage the unannotated data for estimating a representation with useful features for segmentation. The true labels  $y$  of the pixels of the unannotated data are set as real data. The loss function for training the discriminator  $D$  with the unannotated data is defined as follows:

$$\begin{aligned} L_u &= -E_{x \sim p_{data_u}(x)} [\log(D(y|x))] \\ &= -E_{x \sim p_{data_u}(x)} [\log(1 - D(x)_{fake})], \end{aligned} \quad (4)$$

where  $p_{data_u}(x)$  is the distribution of pixels of chest radiographs without annotation;  $D(x)_{fake}$  is the possibility of the pixel belonging to the fake data. In U-shaped GAN, there is no output designed with the label “real data.” We used the  $1 - D(x)_{fake}$  instead.

The generator  $G$  maps a random noise  $z$  to a sample  $G(z)$  that is close to chest radiographs, while the discriminator  $D$  is trained to label the generated sample  $G(z)$  as fake. The true labels  $y$  of the pixels of the generated data are set as fake data. The loss function for the discriminator  $D$  with the generated data is given as

$$L_g = -E_{z \sim p_{noise}(z)} [\log(D(G(z))_{fake})], \quad (5)$$

where  $D(G(z))_{fake}$  is the possibility of the generated pixel belonging to the fake data.

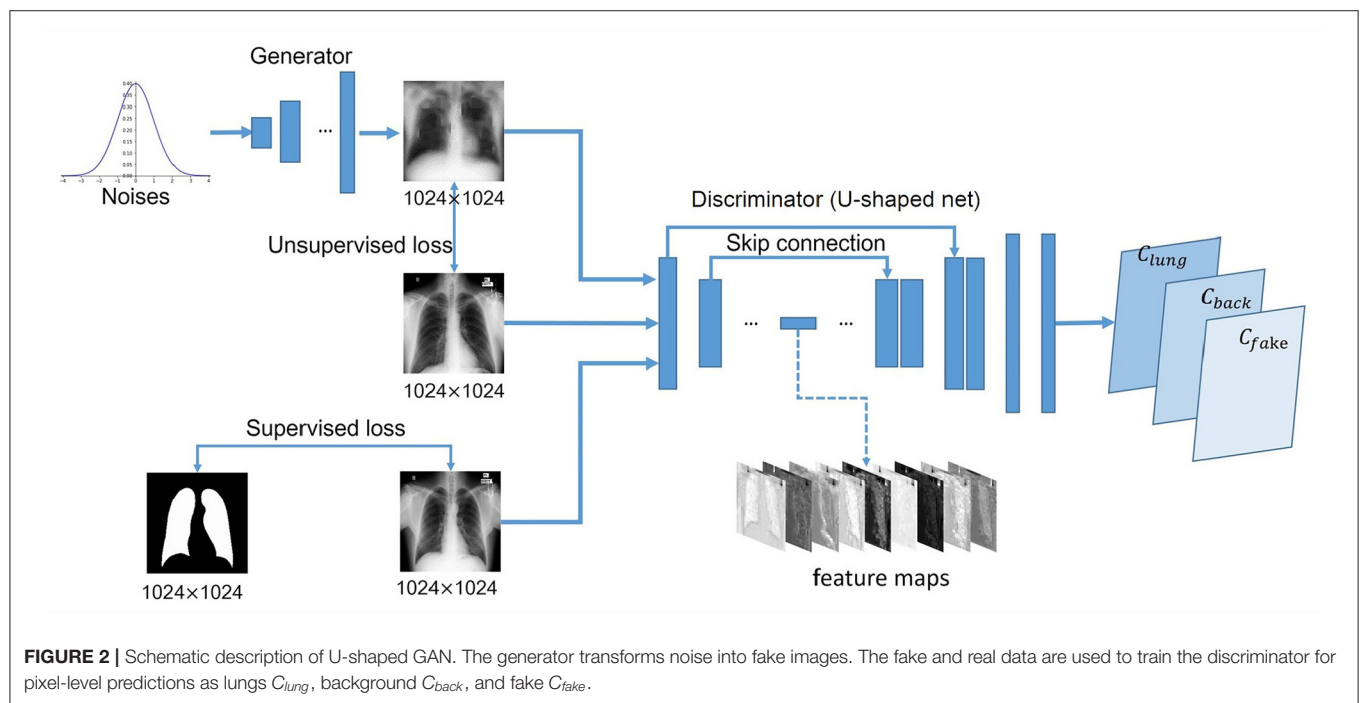
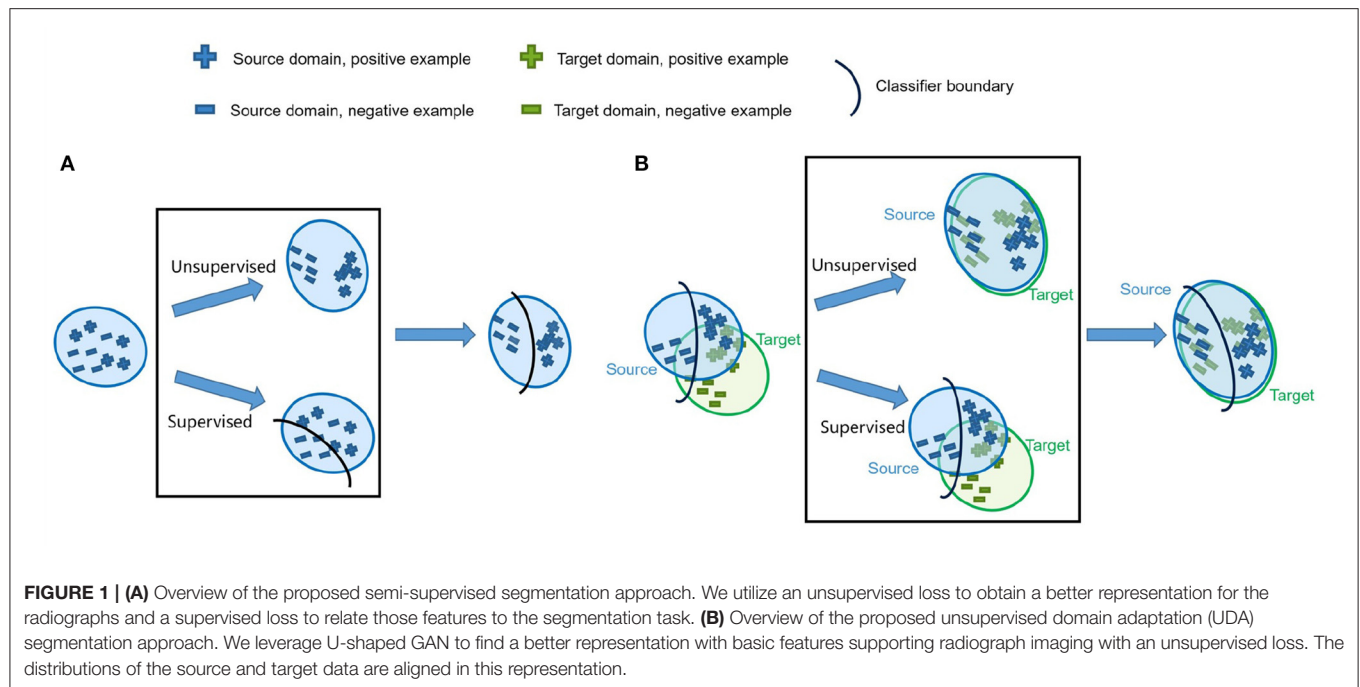
We minimize the pixel-level discriminator loss  $L_D$  with respect to three types of input data, as follows:

$$\begin{aligned} L_D &= -E_{x \sim p_{data_u}(x)} [\log(1 - D(x)_{fake})] \\ &\quad - E_{z \sim p_{noise}(z)} [\log(D(G(z))_{fake})] \\ &\quad - E_{x \sim p_{data_i}(x,y)} [\log(D(y|x))]. \end{aligned} \quad (6)$$

The first and second terms of  $L_D$  are devised for unannotated data as an unsupervised loss to increase the ability of the discriminator to identify the real radiographs from fake ones and to find salient features in the chest radiograph. The third term is devised for annotated data as a supervised loss training  $D$  to find correct relations among these features and the segmentation task. We use a soft maximum over the outputs.  $D(y|x)$  for the annotated data is a component of the  $1 - D(x)_{fake}$ . Increasing the probability  $D(y|x)$  will decrease the probability of  $D(x)_{fake}$ . The third term has the same effect as the first term and acts as an unsupervised loss to increase the ability of the discriminating and salient features finding. We minimize the generator loss  $L_G$  to train the generator  $G$  as follows:

$$L_G = E_{z \sim p_{noise}(z)} [\log(D(G(z))_{fake})]. \quad (7)$$

Because all of the annotated and unannotated data contribute to the discriminating ability of U-shaped GAN,  $G$  generates



a distribution  $p_g(G(z))$  converging to the real distribution  $p(x)$  of the whole dataset consisting of annotated and unannotated radiographs.

### 2.2.2. UDA Approach

The chest radiographs from various sources acquired by the same imaging modality differ in three aspects: image quality, image appearance, and spatial configuration (8). The features relevant

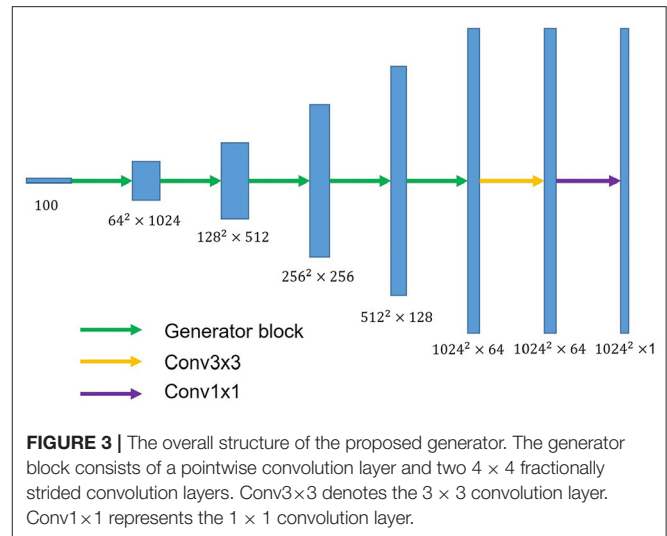
to the basic imaging causes among the radiographs are similar. Based on this property, we aim to develop a UDA approach to find these similar features, filter features of individual domains, and align the source and target domains in a representation, which is similar to our semi-supervised learning approach. U-shaped GAN is extended to UDA with minimal modification and uses nearly the same training process in the semi-supervised learning and UDA approaches. We use a U-shaped GAN to

search for the aforementioned features with the annotated source domain and unannotated target domain, as shown in **Figure 1B**. The target data serves as the unannotated data in the semi-supervised learning approach to generalize the model trained on the source dataset. The source data serves as the annotated data to ensure the accuracy of the segmentation task.

U-shaped GAN is extended to UDA, where only the source data are annotated, and the labels of target data are not available. The source and target domain data are used as the annotated and unannotated data in  $L_D$ , respectively, to train the discriminator. The generator loss function is the same as  $L_G$  in the semi-supervised approach. The generator  $G$  produces the fake data matching the aligned representation of the source and target data in the min-max game of  $G$  and  $D$ . The fake data are close to the real radiographs in both domains, and the discriminator learns better features related to the basic imaging causes of radiographs. U-shaped GAN then finds a suitable representation for the radiographs of the target and source data. The distributions of source and target data are aligned in this representation.  $L_D$  and  $L_G$  encourage domain-invariant detectors to emerge in U-shaped GAN. The annotated source data also guarantees the correct segmentation prediction.

### 2.3. Network Architecture

U-shaped GAN is proposed to label each input image pixel  $y$  as a lung, background, or fake pixel. The U-shaped net is incorporated into the structure of GAN, serving as the discriminator, to label each pixel. A schematic description of U-shaped GAN is shown in **Figure 2**. Training CNNs at a high resolution is effective for lung segmentation predictions; however, it increases the computational burden (20, 21). Therefore, we use a pointwise convolution layer followed by two  $4 \times 4$  fractionally strided convolution layers to form a new generator block (**Figure 3**), instead of the fractionally strided convolution layers used in deep convolutional GANs (25), to achieve a high resolution of  $1,024 \times 1,024$ . The pointwise convolution layer reduces feature dimensions while retaining the salient features and leaves the  $4 \times 4$  layers with fewer parameters. This approach significantly reduces the computational complexity. The model parameters decrease from  $1.34 \times 10^8$  to  $1.35 \times 10^7$  and the floating-point operations (FLOPs) decrease from  $5.84 \times 10^8$  to  $1.53 \times 10^8$ . In addition, we improve U-shaped GAN with a modified U-shaped network as the discriminator, as shown in **Figure 4**. The U-shaped net consists of a feature encoder and decoder modules (24). We replace the encoder with a pretrained ResNet-50 (26), which further reduces the computational burden of training from scratch. Moreover, ResNet-50 solves the degradation problem by adding identity connections to the convolution network (26). The feature decoder module restores the high-level semantic features extracted from the feature encoder module. The modified decoder module comprises four building blocks, as shown in **Figure 4**. The fundamental building block mainly comprises a  $3 \times 3$  convolution layer followed by a  $4 \times 4$  fractionally strided convolution layer. A pointwise convolution layer is used to connect them to reduce relevant parameters. Skip connections take information directly from the encoder to the decoder



layers and recover the information loss due to consecutive pooling and striding convolutional operations (24). We use instance normalization (27) followed by LeakyReLU activation functions (28) between each layer.

## 3. EXPERIMENTS

### 3.1. Datasets

In our experiments, we utilize the Japanese Society of Radiological Technology (JSRT) (29) and Montgomery County (MC) datasets (30, 31). The JSRT dataset contains 247 posterior-anterior (PA) chest radiographs, of which 154 contain lung nodules and 93 have no nodules (29). The ground truth lung masks can be obtained in the Segmentation in Chest Radiographs dataset (32). The MC dataset contains PA chest radiographs collected from the National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. It consists of 80 normal and 58 abnormal cases with manifestations of tuberculosis (30, 31). The ground truth lung masks are also contained in the MC dataset.

### 3.2. Metrics

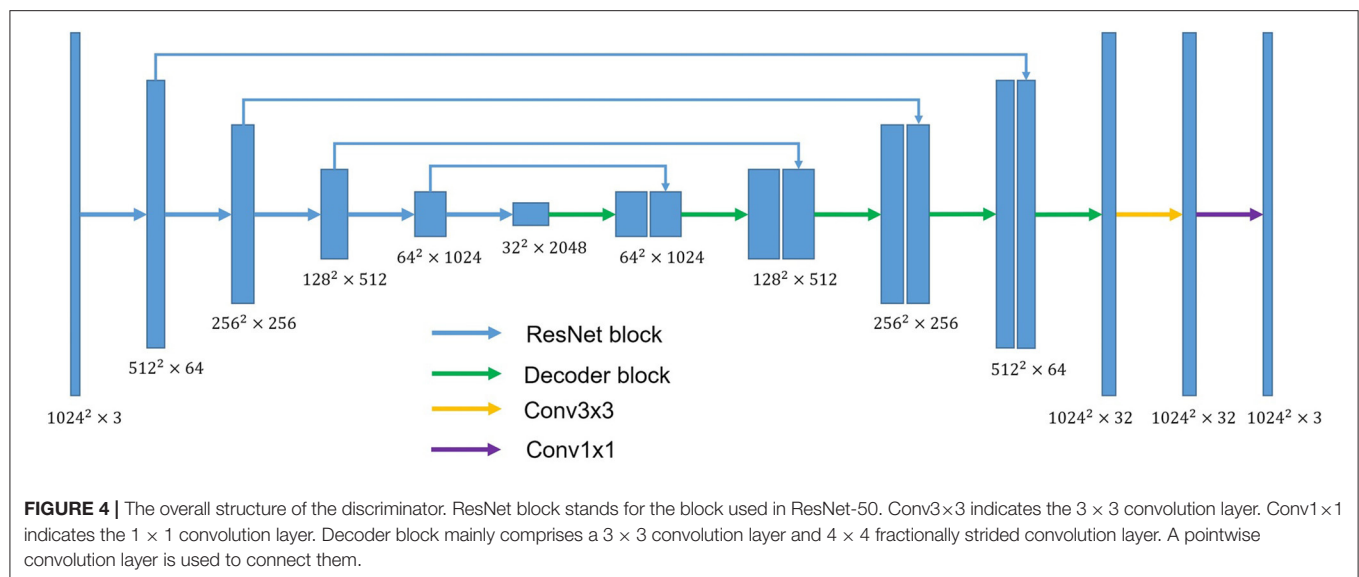
Several algorithms with different evaluation metrics are available in the literature. We used two commonly used methods, the Jaccard index and Dice score metrics, to compare U-shaped GAN with the state-of-the-art models.

(1) The Jaccard index statistic is used for gauging the similarity and diversity of sample sets. It shows the agreement between the ground truth  $B$  and the predicted set of pixels  $A$  and is given as:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (8)$$

(2) The Dice score measures the overlap between the ground truth  $B$  and the predicted set of pixels  $A$  as follows:

$$D(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|}. \quad (9)$$



### 3.3. Implementation Details

U-shaped GAN is implemented in Python using the PyTorch framework. The gray chest radiographs are resized to  $1,024 \times 1,024$  and converted to RGB images compatible with the pre-trained ResNet-50 before placement in the network. The weights of ResNet-50 in our discriminator are pre-trained on the ImageNet dataset (33). The default 5-fold cross-validation is considered to train the semi-supervised model. In the semi-supervised approach, U-shaped GAN is trained with the JSRT or MC dataset with a portion of pixel-wise annotated data and the remainder without pixel-level annotations with 500 epochs. The Adam optimizer (34) is used to train 500 epochs of the generator and discriminator with initial learning rates of 0.001 and 0.0001, respectively, and multiplied by 0.1 after 200 epochs. For semi-supervised learning, we randomly select 12.5, 25, and 50% of the radiographs from the training set as the annotated set, with the remainder forming the unannotated set. Moreover, we train our model with 100% of the training dataset in the supervised approach. An ablation study is performed to discover the performance contribution from the modified architecture of U-shaped GAN and the proposed pixel-level semi-supervised loss. As in Li et al. (35), we train U-shaped GAN *via* the semi-supervised approach and supervised approach with 35 annotated radiographs on the JSRT dataset and with 24 annotated radiographs on the MC dataset. The supervised approach is conducted solely with the same annotated set and segmentation network. Moreover, we explore the effect of U-shaped GAN with the original GAN loss, called the original approach, by adding classification layers to U-shaped GAN paralleling with the decoder of our discriminator. The classification layers are identical to those in ResNet-50. In the original approach, the classification layers discriminate the real radiographs from the fake ones, and our discriminator just works as a segmentation network to predict the probability of belonging to the lungs of each pixel in the original approach. For UDA, we first employ the MC and JSRT datasets as the source and target domains,

**TABLE 1 |** Comparison of U-shaped generative adversarial network (GAN) with the state-of-the-art semi-supervised model.

Dataset	Model	Main method	Dice	IoU
JSRT	Li (35)	Semi-supervised CNN(35)	0.967	-
	U-shaped GAN	Semi-supervised GAN(35)	<b>0.971</b>	<b>0.944</b>

The numbers after the main method are the numbers of the annotated radiographs used in the semi-supervised approach. Numbers in bold indicate the best result among the models.

respectively, and then swap their roles. We randomly split each dataset into 7:1:2 for training, validation, and test sets. We train U-shaped GAN similarly to the semi-supervised approach using the source and target data as the annotated and unannotated data, respectively.

## 4. RESULTS

### 4.1. Semi-Supervised Segmentation

U-shaped GAN is trained on the MC and JSRT datasets independently. The comparison with the state-of-the-art semi-supervised CNN (35) is shown in **Table 1**. As few semi-supervised models on chest radiographs are available, we also compare U-shaped GAN with 1) human observation (32); 2) traditional methods (30, 36); and 3) supervised CNNs (21, 35, 37, 38). The comparison is shown in **Table 2**.

U-shaped GAN trained with 100% annotated data achieves a performance increase of 0.4–10.8% over the state-of-the-art traditional models and supervised CNNs on both the JSRT and MC datasets. The results validate the effectiveness of the design of the segmentation network.

Our semi-supervised model (Dice = 0.975, IoU = 0.951) trained with 25% annotated data outperforms the state-of-the-art supervised models and human observation on the JSRT dataset. Our semi-supervised model (Dice = 0.968, IoU = 0.940) trained

**TABLE 2** | Comparison of U-shaped GAN with other lung segmentation methods for chest radiograph datasets.

Dataset	Model	Main method	Dice	IoU
JSRT	Human (32)	Human observation	-	0.946 ± 0.018
	Candemir (30)	Traditional method	0.967 ± 0.008	0.954 ± 0.015
	U-net (35)	Supervised CNN	0.946	-
	InvertedNet (21)	Supervised CNN	0.974	0.950
	Li (35)	Supervised CNN	0.967	-
	U-shaped GAN	Supervised GAN	<b>0.979 ± 0.001</b>	<b>0.958 ± 0.003</b>
	U-shaped GAN	Semi-supervised GAN(25%)	0.975 ± 0.001	0.951 ± 0.002
MC	Bosdelekidis (36)	Traditional method	0.923	0.862
	Candemir (30)	Traditional method	0.960 ± 0.018	0.941 ± 0.034
	U-net (37)	Supervised CNN	-	0.942 ± 0.046
	Souza (38)	Supervised CNN	0.936	0.881
	U-shaped GAN	Supervised GAN	<b>0.976 ± 0.006</b>	<b>0.955 ± 0.010</b>
	U-shaped GAN	Semi-supervised GAN(25%)	0.968 ± 0.011	0.940 ± 0.019

The number after the main method is the proportion of annotated radiographs from the training set in the semi-supervised approach. Numbers in bold indicate the best result among the models.

with 25% annotated data outperforms the most state-of-the-art supervised models but performs slightly worse than the U-net model (IoU = 0.942) on the MC dataset. Our proposed semi-supervised model achieves outstanding performance with limited annotated datasets. Moreover, U-shaped GAN outperforms the state-of-the-art semi-supervised model (35) in both supervised and semi-supervised settings by 1.2 and 0.8%, respectively. **Figure 5** shows a few examples of semi-supervised results with U-shaped GAN. The ground truth contour of the lungs is shown in green, and the segmentation result of the algorithm is in red.

We evaluate our approach with 12.5, 25, 50, and 100% annotated radiographs (remaining portions consist of unannotated radiographs). The annotated radiographs in the data splits are randomly sampled from the whole dataset. Notably, the approach works well even with 12.5% annotated data, as shown in **Table 3**. For the details, readers are referred to **Supplementary Figures 1, 2**.

We apply U-shaped GAN on a pneumothorax segmentation dataset (39) with a semi-supervised approach. This dataset contains 2,669 radiographs with annotated pneumothorax lesion areas. U-shaped GAN shows promising performance on the pneumothorax segmentation in the semi-supervised approach. Most of the results predict rough areas of the pneumothorax lesion correctly, which provides credible help to the radiologist to find the lesion rapidly. For the details, readers are referred to **Supplementary Figure 3**.

## 4.2. Unsupervised Domain Adaptation

We use the MC and JSRT datasets as the source and target domains, respectively, and then swap their roles for UDA. The performances of our UDA model on the target domains are compared under various settings: 1) the model being trained on source data and tested on the target domain with no DA (T-noDA); 2) UDA model testing on the source domain (S-test); 3) human observation (32); 4) UDA models with CNNs (15, 37).

As shown in **Table 4**, when directly applying the learned source domain model to target data, the model performance significantly degrades, indicating that domain shift would severely impede the generalization performance of CNNs. However, remarkable improvements are achieved by applying the unsupervised loss on the target images. Compared to the T-noDA results, the segmentation predictions increase by 3.3% and 5.3% on the JSRT and MC datasets, respectively, with our UDA approach.

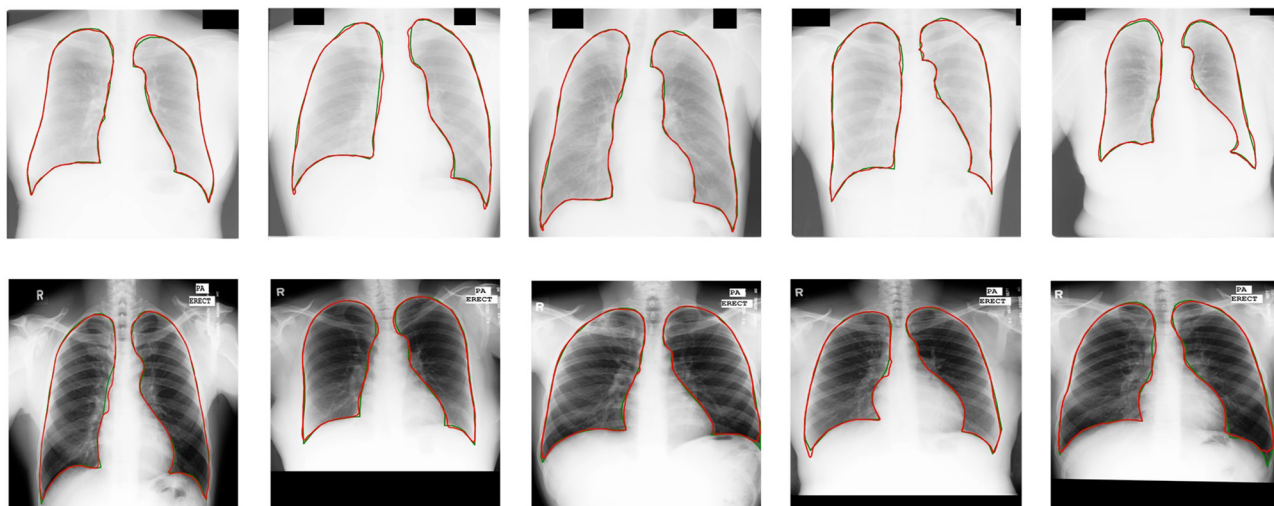
Experimental results demonstrate a significant enhancement in performance compared to other models. Compared with other UDA models based on CNNs (15, 37), U-shaped GAN achieves a significant improvement over MUNIT (5.7%), CyUDA (4.0%), and SeUDA (2.1%). Moreover, U-shaped GAN is even comparable to human observation. **Figure 6** shows a few examples of UDA results with U-shaped GAN. The ground truth contour of the lungs is shown in green, and the segmentation result of the algorithm is in red.

Compared to models trained with 100% annotated data on source domains, the UDA models obtain similar accuracy on the source domains (S-test). Therefore, U-shaped GAN is able to improve results on the target domains while maintaining segmentation performance on the source domains. For the details, readers are referred to **Supplementary Figures 4, 5**.

## 4.3. Ablation Study

For the ablation study, we study the effectiveness of our modified segmentation network, the architecture of U-shaped GAN, and the proposed pixel-level semi-supervised loss. To demonstrate the advantage of U-shaped GAN, we report the scores of U-net (the original U-shaped net), U-shaped GAN trained with the supervised approach, U-shaped GAN trained with the original approach, and U-shaped GAN trained with our pixel-level loss.

First, the results of the comparison of our modified segmentation network and the original U-shaped net are shown in **Table 5**. The segmentation network of U-shaped GAN is



**FIGURE 5 |** Semi-supervised results (25% annotated data) with U-shaped GAN. The radiographs from the Japanese Society of Radiological Technology (JSRT) dataset and Montgomery County (MC) dataset appear at the top and bottom, respectively. Green and red contours indicate the true ground and automatic segmentation results, respectively.

**TABLE 3 |** Comparison of the results of U-shaped GAN trained with different proportions of annotated data.

Dataset	Annotated data (%)	Dice	IoU
JSRT	100	$0.979 \pm 0.001$	$0.958 \pm 0.003$
	50	$0.977 \pm 0.002$	$0.956 \pm 0.003$
	25	$0.975 \pm 0.001$	$0.951 \pm 0.002$
	12.5	$0.964 \pm 0.007$	$0.934 \pm 0.008$
MC	100	$0.976 \pm 0.006$	$0.955 \pm 0.010$
	50	$0.973 \pm 0.007$	$0.949 \pm 0.013$
	25	$0.968 \pm 0.011$	$0.940 \pm 0.019$
	12.5	$0.958 \pm 0.021$	$0.922 \pm 0.034$

designed with high resolution radiographs following the main idea of the U-shaped net (24). It is shown that our modified segmentation network improves the prediction by 3.28% on the JSRT dataset and 2.02% on the MC dataset when trained with the whole annotated datasets.

Second, the effectiveness of the architecture of U-shaped GAN is investigated. The U-shaped net is incorporated into the structure of GAN leveraging unannotated data to assist the segmentation task. By adding an original GAN loss to the supervised approach, the Dice scores increase from 0.968 to 0.970 on the JSRT dataset and from 0.966 to 0.971 on the MC dataset. The architecture of U-shaped GAN is successful in leveraging unannotated data to find a representation for the whole dataset, shown in Table 6.

Third, we show the comparison between results gained by the GAN original loss and our pixel-level GAN loss with semi-supervised training. The pixel-level GAN loss increases the capacity of U-shaped GAN in finding the representation of the

**TABLE 4 |** Comparison of segmentation results among different unsupervised domain adaptation (UDA) methods.

Dataset	Model	Main method	Dice	IoU
MC → JSRT	Human observation (32)	Human observation	-	0.946
	T-noDAg	-	0.934	0.895
	S-test	-	0.981	0.963
	MUNIT (37)	UDA with MUNIT	-	0.882
	CyUDA (15)	UDA with CycleGAN	0.928	-
	SeUDA (15)	UDA with CycleGAN	0.945	-
	Our method	UDA with GAN	<b>0.965</b>	<b>0.932</b>
JSRT → MC	T-noDAg	-	0.918	0.880
	S-test	-	0.980	0.961
	Our method	UDA with GAN	<b>0.967</b>	<b>0.936</b>

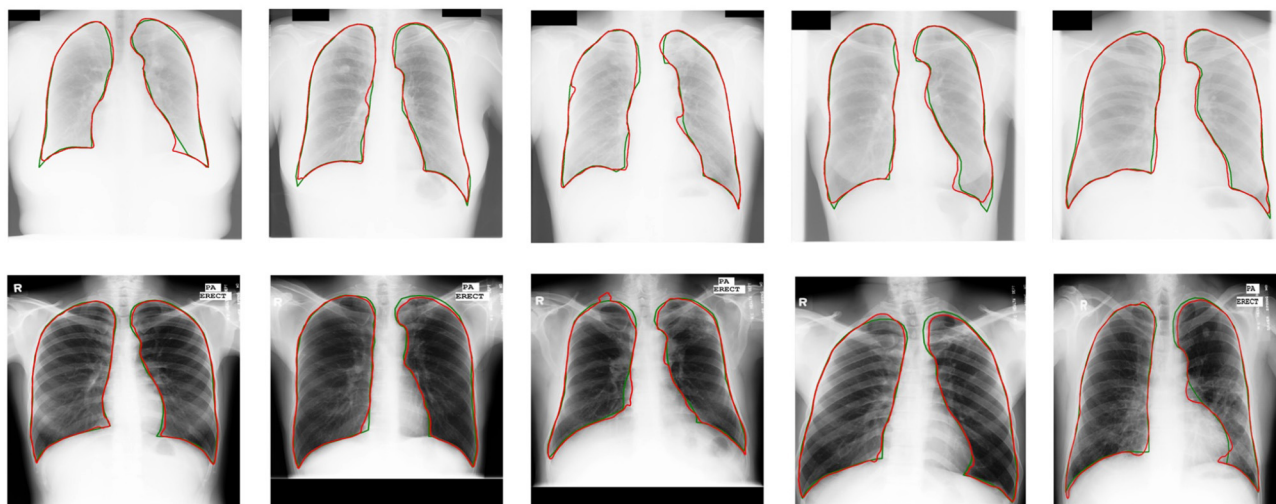
Numbers in bold indicate the best result among the models.

whole dataset. The segmentation results improve from Dice = 0.970, IoU = 0.941 to Dice = 0.971, IoU = 0.944 on the JSRT dataset and Dice = 0.971, IoU = 0.945 to Dice = 0.973, IoU = 0.948 on the MC dataset, shown in Table 6.

Some results of U-shaped GAN with different training approaches on confusing samples are shown in Figure 7. U-shaped GAN seems to be resistant to interference by irrelevant features, such as other organs and lesion areas, by using the GAN architecture and to increase the resistant capability by the pixel-level loss.

## 5. DISCUSSION

In this study, we propose U-shaped GAN to address the scarcity of annotated data and domain shift. U-shaped GAN



**FIGURE 6 |** UDA results with U-shaped GAN. The radiographs from the JSRT dataset with MC → JSRT UDA results are on the top; the radiographs from the MC dataset with JSRT → MC UDA results are on the bottom. Green and red contours indicate the ground truth and automatic segmentation results, respectively.

**TABLE 5 |** Comparison of U-shaped GAN and U-net trained with the whole dataset.

Dataset	Model	Main method	Dice	IoU
JSRT	U-net (35)	Supervised approach	0.946	-
	U-shaped GAN	Supervised approach	<b>0.977</b>	<b>0.955</b>
MC	U-net (37)	Supervised approach	-	0.942
	U-shaped GAN	Supervised approach	<b>0.980</b>	<b>0.961</b>

Numbers in bold indicate the best result among the models.

for radiographs shows strong performance in semi-supervised learning and UDA approaches. To handle datasets from multiple medical centers conveniently and efficiently, we combine semi-supervised learning and UDA in radiograph segmentation into a single method. U-shaped GAN functions similarly with the annotated and unannotated data in the semi-supervised and UDA approaches. The effectiveness of the model is demonstrated through extensive experiments.

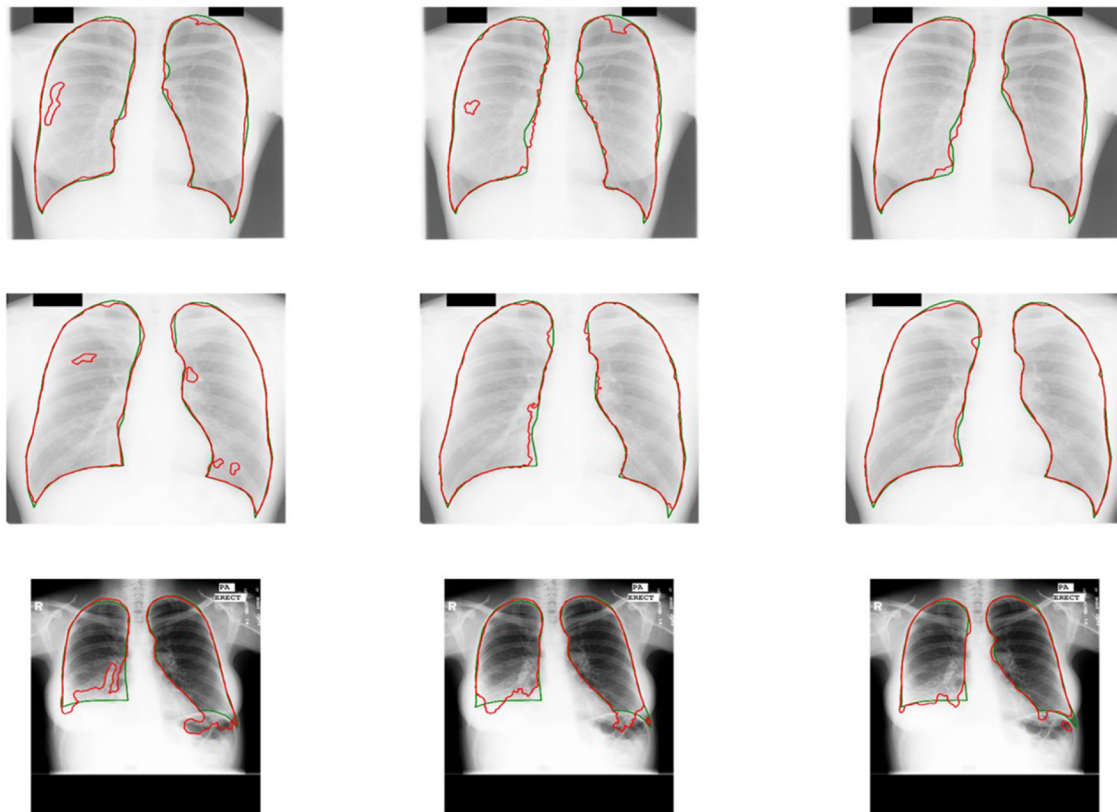
Training models with high image resolution is effective; however, it would increase the computational burden (20, 21). The previous models concentrated on the  $256 \times 256$  or  $512 \times 512$  image resolutions (15, 21, 35, 37, 38). We propose U-shaped GAN for high resolution radiographs ( $1,024 \times 1,024$ ). We use pointwise convolution for dimensionality reduction, decreasing the number of feature maps while retaining their salient features. Moreover, we design the U-shaped net with a pretrained ResNet-50 as encoder, which further reduces the computational burden of the training encoder from scratch. U-shaped GAN trained with the whole annotated data achieves a more accurate performance than the state-of-the-art supervised models as well as the original U-shaped net. This prediction result shows that U-shaped GAN is effective for segmentation prediction.

**TABLE 6 |** Comparison of U-shaped GAN with different training approaches.

Dataset	Model	Main method	Dice	IoU
JSRT	U-shaped GAN	Supervised approach(35)	0.968	0.939
	U-shaped GAN	Original approach(35)	0.970	0.941
	U-shaped GAN	Semi-supervised approach(35)	<b>0.971</b>	<b>0.944</b>
MC	U-shaped GAN	Supervised approach(24)	0.966	0.936
	U-shaped GAN	Original approach(24)	0.971	0.945
	U-shaped GAN	Semi-supervised approach(24)	<b>0.973</b>	<b>0.948</b>

The number after the main method is the number of the annotated radiographs used in training. Numbers in bold indicate the best result among the models.

The previous studies often analyzed semi-supervised learning and UDA problems separately (3, 11, 12, 14–16). In semi-supervised learning, the previous GANs were usually proposed to distinguish between segmentation probability maps and the ground truth (3, 11, 12). The generators produced the segmentation probability maps as the segmentation networks (3, 11, 12). In UDA, the previous GANs were usually proposed to distinguish between source data and target data (14–16). The generators transferred the target domain to the source domain and extra networks were designed for the segmentation (14–16). The annotated and unannotated data may come from either the same or different domains in the dataset collected from multiple centers. Dealing with the two problems separately increases the model complexity. Moreover, separating the dataset to train two segmentation networks decreases the utilization efficiency of collected data. Therefore, we propose a single model to deal with the semi-supervised learning and UDA approaches at the same time. We attribute the model's remarkable generalization capabilities to the effective use of the unannotated data. We use adversarial learning to achieve a representation for lung



**FIGURE 7 |** Comparison of the results of U-shaped GAN with different training approaches on confusing samples. The radiographs in the first row are results of the supervised approach. The second row is the results of the original approach. The third row is the results of the semi-supervised approach. Green and red contours indicate the ground truth and automatic segmentation results, respectively.

segmentation in chest radiographs. In U-shaped GAN, we employ a generator to generate realistic data, which, together with the real data (most of them are unannotated data), force the discriminator to find the most salient features. Our discriminator, which in addition to classifying the pixels into lungs, determines whether a given pixel belongs to the real or generated data.

U-shaped GAN exploits more widely available unannotated data to complement small annotated data with a semi-supervised loss. U-shaped GAN achieves greater performance than the state-of-the-art semi-supervised model. Moreover, it is comparable to the supervised models with 25% annotated data. U-shaped GAN works well even with 12.5% annotated data with Dice scores of 0.964 and 0.958 on JSRT and MC datasets, respectively. Unlike the previous semi-supervised study using the generators to produce the segmentation probability maps as the segmentation networks (3, 11, 12, 25), U-shaped GAN uses the generator to generate realistic data. The realistic data, together with the real data, force our discriminator to learn a better representation for the radiographs. Compared with the result achieved by trained in the supervised approach, U-shaped GAN achieves increased performance in the semi-supervised learning approach with adversarial learning. Instead of discriminating real or

fake labels on image-level (3, 11, 12, 25), a pixel-level loss is proposed to extract more information from the radiographs. The segmentation accuracy is improved when using the proposed loss. For the data from one domain, U-shaped GAN effectively leverages the unannotated data to achieve high segmentation accuracy and reduces the cost of medical image annotation.

U-shaped GAN is extended to UDA to reduce domain shift without the extra expense of annotation on the target domain. Instead of transferring the target domain to the source domain (14–16, 37), our generator generates realistic data. Discriminating the realistic data from the real ones, our discriminator learns a better representation. U-shaped GAN is better than the state-of-the-art UDA models and comparable to human observation. It achieves high accuracy on the target domain while maintaining the accuracy on the source domain (S-test). Thus, the model can be trained with data collected from multiple medical centers. Regardless of whether the unannotated data come from single or multiple domains, the prediction accuracies on their corresponding domains are increased, and the accuracies on other domains are maintained. Because the same networks are used in the two approaches, the datasets are sufficient to train U-shaped GAN and generalize the model among various domains, making it suitable for clinical

applications in a multiple center system. In addition, using the same architecture at multiple medical centers reduces the model complexity.

Results of our evaluation are promising, but U-shaped GAN has only been fully tested with lung segmentation. In the future, we will extend the model to detect a wider range of lung diseases by collecting additional chest radiographs of different diseases from multiple medical centers.

## 6. CONCLUSION

In this study, we propose U-shaped GAN to overcome the crucial problems caused by scarce labeled data and inevitable domain shift. The GAN-based model is designed at a high resolution ( $1,024 \times 1,024$ ) for effective segmentation. The semi-supervised learning approach and UDA approach are modeled into a unified framework for effective radiograph segmentation. We leverage unannotated and annotated data with a pixel-level semi-supervised loss. U-shaped GAN is compatible with varying data distributions of multiple medical centers, with efficient training and optimizing performance. Our experiment results demonstrate that U-shaped GAN achieved more accurate lung segmentation performance as compared with the state-of-the-art models. U-shaped GAN is more appealing to the model development and clinical application by eliminating the need to use two different models to deal with the aforementioned problems.

## REFERENCES

1. Souly N, Spampinato C, Shah M. Semi supervised semantic segmentation using generative adversarial network. In: *Proceedings of the IEEE International Conference on Computer Vision*. Venice: IEEE (2017). p. 5688–96. doi: 10.1109/ICCV.2017.606
2. Hung WC, Tsai YH, Liou YT, Lin YY, Yang MH. Adversarial learning for semi-supervised semantic segmentation. *arXiv preprint arXiv:180207934* (2018).
3. Zheng H, Lin L, Hu H, Zhang Q, Chen Q, Iwamoto Y, et al. Semi-supervised segmentation of liver using adversarial learning with deep atlas prior. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer (2019). p. 148–56.
4. Wang H, Gu H, Qin P, Wang J. CheXLocNet: automatic localization of pneumothorax in chest radiographs using deep convolutional neural networks. *PLoS ONE*. (2020) 15:e0242013. doi: 10.1371/journal.pone.0242013
5. Van Eycke YR, Foucart A, Decaestecker C. Strategies to reduce the expert supervision required for deep Learning-Based segmentation of histopathological images. *Front Med*. (2019) 6:222. doi: 10.3389/fmed.2019.00222
6. Nawrocki T, Maldjian PD, Slasky SE, Contractor SG. Artificial intelligence and radiology: have rumors of the radiologist's demise been greatly exaggerated? *Acad Radiol*. (2018) 25:967–72. doi: 10.1016/j.acra.2017.12.027
7. Jin D, Harrison AP, Zhang L, Yan K, Wang Y, Cai J, et al. Chapter 14 - artificial intelligence in radiology. In: Xing L, Giger ML, Min JK, editors. *Artificial Intelligence in Radiology*. Cambridge, MA: Academic Press (2021). p. 265–89.
8. Zhang L, Wang X, Yang D, Sanford T, Harmon S, Turkbey B, et al. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. *IEEE Trans Med Imaging*. (2020) 39:2531–40. doi: 10.1109/TMI.2020.2973595
9. Briganti G, Le Moine O. Artificial intelligence in medicine: today and tomorrow. *Front Med*. (2020) 7:27. doi: 10.3389/fmed.2020.00027

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

HG and HW conceived the idea for this study. HW worked on the end-to-end implementation of the study. JW provided relevant insights on the clinical impact of the research work and handled the redaction of the paper. PQ managed the project and provided the funding for the research. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Natural Science Foundation of China (grant nos. 61633006 and 81872247) and the Fundamental Research Funds for the Central Universities, China (grant no. DUT21YG118).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2021.782664/full#supplementary-material>

10. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: *Advances in Neural Information Processing Systems*. Montréal (2014). p. 2672–80.
11. Liu X, Cao J, Fu T, Pan Z, Hu W, Zhang K, et al. Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning. *IEEE Access*. (2018) 7:3046–3061. doi: 10.1109/ACCESS.2018.2889321
12. Decourt C, Duong L. Semi-supervised generative adversarial networks for the segmentation of the left ventricle in pediatric MRI. *Comput Biol Med*. (2020) 123:103884. doi: 10.1016/j.compbio.2020.103884
13. Ghahfoorian M, Mehrtash A, Kapur T, Karssemeijer N, Marchiori E, Pesteie M, et al. Transfer learning for domain adaptation in mri: application in brain lesion segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer (2017). p. 516–24.
14. Kamnitsas K, Baumgartner C, Ledig C, Newcombe V, Simpson J, Kane A, et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: *International Conference on Information Processing in Medical Imaging*. Cham: Springer (2017). p. 597–609.
15. Chen C, Dou Q, Chen H, Heng PA. Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation. In: *International Workshop on Machine Learning in Medical Imaging*. Cham: Springer (2018). p. 143–51.
16. Shen R, Yao J, Yan K, Tian K, Jiang C, Zhou K. Unsupervised domain adaptation with adversarial learning for mass detection in mammogram. *Neurocomputing*. (2020) 393:27–37. doi: 10.1016/j.neucom.2020.01.099
17. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI: IEEE (2017). p. 2097–106.

18. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kadom N, Kansagra AP, et al. Deep learning in radiology. *Acad Radiol.* (2018) 25:1472–80. doi: 10.1016/j.acra.2018.02.018
19. Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M. Covid-19 image data collection: prospective predictions are the future. *arXiv preprint arXiv:2006.11988.* (2020).
20. Eslami M, Tabarestani S, Albarqouni S, Adeli E, Navab N, Adjouadi M. Image-to-images translation for multi-task organ segmentation and bone suppression in Chest X-Ray radiography. *IEEE Trans Med Imaging.* (2020) 39:2553–65. doi: 10.1109/TMI.2020.2974159
21. Novikov AA, Lenis D, Major D, Hladvka J, Wimmer M, Bhler K. Fully convolutional architectures for multiclass segmentation in chest radiographs. *IEEE Trans Med Imaging.* (2018) 37:1865–76. doi: 10.1109/TMI.2018.2806086
22. Donahue J, Krähenbühl P, Darrell T. Adversarial feature learning. *arXiv preprint arXiv:1605.09782.* (2016).
23. Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep Learning.* Cambridge, MA: MIT press (2016).
24. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer (2015). p. 234–41. doi: 10.1007/978-3-319-24574-4\_28
25. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434.* (2015).
26. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Las Vegas, NV: IEEE (2016). p. 770–8.
27. Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022.* (2016).
28. Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. In: *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Vol. 30.* Atlanta, GA (2013). p. 3.
29. Shiraishi J, Katsuragawa S, Ikezoe J, Matsumoto T, Kobayashi T, Komatsu Ki, et al. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *Am J Roentgenol.* (2000) 174:71–4. doi: 10.2214/ajr.174.1.1740071
30. Candemir S, Jaeger S, Palaniappan K, Musco JP, Singh RK, Xue Z, et al. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Trans Med Imaging.* (2013) 33:577–90. doi: 10.1109/TMI.2013.2290491
31. Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan F, et al. Automatic tuberculosis screening using chest radiographs. *IEEE Trans Med Imaging.* (2013) 33:233–45. doi: 10.1109/TMI.2013.2284099
32. Van Ginneken B, Stegmann MB, Loog M. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Med Image Anal.* (2006) 10:19–40. doi: 10.1016/j.media.2005.02.002
33. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis.* (2015) 115:211–52. doi: 10.1007/s11263-015-0816-y
34. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980.* (2014).
35. Li D, Yang J, Kreis K, Torralba A, Fidler S. Semantic segmentation with generative models: semi-supervised learning and strong out-of-domain generalization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* (2021) p. 8300–11.
36. Bosdelekidis V, Ioakeimidis NS. Lung field segmentation in chest X-rays: a deformation-tolerant procedure based on the approximation of rib cage seed points. *Appl Sci.* (2020) 10:6264. doi: 10.3390/app10186264
37. Oliveira H, dos Santos J. Deep transfer learning for segmentation of anatomical structures in chest radiographs. In: *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI).* Parana: IEEE (2018). p. 204–11.
38. Souza JC, Diniz JOB, Ferreira JL, da Silva GLF, Silva AC, de Paiva AC. An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks. *Comput Methods Programs Biomed.* (2019) 177:285–96. doi: 10.1016/j.cmpb.2019.06.005
39. Filice RW, Stein A, Wu CC, Arteaga VA, Borstelmann S, Gaddikeri R, et al. Crowdsourcing pneumothorax annotations using machine learning annotations on the NIH chest X-ray dataset. *J Digit Imaging.* (2020) 33:490–6. doi: 10.1007/s10278-019-00299-9

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Gu, Qin and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Cost-Sensitive Uncertainty Hypergraph Learning for Identification of Lymph Node Involvement With CT Imaging

Qianli Ma<sup>1\*</sup>, Jielong Yan<sup>2</sup>, Jun Zhang<sup>3</sup>, Qiduo Yu<sup>1</sup>, Yue Zhao<sup>1</sup>, Chaoyang Liang<sup>1</sup> and Donglin Di<sup>2\*</sup>

<sup>1</sup> Department of Thoracic Surgery, China-Japan Friendship Hospital, Beijing, China, <sup>2</sup> The School of Software, Tsinghua University, Beijing, China, <sup>3</sup> Tencent AI Lab, Shenzhen, China

## OPEN ACCESS

### Edited by:

Kuanquan Wang,  
Harbin Institute of Technology, China

### Reviewed by:

Yuyao Zhang,  
ShanghaiTech University, China  
Li Wang,  
University of North Carolina at  
Chapel Hill, United States

### \*Correspondence:

Qianli Ma  
mars\_qhma@163.com  
Donglin Di  
donglin.ddl@gmail.com

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

**Received:** 21 December 2021

**Accepted:** 17 January 2022

**Published:** 10 February 2022

### Citation:

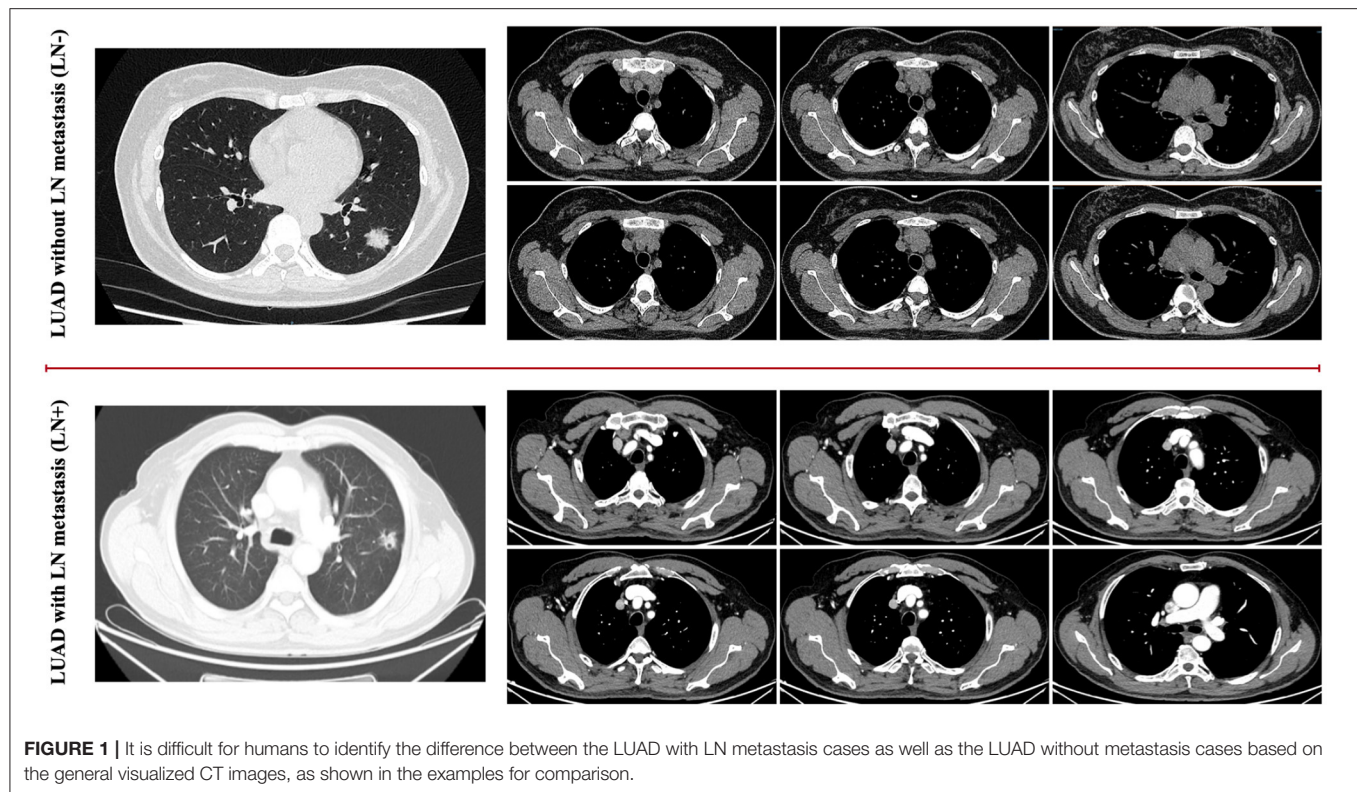
Ma Q, Yan J, Zhang J, Yu Q, Zhao Y,  
Liang C and Di D (2022)  
Cost-Sensitive Uncertainty  
Hypergraph Learning for Identification  
of Lymph Node Involvement With CT  
Imaging. *Front. Med.* 9:840319.  
doi: 10.3389/fmed.2022.840319

Lung adenocarcinoma (LUAD) is the most common type of lung cancer. Accurate identification of lymph node (LN) involvement in patients with LUAD is crucial for prognosis and making decisions of the treatment strategy. CT imaging has been used as a tool to identify lymph node involvement. To tackle the shortage of high-quality data and improve the sensitivity of diagnosis, we propose a Cost-Sensitive Uncertainty Hypergraph Learning (CSUHL) model to identify the lymph node based on the CT images. We design a step named “Multi-Uncertainty Measurement” to measure the epistemic and the aleatoric uncertainty, respectively. Given the two types of attentional uncertainty weights, we further propose a cost-sensitive hypergraph learning to boost the sensitivity of diagnosing, targeting task-driven optimization of the clinical scenarios. Extensive qualitative and quantitative experiments on the real clinical dataset demonstrate our method is capable of accurately identifying the lymph node and outperforming state-of-the-art methods across the board.

**Keywords:** lymph node involvement, CT imaging, hypergraph learning, cost-sensitive, lung cancer

## 1. INTRODUCTION

Lung cancer is the most commonly diagnosed cancer and the leading cause of cancer death worldwide (1, 2). About 2.1 million new lung cancer cases and 1.8 million deaths were predicted in 2018 (3). In 2020, these two numbers rise to 2.2 million and 1.8 million, respectively, (2). Lung adenocarcinoma (LUAD) is the most common type of lung cancer (4–6). The presence of metastasis in the lymph nodes (7) is an important prognostic factor in lung cancer. Accurate identification of lymph node (LN) involvement in patients with LUAD, as shown in **Figure 1**, is crucial for prognosis and treatment strategy decisions (8, 9). Patients without metastatic lymph nodes, or with only intrapulmonary or hilar lymph nodes, are generally considered candidates for straightforward resection. Although the sub-types of LUAD are found related to the predictors of LN metastasis, they are available postoperatively (10). Information of the preoperative LN metastasis is valuable for the adequacy of surgical resection and the decision of the adjuvant therapy (11). The accurate prediction of pathologic stage for patients with lung cancer is of utmost importance. Pathologic tumor stage is considered a pivotal factor relating to survival in NSCLC, and the 5-year survival rates vary from 83% in pathological stage IA to 23% in stage IIIA tumors (12). Computed tomography (CT) is commonly used for the evaluation of pulmonary



**FIGURE 1** | It is difficult for humans to identify the difference between the LUAD with LN metastasis cases as well as the LUAD without metastasis cases based on the general visualized CT images, as shown in the examples for comparison.

nodules (13, 14). Many studies were designed to determine whether pulmonary nodules are benign or malignant. Zhong et al. (15) propose to use relief-based feature method and support vector machine to evaluate the impact of radiomics features in predicting the prognosis of occult mediastinal lymph node metastasis in lung adenocarcinoma. Dai et al. (16) find that lymph node micrometastases are more frequently seen in adenocarcinomas with a micropapillary component, which could give suggestive prognostic information to patients with stage I resected lung adenocarcinoma with a micropapillary component. CT has been widely used as a noninvasive diagnostic modality for diagnosis, clinical staging, survival prediction, surveillance of therapeutic response in lung cancer patients (17, 18). However, few studies have used chest CT to explore whether lymph node metastasis in LUAD (19). Therefore, in order to make a better decision on the prognosis and treatment strategies of lung cancer, as well as more fully grasp the information of lymph nodes, in this work, we utilize CT to predict whether LUAD has lymph node metastasis.

There are two main challenges of identifying the lymph node with CT imaging, listed below, that motivate our approach.

1. Noisy data, due to the collection of clinical CT images using different reconstruction kernels and CT manufacturers, along with possible patient movements;
2. The reliable sensitivity of diagnosis is relatively more important and meaningful than other criteria in the clinical scenario.

For the first challenge, a few current research works are proposed to tackle the issue of clinical data quality, mainly focusing on noise and artifact reduction, super resolution and other aspects (20). Zhang and Yu (21) propose to train their convolutional neural network using virtual metal-inserted CT images, targeting on the noise of metal artifacts. Tan et al. (22) further utilizes the SRGAN neural network to reconstruct super resolution images from the original chest CT images to improve the resolution and ultimately improve the classification results of COVID-19. Due to the scale of available data in this task being limited, we adopt the two uncertainty measurements (23) to improve the quality of pathological representations, *i.e.*, epistemic and the aleatoric uncertainty, respectively, generated by the “Classifier Measuring” and “Statistical Measuring.” In this manner, our model is capable of allocating the different attentional weights combined with the two uncertainty measurements. Due to ignoring the underlying correlation between samples, some machine learning methods such as Random forest, Boosting, or CNN are lacking in effect, but graph learning and further hypergraph learning methods can make up for this deficiency. Hypergraph Learning methods (24–26) perform well on generate the high-order representations for complex data, such as whole-slide images (WSI) (27), CT imaging (23), drug-target interactions (28), *etc.* Therefore, given the data with uncertainty weights, we further propose an uncertainty hypergraph learning to extract the high-order representations from the CT images, which augments the pathological informative features effectively.

With regards to the second challenge, several works have made efforts on lymph node involvement. Zhou et al. (29) studied one or a combination of machine learning methods in Logistic regression, Random forest, XGBoost, and GBDT to construct lymph node metastasis in patients with poorly differentiated intramucosal gastric cancer. Supervised machine learning methods including random forest classifier, artificial neural network, decision tree, gradient boosting decision tree, extreme gradient boosting, and adaptive boosting can also be used to predict central lymph node metastasis in patients with papillary thyroid cancer (30). Besides improving the accuracy of overall prediction, we further focus on boosting the performance on sensitivity by the designed “Two-Stage Cost Sensitive Hypergraph Learning.” One stage is to capture the cost sensitivity of negative cases in the latent feature spaces, which will enable the hypergraph model to allocate the lymph node involvement cases more weights. The other is called “Supervising Cost Sensitivity,” making the loss function supervise the hypergraph model with more attached importance on the patients with LUAD for individual preoperative prediction of LN metastasis. Combining the structures introduced above, the overall framework we proposed, named as “Cost-Sensitive Uncertainty Hypergraph Learning (CSUHL)” has the ability to identify the lymph node accurately and outperform state-of-the-art methods on our collected real clinical dataset.

The main contributions of this paper are summarized as follows:

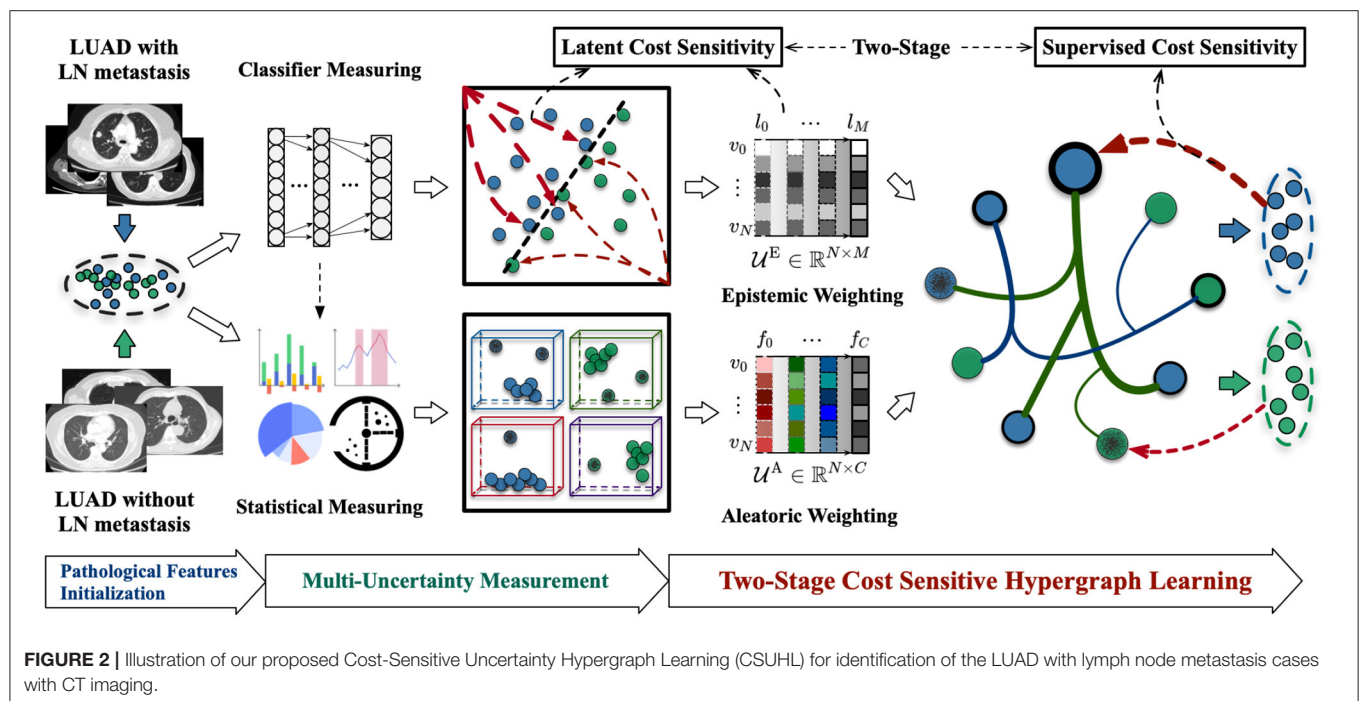
1. We propose a framework—CSUHL to tackle the task of identifying the lymph node, focusing on the uncertainty measurement of clinical CT imaging, as well as the cost sensitive hypergraph learning for identifying.
2. On our collected real clinical dataset, we conduct extensive experiments to demonstrate the proposed method consistently outperforms state-of-the-art methods across-the-board, relatively improving the performance on the accuracy (ACC), sensitivity (SEN), specificity (SPEC), Balance (BAC) by up to 3.90, 8.00, 2.02, and 4.95%, respectively, compared with the previous best method.

## 2. MATERIALS AND METHODS

In this section, we will first introduce the materials we collected and the processing for extracting the initial features in details. Then we will illustrate our proposed framework—“Cost-Sensitive Uncertainty Hypergraph Learning (CSUHL),” as shown in **Figure 2**, which is composed of three steps, *i.e.*, “Pathological Features Initialization,” “Multi-Uncertainty Measurement,” and “Two-Stage Cost Sensitive Hypergraph Learning,” respectively.

### 2.1. Materials and Preprocessing

In this study, a total of 61 CT images were collected, including 35 from lymph node negative patients and the rest 26 from lymph node positive patients. These images were provided by the China-Japan Friendship Hospital. All the cases were acquired from January 2017 to March 2019. The CT scanners used in this study include Aquilion ONE from TOSHIBA, MEDICAL System Revolution from GE, and SOMATOM Definition Flash from SEMENS. The CT protocol here includes: 120KV, reconstructed CT thickness is 1mm, and breath-hold at full inspiration. All images were de-identified before sending for analysis. This study was approved by the Institutional Review Board. Written informed consent was waived due to retrospective nature of the study.



**FIGURE 2 |** Illustration of our proposed Cost-Sensitive Uncertainty Hypergraph Learning (CSUHL) for identification of the LUAD with lymph node metastasis cases with CT imaging.

## 2.2. Cost-Sensitive Uncertainty Hypergraph Learning

### 2.2.1. Pathological Features Initialization

In this stage, we extract the initial features from the patient's CT images, consisting of regional features and radiomics features. We first apply the deep learning pre-trained method, named VB-Net (31), to segment the left/right lung, 5 lung lobes, 18 lung segments and infection lesions for each CT image in the portal software. In the expression of regional features, we generate a feature with a dimension of  $\mathbb{R}^{96}$  for each patient, expressing features such as the count of infected lesions and the mean value of lesion area. When extracting the radiomics features, we generated a feature with a dimension of  $\mathbb{R}^{93}$  for each patient, which means first-order intensity statistics and texture features. In the end, we concatenate the regional features and radiomics features obtained above to obtain an overall feature with a dimension of  $\mathbb{R}^C$  ( $C = 191$ ) representing patient information.

### 2.2.2. Multi-Uncertainty Measurement

As shown in **Figure 2**, there are two types of uncertainty measurement in our method, namely “Epistemic Weighting” and “Aleatoric Weighting,” respectively. The epistemic uncertainty refers to the inability of model for classifying the lymph node involvement in patients with LUAD. We utilize the general Multilayer Perceptron (MLP) Neural Network with the dropout variation inference to classify the data based on the initialized features. Illustrated as the “Epistemic Weighting” module in **Figure 2**, denoted as  $\mathcal{U}^E \in \mathbb{R}^{N \times M}$ , the effect of the dropout can be attributed to imposing a Gaussian distribution on each layer during the inference stage. For  $N$  samples, there are  $M$  layers that, respectively, generate the epistemic uncertainty weights in different levels. For the  $M$  layers, each case with  $\mathbf{x}^*$  features, is predicted for  $K$  times, the final epistemic weight for each case is calculated using the variance of these  $K$  values, formulated as Equation (1):

$$\mathcal{U}_{q(\mathbf{y}^*|\mathbf{x}^*)}^E(\mathbf{y}^*) = \frac{1}{KM} \sum_{k=1}^K \sum_{l=1}^M \hat{\mathbf{y}}_{(l)}^*(\mathbf{x}^*, \omega^k) \quad (1)$$

where  $i$  denotes the  $i$ th sample and  $k$  denotes the  $k$ th test with dropout.  $(l)$  denote the  $l$ th layer of the constructed MLP model.  $\hat{\mathbf{y}}^*$  denotes the corresponding output of the input  $\mathbf{x}^*$ .  $\omega^k = \{\mathbf{W}_i\}^k$  denotes the trainable variables for a model at the  $k$ th time.

We adopt a statistical measuring method to generate the aleatoric uncertainty weights  $\mathcal{U}^A$ . As shown in **Figure 2**, the dimension of aleatoric uncertainty weights is  $\mathbb{R}^{N \times C}$ , where  $N$  and  $C$  denotes the number of samples and the scale of features, respectively. For each feature, we estimate the weights of aleatoric uncertainty by minimizing the Kullback-Leibler (KL) divergence (32–34) between the standard feature distribution and the predicted features. The detailed theoretical derivation and demonstration of calculation can be found in UVHL (23) and the main formulation is following:

$$\mathcal{U}^A(\mathbf{x}_i) = \sigma_{\Theta}^2(\mathbf{x}_i) = \exp(\alpha_{\Theta}(\mathbf{x}_i)) \quad (2)$$

where  $\sigma_{\Theta}^2$  denotes the predicted variance. To avoid the potential division by zero,  $\alpha_{\Theta}(\mathbf{x})$  is the replacement of  $\log \sigma_{\Theta}^2(\mathbf{x})$ . Therefore,  $\alpha_{\Theta}: \mathbb{R}^{191} \mapsto \mathbb{R}^1$  is the module to yield the aleatoric uncertainty score for each case.

### 2.2.3. Two-Stage Cost Sensitive Hypergraph Learning

To identify the LUAD cases with higher sensitivity, we design a two-stage cost sensitive hypergraph learning in the final step of our framework. Given the vertices with initialized pathological features as well as the corresponding two types of uncertainty weights, we sequentially construct the uncertainty-vertex hypergraph and conduct cost-sensitive hypergraph learning.

When constructing the uncertainty-vertex hypergraph, we take each vertex  $v \in \mathcal{V}$  denoting one sample with the corresponding two types of uncertainty weights  $\mathcal{U}^E$  and  $\mathcal{U}^A$ . We use  $\mathcal{V}$  to denote the vertex set,  $\mathcal{E}$  denoting the hyperedges set, and  $\mathbf{W}$  denoting the pre-defined matrix of hyperedge weights. We adopt the k-nearest neighbors algorithm (KNN) to define the relationships for each vertex. There are two groups of hyperedges, respectively, stand for the regional features and radiomics features, denoted as  $\mathcal{E}_{reg}$  and  $\mathcal{E}_{rad}$ , which are represented by the corresponding incidence matrices  $\mathbf{H}_{reg} \in \mathbb{R}^{N \times |\mathcal{E}_{reg}|}$  and  $\mathbf{H}_{rad} \in \mathbb{R}^{N \times |\mathcal{E}_{rad}|}$ . The final combined global incidence matrix  $\mathbf{H} \in \mathbb{R}^{N \times (|\mathcal{E}_{reg}| + |\mathcal{E}_{rad}|)}$  can be formulated as Equation (3).

$$\mathbf{H}(v_i, e_j) = \begin{cases} \mathcal{U}_i^A + \mathcal{U}_i^E, & v_i \in e_j, e_j \in [\mathbf{H}_{reg} \parallel \mathbf{H}_{rad}] \\ 0, & v_i \notin e_j, e_j \in [\mathbf{H}_{reg} \parallel \mathbf{H}_{rad}] \end{cases} \quad (3)$$

where  $[\parallel]$  denotes the concatenating operation between two matrices. Finish constructing the hyperedges, the uncertainty-vertex hypergraph can be denoted as  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{H}, \mathbf{W}, \mathbf{U})$ , where  $\mathbf{U}$  is the summary matrix of  $\mathcal{U}^A$  and  $\mathcal{U}^E$ .

There are two stages of operating the cost sensitivity, namely latent cost sensitivity and supervising cost sensitivity. When measuring the epistemic uncertainty weights in the stage of “Multi-Uncertainty Measurement,” we design the first latent cost sensitivity by the modified cross-entropy loss function, formulated as follows:

$$\mathcal{L} = \frac{1}{N} \sum_i^N - [\lambda \cdot y_i \cdot \log(p_i) + (1 - \lambda) \cdot (1 - y_i) \cdot \log(1 - p_i)] \quad (4)$$

where  $y_i$  denotes the label of  $i$ th sample, whose value is 0 or 1 for negative case and positive case.  $\lambda \in (0, 1)$  is the parameter to represent the degree of cost sensitivity, whose value larger the more sensitivity. The other cost sensitivity for supervising is designed in the procedure of hypergraph learning, formulated as:

$$\mathcal{Q}_U(\mathbf{F}) = \arg \min_{\mathbf{F}} \{\Omega(\mathbf{F}) + \psi \tilde{\mathcal{R}}_{emp}(\mathbf{F})\} \quad (5)$$

where  $\Omega(\cdot)$  and  $\tilde{\mathcal{R}}_{emp}(\cdot)$  denote the smoothness regularizer function and the cost-sensitive empirical loss term, respectively. The hypergraph Laplacian matrix is

$\Theta_U = D_v^{-\frac{1}{2}} H W D_e^{-1} H^T D_v^{-\frac{1}{2}}$ . The smoothness regularizer function is formulated as:

$$\Omega(F, V, U, E, W) = \text{tr}(F^T (U^T - U^T \Theta_U U) F) \quad (6)$$

The cost-sensitive empirical loss term is designed as:

$$\begin{aligned} \tilde{\mathcal{R}}_{emp}(F, U) = \sum_{k=1}^K & \left[ \lambda \|F(:, k) - Y(:, k)\|_{y=1}^2 \right. \\ & \left. + (1 - \lambda) \|F(:, k) - Y(:, k)\|_{y=0}^2 \right] \end{aligned} \quad (7)$$

where  $F(:, k)$  is the  $k_{th}$  column of  $F$ . The  $\lambda$  in Equation (7) is same with the role in Equation (1).

The uncertainty vertex-weighted hypergraph loss function  $\mathcal{R}_{emp}(\cdot)$  can be further rewritten as:

$$\tilde{\mathcal{R}}_{emp}(F, U, \lambda) = \lambda \triangleright \left[ \text{tr}(F^T U^T U F + Y^T U^T U Y - 2F^T U^T U Y) \right] \quad (8)$$

Therefore, the target label matrix  $F$  can be obtained as:

$$F = \lambda \triangleright \left[ \psi(U^T - U^T \Theta_U U + \lambda U^T U)^{-1} U^T U Y \right] \quad (9)$$

where  $\triangleright$  denotes the degree of cost-sensitive  $\lambda$  operating on the following item, referring the effect in Equations 1 and 7. With the generated label matrix  $F \in \mathbb{R}^{n \times K}$  ( $K = 2$  in our task), the new coming testing case can be identified as LUAD or normal case accordingly.

### 3. EXPERIMENT

In this section, we will elaborate on the dataset, evaluation metrics, implementation, comparison methods, experimental results, and discussion.

#### 3.1. Evaluation Metrics

In the experiment, we adopt six metrics to evaluate the accuracy of the model.

(1). Accuracy (ACC) represents the proportion of correct predictions by the model and can be calculated as  $ACC = \frac{TP+TN}{TP+TN+FP+FN}$ . (2). Sensitivity (SEN), (3). Specificity (SPEC), (4). Positive Predictive Value (PPV), and (5). Negative Predictive Value (NPV), respectively, represent the proportion of correct predictions among the positive sample values, negative sample values, positive predicted values, and negative predicted values. The calculation formulas can be found in Table 1. (6). Balance (BAC) represents the mean value of SEN and SPEC.

Sensitivity, known as true positive rate, represents the proportion of patients with lymph node involvement that are successfully detected in the task. Specificity represents the possibility of patients without lymph node involvement that are excluded. Ideally, the model with both high sensitivity and high specificity is what we most hope for, but in practice, there is a trade-off between these two indicators. Compared with specificity, higher sensitivity basically possesses greater practical value.

**TABLE 1 |** The definition of the confusion matrix for identification of lymph node involvement.

	Classify as lymph node involvement	Classify as non involvement	
Lymph node involvement	True Positive (TP)	False Negative (FN)	$SEN = \frac{TP}{TP+FN}$
Non involvement	False Positive (FP)	True Negative (TN)	$SPEC = \frac{TN}{TN+FP}$
	$PPV = \frac{TP}{TP+FP}$	$NPV = \frac{TN}{TN+FN}$	

#### 3.2. Implementation

The entire dataset contains 61 CT images, of which 26 are lymph node involvement and the remaining 35 are on the contrary, are randomly partitioned into 10 subsets when comparing our model with the comparison models. In the task, the cross-validation process is performed 10 times, each time a subset is selected as the validation set, and the rest as the training set. To reduce the impact of random data on the results, the value of each metric in the experiment is an average of 10 times, and the standard deviation is reported as a comparison. To prevent inductive bias, each dimension of the training set features is normalized to [0, 1] using its own mean and variance and samples in validation set utilize the same parameters to normalize.

The uncertainty score  $\mathcal{U}_i$  of each sample and the uncertainty measurement model are generated by the overall training set. The algorithm used to construct the incidence matrix of hypergraph is K-nearest neighbors (KNN), which leads the choice of parameter  $\mathcal{K}$  to affect the effect. However, choosing  $\mathcal{K}$  is not a easy job. A hyperedge from a large  $\mathcal{K}$  connects too many nodes, which may over-describe the relationship between the data and generate noise. On the contrary, a hyperedge with small  $\mathcal{K}$  means that the number of connected nodes is small, which limits the exploration ability of the high-order relationship of hypergraph and not obtain full information. We conduct a strategy to learn the proper parameter  $\mathcal{K}$  automatically here. We put 2 to 20 in the candidate pool of  $\mathcal{K}$  to select the applicable  $\mathcal{K}$  for the task. In one training and testing, we cross-validate the training data 10 times for each  $\mathcal{K}$  in the candidate pool to obtain prediction values on different  $\mathcal{K}$ . The  $\mathcal{K}$  with the highest predicted score will be used in testing, so the whole process is the automatic selection of  $\mathcal{K}$ . The total training time for each fold is about 1 h, while the testing time is extremely fast, only taking about 5 s.

#### 3.3. Comparison Methods

The following methods are compared by our experiment:

1. Support Vector Machine (SVM) (35): It is a linear classifier that uses supervised learning to perform two-class classification of data. It relies on the convex quadratic programming problem to separate the samples correctly and away from the classification hyperplane.
2. Transductive Hypergraph Learning(tHL) (36): It is an algorithm for hypergraph embedding and transduction inference, mainly extending the spectral clustering technology of graphs to hypergraphs.

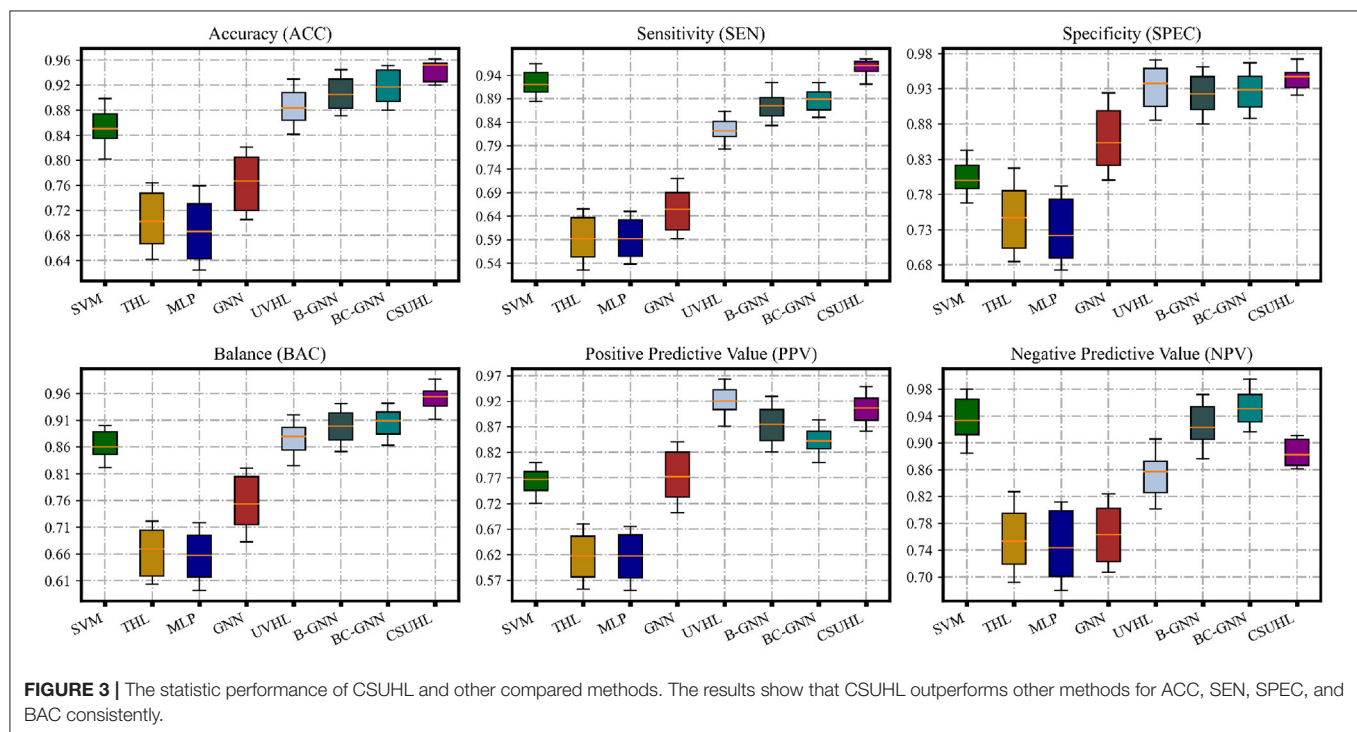
3. Multilayer Perceptron (MLP) Neural Network (37, 38): It contains a multi-layer feedforward neural network that maps multiple inputs to a single output.
4. GNN (39): It is a basic graph neural network that applies the local first-order approximation of the spectral graph convolution to determine the convolutional network structure for semi-supervised classification.
5. UVHL (23): It is an uncertainty vertex-weighted hypergraph learning method, which can reduce the problems caused by noisy data and confusing cases with clinical or imaging features.
6. B-GNN (40): It is a binary graph convolutional network, in which some floating-point operations are replaced by binary operations to achieve inference acceleration. A back-propagation method based on gradient approximation is used to train the binarized graph convolutional layer.

7. BC-GNN (40): It is an improved version of B-GNN, adding a cost-sensitive loss function.

### 3.4. Experimental Results

The experimental mean results and phenomena of our model and compared models can be seen in **Figure 3**; **Table 2**, and the following can be observed:

1. Comparing all methods, our proposed model is in a leading position in various metrics. Compared with non-graph-based methods (SVM and MLP), our models has a great lead. There are 12.04 and 38.89% improvements in ACC metric for the two, respectively, which shows that hypergraph has the ability to describe the correlation and handle this task.
2. In the GNN-based methods, compared with GNN, B-GNN, BC-GNN in ACC, the improvement is 24.22, 5.26, and 3.90%,



**TABLE 2 |** Prediction accuracy comparison of different methods on our collected LUAD dataset.

Methods		ACC		SEN		SPEC		BAC		PPV		NPV
SVM	( <i>p-value</i> )	0.85000	2.324e-5	0.92000	5.624e-4	0.80000	1.824e-4	0.86000	6.815e-5	0.76667	0.0498e-5	0.93333
THL	( <i>p-value</i> )	0.70238	1.173e-5	0.59167	1.438e-6	0.74667	4.235e-4	0.66917	1.037e-4	0.61667	0.1237e-5	0.75333
MLP	( <i>p-value</i> )	0.68571	6.734e-4	0.59167	3.568e-5	0.72167	8.967e-3	0.65667	2.358e-4	0.61667	8.845e-4	0.74333
GNN	( <i>p-value</i> )	0.76667	4.891e-4	0.65385	6.784e-4	0.85294	3.578e-4	0.75339	3.567e-4	0.77273	9.487e-4	0.76316
UVHL	( <i>p-value</i> )	0.88333	2.346e-3	0.82143	7.624e-4	0.93750	6.78e-4	0.87946	1.895e-3	<b>0.92000</b>	-	0.85714
B-GNN	( <i>p-value</i> )	0.90480	7.823e-3	0.87500	2.135e-3	0.92300	7.895e-3	0.89900	8.233e-3	0.87500	9.356e-3	0.92300
BC-GNN	( <i>p-value</i> )	0.91667	4.721e-2	0.88889	1.468e-3	0.92857	2.568e-2	0.90873	9.134e-3	0.84211	7.804e-3	<b>0.95122</b>
CSUHL	( <i>std</i> )	<b>0.95238†</b>	±0.0346	<b>0.96000†</b>	±0.0596	<b>0.94737†</b>	±0.0277	<b>0.95368†</b>	±0.0150	0.90654	±0.0286	0.88235

For each 10-fold, we compute the accuracy of the proposed method on testing data, and compare them with those of CSUHL via paired t-test to generate the *p-values* for each metric. ("†" denotes significance level is reached as  $p - value < 0.05$ ). The bold values represent the best values of the indicators in each set of experiments.

respectively, which proves that our methods can describe complex associations better.

3. Compared with hypergraph-based methods such as THL and UVHL, the model has 35.59 and 7.82% rises on the acc, respectively, which benefits from the uncertainty measurement and multiple loss.
4. Except for ACC, our method is the only one that exceeds and close to 95% on SEN and SPEC, respectively, which has practical value for actual medical diagnosis.
5. From **Figure 3**, it can be observed from the standard deviation that our model has more stable results compared to other comparison methods, which shows that our model provides more reliable and robust prediction results.

## 4. DISCUSSION

To evaluate the effectiveness of different uncertainty and different hypergraphs and different cost sensitivity, in this section, we conduct ablation experiments, respectively, to determine the contribution of each component.

### 4.1. Study on Multi-Uncertainty

To evaluate the effectiveness of different uncertainty, we conduct an ablation study, which uses aleatoric uncertainty or epistemic uncertainty, respectively.

#### 4.1.1. Aleatoric Uncertainty

Denoted as  $\mathcal{U}^A$ , the results of using aleatoric uncertainty individually are shown in row 1 of **Table 3**.

We can find out that the use of both types of uncertainty brings about 7.58, 4, and 10.53% growth in ACC, SEN, and SPEC, respectively, than using aleatoric uncertainty. It is worth mentioning that even with only aleatoric uncertainty, the model is still higher than most comparison methods in ACC and better than other methods in SEN according to **Table 2**.

More specifically, the pathological features with higher aleatoric uncertainty weights are consistent with the clinical experience, such as the distribution of different nodules, *i.e.*, lobulated nodules, spiculate nodules, and globular nodules.

#### 4.1.2. Epistemic Uncertainty

Denoted as  $\mathcal{U}^E$ , the results of using epistemic uncertainty individually are shown in row 2 of **Table 3**.

It can be observed using epistemic uncertainty alone lags behind 10.49, 7.85, and 12.54 than CSUHL in acc, sen, and spec metrics, respectively. Compared with the results of using aleatoric uncertainty, the indicators using epistemic uncertainty are lower, indicating that aleatoric uncertainty plays a greater role in our model. The combination of the two uncertainties is better than the single-use, which proves the effectiveness of multi-uncertainty.

## 4.2. Study on Types of Hypergraph

To evaluate the effectiveness of different hypergraphs, we conduct an ablation study, using regional hypergraph or radiomics hypergraph, which use regional features and radiomics features from CT, respectively.

#### 4.2.1. Regional Hypergraph

Denoted as  $\mathcal{G}_{reg}$ , the results of using regional hypergraph individually are shown in row 4 of **Table 3**.

The regional hypergraph only has an accuracy rate of about 80%, and the same for SEN and SPEC, indicating that only extracting the regional features of CT has little effect and the regional hypergraph cannot provide accurate correlation information.

#### 4.2.2. Radiomics Hypergraph

Denoted as  $\mathcal{G}_{rad}$ , the results of using radiomics hypergraph individually are shown in row 5 of **Table 3**.

The results of radiomics hypergraph are much better than the former, with ACC and SEN exceeding 90%, although there is still a gap in the combination of two hypergraphs. It can be found that the results in radiomics hypergraph have better sensitivity than specificity, which proves that the radiomic hypergraph has more advantages in identifying lymph node involvement. The combined hypergraph is higher in all indicators than when used alone, showing that it has the ability to utilize a variety of different features.

**TABLE 3** | Prediction accuracy comparison of different methods on our collected LUAD dataset.

Methods		ACC	SEN	SPEC	BAC	PPV	NPV
<b>1) Aleatoric Uncertainty (<math>\mathcal{U}^A</math>)</b>	(std)	0.88525 ±0.1845	0.92308 ±0.2451	0.85714 ±0.1684	0.89011 ±0.1795	0.82759 ±0.0781	<b>0.93750</b> ±0.1864
<b>2) Epistemic Uncertainty (<math>\mathcal{U}^E</math>)</b>	(std)	0.85246 ±0.0351	0.88462 ±0.0763	0.82857 ±0.0374	0.85659 ±0.0746	0.79310 ±0.0890	0.90625 ±0.0785
<b>3) CSUHL (<math>\mathcal{U}^A + \mathcal{U}^E</math>)</b>	(std)	<b>0.95238<sup>†</sup></b> ±0.0346	<b>0.96000<sup>†</sup></b> ±0.0596	<b>0.94737<sup>†</sup></b> ±0.0277	<b>0.95368<sup>†</sup></b> ±0.0150	<b>0.90654</b> ±0.0286	0.88235 ±0.0735
<b>4) Regional Hypergraph (<math>\mathcal{G}_{reg}</math>)</b>	(std)	0.80328 ±0.0567	0.80769 ±0.978	0.80000 ±0.1643	0.80385 ±0.0776	0.75000 ±0.1347	0.84848 ±0.1613
<b>5) Radiomics Hypergraph (<math>\mathcal{G}_{rad}</math>)</b>	(std)	0.90164 ±0.0891	0.92308 ±0.0346	0.88571 ±0.0917	0.90440 ±0.0176	0.85714 ±0.0783	<b>0.93939</b> ±0.0635
<b>6) CSUHL (<math>\mathcal{G}_{reg} + \mathcal{G}_{rad}</math>)</b>	(std)	<b>0.95238<sup>†</sup></b> ±0.0346	<b>0.96000<sup>†</sup></b> ±0.0596	<b>0.94737<sup>†</sup></b> ±0.0277	<b>0.95368<sup>†</sup></b> ±0.0150	<b>0.90654</b> ±0.0286	0.88235 ±0.0735
<b>7) Latent Cost Sensitivity</b>	(std)	0.86885 ±0.0678	0.88462 ±0.0341	0.85714 ±0.0867	0.87088 ±0.1456	0.82143 ±0.1034	0.90909 ±0.0918
<b>8) Supervising Cost Sensitivity</b>	(std)	0.91803 ±0.0451	0.92308 ±0.0813	0.91429 ±0.0561	0.91868 ±0.0971	0.88889 ±0.0936	<b>0.94118</b> ±0.0771
<b>9) CSUHL</b>	(std)	<b>0.95238<sup>†</sup></b> ±0.0346	<b>0.96000<sup>†</sup></b> ±0.0596	<b>0.94737<sup>†</sup></b> ±0.0277	<b>0.95368<sup>†</sup></b> ±0.0150	<b>0.90654</b> ±0.0286	0.88235 ±0.0735

For each 10-fold, we compute and report the average performance of the proposed method on testing data. The bold values represent the best values of the indicators in each set of experiments.

### 4.3. Study on Cost Sensitivity

To evaluate the effectiveness of different cost sensitivity, we conduct an ablation study, which uses latent cost sensitivity or supervising cost sensitivity, respectively.

#### 4.3.1. Latent Cost Sensitivity

The results of using latent cost sensitivity individually are shown in row 7 of **Table 3**.

When only latent cost sensitivity is used, it is equivalent to not using supervising cost sensitivity. As a result, the hypergraph information cannot be captured in order that various indicators are significantly reduced.

#### 4.3.2. Supervising Cost Sensitivity

The results of using supervising cost sensitivity individually are shown in row 8 of **Table 3**.

Compared with the combination of the two cost sensitivity, the supervising cost sensitivity has an accuracy disparity of about 3.61%, but it is higher than latent cost sensitivity. It should be noticed that cost sensitivity of using only supervising is the highest among the three on NPV, indicating that the true label is mostly negative in the samples identified as non lymph node involvement. In general, using two cost sensitivities together is better than using one of them, proving the effectiveness of cost sensitivity component.

## 5. CONCLUSION

In this paper, we propose a cost-Sensitive Uncertainty Hypergraph Learning (CSUHL) to identify lung adenocarcinoma (LUAD) cases with lymph node (LN) metastasis from the cases without lymph node (LN) metastasis. Confronting the challenging issues from the shortage of high-quality data and unreliable sensitivity of diagnosis, our proposed method employs three stages, namely “Pathological Features Initialization,” “Multi-Uncertainty Measurement,” and “Two-Stage Cost Sensitive Hypergraph Learning” to represent the

complex clinical information and formulate the high-order data correlation among the known LUAD with LN metastasis cases and the LUAD without LN metastasis cases. Through the epistemic and aleatoric uncertainty as well as the two types of cost sensitivity (latent and supervising), our method is capable of outperforming state-of-the-art methods on our collected LUAD dataset across the board.

In future work, we will further investigate the practical limitations on the computer-aid-diagnosis (CAD), such as enhancing the speed of inference, transferring the model to learn, and predicting the other related downstream clinical tasks.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Clinical Research Ethics Committee of China-Japan Friendship Hospital. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

QM and DD: study design. QM, QY, YZ, and CL: data collection. JY, JZ, and DD: data analysis. JZ and DD: supervision. JY and DD: manuscript writing. All authors contributed to the article and approved the submitted version.

## FUNDING

This study was supported by the Peking science and technology fund (No. Z191100006619008) and Elite Medical Professionals project of China-Japan Friendship Hospital (No. ZRJY2021-GG02).

## REFERENCES

1. Planchard D, Popat S, Kerr K, Novello S, Smit E, Faivre-Finn C, et al. Metastatic non-small cell lung cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* (2018) 29:iv192–iv237. doi: 10.1093/annonc/mdy275
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2021) 71:209–49. doi: 10.3322/caac.21660
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2018) 68:394–424. doi: 10.3322/caac.21492
4. Travis WD. Pathology of lung cancer. *Clin Chest Med.* (2011) 32:669–92. doi: 10.1016/j.ccm.2011.08.005
5. Relli V, Trerotola M, Guerra E, Alberti S. Abandoning the notion of non-small cell lung cancer. *Trends Mol Med.* (2019) 25:585–94. doi: 10.1016/j.molmed.2019.04.012
6. Zhang L, Zhang Z, Yu Z. Identification of a novel glycolysis-related gene signature for predicting metastasis and survival in patients with lung adenocarcinoma. *J Transl Med.* (2019) 17:1–13. doi: 10.1186/s12967-019-02173-2
7. Vansteenkiste JF, Stroobants SG, De Leyn P, Dupont PJ, Bogaert J, Maes A, et al. Lymph node staging in non-small-cell lung cancer with FDG-PET scan: a prospective study on 690 lymph node stations from 68 patients. *J Clin Oncol.* (1998) 16:2142–9.
8. De Leyn P, Lardinois D, Van Schil PE, Rami-Porta R, Passlick B, Zielinski M, et al. ESTS guidelines for preoperative lymph node staging for non-small cell lung cancer. *Elsevier Science BV.* (2007) 32:1–8. doi: 10.1016/j.ejcts.2007.01.075
9. Chen D, Ding Q, Wang W, Wang X, Wu X, Mao Y, et al. Characterization of extracapsular lymph node involvement and its clinicopathological characteristics in stage II–IIIA lung adenocarcinoma. *Ann Surg Oncol.* (2021) 28:2088–98. doi: 10.1245/s10434-020-09154-6
10. Park JK, Kim JJ, Moon SW, Lee KY. Lymph node involvement according to lung adenocarcinoma subtypes: lymph node involvement is influenced by lung adenocarcinoma subtypes. *J Thor Dis.* (2017) 9:3903. doi: 10.21037/jtd.2017.08.132
11. Dietel M, Bubendorf L, Dingemans AMC, Doooms C, Elmberger G, García RC, et al. Diagnostic procedures for non-small-cell lung cancer (NSCLC):

- recommendations of the European Expert Group. *Thorax*. (2016) 71:177–84. doi: 10.1136/thoraxjnl-2014-206677
12. León-Atance P, Moreno-Mata N, González-Aragoneses F, Cañizares-Carretero MÁ, García-Jiménez MD, Genovés-Crespo M, et al. Multicenter analysis of survival and prognostic factors in pathologic stage I non-small-cell lung cancer according to the new 2009 TNM classification. *Archivos de Bronconeumología (English Edition)*. (2011) 47:441–6. doi: 10.1016/j.arbres.2011.04.004
  13. Choe J, Lee SM, Do KH, Lee G, Lee JG, Lee SM, et al. Deep learning-based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. *Radiology*. (2019) 292:365–73. doi: 10.1148/radiol.2019181960
  14. Palumbo B, Bianconi F, Palumbo I, Fravolini ML, Ministrini M, Nuvoli S, et al. Value of shape and texture features from 18F-FDG PET/CT to discriminate between benign and malignant solitary pulmonary nodules: an experimental evaluation. *Diagnostics*. (2020) 10:696. doi: 10.3390/diagnostics10090696
  15. Zhong Y, Yuan M, Zhang T, Zhang YD, Li H, Yu TF. Radiomics approach to prediction of occult mediastinal lymph node metastasis of lung adenocarcinoma. *Am J Roentgenol*. (2018) 211:109–113. doi: 10.2214/AJR.17.19074
  16. Dai C, Xie H, Kadeer X, Su H, Xie D, Ren Y, et al. Relationship of lymph node micrometastasis and micropapillary component and their joint influence on prognosis of patients with stage I lung adenocarcinoma. *Am J Surg Pathol*. (2017) 41:1212–20. doi: 10.1097/PAS.0000000000000901
  17. Shim SS, Lee KS, Kim BT, Chung MJ, Lee EJ, Han J, et al. Non-small cell lung cancer: prospective comparison of integrated FDG PET/CT and CT alone for preoperative staging. *Radiology*. (2005) 236:1011–9. doi: 10.1148/radiol.2363041310
  18. Yu W, Tang C, Hobbs BP, Li X, Koay EJ, Wistuba II, et al. Development and validation of a predictive radiomics model for clinical outcomes in stage I non-small cell lung cancer. *Int J Radiat Oncol Biol Phys*. (2018) 102:1090–7. doi: 10.1016/j.ijrobp.2017.10.046
  19. Tsutani Y, Miyata Y, Nakayama H, Okumura S, Adachi S, Yoshimura M, et al. Prediction of pathologic node-negative clinical stage IA lung adenocarcinoma for optimal candidates undergoing sublobar resection. *J Thor Cardiovas Surg*. (2012) 144:1365–71. doi: 10.1016/j.jtcvs.2012.07.012
  20. Higaki T, Nakamura Y, Tatsugami F, Nakaura T, Awai K. Improvement of image quality at CT and MRI using deep learning. *Japanese J Radiol*. (2019) 37:73–80. doi: 10.1007/s11604-018-0796-2
  21. Zhang Y, Yu H. Convolutional neural network based metal artifact reduction in x-ray computed tomography. *IEEE Trans Med Imag*. (2018) 37:1370–81. doi: 10.1109/TMI.2018.2823083
  22. Tan W, Liu P, Li X, Liu Y, Zhou Q, Chen C, et al. Classification of COVID-19 pneumonia from chest CT images based on reconstructed super-resolution images and VGG neural network. *Health Inf Sci Syst*. (2021) 9:1–12. doi: 10.1007/s13755-021-00140-0
  23. Di D, Shi F, Yan F, Xia L, Mo Z, Ding Z, et al. Hypergraph learning for identification of COVID-19 with CT imaging. *Med Image Anal*. (2021) 68:101910. doi: 10.1016/j.media.2020.101910
  24. Feng Y, You H, Zhang Z, Ji R, Gao Y. Hypergraph neural networks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33 Honolulu, HI (2019). p. 3558–65.
  25. Fan H, Zhang F, Wei Y, Li Z, Zou C, Gao Y, et al. Heterogeneous hypergraph variational autoencoder for link prediction. *IEEE Trans Pattern Anal Mach Intell*. (2021). doi: 10.1109/TPAMI.2021.3059313. [Epub ahead of print].
  26. Gao Y, Zhang Z, Lin H, Zhao X, Du S, Zou C. Hypergraph learning: methods and practices. *IEEE Trans Pattern Anal Mach Intell*. (2020). doi: 10.1109/TPAMI.2020.3039374
  27. Di D, Li S, Zhang J, Gao Y. Ranking-based survival prediction on histopathological whole-slide images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Lima: Springer (2020). p. 428–38.
  28. Ruan D, Ji S, Yan C, Zhu J, Zhao X, Yang Y, et al. Exploring complex and heterogeneous correlations on hypergraph for the prediction of drug-target interactions. *Patterns*. (2021) 2:100390. doi: 10.1016/j.patter.2021.100390
  29. Zhou CM, Wang Y, Ye HT, Yan S, Ji M, Liu P, et al. Machine learning predicts lymph node metastasis of poorly differentiated-type intramucosal gastric cancer. *Sci Rep*. (2021) 11:1–7. doi: 10.1038/s41598-020-80582-w
  30. Wu Y, Rao K, Liu J, Han C, Gong L, Chong Y, et al. Machine learning algorithms for the prediction of central lymph node metastasis in patients with papillary thyroid cancer. *Front. Endocrinol*. (2020) 11:816. doi: 10.3389/fendo.2020.577537
  31. Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, et al. Lung infection quantification of COVID-19 in CT images with deep learning. *arXiv preprint arXiv:200304655*. (2020). doi: 10.1002/mp.14609
  32. Van Erven T, Harremoës P. Rényi divergence and Kullback-Leibler divergence. *IEEE Trans Inf Theory*. (2014) 60:3797–820. doi: 10.1109/TIT.2014.2320500
  33. Hershey JR, Olsen PA. Approximating the Kullback-Leibler divergence between Gaussian mixture models. In: *IEEE International Conference on Acoustics*. Vol. 4 (2007). Honolulu, HI p. IV–317.
  34. Moreno PJ, Ho PP, Vasconcelos N. A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications. In: *Advances in Neural Information Processing Systems*. (2004). p. 1385–1392.
  35. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. (1995) 20:273–97.
  36. Zhou D, Huang J, Schölkopf B. Learning with hypergraphs: clustering, classification, and embedding. In: *Advances in Neural Information Processing Systems*. CiteSeer (2007). p. 1601–1608.
  37. Thimm G, Fiesler E. High-order and multilayer perceptron initialization. *IEEE Trans Neural Netw*. (1997) 8:349–59.
  38. Orhan U, Hekim M, Ozer M. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Exp Syst Appl*. (2011) 38:13475–81. doi: 10.1016/j.eswa.2011.04.149
  39. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:160902907*. (2016).
  40. Wang J, Wang Y, Yang Z, Yang L, Guo Y. Bi-gcn: Binary graph convolutional network. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021). p. 1561–70.

**Conflict of Interest:** JZ is employed by Tencent AI Lab, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ma, Yan, Zhang, Yu, Zhao, Liang and Di. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Progression-Free Survival Prediction in Small Cell Lung Cancer Based on Radiomics Analysis of Contrast-Enhanced CT

Ningxin Chen<sup>1†</sup>, Ruikun Li<sup>2†</sup>, Mengmeng Jiang<sup>3</sup>, Yixian Guo<sup>3</sup>, Jiejun Chen<sup>3</sup>, Dazhen Sun<sup>2</sup>, Lisheng Wang<sup>2\*</sup> and Xiuzhong Yao<sup>3\*</sup>

<sup>1</sup> Department of Radiology, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, China, <sup>2</sup> Department of Automation, Shanghai Jiao Tong University, Shanghai, China, <sup>3</sup> Department of Radiology, Shanghai Institute of Medical Imaging, Zhongshan Hospital of Fudan University, Shanghai, China

## OPEN ACCESS

### Edited by:

Kuanquan Wang,  
Harbin Institute of Technology, China

### Reviewed by:

Runnan He,  
Peng Cheng Laboratory, China  
Bing Zhang,  
Nanjing Drum Tower Hospital, China

### \*Correspondence:

Xiuzhong Yao  
zsyxyz@163.com  
Lisheng Wang  
lswang@sjtu.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

Received: 11 December 2021

Accepted: 25 January 2022

Published: 24 February 2022

### Citation:

Chen N, Li R, Jiang M, Guo Y, Chen J,  
Sun D, Wang L and Yao X (2022)  
Progression-Free Survival Prediction  
in Small Cell Lung Cancer Based on  
Radiomics Analysis of  
Contrast-Enhanced CT.  
Front. Med. 9:833283.  
doi: 10.3389/fmed.2022.833283

**Purposes and Objectives:** The aim of this study was to predict the progression-free survival (PFS) in patients with small cell lung cancer (SCLC) by radiomic signature from the contrast-enhanced computed tomography (CT).

**Methods:** A total of 186 cases with pathological confirmed small cell lung cancer were retrospectively assembled. First, 1,218 radiomic features were automatically extracted from tumor region of interests (ROIs) on the lung window and mediastinal window, respectively. Then, the prognostic and robust features were selected by machine learning methods, such as (1) univariate analysis based on a Cox proportional hazard (CPH) model, (2) redundancy removing using the variance inflation factor (VIF), and (3) multivariate importance analysis based on random survival forests (RSF). Finally, PFS predictive models were established based on RSF, and their performances were evaluated using the concordance index (C-index) and the cumulative/dynamic area under the curve (C/D AUC).

**Results:** In total, 11 radiomic features (6 for mediastinal window and 5 for lung window) were finally selected, and the predictive model constructed from them achieved a C-index of 0.7531 and a mean C/D AUC of 0.8487 on the independent test set, better than the predictions by single clinical features (C-index = 0.6026, mean C/D AUC = 0.6312), and single radiomic features computed in lung window (C-index = 0.6951, mean C/D AUC = 0.7836) or mediastinal window (C-index = 0.7192, mean C/D AUC = 0.7964).

**Conclusion:** The radiomic features computed from tumor ROIs on both lung window and mediastinal window can predict the PFS for patients with SCLC by a high accuracy, which could be used as a useful tool to support the personalized clinical decision for the diagnosis and patient management of patients with SCLC.

**Keywords:** small cell lung cancer (SCLC), radiomics analysis, progression-free survival (PFS), contrast-enhanced CT, mediastinal window

## INTRODUCTION

Small cell lung cancer (SCLC) is an aggressive pulmonary neuroendocrine tumor that causes about 250,000 deaths worldwide yearly (1, 2). SCLC accounts for approximately 13% of lung cancers (3). SCLC is sensitive to both chemotherapy and radiotherapy, but has poor treatment performance with rapid recurrence, early metastatic dissemination, and poor prognosis (1, 4). Particularly, the overall survival (OS) rate and mortality of patients with SCLC have remained the same during the past several decades (5, 6). SCLC typically appears as a central tumor with hilar/mediastinal lymphadenopathy and distant metastases (7). The prognosis of SCLC strongly depends on the tumor stage. For example, the 5-year survival rate of limited stage is 20–25%, while only 2% for an extensive-stage (ES) disease (8).

In clinical practice, CT imaging is widely used for the clinical diagnosis of SCLC. However, CT-based SCLC diagnosis greatly depends on the experience and knowledge of radiologists. They can provide some qualitative analysis for SCLC prognosis by their experience, but the quantitative analysis of SCLC prognosis is generally lacked (9). In the clinical treatment of SCLC, etoposide–cisplatin or irinotecan–platinum remains the first choice for the first-line treatment of SCLC. Clinicians have observed that some patients made poor response to the chemotherapy (10, 11). It is clinically important to predict which patients with SCLC will have poor response to the chemotherapy, but this prognosis problem is seldom studied. In this paper, we try to construct a prediction model to quantitatively predict outcomes for patients with SCLC by their CT images scanned before treatment.

Recently, radiomics analysis is increasingly popular in tumor field, which is used for the diagnosis, stage, histology, prognosis, and treatment response assessment of various tumors (12, 13). By extracting a large number of image features from tumor regions, which depict the spatial heterogeneity in tumors, radiomics analysis is able to quantitatively characterize the tumor genomics and phenotypes, and identify tumor attributes that may be relevant to the tumor prognosis (14–16). However, due to the relatively low incidence of SCLC in all lung cancer subtypes, there were few reports focusing on the prognostic ability of radiomic features in SCLC. In regular diagnostic radiology, contrast-enhanced CT images are usually scanned for SCLC diagnosis. Radiologists observed CT images on both the lung window and mediastinal window. Here, an enhanced mediastinal window can provide more information for lung cancer diagnosis by providing clear tumor margin and enhancing pattern and intensity. It has been reported that the contrast-enhanced lung CT scan could influence the accuracy of pulmonary nodule classification (17, 18), and could independently predict the pathologic grade of lung adenocarcinoma (19). Additionally, researchers have mentioned the advantages of radiomics analysis in the prediction of response to chemotherapy in patients with SCLC (20), but radiomics analysis is performed only on the lung window of CT images. In such case, important radiomic features from the mediastinal window might be lost. This paper will study the progression-free survival (PFS) analysis in patients with SCLC by radiomics analysis on both the lung window and contrast-enhanced mediastinal window of CT images. By screening risk

factors for SCLC prognosis and constructing a high-accuracy prediction model for outcomes of patients with SCLC, we will provide a useful tool for quantitatively predicting the PFS in patients with SCLC.

## MATERIALS AND METHODS

### Patients

We retrospectively search the database from November 2012 to May 2019 under an active institutional review board, and written informed consents were acquired. The inclusion criteria were as follows: (1) patients pathologically proved SCLC by biopsy or operation. (2) Patients underwent chest CT scans. (3) Patients received standard Etoposide–Cisplatin combination chemotherapy. The exclusion criteria were as follows: (1) patients with treatment history before baseline CT scans; (2) patients lack of stage before treatment; (3) patients losing contact or die before progression; (4) patients without progression until deadline; (5) patients without contrast enhanced CT imaging; and (6) insufficient CT imaging. Finally, 186 cases were included in this study and divided into a training set (130 cases) and an independent test set (56 cases). The clinical information, such as age, gender, and stage were recorded.

### CT Imaging Parameters

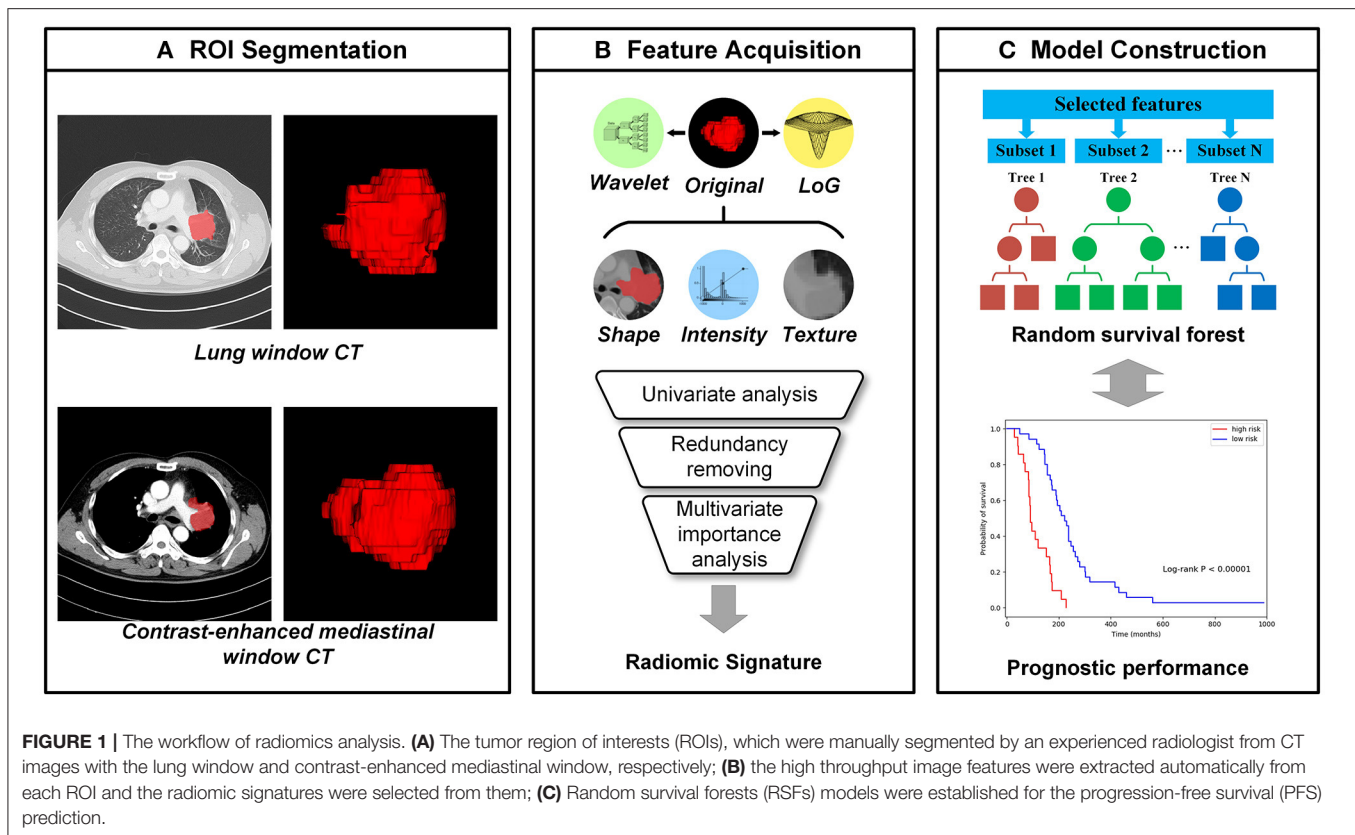
All patients underwent contrast-enhanced chest CT on 64-slice multidetector row CT scanner (LightSpeed 64; GE Medical Systems, Milwaukee, WI, USA) with the following acquisition and reconstruction parameters: tube voltage of 120 kV; tube current of intelligent mAs; section thickness of 5 mm, and reconstruction interval of 5 mm. The contrast enhanced CT was administered intravenously with an amount of 60–70 ml of iohexol (Omnipaque 300; Amersham, Shanghai, China) followed by a saline flush of 20 ml, by using a power injector (LF CT 9000; Liebel-Flarsheim, Cincinnati, OH, USA) at a flow rate of 2.5–3.0 ml/s.

### Tumor Segmentation

A radiologist with 10 years of experience segmented each tumor region manually slice-by-slice on the axial CT images using the ITK-SNAP software (<http://www.itksnap.org/pmwiki/pmwiki.php>). Each tumor was segmented two times, first on the mediastinal window for consistency and then on the lung window, respectively. The lung window level (1600 and –500 HU) and the mediastinal window level (300 and 40 HU) were used during the tumor segmentation. The vessels and air regions were carefully excluded from the segmented tumor regions.

### Radiomic Feature Extraction

As shown in **Figure 1**, high through image features were automatically calculated from each tumor region using PyRadiomics package (21), including features describing shape, intensity, texture, etc. (22) Shape features reflect geometric properties of tumor regions. Intensity features were calculated using first-order statistics to depict the distribution of voxel intensities. Texture features could quantify the tumor heterogeneity and were calculated based on different texture



matrixes. Additionally, wavelet filters and Laplacian of Gaussian (LoG) filters are applied to the original images for richer feature extraction (23). Wavelet features were extracted based on wavelet decomposition, which performed the multi-scale analysis of intensity and texture information. The LoG filters could enhance edge information, and 5 different sigma values were used to emphasize tumor textures of different coarseness. In total, 1,218 image features were extracted from each tumor region on the lung window and mediastinal window independently, and a total of 2,436 comprehensive CT image features were extracted.

## Feature Selection

Feature selection was conducted in the training set in three steps. First, the univariate prognostic ability of each feature was evaluated using a Cox proportional hazard (CPH) model (24). Features with concordance index (C-index) less than 0.5 were removed, because their prognostic abilities are worse than a random model:

$$F_{uni} = \{f | c(CPH(f, y)) > 0.5, f \in F\} \quad (1)$$

Where,  $F$  is the unselected feature set,  $c(\bullet)$  is the C-index of the CPH model which can be calculated by Equation 4. Second, to remove redundancy from image features, the variance inflation factor (VIF) was applied to quantify the collinearity between features (25). This procedure was performed iteratively until VIF values of all remaining features were less than the certain threshold, and the feature with the highest VIF value

was removed in each iteration. The iteration process can be defined as:

$$F_{uni} = F_{uni} - (f | VIF(f, F_{uni} - f) > T_{vif}) \quad (2)$$

Third, we used random survival forests (RSFs) for multivariate analysis to further simplify the features (26). An RSF model is an ensemble of tree-based learners which has powerful non-linear analysis capability. For each feature, an importance score was calculated based on the evaluation of relevance and prognosis of all features by RSF. Features with important scores above a certain threshold were finally selected and used to generate the radiomic signature:

$$F_{mul} = \{f | Score(RSF(f, F_{uni}, y)) > T_{rsf}, f \in F_{uni}\} \quad (3)$$

In above procedures, statsmodels (27), scikit-learn (28), and scikit-survival (29–31) packages were used for the implementation and grid search was performed to determine the parameters and thresholds. To compare the prognosis of different CT windows, the feature selection was performed on lung window features and mediastinal window features, respectively.

## Prognostic Model Establishment

Based on the selected radiomic features and three clinical features (gender, age, and stage), the random survival forests can be used to establish their corresponding prognostic model. For clinical features, selected lung window features and mediastinal window

features, three different prognostic models can be generated, respectively. By combining the two classes of selected radiomic features into a feature set and simplifying them according to the Pearson's correlation (32), the corresponding prognostic models can be generated. By combining selected radiomic features and clinical features, the prognostic model can be constructed. For all above models, the 3-fold cross-validation was performed to determine model parameters in the training set.

## Statistical Analysis

A univariate analysis was used to assess the statistical significance of the clinical characteristics of the patients, and the independent sample  $t$ -test and  $\chi^2$  test were performed for continuous and categorical variables, respectively. The correlation matrix of selected features was calculated based on Pearson's correlation coefficient and illustrated in a heatmap. The predictive capacity of radiomic signature was evaluated using Kaplan–Meier analysis (33). The performance of each prognostic model was evaluated on the independent test set with the C-index which is most frequently used in the survival analysis and assesses the overall prognostic ability of the model and can be defined as:

$$c = \frac{1}{num} \sum_{i: \delta_i=1} \sum_{j: y_i < y_j} I[r_i > r_j] \quad (4)$$

Where,  $i, j \in \{1, \dots, N\}$ ,  $num$  denotes the number of all comparable pairs,  $\delta_i$  is the binary event indicator,  $I[\bullet]$  is the indicator function, and  $r$  is the risk predicted by the model. In addition, the time-dependent cumulative/dynamic AUC was calculated to measure the performance in a specific time range, which is an extension of the area under the receiver operating characteristic (ROC) curve (AUC) in survival data (34, 35). The cumulative/dynamic AUC (C/D AUC) at time  $t$  can be defined as:

$$A\hat{U}C(t) = \frac{\sum_{i=1}^n \sum_{j=1}^n I[y_j > t] I[y_i < t] \omega_i I[r_i > r_j]}{(\sum_{i=1}^n I[y_i > t]) (\sum_{i=1}^n I[y_i \leq t] \omega_i)} \quad (5)$$

Where,  $\omega_i$  is inverse probability of censoring weights (IPCW). All statistical analyses were two-sided with the statistical significance level of 0.05, and performed with statsmodels, scikit-learn, and scikit-survival packages in Python 3.6.

## RESULTS

### Clinical Characteristics

A total of 186 cases were enrolled and divided into a training set and an independent test set. As shown in **Table 1**, the median (range) of age in the two sets were 62 (37–80) and 62 (43–78). In training set, 109 (83.8%) cases were men and 21 (16.2%) cases were women, with 47 (36.2%) cases were limited stage and 83 (63.8%) cases were extensive stage. In the test set, there were 50 (89.3%) male cases and 6 (10.7%) female cases, and 19 (33.9%) and 37 (66.1%) cases were limited and extensive stage, respectively. There were no significant differences for all clinical characteristics between the training and test sets ( $p = 0.334$ – $0.771$ ). In addition, through the univariate survival analysis, stage

**TABLE 1 |** Clinical characteristics.

Characteristics	Training set	Test set	<i>p</i> -value
Number	130	56	
Gender			
Male	109 (83.8%)	50 (89.3%)	0.334
Female	21 (16.2%)	6 (10.7%)	
Age	62 (37–80)	62 (43–78)	0.635
Stage			
Limited stage	47 (36.2%)	19 (33.9%)	0.771
Extensive stage	83 (63.8%)	37 (66.1%)	

( $p = 0.007$ ) and sex ( $p = 0.039$ ) were statistically relevant to the survival, and there was no significant difference regarding to age ( $p = 0.698$ ).

During the follow-up, different types of progression were observed. Among 186 cases, there were 85 (45.7%) cases of primary progression, 23 (12.4%) cases of original metastases progression, 13 (7.0%) cases of primary progression and newly metastases, 11 (5.9%) cases of both the primary and metastasis progression, 16 (8.6%) cases of newly brain metastases, 14 (7.5%) cases of newly bone metastases, 8 (4.3%) cases of newly lung metastases, 6 (3.2%) cases of newly liver metastases, 1 (0.5%) case of cardiac metastases, 1 (0.5%) cases of pancreas metastases, 3 (1.6%) cases of adrenal metastases, and 5 (2.7%) cases of newly multiple metastases.

### Important Radiomic Feature Selection

In total, 1,218 radiomic features were extracted from CT images on the lung window and mediastinal window, respectively, and the feature selection was performed independently on both windows. After the univariate analysis, 724 and 895 features were remained, respectively. Then, with an iterative collinearity elimination based on VIF, a mass of redundant features was removed, resulting in 17 and 16 remained features for the lung window and mediastinal window, respectively. Finally, an RSF-based multivariate analysis was performed on the remaining features, and 6 features for the lung window (f1, f2, f3, f4, f5, and f6) and 6 features for the mediastinal window (f7, f8, f9, f10, f11, and f12) were selected as radiomic signature, as shown in **Table 2** and **Figures 2, 3**. The correlation heatmaps in **Figures 2, 3** indicated that the selected features are relatively independent. Definitions of all selected features are in compliance with the Imaging Biomarker Standardization Initiative (IBSI) (36).

### Performance of Prognostic Model

Based on 3 clinical, 6 lung window, and 6 mediastinal window features, three basic prognostic models can be established, respectively, namely, Model\_C3, Model\_L6, and Model\_M6. For these basic models, the optimal risk cut-off was determined using log-rank test on the training set, and patients in the training set and test set were stratified into high-risk and low-risk groups using the same cut-off, respectively. In **Figure 4**, the Kaplan–Meier curves of lung window (cut-off = 112) and mediastinal window (cut-off = 117) features revealed the

**TABLE 2 |** Description of the selected radiomic features.

Index	Feature	Description
F1	Lung_log-sigma-1-0-mm-3D_glcml_Correlation	A Measure of the linear dependency of gray level values to their respective voxels in the GLCM
F2	Lung_log-sigma-3-0-mm-3D_glcml_ClusterShade	A measure of the skewness and uniformity of the GLCM
F3	Lung_original_firstorder_90 Percentile	The 90th percentile of the voxels included in the ROI
F4	Lung_wavelet-LHH_glcml_ClusterShade	A measure of the skewness and uniformity of the GLCM
F5	Lung_log-sigma-4-0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis	A measure of the proportion of the joint distribution of smaller size zones with lower gray-level values
F6	Lung_original_shape_Flatness	A measure of the relationship between the largest and smallest principal components in the ROI shape
F7	Mediastinal_log-sigma-4-0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis	A measure of the proportion of the joint distribution of smaller size zones with lower gray-level values
F8	Mediastinal_wavelet-HHL_firstorder_Skewness	A measure of the asymmetry of the distribution of values about the Mean value
F9	Mediastinal_original_shape_Flatness	A measure of the relationship between the largest and smallest principal components in the ROI shape
F10	Mediastinal_log-sigma-5-0-mm-3D_glszm_GrayLevelNonUniformityNormalized	A measure of the variability of gray-level intensity values in the image
F11	Mediastinal_wavelet-HLH_firstorder_Skewness	A measure of the asymmetry of the distribution of values about the Mean value
F12	Mediastinal_log-sigma-3-0-mm-3D_glcml_InverseVariance	A measure of inverse variance

significant difference in PFS between high-risk and low-risk groups, which is better than clinical features (cut-off = 112), demonstrating the predictive capabilities of the two different radiomic signatures.

By combining the lung window and mediastinal window radiomic features into a feature set, the combined model based on all 12 radiomic features (Model\_L6+M6) was established. Additionally, to simplify radiomic features and reduce the risk of overfitting, we removed features with the higher correlation between two feature sets (f6, which has a Pearson's correlation coefficient of 0.75 with f9 and a lower importance score). Then, 11 remained radiomic features were used to generate a simplified model (Model\_L5+M6). The overall prognostic ability of each prognostic model was evaluated with C-index and illustrated in **Table 3**. The C-index values of Model\_C3, Model\_L6, Model\_M6, Model\_L6+M6, and Model\_L5+M6 are 0.6426, 0.7455, 0.7728, 0.7927, and 0.8033 for the training set and 0.6026, 0.6951, 0.7192, 0.7362, and 0.7531 for the test set, respectively. Model\_L5+M6 attains the highest C-index among the models. In addition, by combining the selected

radiomic features and clinical features, their corresponding prognostic models have better performances or the performances comparable with the radiomic features-based models.

The C/D AUC can evaluate the model performance from time level on the test set. The C/D AUC\_90 evaluates how well these models can distinguish the patients progressing before and after 90 days, which is an important time for the prognosis of SCLC. As illustrated in **Table 3**, Model\_L5+M6 achieved the best C/D AUC\_90 of 0.8902. Furthermore, the restricted mean C/D AUC (37), which is a summary measure of discrimination ability of each model, was calculated. The mean C/D AUC of Model\_C3, Model\_L6, Model\_M6, Model\_L6+M6, and Model\_L5+M6 are 0.6312, 0.7836, 0.7964, 0.8387, and 0.8487, respectively. These results are consistent with C-index results of these models.

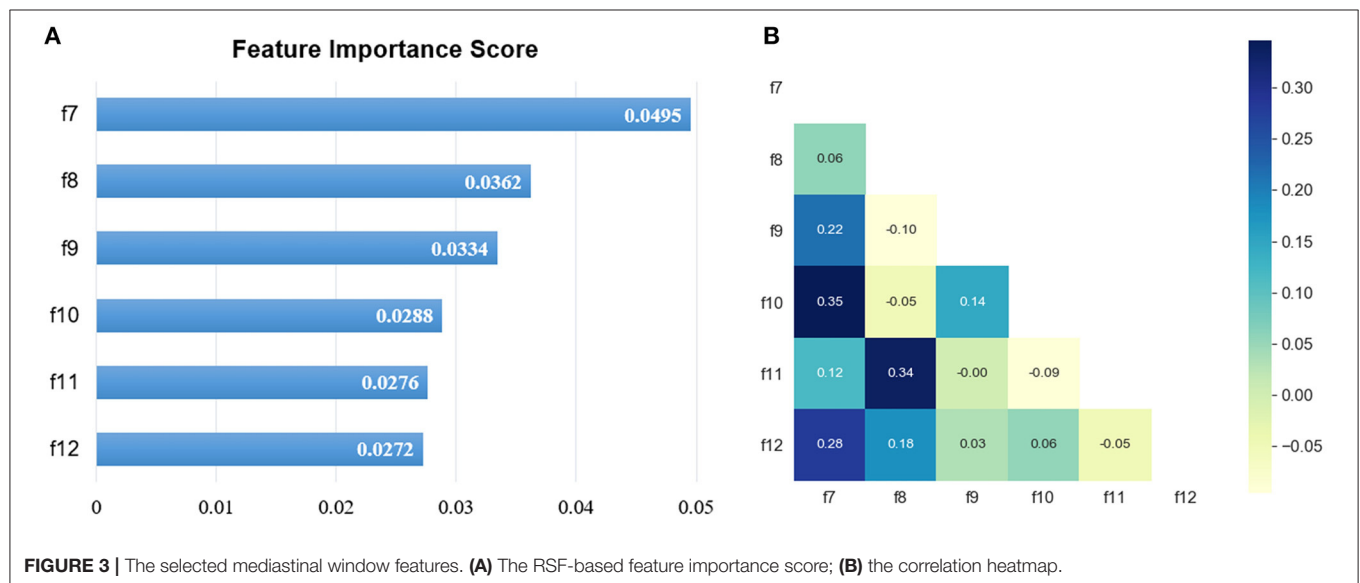
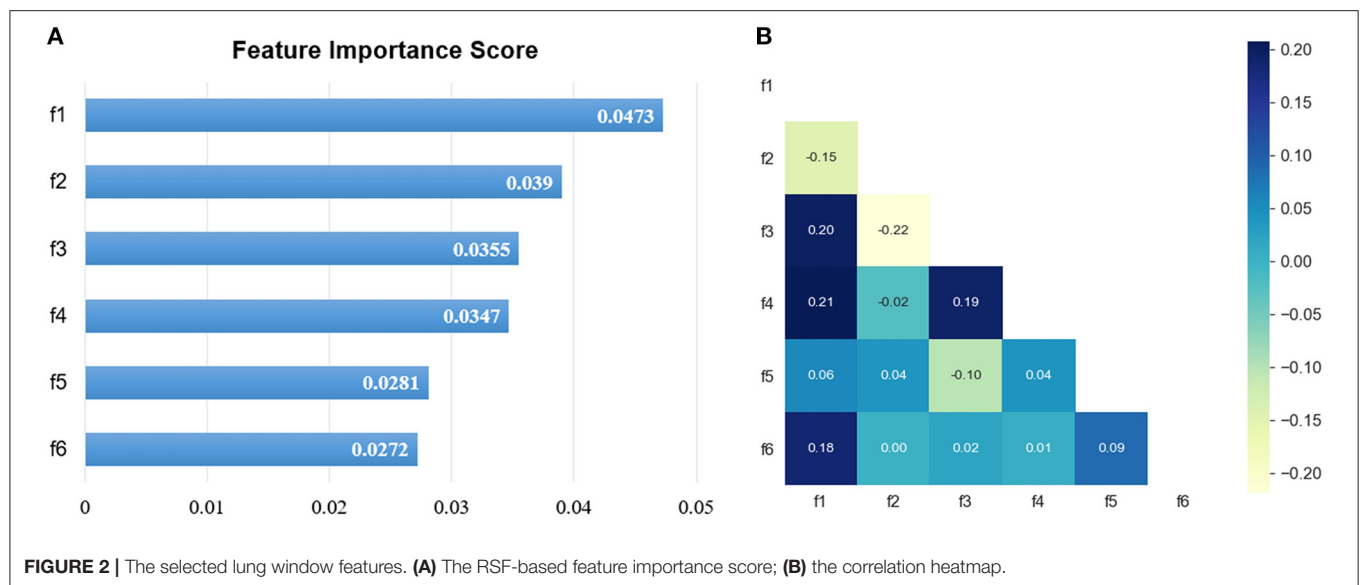
Based on our survival models, a survival function can be generated and used to analyze the prognosis for each patient. Three typical cases with different survival time were illustrated in **Figure 5**. The survival function gives the progression-free probability at different times, which can provide clinicians with a more intuitive and reliable prognostic prediction and is helpful for the personalized clinical decision-making.

Two typical cases which are challenging and hard to predict the prognosis are illustrated in **Figure 6**. For case A, a pulmonary tumor with mediastinal and lateral hilar lymph node metastases was found on CT images. No distant metastasis was detected. This 62-year-old man was diagnosed with a good prognosis based on the visual analysis and limited stage. However, the actual PFS was only 90 days, and our model correctly predicted it with a high risk of 157.0. For case B, in a 67-year-old man, bone metastases and lymph node metastases were found on the primary radiological scans. The primary pulmonary tumor of case B, with obstructive pneumonia, was larger than the case A. This patient was diagnosed as a poor prognosis with extensive stage. However, the actual PFS was over a year (431 days), and our model correctly predicted it with a low risk of 55.0.

## DISCUSSION

In this study, a PFS predictive model was proposed by integrating the radiomic features extracted from lung window and mediastinal window CT images in patients with SCLC. Our main findings indicated the contrast enhanced-mediastinal window radiomic features as an independent reliable prognostic factor. Our model effectively separated the groups of high risk and low risk, gave better predictive performance than the typical clinical visual analysis, and generate a survival function to analyze the prognosis for each patient. Our radiomics-based model offers more reliable PFS predictions which could support the personalized clinical decision-making.

Radiomics can extract more quantitative information than bare eye to guide clinical decisions, usually from non-enhanced CT images. Unenhanced CT images reflect the tumor heterogeneity and microenvironment, demonstrating the prognostication and treatment response. Fried et al. (38) and Ganeshan et al. (39) indicated that texture features from pretreatment non-contrast CT scans may provide the

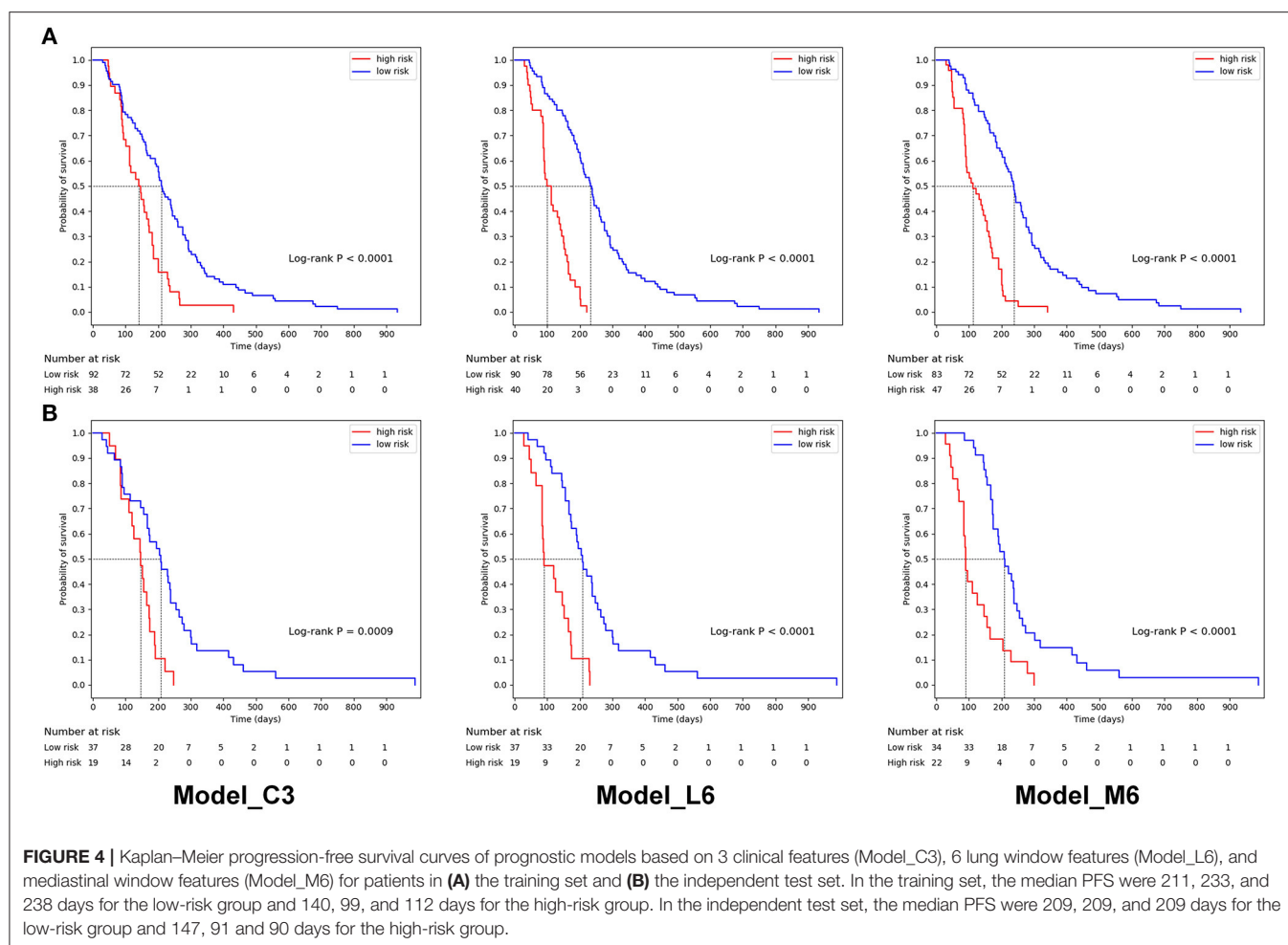


prognostic information for patients with non-small cell lung cancer (NSCLC). Mohammadhadi Khorram et al. (40) stated that radiomics was useful for predicting the early-stage NSCLC recurrence, progression, and recurrence free survival. For patients with SCLC, Haifeng Wei et al. (10) revealed radiomic texture characteristics may be an independent predictor of the efficacy of chemotherapy and help clinical guidance. Additional enhanced contrast mediastinal images may provide more image information about the grade of enhancement and the heterogeneity of the tumor, which may be due to the presence of different tumor vascularization (41). Radiomic texture analysis on the contrast-enhanced CT could be a good predictor of the survival and treatment response in patients with NSCLC (42, 43). Another study stated that the texture analysis of

CECT images provides the predicted pathologic grading of lung adenocarcinoma (19).

In this study, we investigated the value of radiomic signature in predicting the prognostication and treatment response in patients with SCLC. Two cases with different prognosis are shown in **Figure 6**. The prognosis was contrary to the expectation with clinical visual analysis. Compared with the inadequate clinical outcomes, our radiomic model with the contrast-enhanced mediastinal window predicted an accurate outcome with little difference.

Our radiomic models showed a good performance of C-index in predicting the PFS with conventional lung window features and enhanced mediastinal window features respectively, with a lowest C-index of clinical model. Furthermore, the predictive



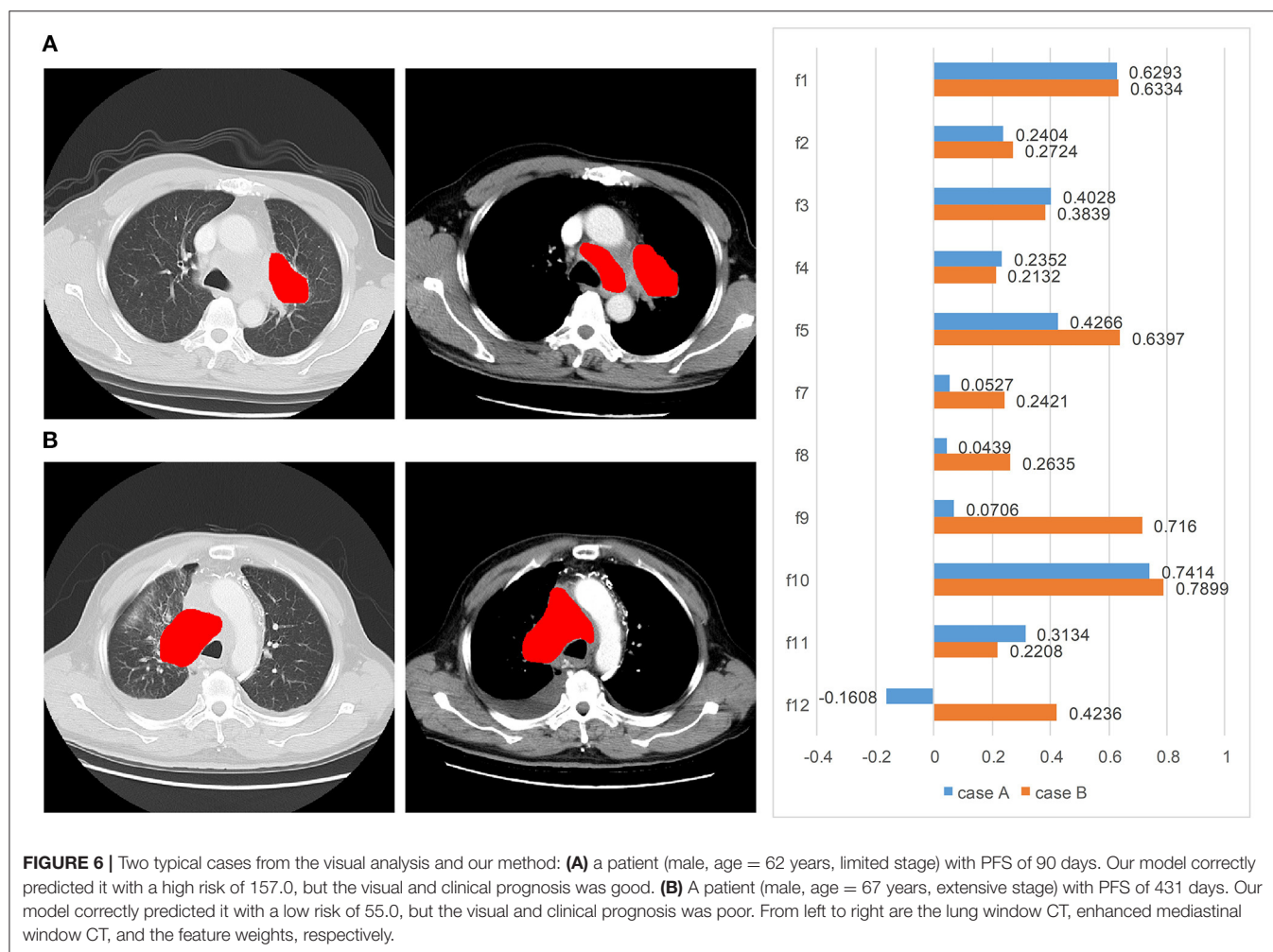
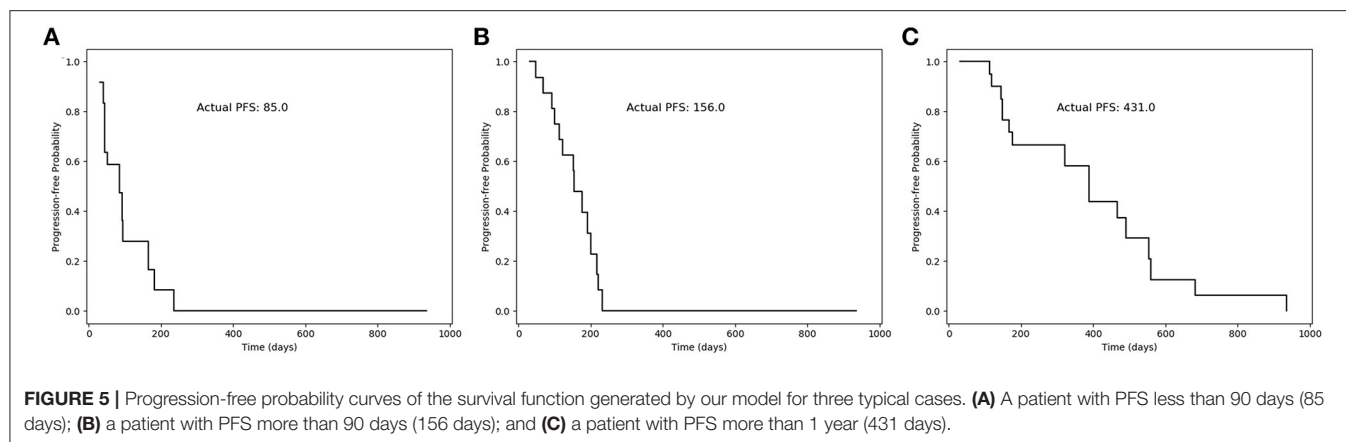
performance of radiomic features in an enhanced mediastinal window was superior to the lung window, indicating that enhanced mediastinal window features played an important role in predicting the PFS in patients with SCLC. Some previous studies showed that the predictive performance of quantitative texture features were similar between non-contrast and contrast-enhanced CT in diagnosing lung nodule (17, 44). Another study reported that the unenhanced CT was better than the contrast-enhanced CT on the predictive performance. However, those studies focused on lung adenocarcinoma and did not include SCLC cases which may cause result deviation (45). Our study results were generally in agreement with some other studies. Linning et al. (18) and Liu et al. (19) investigated that the contrast-enhanced CT were useful predictors of survival and treatment response, which may be related to the tumor heterogeneity. After contrast administration, tumoral vascularity may reflect local spatial variations in image brightness, and then result in the variability of radiomic features (18, 46).

For further verification, we analyzed the time-dependent cumulative/dynamic ROC curve and calculated the

**TABLE 3 |** Prognostic performance of different survival prediction models.

Model	C-index		C/D AUC_90	Mean C/D AUC
	Training set	Test set		
Basic models				
Model_C3	0.6426	0.6026	0.5218	0.6312
Model_L6	0.7455	0.6951	0.7727	0.7836
Model_M6	0.7728	0.7192	0.8646	0.7964
Combined models based on radiomic features				
Model_L6+M6	0.7927	0.7362	0.8769	0.8387
<b>Model_L5+M6</b>	0.8033	<b>0.7531</b>	<b>0.8902</b>	<b>0.8487</b>
Combined models based on radiomic and clinical features				
Model_L6+C3	0.7500	0.7316	0.7898	0.8206
Model_M6+C3	0.7933	0.7440	0.8485	0.8367
Model_L6+M6+C3	0.7961	0.7459	0.8523	0.8413
Model_L5+M6+C3	<b>0.8276</b>	0.7518	0.8258	0.8441

The bold values indicate the highest score in each performance metric. C3 means 3 clinical features, L6 means 6 selected lung window radiomic features, M6 means 6 selected mediastinal window radiomic features, L5 means the remained 5 lung window radiomic feature by removing the most correlated feature (f6) among the 12 radiomic features according to Pearson's correlation on the basis of L6.



time-dependent AUC. The time-dependent cumulative/dynamic ROC curve analysis defined a marker value updated at each time point during the disease status individually, allowed to compare the marker's predictive ability and may give guidance for medical decisions (47). In our study, we achieved similar

results with the C-index, suggesting the importance of enhanced-mediastinal window in predicting the PFS. Our study illustrated an important time of 90 days for the prognosis of SCLC and the median survival time was around 200 days in high-risk groups from Kaplan–Meier curves.

There were numerous texture features in CT images, which may provide different anatomical and biological information in tumor. Thus, the selection of texture features was meaningful and time-saving for model building. During feature extraction, original images were transformed to derived images with the wavelet filter and LoG filter to extract more radiomic features. After feature selection, the most important features for lung window and mediastinal window were selected respectively. For both lung window and mediastinal window, 50% (3/6) selected features were LoG-based, including the top-ranked features. These results illustrated that LoG-based features play a more important role in the survival analysis of SCLC. Actually, a LoG filter could enhance the edge information by emphasizing the areas of gray level change. The fine textures and coarse textures of the tumor were both taken into consideration by high and low sigma parameter. During the clinical diagnosis of SCLC based on CT images, the texture feature of the tumor edge is always an important indicator, which also proves the rationality of extracted radiomic features.

There were several limitations in our study. First of all, due to the relatively low incidence of SCLC in Asian race, the population of this study is relatively small. Second, previous studies indicated that the suitable section thickness may be as thin as 1.25 or 2.5 mm. But the section thickness was 5 mm in this retrospective study, which may reduce the predictive accuracy of lung window. Third, smoking is an important risk factor in patients with SCLC especially in women, which did not calculate into the clinical features in our study.

## CONCLUSION

In summary, our study revealed that the textual features extracting from the contrast-enhanced mediastinal window were useful for predicting the PFS. The integration of textual features from the lung window and contrast-enhanced

mediastinal window provided the more valuable information in survival prediction in comparison with the conventional visual assessment, which could be applied to support personalized clinical decision for the diagnosis and patient management in patients with SCLC.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Zhongshan Hospital of Fudan University. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

XY and LW designed and conducted the study. NC, RL, XY, and LW contributed to the manuscript writing process. NC and RL prepared the first draft of the entire manuscript. NC performed tumor segmentation. RL, DS, and LW contributed to the radiomic feature extraction, feature selection, and data analysis. NC, MJ, YG, and JC contributed to data interpretation. All authors provided the final approval of the version submitted for publication.

## FUNDING

NC reports funding from the Medical and Health Foundation for Young Scientists of Fujian Province (Grant No. 2020QNA059).

## REFERENCES

- Rudin CM, Poirier JT. Small-cell lung cancer in 2016: shining light on novel targets and therapies. *Nat Rev Clin Oncol.* (2017) 14:75–6. doi: 10.1038/nrclinonc.2016.203
- Gazdar AF, Bunn PA, Minna JD. Small-cell lung cancer: what we know, what we need to know and the path forward. *Nat Rev Cancer.* (2017) 17:725–37. doi: 10.1038/nrc.2017.87
- Noone AM, HN, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, et al. SEER Cancer Statistics Review (CSR) 1975–2015, National Cancer Institute. In: Bethesda, MD. *SEER Data Submission.* Available online at: [https://seer.cancer.gov/csr/1975\\_2015/](https://seer.cancer.gov/csr/1975_2015/) (accessed April 2018)
- Bunn PA, Jr., Minna JD, Augustyn A, Gazdar AF, Ouadah Y, Krasnow MA, et al. Small cell lung cancer: can recent advances in biology and molecular biology be translated into improved outcomes? *J Thorac Oncol.* (2016) 11:453–74. doi: 10.1016/j.jtho.2016.01.012
- Fruh M, De Ruyscher D, Popat S, Crino L, Peters S, Felip E, et al. Small-cell lung cancer (SCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* (2013) 24:vi99–105. doi: 10.1093/annonc/mdt178
- Zhao H, Ren D, Liu H, Chen J. Comparison and discussion of the treatment guidelines for small cell lung cancer. *Thorac Cancer.* (2018) 9:769–74. doi: 10.1111/1759-7714.12765
- Carter BW, Glisson BS, Truong MT, Erasmus JJ. Small cell lung carcinoma: staging, imaging, and treatment considerations. *Radiographics.* (2014) 34:1707–21. doi: 10.1148/rg.346140178
- Zimmerman S, Das A, Wang S, Julian R, Gandhi L, Wolf J. 2017–2018 scientific advances in thoracic oncology: small cell lung cancer. *J Thorac Oncol.* (2019) 14:768–83. doi: 10.1016/j.jtho.2019.01.022
- Byers LA, Rudin CM. Small cell lung cancer: where do we go from here? *Cancer.* (2015) 121:664–72. doi: 10.1002/cncr.29098
- Wei H, Yang F, Liu Z, Sun S, Xu F, Liu P, et al. Application of computed tomography-based radiomics signature analysis in the prediction of the response of small cell lung cancer patients to first-line chemotherapy. *Exp Ther Med.* (2019) 17:3621–9. doi: 10.3892/etm.2019.7357
- Pietanza MC, Byers LA, Minna JD, Rudin CM. Small cell lung cancer: will recent progress lead to improved outcomes? *Clin Cancer Res.* (2015) 21:2244–55. doi: 10.1158/1078-0432.CCR-14-2958
- Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology.* (2016) 281:947–57. doi: 10.1148/radiol.2016152234
- Coroller TP, Agrawal V, Huynh E, Narayan V, Lee SW, Mak RH, et al. Radiomic-based pathological response prediction from primary tumors and lymph nodes in NSCLC. *J Thorac Oncol.* (2017) 12:467–76. doi: 10.1016/j.jtho.2016.11.2226

14. Chen BT, Chen Z, Ye N, Mambetsariev I, Fricke J, Daniel E, et al. Differentiating peripherally-located small cell lung cancer from non-small cell lung cancer using a CT radiomic approach. *Front Oncol.* (2020) 10:593. doi: 10.3389/fonc.2020.00593
15. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
16. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
17. Wu W, Pierce LA, Zhang Y, Pipavath SNJ, Randolph TW, Lastwika KJ, et al. Comparison of prediction models with radiological semantic features and radiomics in lung cancer diagnosis of the pulmonary nodules: a case-control study. *Eur Radiol.* (2019) 29:6100–8. doi: 10.1007/s00330-019-06213-9
18. Linning E, Lu L, Li L, Yang H, Schwartz LH, Zhao B. Radiomics for classifying histological subtypes of lung cancer based on multiphase contrast-enhanced computed tomography. *J Comput Assist Tomogr.* (2019) 43:300–6. doi: 10.1097/RCT.0000000000000836
19. Liu Y, Liu S, Qu F, Li Q, Cheng R, Ye Z. Tumor heterogeneity assessed by texture analysis on contrast-enhanced CT in lung adenocarcinoma: association with pathologic grade. *Oncotarget.* (2017) 8:53664–74. doi: 10.18632/oncotarget.15399
20. Jain P, Khorrami M, Gupta A, Rajiah P, Bera K, Viswanathan VS, et al. Novel non-invasive radiomic signature on CT scans predicts response to platinum-based chemotherapy and is prognostic of overall survival in small cell lung cancer. *Front Oncol.* (2021) 11:744724. doi: 10.3389/fonc.2021.744724
21. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* (2017) 77:e104–e7. doi: 10.1158/0008-5472.CAN-17-0339
22. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* (2014) 5:4006. doi: 10.1038/ncomms5006
23. Wang L, Dong T, Xin B, Xu C, Guo M, Zhang H, et al. Integrative nomogram of CT imaging, clinical, and hematological features for survival prediction of patients with locally advanced non-small cell lung cancer. *Eur Radiol.* (2019) 29:2958–67. doi: 10.1007/s00330-018-5949-2
24. Cox DR. Regression Models and Life-Tables. *J Royal Statist Soc Series B Methodol.* (1972) 34:187–220. doi: 10.1111/j.2517-6161.1972.tb00899.x
25. O'brien RM. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity.* (2007) 41:673–90. doi: 10.1007/s11135-006-9018-6
26. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat.* (2008) 2:841–60. doi: 10.1214/08-AOAS169
27. Seabold S, PJS. Econometric and statistical modeling with python. In: *Proceedings of the 9th Python in Science Conference.* (2010). vol. 57. p. 92–6. doi: 10.25080/Majora-92b1922-011
28. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* (2011) 12:2825–30.
29. Pölsterl S, Navab N, Katouzian A. Fast training of support vector machines for survival analysis. In: *European Conference on Machine Learning.* (2015). p. 243–59. doi: 10.1007/978-3-319-23525-7\_15
30. Pölsterl S, Navab N, Katouzian A. An efficient training algorithm for kernel survival support vector machines. arXiv preprint arXiv:1611.07054. (2016).
31. Pölsterl S, Gupta P, Wang L, Conjeti S, Katouzian A, Navab N. Heterogeneous ensembles for predicting survival of metastatic, castrate-resistant prostate cancer patients. *F1000Research.* (2016) 5. doi: 10.12688/f1000research.8231.1
32. Benesty J, Chen J, Huang Y, Cohen I. *Pearson Correlation Coefficient: Noise Reduction in Speech Processing.* Springer Topics in Signal Processing. (2009). p. 1–4. doi: 10.1007/978-3-642-00296-0\_5
33. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc.* (1958) 53:457–81. doi: 10.1080/01621459.1958.10501452
34. Uno H, Cai T, Tian L, Wei LJ. Evaluating prediction rules for year survivors with censored regression models. *J Am Stat Assoc.* (2007) 102:527–37. doi: 10.1198/01621450700000149
35. Hung H, Chiang C. Estimation methods for time-dependent AUC models with survival data. *Canad J Statist.* (2009) 38:8–26. doi: 10.1002/cjs.10046
36. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative - feature definitions. arXiv preprint arXiv:1612.07003. (2016).
37. Lambert J, Chevet S. Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent ROC curves. *Stat Methods Med Res.* (2016) 25:2088–102. doi: 10.1177/0962280213515571
38. Fried DV, Tucker SL, Zhou S, Liao Z, Mawlawi O, Ibbott G, et al. Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer. *Int J Radiat Oncol Biol Phys.* (2014) 90:834–42. doi: 10.1016/j.ijrobp.2014.07.020
39. Ganesan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. *Eur Radiol.* (2012) 22:796–802. doi: 10.1007/s00330-011-2319-8
40. Khorrami M, Bera K, Leo P, Vaidya P, Patil P, Thawani R, et al. Stable and discriminating radiomic predictor of recurrence in early stage non-small cell lung cancer: multi-site study. *Lung Cancer.* (2020) 142:90–7. doi: 10.1016/j.lungcan.2020.02.018
41. Chen L, Wang H, Zeng H, Zhang Y, Ma X. Evaluation of CT-based radiomics signature and nomogram as prognostic markers in patients with laryngeal squamous cell carcinoma. *Cancer Imaging.* (2020) 20:28. doi: 10.1186/s40644-020-00310-5
42. Ravanelli M, Farina D, Morassi M, Roca E, Cavalleri G, Tassi G, et al. Texture analysis of advanced non-small cell lung cancer (NSCLC) on contrast-enhanced computed tomography: prediction of the response to the first-line chemotherapy. *Eur Radiol.* (2013) 23:3450–5. doi: 10.1007/s00330-013-2965-0
43. Ahn SY, Park CM, Park SJ, Kim HJ, Song C, Lee SM, et al. Prognostic value of computed tomography texture features in non-small cell lung cancers treated with definitive concomitant chemoradiotherapy. *Invest Radiol.* (2015) 50:719–25. doi: 10.1097/RLI.0000000000000174
44. Fan L, Fang M, Li Z, Tu W, Wang S, Chen W, et al. Radiomics signature: a biomarker for the preoperative discrimination of lung invasive adenocarcinoma manifesting as a ground-glass nodule. *Eur Radiol.* (2019) 29:889–97. doi: 10.1007/s00330-018-5530-z
45. He L, Huang Y, Ma Z, Liang C, Liang C, Liu Z. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci Rep.* (2016) 6:34921. doi: 10.1038/srep34921
46. Bezy-Wendling J, Kretowski M, Rolland Y, Le Bidon W. Toward a better understanding of texture in vascular CT scan simulated images. *IEEE Trans Biomed Eng.* (2001) 48:120–4. doi: 10.1109/10.900272
47. Bansal A, Heagerty PJ, A. comparison of landmark methods and time-dependent ROC methods to evaluate the time-varying performance of prognostic markers for survival outcomes. *Diagn Progn Res.* (2019) 3:14. doi: 10.1186/s41512-019-0057-6

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chen, Li, Jiang, Guo, Chen, Sun, Wang and Yao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Feasibility Study of Intelligent Three-Dimensional Accurate Liver Reconstruction Technology Based on MRI Data

Shaodong Cao<sup>1\*†</sup>, Huan Li<sup>1†</sup>, Suyu Dong<sup>2</sup> and Zhenxuan Gao<sup>3</sup>

<sup>1</sup> Medical Imaging Department of the Fourth Affiliated Hospital of Harbin Medical University, Harbin, China, <sup>2</sup> The School of Information and Computer Engineering, Northeast Forestry University, Harbin, China, <sup>3</sup> Neurosurgery Department of the Fourth Affiliated Hospital of Harbin Medical University, Harbin, China

## OPEN ACCESS

### Edited by:

Yong Xu,  
Harbin Institute of Technology,  
Shenzhen, China

### Reviewed by:

Xiaowei Han,  
Nanjing Drum Tower Hospital, China  
Dan Wang,  
Shanghai Jiao Tong University, China

### \*Correspondence:

Shaodong Cao  
shaodong\_cao@163.com

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

**Received:** 13 December 2021

**Accepted:** 07 February 2022

**Published:** 17 March 2022

### Citation:

Cao S, Li H, Dong S and Gao Z (2022)  
Feasibility Study of Intelligent  
Three-Dimensional Accurate Liver  
Reconstruction Technology Based on  
MRI Data. *Front. Med.* 9:834555.  
doi: 10.3389/fmed.2022.834555

Intelligent three-dimensional (3D) reconstruction technology plays an important role in the diagnosis and treatment of diseases. It has been widely used in assisted liver surgery. At present, the 3D reconstruction information of liver is mainly obtained based on CT enhancement data. It has also been commercialized. However, there are few reports on the display of 3D reconstruction information of the liver based on MRI. The purpose of this study is to propose a new idea of intelligent 3D liver reconstruction based on MRI technology and verify its feasibility. Two different liver scanning data (CT and MRI) were selected from the same batch of patients at the same time (patients with a time interval of no more than two weeks and without surgery). The results of liver volume, segmentation, tumor, and simulated surgery based on MRI volume data were compared with those based on CT data. The results show that the results of 3D reconstruction based on MRI data are highly consistent with those based on CT 3D reconstruction. At the same time, in addition to providing the information provided by CT 3D reconstruction, it also has its irreplaceable advantages. For example, multi-phase (early, middle and late arterial, hepatobiliary, etc.) scanning of MRI technology can provide more disease information and display of biliary diseases. In a word, MRI technology can be used for 3D reconstruction of the liver. Hence, a new feasible and effective method to show the liver itself and its disease characteristics is proposed.

**Keywords:** 3D precise reconstruction technology, liver, cancer, segmentation, simulated liver surgery

## INTRODUCTION

With the deepening of medical development and the proposal of personalized and accurate treatment schemes, clinicians are no longer satisfied with the information provided by traditional imaging examination. The promotion of three-dimensional (3D) reconstruction technology greatly meets the needs of clinicians to a certain extent. This technology has been widely used in the human body, especially in hepatobiliary surgery, breast surgery, maxillofacial surgery, and bone surgery (1–3). In 1990, the earliest application in the liver was proposed by Hashimoto et al. (4). Using CT scanning data and computer software, 3D renderings of liver vessels and tumors were reconstructed to guide the surgical treatment of liver cancer. The 3D reconstruction can observe the characteristics of lesions from multiple angles, determine reasonable and effective surgical methods,

ensure perioperative safety, and greatly meet the requirements of surgery for accuracy and safety (5). At present, the widely used 3D liver reconstruction technology is based on three-phase CT enhancement to show the anatomical structure of liver parenchyma, important liver vessels and biliary system, and clarify the relationship between liver and tumor and the range of tumor involvement when tumor occurs (6–8). However, some patients choose magnetic resonance liver enhancement examination because of iodine allergy and selective use of magnetic resonance liver-specific contrast agents. Whether these patients can carry out 3D liver reconstruction based on MRI data is a clinical concern. The purpose of this study is to determine the feasibility of accurately obtaining 3D liver reconstruction information based on MRI technology and to explore the clinical significance of 3D reconstruction based on MRI liver enhancement data in guiding clinical selection of accurate and personalized treatment, which plays a positive guiding role in both operation and interventional therapy, so as to reduce complications. Finally, our study provides a new idea for 3D liver reconstruction.

## DATA AND METHODS

### Case Data

Fifty-two patients aged from 31 to 75 years (average age 56.06 years), suspected of liver space occupying, and came to the Fourth Affiliated Hospital of Harbin Medical University from October 2020 to May 2021 were selected as the research subjects. All patients underwent plain CT scan combined with enhancement and plain MRI scan combined with enhancement within one week.

### Inclusion and Exclusion Criteria

Inclusion criteria were as follows: ① All patients received CT and MRI enhancement scanning at the same time within two weeks; ② No corresponding liver treatment was performed between the two examinations; and ③ The image data are relatively complete. Exclusion criteria were as follows: the patient's breath holding was poor, and the image quality affected the diagnosis. All patients signed an informed consent and obtained the approval of the hospital Ethics Committee.

### Scanning Method and Sequence Information

Before examination, the patient was fasted of water for more than 4 h and asked to practice breathing and breath holding to reduced the artifacts caused by respiratory movement as much as possible before being asked to lie on his/her back. The scanning range was from the top of diaphragm to the lower edge of liver. A Canon medical Aquilion ONE 320 slice CT Scanner was used for plain scanning of the liver, gallbladder, spleen, and pancreas. After the plain scanning, a dynamic three-phase enhanced scanning was performed. The contrast agent was administered through the injection of iopanol through elbow vein. The injection rate was 3.5 ml/s, the injection amount was 0.1 mol/kg, and the reconstruction layer thickness was 1 mm. MRI examination was performed with a Philips ingenia 3.0T MRI scanner (Philips Medical Technology) in the Netherlands and

an abdominal phased array coil. The scanning sequence was a plain MRI scan, diffusion weighted imaging (DWI) scan, and then a dynamic enhancement scan. The enhancement sequence was performed by an mdixon 3D volume scanning (multi-phase arterial phase, portal phase, delayed phase, and hepatobiliary phase). The contrast agent, an injection amount of 0.1 mol/kg of multihance or promethicin, was injected through an elbow vein at the injection rate of 2 ml/s. Sequence information were as follows: 1) Axial T2WI is TSESPAIR sequence (TR 1000 ms, TE 70 ms), imaging field of vision is  $300 \times 350 \times 191$  mm, voxel is  $1.6 \times 1.8$  mm, Coronal T2WI is TSE sequence (TR 626 ms, TE 80 ms), imaging field of view is  $300 \times 402 \times 167$  mm, voxel is  $1.6 \times 1.83$  mm; 2) Axial DWI ( $b = 800$  s/mm<sup>2</sup>) is sequence with TR 3000 ms and TE 71 ms, imaging field of view is  $380 \times 297 \times 209$  mm, voxel is  $3 \times 3$  mm; 3) Axial dynamic enhanced scanning is mdixon-w sequence (TR 3.7 ms, TE 1.31 ms), imaging field of view is  $320 \times 390 \times 225$  mm, voxel is  $1.75 \times 1.76$  mm; 4) Coronal enhanced scanning is mdixon-w sequence (TR 5.5 ms, TE 1.34 ms), imaging field of view is  $350 \times 398 \times 150$  mm, voxel is  $1.7 \times 1.7$  mm. The layer thickness was 5 mm and the layer spacing was  $-2.5$  mm. The contrast agent was injected with MODIS or promethicin through elbow vein. The injection rate was 2 ml/s and the injection volume was 0.1 mol/kg.

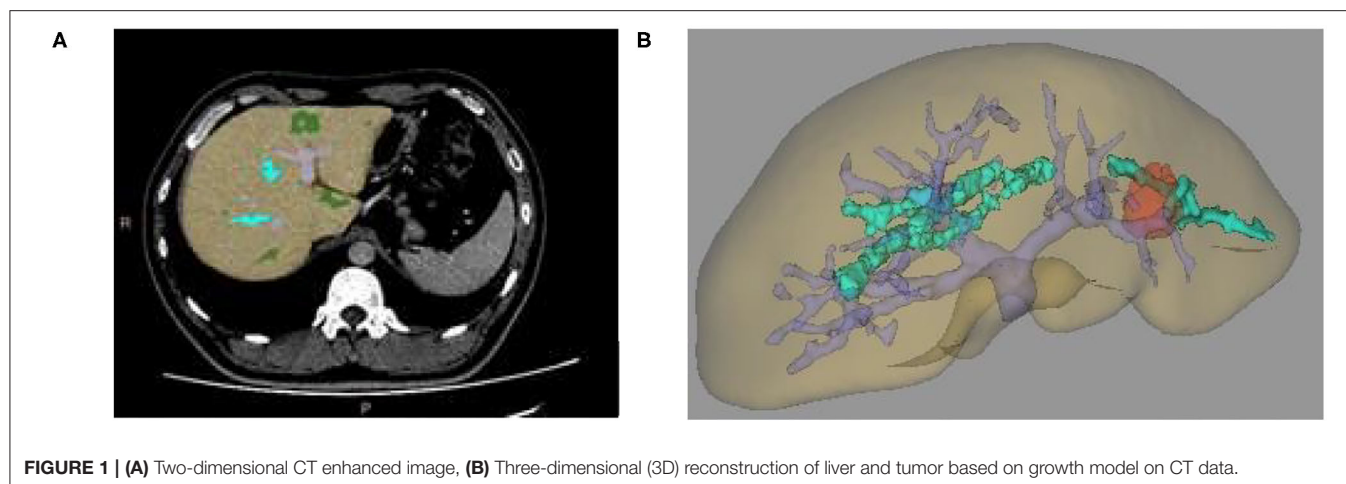
### Image Analysis and Post-Processing

Computed tomography (CT) and MRI scanning images were evaluated by two experienced senior doctors. When the evaluation of the two doctors were inconsistent, the image rating was carried out through negotiation. Since liver segmentation and simulated surgery are based on the portal and hepatic vein, the development of portal vein is basically good. This is because the development quality of hepatic vein directly determines the accuracy of 3D liver reconstruction. Therefore, one radiologist (observer 1) is familiar with abdominal imaging diagnosis and one hepatobiliary surgeon (observer 2) are responsible for Hepatic veins (HVS) in each group. The roughest Hepatic vein (HV) in the image was evaluated independently. The scoring criteria were as follows: no display of HVS was recorded as 0 points; blurred, clear, and sharp pipe wall were recorded as 1, 2, and 3 points, respectively; and slightly higher, higher, and significantly higher lumen density were recorded as 1 point, 2 points, and 3 points, respectively. The scores were divided into three levels: excellent (5–6 points), average (3–4 points), and poor (0–2 points).

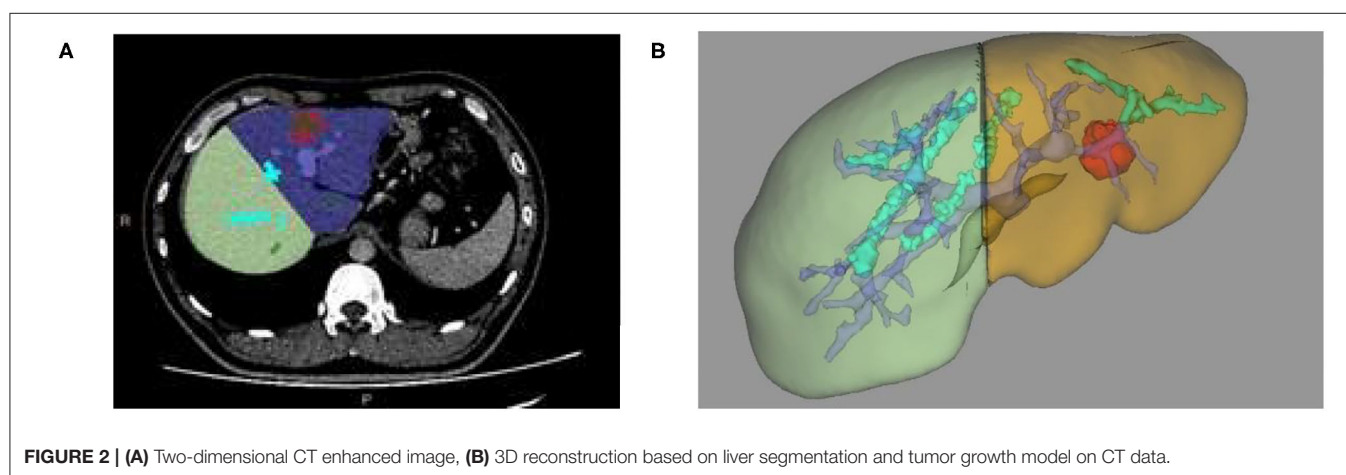
The qualified image is transferred to the Philips Nebula workstation and the surface rendering algorithm was automatically used for 3D reconstruction. The liver volume and the left and right lobes of the liver are represented based on 3D reconstruction. The measurement of volume for each liver was repeated by three times, in which various values are recorded, and the average value is taken.

## 3D LIVER MODEL RECONSTRUCTION AND MEASUREMENT

The digitized liver model obtained by CT 3D reconstruction and the fourth level portal vein vessels were clearly displayed. The original data of plain CT scan phase, arterial phase,



**FIGURE 1 | (A)** Two-dimensional CT enhanced image, **(B)** Three-dimensional (3D) reconstruction of liver and tumor based on growth model on CT data.



**FIGURE 2 | (A)** Two-dimensional CT enhanced image, **(B)** 3D reconstruction based on liver segmentation and tumor growth model on CT data.

portal phase, and delayed phase are imported into the Philips Nebula workstation (Philips Medical Technology) to display the information of transverse axis, sagittal position, and coronal position of the original data.

## The Liver Reconstruction and Measurement Based on CT Image

The liver model is established by using the 3D reconstruction organ cutting method based on portal phase volume data. Then, the volume data on liver CT-enhanced portal phase was reconstructed based on Marching Cube surface rendering algorithm:

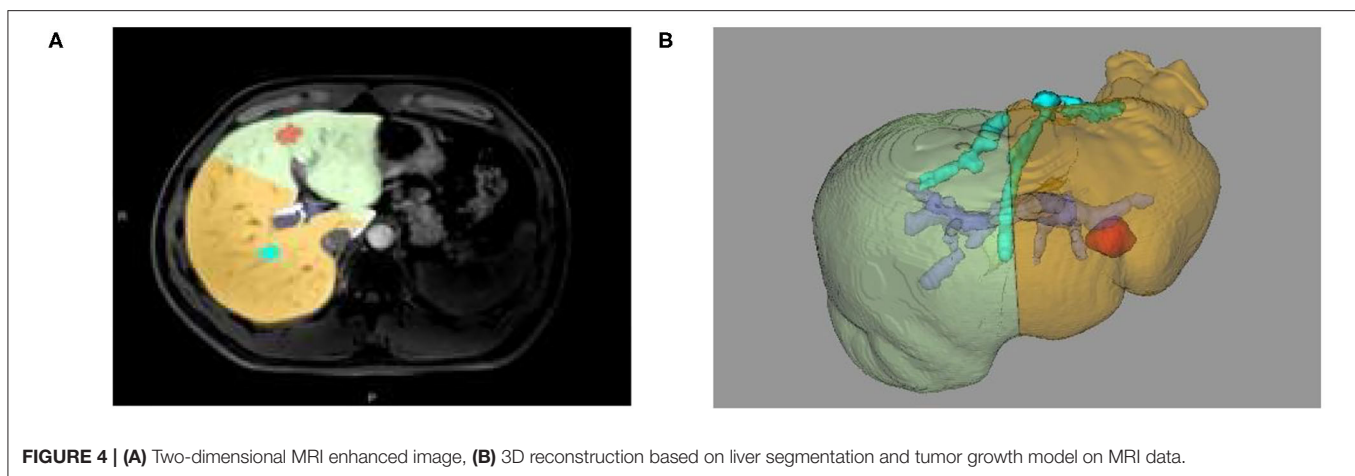
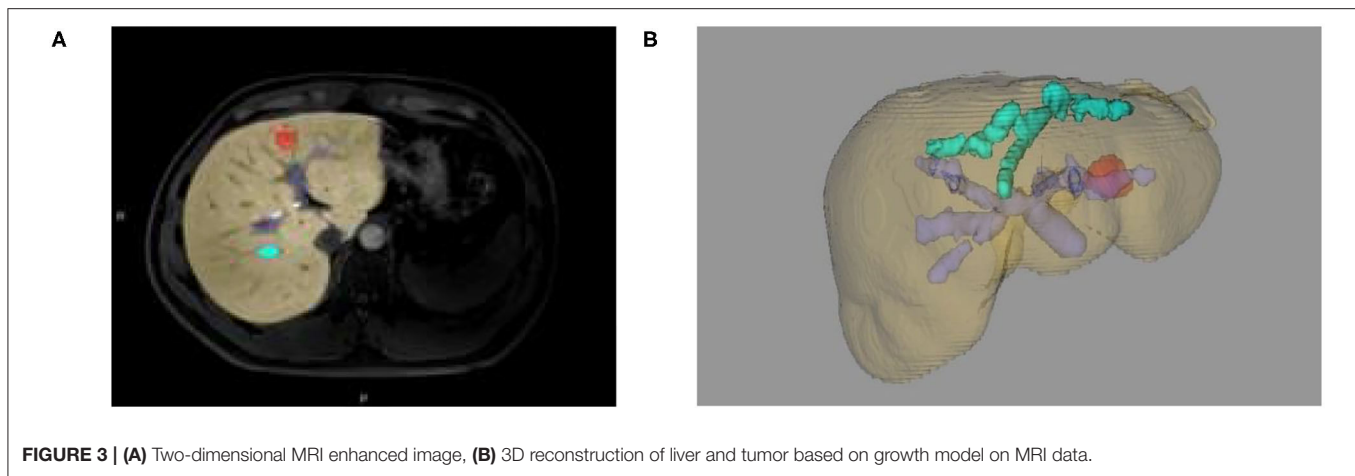
- (1) We automatically reconstructed the liver model according to the liver position and CT value based on the surface rendering algorithm. At the same time, according to the difference between vascular CT value and liver CT value, the portal vein and hepatic vein were automatically segmented and different sequences (coronal, sagittal, and transverse axial) are compared. Whether the automatic selection contained the liver model was checked. If the automatic

segmentation was wrong, several seed points were manually constructed and the threshold range was appropriately adjusted to induce all seed points to fill and grow to contain liver tissue and eliminate non-liver tissue. Finally, a standard liver model was made (**Figure 1A**).

- (2) At the same time, according to the difference of CT values, seed points were manually constructed to fill the growing tumor model (**Figure 1B**).
- (3) According to the position of blood vessels such as portal vein and hepatic vein, the liver is automatically segmented to obtain the liver volume of each segment of the left and right lobes. At the same time, simulated hepatectomy is carried out according to clinical needs to calculate the residual liver volume to determine whether surgical treatment can be carried out directly in a clinic (as shown in **Figures 2A,B**).

## Liver Reconstruction and Measurement Based on MRI Image

The establishment of the liver model was based on the 3D reconstruction organ cutting method based on the volume data



of the portal vein phase (the signal of portal vein and hepatic vein is higher than the signal of the liver parenchyma) or hepatobiliary phase (the signal of portal vein and hepatic vein is lower than the signal of the liver parenchyma). The liver MRI volume data was reconstructed based on the Marching Cube surface-rendering algorithm from Philips Nebula workstation. Since the MRI data cannot be directly transferred into the liver post-processing software, the volume data was first transferred into the AVA vascular processing software. Only then was the secondary post-processing software selected to enter the liver software. Because the liver in the MRI data cannot be automatically identified, it is manually identified as follows:

- (1) Several seed points were manually constructed on the liver. After the threshold range was properly adjusted, all seed points were induced to fill the growing liver tissue. After comparing the data from different sequences (coronal position, sagittal position, and transverse axis position), we analyzed whether the seed points contained the liver model. Then, we reconstructed the 3D liver model based on Marching Cube algorithm to make the liver model standard and complete (as shown in **Figure 3A**);
- (2) The reconstruction of the 3D models of liver space of which the portal vein, hepatic vein, and liver segmentation occupy

is realized. Likewise, simulated liver surgery, via the same method in liver reconstruction and measurement based on CT image, is made possible (as shown in **Figures 3B, 4A,B**).

## Statistical Analysis

The total liver volume, right liver volume, and the clarity of hepatic vein from the two groups of patients were measured at the same time and compared between the two groups. The SPSS24.0 software was used for data analysis. The counting data is expressed as *n*. The data measurement was based on the normal distribution with parameters  $X \pm S$  index. Two independent samples were analyzed by univariate analysis using *t*-test,  $p < 0.05$  indicated that there is a statistical significance.

## RESULTS

All 52 patients were divided into a CT group and an MRI group according to the scanning method. We compared and analyzed the data from the two groups on the aspect of functional volume, total volume, right liver volume, and hepatic vein score. It can clearly be seen that the 3D reconstruction based on MRI volume data is consistent with the 3D reconstruction results based on CT volume data ( $p > 0.05$ ) (**Tables 1, 2**).

**TABLE 1** | Liver volume measured by CT and MRI reconstruction.

Project	CT group	MRI group	Inspection value	P value
Functional volume (cm <sup>3</sup> , $\bar{x} \pm s$ )	1270.5 $\pm$ 308.6	1257.0 $\pm$ 314.0	−0.739	0.464
Total volume (cm <sup>3</sup> , $\bar{x} \pm s$ )	1305.8 $\pm$ 315.0	1284.4 $\pm$ 323.4	−1.145	0.257

**TABLE 2** | Right lobe volume measured by CT and MRI reconstruction.

Project	CT group	MRI group	Inspection value	P value
Right liver volume (cm <sup>3</sup> , $\bar{x} \pm s$ )	824.6 $\pm$ 246.9	826.2 $\pm$ 248.3	0.110	0.913

**TABLE 3** | Clarity analysis of hepatic vein reconstructed by CT and MRI.

Project	CT group	MRI group	T-test	P value
Hepatic vein score [score, M (P <sub>25</sub> , P <sub>75</sub> )]	4 (2~6)	5 (5~6)	−3.963	<0.01

Moreover, the clarity of hepatic vein display in MRI group was significantly better than that in CT group ( $p < 0.05$ ) (as shown in Table 3).

## DISCUSSION

### The Necessity of Selecting 3D Reconstruction Technology Based on MRI Data

It is well-known that MRI has certain advantages over CT. Firstly, MRI is non-invasive and non-radiographic and hence will not bring biological damage. Moreover, MRI has advantages in sensitivity and spatial resolution, which is conducive to the comparative observation of soft tissue (9). On the other hand, some people are allergic to iodine contrast agent in CT examination and passively give up CT examination. Also, some patients directly choose a liver-specific MRI contrast agent for enhancement in physical examination ultrasound, while other patients only choose MRI examination due to various concerns and refuse CT examination. Gadolinium-DTPA (GD-DTPA) is the most commonly used contrast agent for liver MRI enhancement, while hepatobiliary MRI can provide more lesion details and better qualitative analysis of disease characteristics. The most widely used clinical contrast agents are gadolinium ethoxybenzyl diethylenetriamine pentaacetic acid (GD-EOB-DTPA) whose trade names are prometaxel and gadobenate dimeglumine (GD-BOPTA). It is found (10, 11) that both MODIS and promethicin

can be ingested by hepatocytes and excreted into the biliary system, which can significantly improve the sensitivity and specificity of related liver cancer. They are mainly used for the diagnosis of focal lesions in the liver, especially for the detection of small lesions and the identification of benign and malignant nodules (12). In addition, some literatures suggest that the uptake of GD-EOB-DTPA by hepatocytes in patients with liver cirrhosis is delayed. Therefore, it can also be used for the staging of liver fibrosis and the non-invasive evaluation of liver function. With the clinical popularization of hepatobiliary specific contrast agents, enhanced MRI has more advantages in the detection of early liver cancer. Without the assistance of MRI 3D reconstruction technology, it is difficult to the accurate preoperative evaluation and simulated operation to a certain extent. Does it mean that the assistance of 3D reconstruction is lost without the addition of CT examination? The results of this study prove that 3D reconstruction technology based on MRI enhancement data is feasible (13).

### Advantages of MRI 3D Reconstruction in Analyzing Liver Segmentation

Computed tomography (CT) 3D reconstruction is based on the portal phase. Through the display of hepatic vein and portal vein, the four-level vascular supply area of the liver is analyzed and segmented. According to the results of this experiment, the clarity of the hepatic vein in the MRI group was significantly better than that in CT group ( $p < 0.05$ ). In addition, accurate 3D reconstruction technology based on MRI can not only be based on portal phase, but also refer to hepatobiliary phase. Compared with CT portal phase images, MRI has certain advantages in terms of vascular imaging uniformity and contrast with liver parenchyma. MRI makes the clear display of biliary system in hepatobiliary phase after using hepatobiliary specific contrast agent—a display which cannot be provided by CT data. Conventional magnetic resonance cholangiopancreatography shows the morphological characteristics of bile duct tree and pancreatic duct by inhibiting the soft tissue and bone in all backgrounds. It cannot reflect the functional characteristics of bile duct. Therefore, some patients with ascites will affect its display. Under the action of hepatobiliary-specific contrast agents, cholangiography can be clearly displayed by T1-weighted images. MRI cholangiography, which is widely studied at present, is very helpful for the localization and qualitative diagnosis of biliary diseases. Hepatobiliary-specific contrast-enhanced MRI has become a new method to evaluate liver function (14). Hepatobiliary phase images using hepatobiliary-specific magnetic resonance contrast agents can provide more detailed information of liver lesions, including qualitative analysis of lesions and analysis of hepatocyte function (15). This can not only evaluate liver function and reserve function, but also predict postoperative functional liver volume more accurately combined with 3D reconstruction technology.

## Analysis of Liver Volume in 3D Reconstruction Technology of CT and MRI

The results of this experimental study show that there is no significant difference between MRI 3D reconstruction technology and CT 3D reconstruction technology. From the MR 3D reconstruction image, we can also see that the boundary of the liver is slightly blurred compared with the edge of the CT 3D reconstruction model. Hence, it may slightly expand the resection range of liver lesions to a certain extent. This presentation of blurred boundaries may be related to the layer thickness of MRI liver scanning. Therefore, the relevant influencing factors are expected to be corrected by adjusting the layer thickness and other parameters in a follow-up study. However, MRI has many parameters and contains a large amount of information. The combination of multiple sequences and the application of special imaging parameters has important research value and research potential. Many patients with liver cancer have a history of fatty liver. The comparison of the same inverse bitmap can help the qualitative diagnosis of fatty liver. Moreover, the application of quantitative magnetic susceptibility weighted map sequence makes it possible to quantitatively diagnose liver fat deposition and clarify the abandoned functional liver volume with different texture and reserve function to more accurately and effectively calculate the functional liver volume. Hence, reducing the incidence of postoperative liver failure.

## The Significance of 3D Reconstruction Technology of CT and MRI for Surgery

Many patients with primary liver cancer are complicated with hepatitis and cirrhosis. Liver failure may occur after hepatectomy, resulting in a low expression of hepatocyte growth factor, a reduction of the volume of liver regeneration, and an increase of the risk of postoperative mortality in patients with primary liver cancer (16). Liver failure after hepatectomy is the main cause of death of perioperative patients. The main reason is that the residual liver function is not enough to meet the needs of human body (17). Therefore, it is very important to accurately evaluate the liver reserve function beforehand. According to the experimental results, the results of 3D reconstruction technology of whole liver volume and right liver volume are consistent. Since the outline of MRI data is based on points, the total volume and segmented volume data are consistent. It can be seen that the measurement of other volumes is also consistent. According to the current research (7, 18–20), CT 3D reconstruction technology can accurately predict the scope of hepatectomy, evaluate the residual liver volume by simulating the residual liver volume after hepatectomy, evaluate the liver reserve function, guide the formulation of surgical scheme and assist in the discrimination of complex surgery so as to accurately judge the prognosis and outcome of liver surgery, and give full play to the advantages and characteristics of surgery. Therefore, the accuracy of 3D

reconstruction simulation surgery based on MRI data is of great clinical value.

## DEFICIENCIES AND PROSPECTS

This study is a retrospective study. Magnetic Resonance (MR) data are conventional data without prospective design and optimization. In the next study, the impact of scanning parameter optimization on 3D reconstruction model can be studied. In addition, all 3D reconstruction models based on MRI are manually sketched by using the 3D post-processing software of CT. This brings some inconvenience and factors to the operation process. A study on effective semi-automatic or even full-automatic 3D post-processing method based on plain MRI scan and hepatobiliary phase may further assist the application of clinical work.

## CONCLUSION

It is feasible to accurately obtain the 3D reconstruction information of liver based on MRI technology. In addition, the 3D reconstruction based on MRI data has its irreplaceable advantages in addition to the information provided by CT 3D reconstruction.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Fourth Affiliated Hospital of Harbin Medical University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

SC and HL conceived and designed most of the study. SD performed the simulations, analyses, and wrote most of the manuscript. ZG contributed to the figure design and manuscript writing. SC supervised the project. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Horizontal Project: The Application of Image Data Quality Control in 3D Printing Technology, Project No. 2020-19, by China Postdoctoral Science Foundation under Grant 2021M690574, and by Interdisciplinary Research Foundation of HIT under Grant IR2021230.

## REFERENCES

- Smelt J, Pontiki A, Jahangiri M, Rhode K, Nair A, Bille A. 3D printing for chest wall reconstruction in thoracic surgery: building on experience. *Thorac Cardiovasc Surg.* (2020) 68:352–6. doi: 10.1055/s-0039-1678611
- Breik O, Idle M, Martin T, Praveen P, Parmar S. 3D computer-assisted surgical planning and manufacturing in complex maxillary reconstruction. *Atlas Oral Maxillofac Surg Clin North Am.* (2020) 28:151–64. doi: 10.1016/j.cxom.2020.05.008
- Mehta S, Byrne N, Karunanithy N, Farhadi J. 3D printing provides unrivalled bespoke teaching tools for autologous free flap breast reconstruction. *J Plast Reconstr Aesthet Surg.* (2016) 69:578–80. doi: 10.1016/j.bjps.2015.12.026
- Hashimoto D, Dohi T, Tsuzuki M, Horiuchi T, Ohta Y, Chinzei K, et al. Development of a computer-aided surgery system: 3D graphic reconstruction for treatment of liver cancer. *Surgery.* (1991) 109:589–96.
- Yamanaka J, Okada T, Saito S, Kondo Y, Yoshida Y, Suzumura K, et al. Minimally invasive laparoscopic liver resection: 3D MDCT simulation for preoperative planning. *J Hepatobiliary Pancreat Surg.* (2009) 16:808–15. doi: 10.1007/s00534-009-0112-8
- Cai W, Fan Y, Hu H, Xiang N, Fang C, Jia F. Postoperative liver volume was accurately predicted by a medical image three dimensional visualization system in hepatectomy for liver cancer. *Surg Oncol.* (2017) 26:188–94. doi: 10.1016/j.suronc.2017.03.006
- He YB, Bai L, Jiang Y, Ji XW, Tai QW, Zhao JM, et al. Application of a 3D reconstruction technique in liver autotransplantation for end-stage hepatic alveolar echinococcosis. *J Gastrointest Surg.* (2015) 19:1457–65. doi: 10.1007/s11605-015-2842-z
- Chinese SODM, Liver CCOCMDA, Clinical PMCOCMDA, Digital ISOCORHA. [Clinical practice guidelines for precision diagnosis and treatment of complex liver tumor guided by 3D visualization technology (version 2019)]. *Nan Fang Yi Ke Da Xue Xue Bao.* (2020) 40:297–307. doi: 10.12122/j.issn.1673-4254.2020.03.01
- Usman S, Smith L, Brown N, Major V. Diagnostic accuracy of magnetic resonance imaging using liver tissue specific contrast agents and contrast enhanced multi detector computed tomography: a systematic review of diagnostic test in hepatocellular carcinoma (HCC). *Radiography (Lond).* (2018) 24:e109–14. doi: 10.1016/j.radi.2018.05.002
- Li XQ, Wang X, Zhao DW, Sun J, Liu JJ, Lin DD, et al. Application of Gd-EOB-DTPA-enhanced magnetic resonance imaging (MRI) in hepatocellular carcinoma. *World J Surg Oncol.* (2020) 18:219. doi: 10.1186/s12957-020-01996-4
- Lebert P, Adens-Fauquembergue M, Azahaf M, Gnemmi V, Behal H, Luciani A, et al. MRI for characterization of benign hepatocellular tumors on hepatobiliary phase: the added value of in-phase imaging and lesion-to-liver visual signal intensity ratio. *Eur Radiol.* (2019) 29:5742–51. doi: 10.1007/s00330-019-06210-y
- Kierans AS, Kang SK, Rosenkrantz AB. The diagnostic performance of dynamic contrast-enhanced MR imaging for detection of small hepatocellular carcinoma measuring up to 2 cm: a meta-analysis. *Radiology.* (2016) 278:82–94. doi: 10.1148/radiol.2015150177
- Wu JW, Yu YC, Qu XL, Zhang Y, Gao H. Optimization of hepatobiliary phase delay time of Gd-EOB-DTPA enhanced magnetic resonance imaging for identification of hepatocellular carcinoma in patients with cirrhosis of different degrees of severity. *World J Gastroenterol.* (2018) 24:415–23. doi: 10.3748/wjg.v24.i3.415
- Yoon JH, Lee JM, Kang HJ, Ahn SJ, Yang H, Kim E, et al. Quantitative assessment of liver function by using gadoxetic acid-enhanced MRI: hepatocyte uptake ratio. *Radiology.* (2019) 290:125–33. doi: 10.1148/radiol.2018180753
- Li XM, Chen Z, Xiao EH, Shang QL, Ma C. Diagnostic value of gadobenate dimeglumine-enhanced hepatocyte-phase magnetic resonance imaging in evaluating hepatic fibrosis and hepatitis. *World J Gastroenterol.* (2017) 23:3133–41. doi: 10.3748/wjg.v23.i17.3133
- Ma JL, He LL, Li P, Jiang Y, Hu JL, Zhou YL, Liang XX, et al. Clinical features and outcomes of repeated endoscopic therapy for esophagogastric variceal hemorrhage in cirrhotic patients: ten-year real-world analysis. *Gastroenterol Res Pract.* (2020) 2020:5747563. doi: 10.1155/2020/5747563
- Wang YY, Zhong JH, Su ZY, Huang JF, Lu SD, Xiang BD, et al. Albumin-bilirubin versus Child-Pugh score as a predictor of outcome after liver resection for hepatocellular carcinoma. *Br J Surg.* (2016) 103:725–34. doi: 10.1002/bjs.10095
- Zhang J, Qiao QL, Guo XC, Zhao JX. Application of 3D visualization technique in preoperative planning of progressive hilar cholangiocarcinoma. *Am J Transl Res.* (2018) 10:1730–5.
- Wang P, Que W, Zhang M, Dai X, Yu K, Wang C, et al. Application of 3-dimensional printing in pediatric living donor liver transplantation: a single-center experience. *Liver Transpl.* (2019) 25:831–40. doi: 10.1002/lt.25435
- Wu TC, Fang CH, Liu WY, Cai W, Fan YF, Yang J, et al. 3D reconstruction aids surgery for complicated hepatolithiasis. *Hepatogastroenterology.* (2014) 61:613–22.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Cao, Li, Dong and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# RDHCformer: Fusing ResDCN and Transformers for Fetal Head Circumference Automatic Measurement in 2D Ultrasound Images

Chaoran Yang<sup>1,2</sup>, Shanshan Liao<sup>3</sup>, Zeyu Yang<sup>4</sup>, Jiaqi Guo<sup>1</sup>, Zhichao Zhang<sup>1</sup>, Yingjian Yang<sup>1,2</sup>, Yingwei Guo<sup>1,2</sup>, Shaowei Yin<sup>3</sup>, Caixia Liu<sup>3</sup> and Yan Kang<sup>1,2,5\*</sup>

<sup>1</sup> College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China, <sup>2</sup> Medical Device Innovation Center, Shenzhen Technology University, Shenzhen, China, <sup>3</sup> Department of Obstetrics, Shengjing Hospital of China Medical University, Shenyang, China, <sup>4</sup> Department of Ultrasound, Shengjing Hospital of China Medical University, Shenyang, China, <sup>5</sup> Engineering Research Centre of Medical Imaging and Intelligent Analysis, Ministry of Education, Shenyang, China

## OPEN ACCESS

### Edited by:

Kuanquan Wang,  
Harbin Institute of Technology, China

### Reviewed by:

Yili Zhao,  
University of California, San Francisco,  
United States  
Xueju Wang,  
Peking University Third Hospital, China  
Li Ke,  
Shenyang University of  
Technology, China

### \*Correspondence:

Yan Kang  
kangyan@bmie.neu.edu.cn

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

**Received:** 05 January 2022

**Accepted:** 07 March 2022

**Published:** 29 March 2022

### Citation:

Yang C, Liao S, Yang Z, Guo J, Zhang Z, Yang Y, Guo Y, Yin S, Liu C and Kang Y (2022) RDHCformer: Fusing ResDCN and Transformers for Fetal Head Circumference Automatic Measurement in 2D Ultrasound Images. *Front. Med.* 9:848904. doi: 10.3389/fmed.2022.848904

Fetal head circumference (HC) is an important biological parameter to monitor the healthy development of the fetus. Since there are some HC measurement errors that affected by the skill and experience of the sonographers, a rapid, accurate and automatic measurement for fetal HC in prenatal ultrasound is of great significance. We proposed a new one-stage network for rotating elliptic object detection based on anchor-free method, which is also an end-to-end network for fetal HC auto-measurement that no need for any post-processing. The network structure used simple transformer structure combined with convolutional neural network (CNN) for a lightweight design, meanwhile, made full use of powerful global feature extraction ability of transformer and local feature extraction ability of CNN to extract continuous and complete skull edge information. The two complement each other for promoting detection precision of fetal HC without significantly increasing the amount of computation. In order to reduce the large variation of intersection over union (IOU) in rotating elliptic object detection caused by slight angle deviation, we used soft stage-wise regression (SSR) strategy for angle regression and added KLD that is approximate to IOU loss into total loss function. The proposed method achieved good results on the HC18 dataset to prove its effectiveness. This study is expected to help less experienced sonographers, provide help for precision medicine, and relieve the shortage of sonographers for prenatal ultrasound in worldwide.

**Keywords:** prenatal ultrasound, fetal head circumference, rotating object detection, transformers, convolutional neural network

## INTRODUCTION

Prenatal ultrasound is one of the most important examination methods during pregnancy due to its fast, low-risk and non-invasive characteristics. Fetal head circumference (HC) is one of the most essential biological indexes in accurate assessment of fetal development, which provides a method for monitoring fetal growth, estimating gestational age, and determining delivery mode. It is of

paramount importance to ensure the continued wellbeing of mothers and newborns both during and after pregnancy. In prenatal ultrasound screening, the fetal head circumference is measured on standard plane of thalamus according to the obstetric ultrasound guidelines (1, 2), and the circumference of ellipse can be identified as the fetal HC since the contour of skull is similar to an ellipse. During measurement, the contour of skull is marked by sonographers, and the HC can be calculated by ellipse parameters which is obtained through fitting on post-processing software embedded in ultrasound equipment. Some semi-automated HC measurement is available on newer OB ultrasound machines, like GE Voluson E10 (SonoBiometry). The measurement results of semi-automated methods are directly affected by the accuracy in performing segmentation.

However, it is challenging for AI models to measure HC due to blurred or incomplete skull edge in ultrasound images. Accurate measurement can provide an important reference for the evaluation of fetal growth and development. Therefore, in order to improve efficiency, reliability, and reduce the workload of doctors, clinical practice puts forward high requirements for automatic segmentation (3). It is of great significance to develop an efficient and accurate method for automatic measurement of fetal HC.

In this paper, a lightweight detection network that combined with Transformer and Convolutional neural network (CNN) is proposed to detect the position of the fetal head, regress the parameters of ellipse, and then solve the head circumference value through the parameters. For automatic measurement tasks of HC, it is a one-stage network of detection. The process does not require any post-processing, such as edge extraction or ellipse fitting, and the process comparison between our method and general detection method is shown in **Figure 1**. This work makes the following contributions:

- 1) To our knowledge, our method is the first to apply the rotating ellipse detection method to the skull edge detection task. This is a one-stage network based on anchor-free method;
- 2) Taking Res\_DCN as baseline, Deformable Convolutional Networks (DCN) combined with ResNet can learn the features of irregular boundary better and promote capability of local feature extraction. Meanwhile, powerful global feature extraction ability of Transformer is used to obtain more abundant continuous features of boundary from the global view. The proposed approach combines simple Transformer structure with CNN to obtain complete and accurate elliptical information as much as possible without significantly increasing the amount of computation;
- 3) Soft Stagewise Regression (SSR) strategy is used to map angle regression problems into classification problems. Firstly, the angle is roughly classified, and then the dynamic range is introduced to make every bin can do translation and scaling for fine classification. Classify the angles from coarse to fine to make angle regression accuracy higher;
- 4) Kullback-Leibler Divergence (KLD) loss that is similar to IOU loss is added into total loss function to solve the problem that intersection over union (IOU) between ground truth (GT) and prediction changes greatly caused by small angle deviation or center point deviation of the rotating target, as the IOU of rotate target is difficult to calculate. KLD loss can further improve the regression accuracy of elliptic parameters;
- 5) The proposed method gets good results compared with other existing HC measurement methods in open data set of HC18. It is noteworthy that the method is simple and efficient without requiring any post-processing.

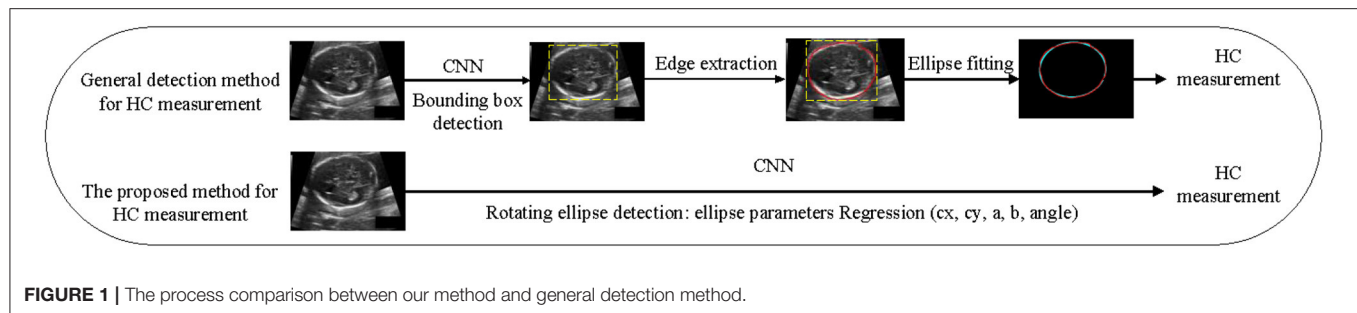
## RELATED WORK AND MOTIVATION

### Related Work

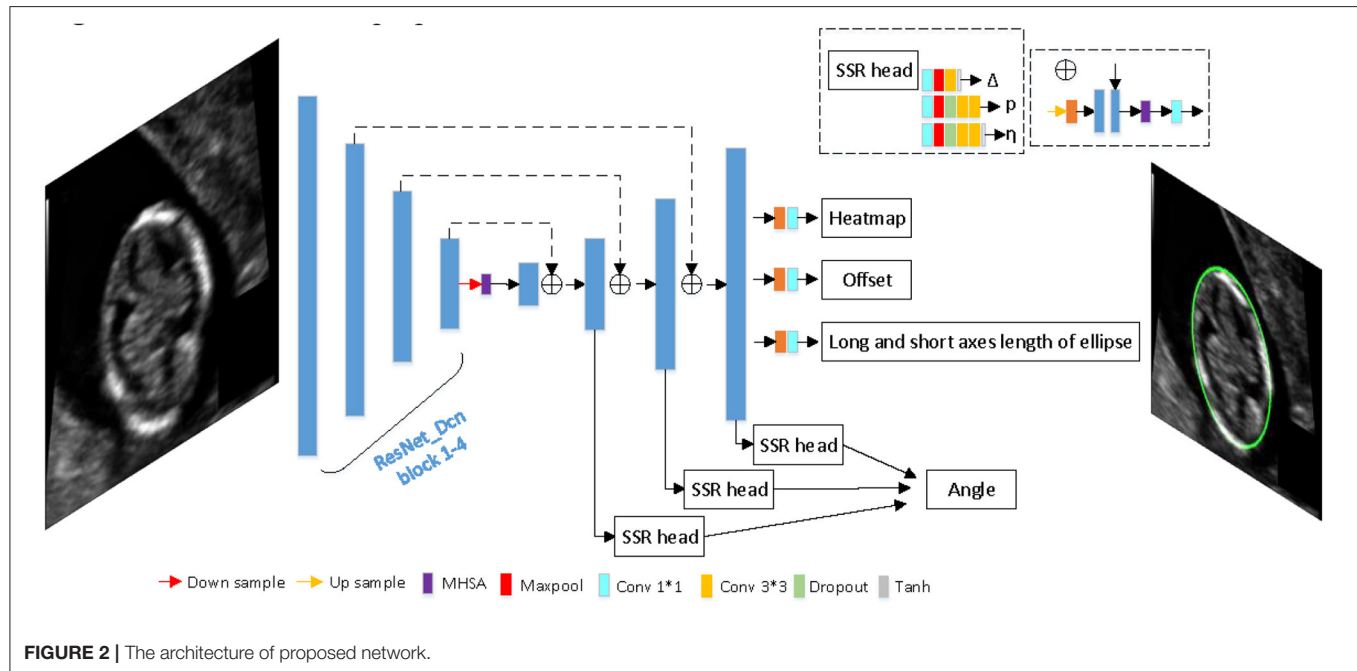
In the past research, many methods based on machine learning have been used to extract skull edge features, such as Haar-like features combined with different classifiers (4–9). There are also some methods based on gradient (10), threshold (11), active boundary model (12), contour fragment model (13), multi-group filters mixing (14) to extract features of skull region or boundary. After the skull features were extracted, different methods such as Hough transform (15) and ElliFit (16) were used to fit the elliptic skull boundary and further measure HC. Although some good results have been achieved by above methods, they all require prior knowledge or artificially designed features with poor robustness and large amount of calculation.

In recent years, CNN have been widely used in medical image segmentation (17, 18), Sinclair et al. (19) and Wu et al. (20) used the cascaded Fully Convolutional Network (FCN) to segment the skull region. U-net and its extended form have a symmetrical structure and extract rich features by using the fusion of different feature layers (21, 22). There are also some methods with different understanding of tasks, such as multi-organ segmentation (23), segmentation and regression multi-task methods (24), which are widely used in skull region or boundary segmentation. Skull boundary detection based on CNN segmentation method has excellent performance in regional segmentation, after predicting the skull region, a series of complex post-processing such as expansion, corrosion and edge extraction are carried out to obtain the skull boundary pixels, and then ellipse parameter fitting is carried out to solve HC, therefore, these methods have huge networks and cumbersome process. The measurement accuracy of head circumference depends on the segmentation result heavily, and the effect is not good for the ultrasonic image with unclear or incomplete boundary.

Object detection technology based on anchor method has good detection results for standard rectangular frame targets (25, 26), but there are no relevant studies on rotating elliptic object detection (i.e., skull edge detection task). The method based on anchor need to preset size of anchor according to IOU, and an appropriate number of anchors are selected with a certain threshold value (such as 0.5) as positive samples for regression distribution of objects. But this leads to two problems in rotating object detection: first, further aggravating the positive and negative sample imbalance. Angle prior should be added to the preset rotating anchor, doubling the number of preset anchors. In addition, rotating anchor angle slightly deviated from GT will lead to sharp decline of IOU. Second, classification is inconsistent with regression. Many studies have discussed this



**FIGURE 1 |** The process comparison between our method and general detection method.



**FIGURE 2 |** The architecture of proposed network.

problem (27), that is, the classification score of predicted results is inconsistent with the positioning accuracy, so inaccurate positioning may be selected when passing the NMS stage or selecting detection results according to the classification score, while the well-positioned anchor is omitted or suppressed.

## Motivation

For skull edge detection task, due to factors such as fetuses at different gestational ages and different positions, the skull edge presents elliptic shapes of different sizes. The detection method based on Anchor needs to design sizes of different proportions according to prior, which is a very complicated process. In addition, IOU between GT and prediction changes greatly because of small angle deviation or center point deviation of the rotating target. Recently, the object detection method based on anchor-free has been greatly developed. CenterNet (28) detects the center point of the object first, and then directly regress the width and height of the object. Of course, we can directly regress a rotation angle to expand CenterNet to rotating object detection. However, the size and angle actually depend on different rotating coordinate systems, so it is difficult to directly

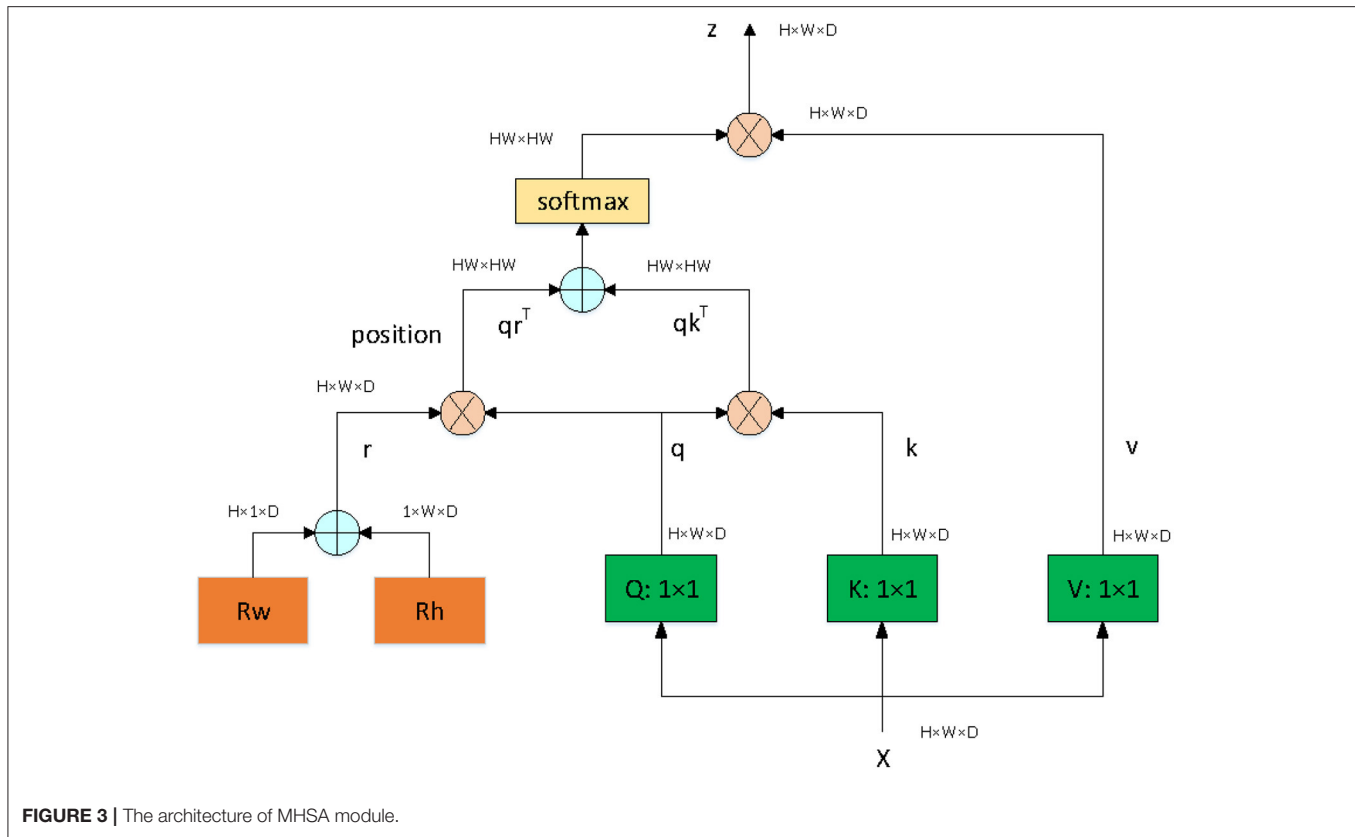
regress parameters. To sum up, we are committed to studying a lightweight and high-precision rotating ellipse detection network for skull edge detection. The proposed method is a one-stage method based on anchor-free to solve the above problems.

## METHODS

In this section, we first describe the overall architecture of the proposed method, and then explain the Gaussian distribution of GT, output maps, and KLD loss function in detail. The output maps are used to generate the oriented ellipse of the objects.

### Architecture

Since our goal is to build a lightweight network, we didn't choose backbone which is too complicated. The proposed network is based on an asymmetric U-shaped architecture (see Figure 2). We use the block 1–5 of ResNet\_DCN as the backbone, simple Multi-head-self-attention [MHSA, see Figure 3, details in reference (29)] is used in encoder's last bottleneck module and the whole up-sampling process. Deformable convolution and self-attention mechanism are used to improve the access



to local information and the continuity of irregular boundary. In decoder, output features of encoder are up-sampled to 1/4 of the input image (scale  $s = 4$ ), we combine a deep layer with a shallow layer through skip connections to share both the high-level semantic information and low-level finer details. In particular, we first up-sample a deep layer to the same size of the shallow layer through bilinear interpolation. The up-sampled features map is refined through a  $3 \times 3$  convolutional layer. The refined feature map is then concatenated with the shallow layer, followed by a  $1 \times 1$  convolutional layer to refine the channel-wise features. In the end, four detection heads are used for ellipse parameters regression (heatmap, center offset, long and short axes of ellipse, angle).

### Gaussian Distribution of GT

The proposed method locates the target based on free-anchor, we need to map GT of keypoints to a 2D Gaussian distribution on the heatmap. The mapping method in reference (28) is not friendly to targets with a large aspect ratio, especially for ellipse, so we modified the mapping method. The GT of an ellipse is  $(c_x, c_y, a, b, \text{angle})$ , where  $(c_x, c_y)$  is the center point of an oriented ellipse,  $a$  and  $b$  are long and short axes of ellipse, respectively, angle is the angle between the short axis and the vertical direction. We generate the smallest horizontal enclosing rectangular box of the ellipse  $(b_x, b_y, w, h)$ ,  $(b_x, b_y)$  is the center point of smallest horizontal enclosing rectangular box of the ellipse,  $w$  and  $h$  are width and height of rectangular box, respectively. We map GT

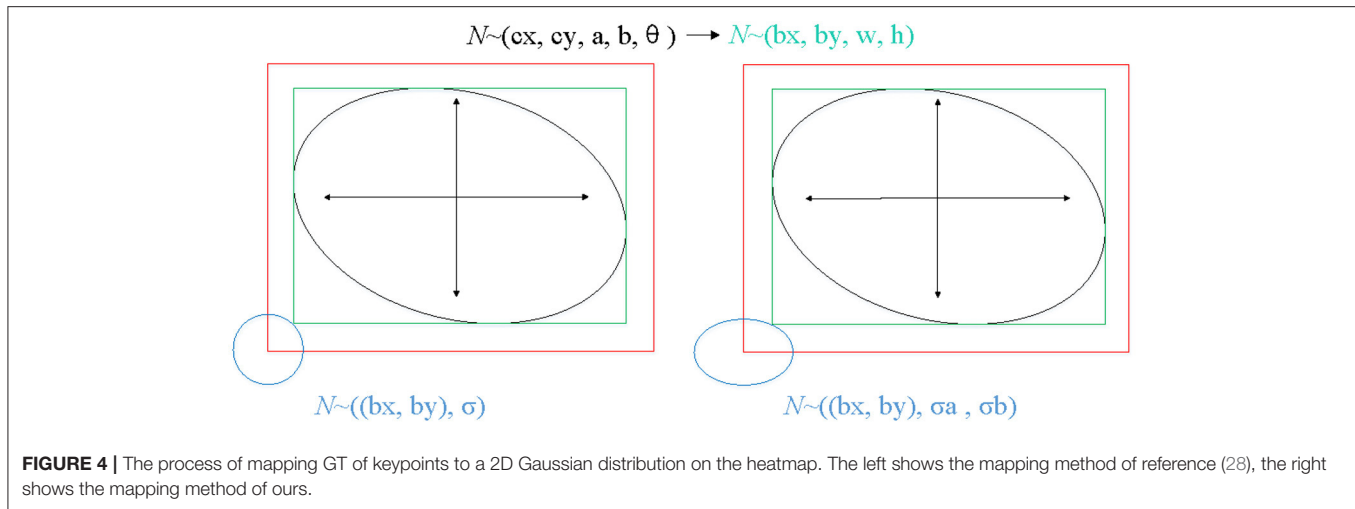
$(b_x, b_y, w, h)$  which can be predicted as a positive sample to 2D Gaussian distribution  $\exp(-(\frac{(p_x - b_x)^2}{2\sigma_a^2} + \frac{(p_y - b_y)^2}{2\sigma_b^2}))$ , where  $\sigma$  is a box size-adaptive standard deviation, (see Figure 4). The heatmap  $P \in R^{c \times \frac{H}{s} \times \frac{W}{s}}$ ,  $H$  and  $W$  are height and width of input image respectively,  $c$  is set to 1 in this work.

### Output Maps Heatmap

In this work, we use the heatmap  $\rho \in R^{c \times \frac{H}{s} \times \frac{W}{s}}$  to detect the center points of arbitrarily oriented objects, where  $c$  is corresponding to one object category. The predicted heatmap value at a particular center point is regarded as the confidence of the object detection. We use the variant focal loss to train the heatmap:

$$L_h = -\frac{1}{N} \sum_i \begin{cases} (1 - \rho_i)^\alpha \log(\rho_i) & \text{if } P_i = 1 \\ (1 - P_i)^\beta \rho_i^\alpha \log(1 - \rho_i) & \text{otherwise} \end{cases} \quad (1)$$

where  $P$  and  $\rho$  refer to the ground-truth and the predicted heatmap values,  $i$  indexes the pixel locations on the feature,  $N$  is the number of objects.  $\alpha$  and  $\beta$  are the hyper-parameters that control the contribution of each point.  $\alpha$  is set to 2 and  $\beta$  is set to 4.



### Center Offset

We transform GT of center point by down-sample from the input image, which is a floating-point type  $C = (\frac{c_x}{s}, \frac{c_y}{s})$ . However, the predicted center point is an integer. To compensate for the discretization error between the floating center point and the integer center point caused by the output stride, we predict an offset map  $O \in R^{2 \times \frac{H}{s} \times \frac{W}{s}}$  can be defined as:

$$O = \left( \frac{c_x}{s} - \left\lfloor \frac{c_x}{s} \right\rfloor, \frac{c_y}{s} - \left\lfloor \frac{c_y}{s} \right\rfloor \right) \quad (2)$$

The offset is optimized with a smooth  $L_1$  loss:

$$L_O = \frac{1}{N} \sum_{k=1}^N \text{Smooth } L_1(O_k - o_k) \quad (3)$$

where  $N$  is the total number of objects,  $o$  refers to the ground-truth offsets,  $k$  indexes the objects.

The smooth  $L_1$  loss can be expressed as:

$$\text{Smooth } L_1(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \quad (4)$$

### Long and Short Axes of Ellipse

We regress to long and short axes of ellipse for each object,  $B = (a, b) \in R^{2 \times \frac{H}{s} \times \frac{W}{s}}$ , where  $a$  is long axes length,  $b$  is short axes length. It can be optimized with a smooth  $L_1$  loss:

$$L_O = \frac{1}{N} \sum_{k=1}^N \text{Smooth } L_1(B_k - b_k) \quad (5)$$

where  $B$  and  $b$  are the ground-truth and the predicted ellipse parameters, respectively.

### Angle

Accurate angle regression is very important for rotating object detection, a small angle variation has marginal influence on the total loss in training, but it may induce a large IOU difference

between the predicted ellipse and the ground-truth ellipse. Because of the symmetry of the ellipse, the rotation angles  $\theta \in [0, 180)$ . Soft-stagewise regression strategy is adopted for angle regression, which takes angle regression as a multi-classification task. We set it as a three-stage classification task ( $S_1 = 18$ ,  $S_2 = 10$ ,  $S_3 = 10$ ). In the first stage, the angle  $\theta \in [0, 180)$  is divided into  $S_1$  parts with a span of  $180 / S_1$ . In the second stage,  $[0, 180 / S_1]$  is divided into  $S_2$  parts with a span of  $180 / S_1 / S_2$ . The third stage is similar, as shown in the **Figure 5**. In each stage, it is a multi-classification task, the sum of the probability of each class and the representative angle of the current class is taken as the final prediction value. The angle is predicted by the following formula for soft-stagewise regression:

$$\theta = \sum_{k=1}^K \sum_{i=0}^{S_k-1} p_i^{(k)} i \left( \frac{V}{\prod_{j=1}^k s_j} \right) \quad (6)$$

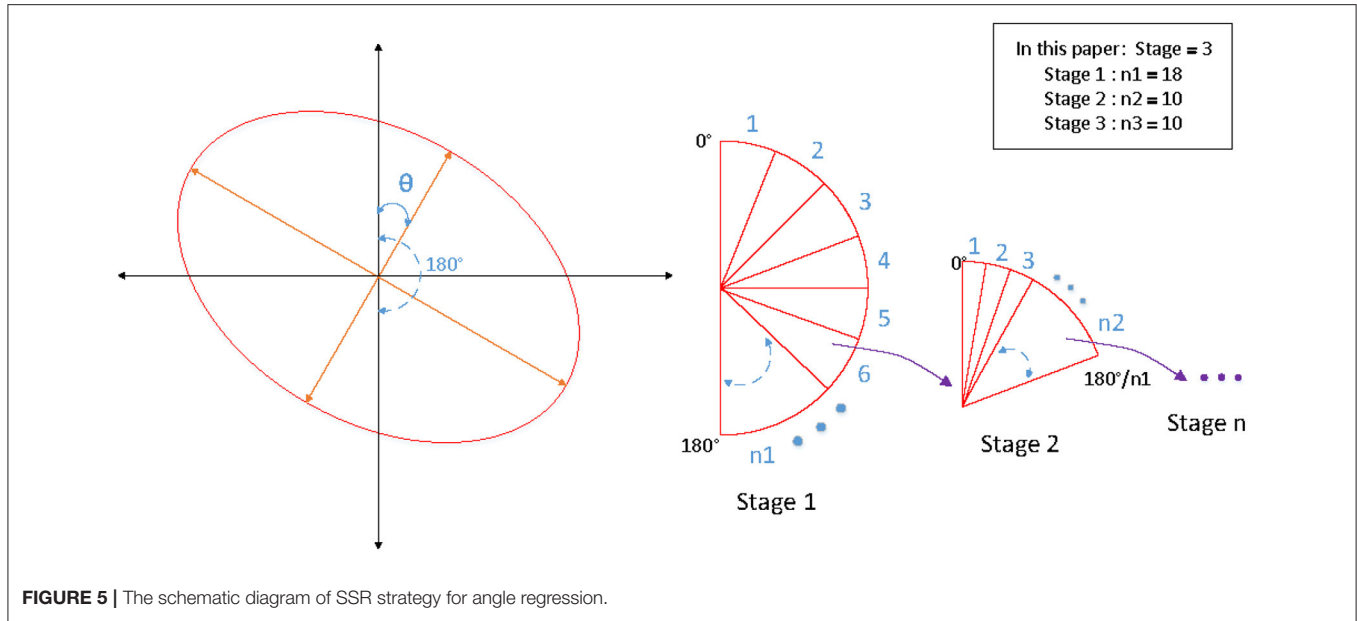
where  $V \in [0, 180)$ ,  $p_i^{(k)}$  refers to the probability of each class for each stage, The last term in the above equation is the bin width  $\omega_k = \frac{V}{\prod_{j=1}^k s_j}$  for the  $k$ -th stage and  $i$  is the bin index. Reference (30) introduced a dynamic range for each bin, that is, it allowed each bin to be shifted and scaled according to the input image. For adjusting the bin width  $\omega_k$  at the  $k$ -th stage, SSR introduce a term  $\Delta k$  to modify  $s_k$  into  $s_k^*$  as follows:

$$s_k^* = s_k(1 + \Delta k) \quad (7)$$

where  $\Delta k$  is the output of a regression network given the input image. For shifting bins, we add an offset term  $\eta$  to each bin index  $i$ . The bin index  $i$  is modified as follows:

$$i = i + \eta_i^{(k)} \quad (8)$$

Thus, the output of SSR head are  $p_i^{(k)}$ ,  $\Delta k$ , and  $\eta_i^{(k)}$ , the angle is regressed from coarse to fine by introducing dynamic range



**FIGURE 5 |** The schematic diagram of SSR strategy for angle regression.

that each bin can do translation and scaling so as to improve the precision of angle regression and reduce the error as much as possible. Angle regression can be optimized with a smooth  $L_1$  loss:

$$L_\theta = \frac{1}{N} \sum_{k=1}^N \text{Smooth } L_1(\theta_k - \theta_k^*) \quad (9)$$

where  $\theta_k$  and  $\theta_k^*$  are the ground-truth and the predicted ellipse parameters, respectively.

## Kullback-Leibler Divergence Loss

The IOU of rotate object is difficult to be calculated, we use KLD to measure the similarity of the two distributions to approximate the IOU. KLD loss also has some advantages when optimizing parameters. When one of the parameters is optimized, the other parameters will be used as its weight to dynamically adjust the optimization rate. In other words, the optimization of parameters is no longer independent, that is, optimizing one parameter will also promote the optimization of other parameters. The optimization of this virtuous circle is the key to KLD as an excellent rotation regression loss. Reference (31) proved its derivability and advantages. Convert GT of the ellipse  $(c_x, c_y, a, b, \theta)$  into a 2-D Gaussian  $N(\mu, \varepsilon)$ , (see **Figure 6**). Specifically, the conversion is:

$$\mu = (c_x, c_y)^T$$

$$\varepsilon^{1/2} = \begin{pmatrix} \frac{b}{2} \cos^2 \theta + \frac{a}{2} \sin^2 \theta & \frac{b-a}{2} \cos \theta \sin \theta \\ \frac{b-a}{2} \cos \theta \sin \theta & \frac{b}{2} \sin^2 \theta + \frac{a}{2} \cos^2 \theta \end{pmatrix} \quad (10)$$

$X_p \sim N_p(\mu_p, \varepsilon_p)$  and  $X_t \sim N_t(\mu_t, \varepsilon_t)$ , the KLD between two 2-D Gaussian is:

$$D_{kl}(N_p \| N_t) = \frac{1}{2} (\mu_p - \mu_t)^T \varepsilon_t^{-1} (\mu_p - \mu_t) + \frac{1}{2} \text{Tr}(\varepsilon_t^{-1} \varepsilon_p) + \frac{1}{2} \ln \frac{|\varepsilon_t|}{|\varepsilon_p|} - 1 \quad (11)$$

The KLD loss is:

$$L_{reg} = 1 - \frac{1}{1 + \log(D_{kl}(N_p \| N_t) + 1)} \quad (12)$$

The final regression loss is:

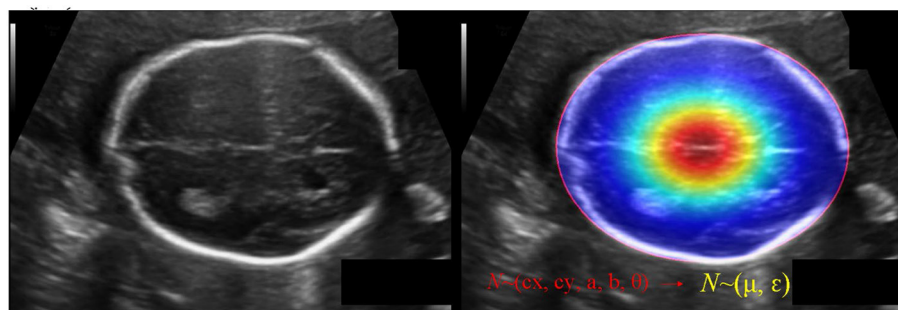
$$L_{total} = L_h + L_O + L_\theta + L_{reg} \quad (13)$$

## EXPERIMENTS

### Datasets and Implementation Details

Dataset is from the HC18 grand-challenge<sup>1</sup> which provided 1334 2D ultrasound images from standard planes, a training set with 999 images and a test set with 335 images. Manual annotations of HC were made by senior experts. Since the data set only provides standard planes, that is to say, each image has a target, and the target accounts for a large proportion of the image. In order to balance the positive and negative samples, we used two ways to generate negative samples, one way is to remove the target in the image and fill it with surrounding information, the other way is to randomly crop the image into patches, and then resize them to the size of the network input. If the IOU with GT is  $< 0.3$ , it will be considered as a negative sample. The size of each 2D ultrasound image is approximately  $540 \times 800$  with the pixel size ranging from 0.052 to 0.6 mm. Data augmentation is

<sup>1</sup> Available online at: <https://hc18.grand-challenge.org>.

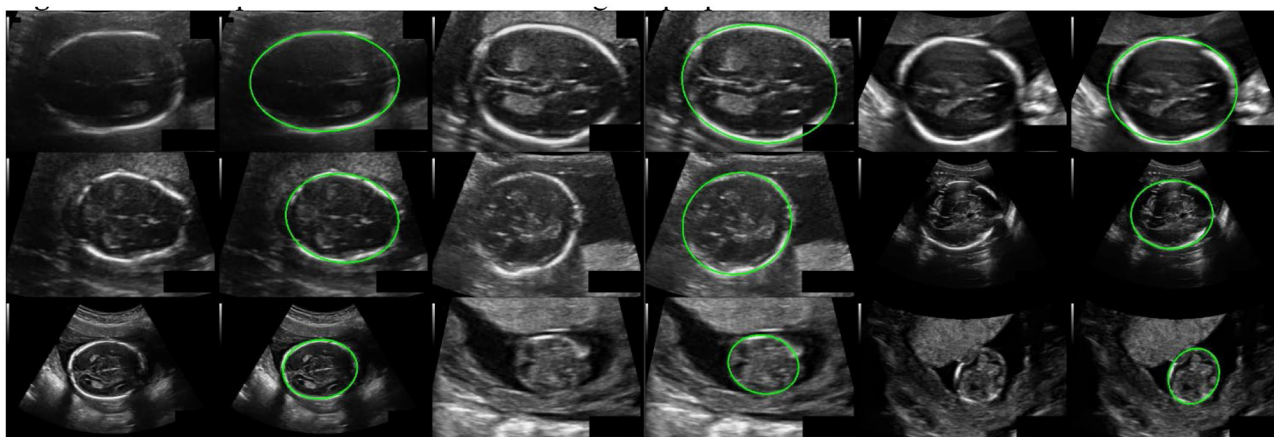


**FIGURE 6** | The schematic diagram of converting GT of the ellipse  $(c_x, c_y, a, b, \theta)$  into a 2-D Gaussian  $N(\mu, \epsilon)$ .

**TABLE 1** | The comparison results between our method and other common methods base on segmentation algorithm.

Model	MAE $\pm$ std (mm)	ME $\pm$ std (mm)	Params (M)	AP
U-Net (33)	2.36 $\pm$ 5.60	0.41 $\pm$ 2.91	31.042	-
U-Net++ (34)	2.29 $\pm$ 2.33	0.27 $\pm$ 2.73	9.163	-
CE-Net (35)	2.24 $\pm$ 2.28	0.16 $\pm$ 2.12	29.003	-
SE-Unet (36)	2.27 $\pm$ 3.61	0.09 $\pm$ 2.33	-	-
HC18 challenge best	<b>1.72 <math>\pm</math> 1.60</b>	<b>0.04 <math>\pm</math> 2.35</b>	-	-
Ours	1.97 $\pm$ 1.89	0.11 $\pm$ 2.71	<b>24.31</b>	<b>84.45</b>

MAE, Mean Absolute Error; ME, Mean Error; std, standard deviation. Param, the size of model parameters; M, Mbyte; AP, Average precision. Bold values represent the best value of each indicator.



**FIGURE 7** | Some examples of detection results using the proposed method.

essential to make model more robust. The data augmentation strategy was as follows: Rotation: rotation angle is  $[-30^\circ, 30^\circ]$ , and the interval is  $10^\circ$ . Scale transformation: the scaling ratio is  $[0.85, 1.15]$ , and the interval is 0.05. Gamma transformation: gamma factor is  $[0.5, 1.5]$ , and the interval is 0.1. Flip: the input image is flipped randomly. After data augmentation, training set is expanded from 999 to 12,999, of which 200 are used as validation set and the rest are used as a new training set. The Stochastic Gradient Descent (SDG) optimizer is selected, the initial learning rate is set to 0.005, the momentum is 0.9, the dropout rate is 0.1, and the batchsize is set to 16. The training

procedure is completed on two NVIDIA GeForce RTX 2080TI graphics cards.

## Evaluation Metrics

In order to comprehensively evaluate the performance of the model and conduct comparative analysis, regression Average Precision (AP), Mean Absolute Error (MAE), Mean Error (ME) of head circumference are adopted as the evaluation metrics of the model in this paper. HC can be calculated as follows (32):

$$HC = \pi \left[ 3(a + b) - \sqrt{(3a + b)(a + 3b)} \right] \quad (14)$$

where  $a$  and  $b$  are parameters of semi-long axis and semi-short axis of the ellipse. Mean Absolute Error of fetal HC is defined as:

$$MAE_{hc} = \frac{1}{N} \sum_{i=1}^N |\widehat{HC}_i - HC_i| \quad (15)$$

Mean Error of fetal HC is defined as:

$$ME_{hc} = \frac{1}{N} \sum_{i=1}^N (\widehat{HC}_i - HC_i) \quad (16)$$

Where  $\widehat{HC}$  and  $HC$  denote the HC measured by the proposed method and the real value of fetal HC, respectively.

Average Precision is a commonly used evaluation metric in object detection that obtained by calculating the area of Precision-Recall curve.

**TABLE 2** | Ablation experiments results.

Backbone	Method	AP (with MHSA/ not)
Res_DCN-50	Smooth L1 (center, a, b, angle)	81.83/77.33
	Smooth L1 (center, a, b) + SSR (angle)	83.25/79.97
	Smooth L1 (center, a, b) + SSR (angle) + KLD	<b>84.45/81.71</b>

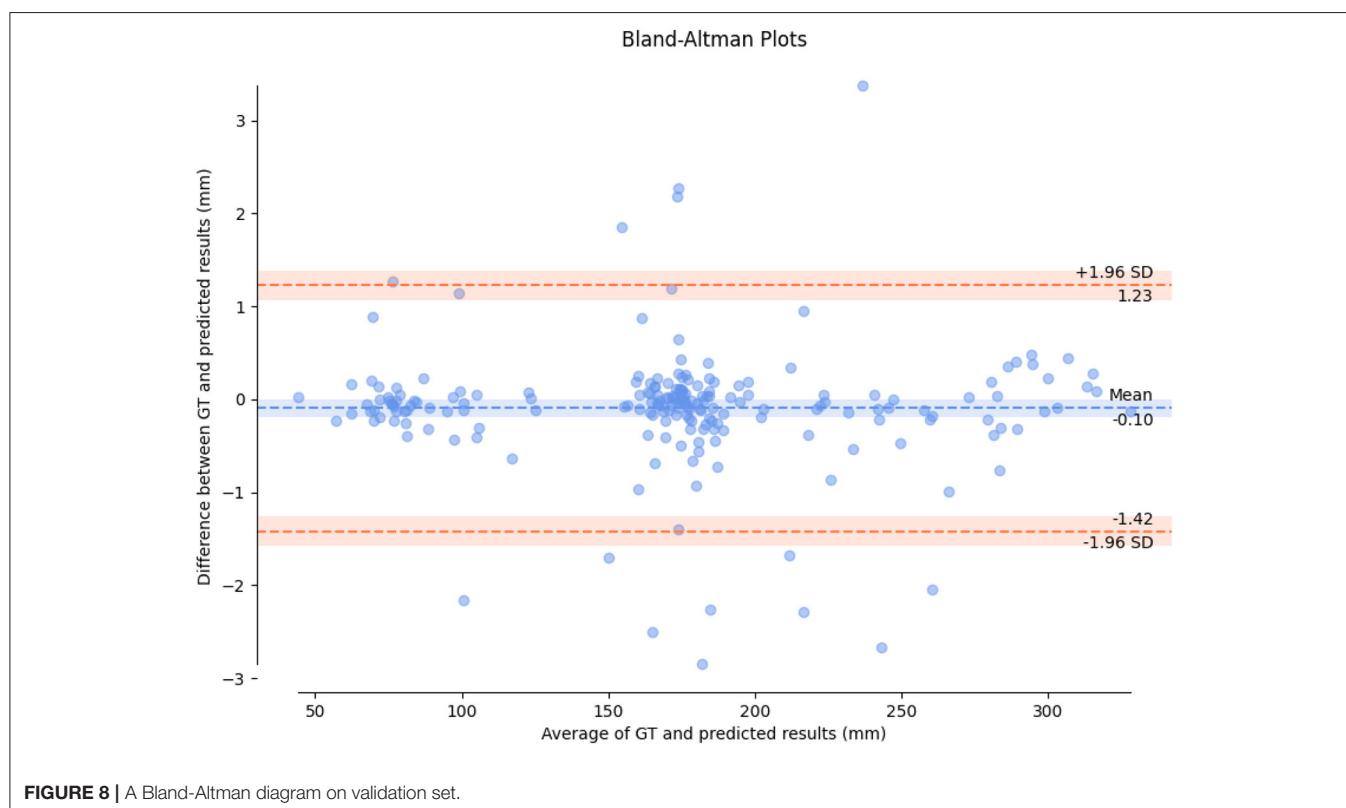
Comparison of Res\_DCN-50 with or without MHSA module, SSR, and KLD loss. Bold values represent the best value of each indicator.

## RESULTS

Our method has achieved good results on HC18 dataset, Average precision (AP) is 84.45%. MAE  $\pm$  std (mm) is  $1.97 \pm 1.89$ , ME  $\pm$  std (mm) is  $0.11 \pm 2.71$ , the parameter size of the proposed model is 24.31M. **Table 1** shows the comparison results between our method and other common methods base on segmentation algorithm. It can be seen that our method has achieved good skull edge detection results without significantly increasing the amount of model parameters, and it can be comparable to the state-of-the-art method. It is worth noting that our method is simple and efficient. Unlike methods based on segmentation algorithm, our method do not need any complicated post-processing, which is an end-to-end network strictly for head circumference detection task. There are some examples of detection results in **Figure 7**.

## Ablation Analysis

Some ablation experiments were conducted to prove the effectiveness of each module design in our algorithm. Taking Res\_DCN-50 as the backbone as an example, the experimental results are as shown in **Table 2**. It can be seen that using normal Smooth L1 function and without MHSA module achieved the AP: 77.33%, while adding the MHSA module AP: 81.83%, increased 4.5%, it indicated that the MHSA module has a significant improvement for the task. With the addition of MHSA module, AP was increased by 1.42-83.25% by using SSR detection head for angle, then after adding KLD Loss, AP was increased



**FIGURE 8** | A Bland-Altman diagram on validation set.

by 1.2–84.45% further, at this time, compared with no MHSA module AP: 81.71%, it is an increase of 2.74%. This indicated that the excellent global feature extraction ability of MHSA module improves the model's ability to extract skull edge continuity features, it is helpful for each module of the network.

## Consistency Analysis

In order to evaluate the consistency between HC measured by the proposed method and real value of HC, we draw a Bland-Altman diagram on validation set, as shown in **Figure 8**. Compared with the real value of HC, the Mean Difference of HC measurement is  $-0.10$  mm with a 95% confidence interval and the error ranges from  $-1.42$ – $1.23$  mm. It indicated the HC measured by the proposed method has a good consistency with the real value.

## CONCLUSION

A new fetal head circumference auto-measurement method based on rotating ellipse detection has been proposed in this paper, which is a strictly end-to-end detection method without any post-processing for the task. As far as we know, this is the first application of end-to-end detection network to measure fetal head circumference directly. We combine transformer and CNN because convolution operations can extract rich context features in local area and transformer (MHSA) module can capture long-distance feature relationship benefitting from its ability of global and dynamic receptive fields. The two complement each other for promoting detection precision of fetal HC without significantly increasing the amount of computation. For the task of rotating elliptic object detection, the precision of angle regression is very important. Slight angle deviation will bring large changes in IOU. Therefore, we used SSR strategy for angle regression and added KLD that is approximate to IOU loss into total loss function. These methods significantly improve the detection precision. This study is expected to help less experienced sonographers, provide help for precision medicine, and relieve the shortage of sonographers for prenatal ultrasound in worldwide. There are also some shortcomings in our work, a little deviation can be allowed in predicting the location of target

center point in the inference stage (that is, positive sample can be determined if the IOU is greater than a certain threshold), therefore, in order to facilitate calculation, we conducted pre-processing operation in the process of mapping the center point of ellipse to 2D Gaussian distribution on the heatmap. We generated the smallest horizontal enclosing rectangle of the ellipse, and used center point of rectangle as the new center point for mapping. There is a slight error with the center point of the ellipse, which may affect the precision of the detection results. This is also the study direction that we need to improve in the future.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: <https://hc18.grand-challenge.org>.

## AUTHOR CONTRIBUTIONS

CY: method and experiment design and manuscript writing. SL, ZY, and SY: data analysis and interpretation. JG, ZZ, YY, and YG: method investigation and experiment analysis. YK and CL: supervision. All authors contributed to the article and approved the submitted version.

## FUNDING

This study was supported by funding from the National Natural Science Foundation of China (grant numbers 81401143), National Key Research and Development Program of China (grant numbers 2018YFC1002900), the Stable Support Plan for Colleges and Universities in Shenzhen, China (grant numbers SZWD2021010), and the National Natural Science Foundation of China (grant numbers 62071311).

## ACKNOWLEDGMENTS

The authors thank HC18 challenge in grand-challenge for providing an open source data.

## REFERENCES

- Chudleigh T. The 18+020+6 weeks fetal anomaly scan national standards. *Ultrasound*. (2010) 18:92–8. doi: 10.1258/ult.2010.010014
- Sanders RC, James AE. The principles and practice of ultrasonography in obstetrics and gynecology. *JAMA*. (1981) 245:80. doi: 10.1001/jama.1981.03310260058043
- Sarris I, Ioannou C, Chamberlain P, Ohuma E, Roseman F, Hoch L. Intra- and interobserver variability in fetal ultrasound measurements. *Ultrasound Obstet Gynecol*. (2012) 39:266–73. doi: 10.1002/uog.10082
- Nadiyah P, Rofiqah N, Firdaus Q, Sigit R, Yuniarti H. Automatic detection of fetal head using haar cascade and fit ellipse. *International Seminar on Intelligent Technology and Its Applications*. Surabaya: IEEE. (2019).
- Jatmiko W, Habibie I, Ma'sum MA, Rahmatullah R, Satwika PI. Automated telehealth system for fetal growth detection and approximation of ultrasound images. *Int J Smart Sens Intell Syst*. (2015) 8:697–719. doi: 10.21307/ijssis-2017-779
- Namburete AIL, Noble JA. Fetal cranial segmentation in 2D ultrasound images using shape properties of pixel clusters. In: *IEEE International Symposium on Biomedical Imaging* (2013).
- van den Heuvel TLA, de Bruijn D, de Korte CL, Ginneken, BV. Automated measurement of fetal head circumference using 2D ultrasound images. *PLoS ONE*. (2018) 13:e0200412. doi: 10.1371/journal.pone.0200412
- Yaqub M, Kelly B, Papageorgiou AT, Noble JA. Guided random forests for identification of key fetal anatomy and image categorization in ultrasound scans. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich. (2015).
- Carneiro G, Georgescu B, Good S, Comaniciu D. Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. *IEEE Trans Med Imag*. (2008) 27:1342–55. doi: 10.1109/TMI.2008.928917
- Irene K, Haidi H, Faza N, Chandra W. *Fetal Head and Abdomen Measurement Using Convolutional Neural Network, Hough Transform, and Difference of Gaussian Revolved along Elliptical Path (Dogell) Algorithm*. arXiv (2019) arXiv:1911.06298.

11. Ponomarev GV, Gelfand MS, Kazanov MD. *A Multilevel Thresholding Combined With Edge Detection and Shape-Based Recognition for Segmentation of Fetal Ultrasound Images*. Proc. Chall. US biometric Meas. from fetal ultrasound images, ISBI (2012).
12. Rahayu KD, Sigit R, Agata D, Pambudi A, Istiqomah N. Automatic gestational age estimation by femur length using integral projection from fetal ultrasonography. *International Seminar on Application for Technology of Information and Communication*. Semarang: IEEE. (2018).
13. Stebbing RV, McManigle JE. *A Boundary Fragment Model for Head Segmentation in Fetal Ultrasound*. Proc. Chall. US Biometric Meas. from Fetal Ultrasound Images, ISBI (2012).
14. Zhang L, Dudley NJ, Lambrou T, Allinson N, Ye X. Automatic image quality assessment and measurement of fetal head in two-dimensional ultrasound image. *J Med Imaging*. (2017) 4:024001. doi: 10.1117/1.JMI.4.2.024001
15. Hough PVC. *Method and Means for Recognizing Complex Patterns*. U.S. Patent. Washington, DC: Patent and Trademark Office. (1962).
16. Prasad DK, Leung MKH, Quek C. ElliFit: an unconstrained, non-iterative, least squares based geometric ellipse fitting method. *Pattern Recognit.* (2013) 46:1449–65. doi: 10.1016/j.patcog.2012.11.007
17. Altaf F, Islam SM, Akhtar N, Janjua NK. Going deep in medical image analysis: concepts, methods, challenges and future directions. *IEEE Access*. (2019) 7:99540–72. doi: 10.1109/ACCESS.2019.2929365
18. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (2014) 580–587. doi: 10.1109/CVPR.2014.81
19. Sinclair M, Baumgartner CF, Matthew J, Bai WJ, Martinez JC, Li YW, et al. Human-level Performance On Automatic Head Biometrics In Fetal Ultrasound Using Fully Convolutional Neural Networks. *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. (2018) 714–717. doi: 10.1109/EMBC.2018.8512278
20. Wu L, Xin Y, Li S, Wang T, Heng PA, Ni D. Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation. In: *International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, Melbourne. (2017).
21. Sobhaninia Z, Emami A, Karimi N, Samavi S. Localization of fetal head in ultrasound images by multiscale view and deep neural networks. In: *International Computer Conference, Computer Society of Iran (CSICC)*. IEEE, Tehran. (2020).
22. Skeika EL, Da Luz MR, Fernandes BJT, Siqueira HV, De Andrade MLSC. Convolutional neural network to detect and measure fetal skull circumference in ultrasound imaging. *IEEE Access*. (2020) 8:191519–29. doi: 10.1109/ACCESS.2020.3032376
23. Aji CP, Fatoni MH, Sardjono TA. Automatic measurement of fetal head circumference from 2-dimensional ultrasound. In: *International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*. IEEE, Surabaya. (2019).
24. Sobhaninia Z, Rafiei S, Emami A, Karimi N, Najarian K, Samavi S, et al. Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning. In: *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Berlin. (2019).
25. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV. (2016).
26. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al. Ssd: single shot multibox detector. In: *European Conference on Computer Vision*. Springer, Cham. (2016).
27. Ming Q, Zhou Z, Miao L, Zhang H, Li L. *Dynamic Anchor Learning for Arbitrary-Oriented Object Detection*. New York: AAAI. (2020).
28. Zhou X, Wang D, Krähenbühl P. Objects as Points. *arXiv*. (2019) arXiv:1904.07850.
29. Srinivas A, Lin TY, Parmar N, Shlens J, Abbeel P, Vaswani A. Bottleneck transformers for visual recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021).
30. Yang TY, Huang YH, Lin YY, Hsiu PC, Chuang YY. SSR-Net: a compact soft stagewise regression network for age estimation. *IJCAI*. (2018) 5:7. doi: 10.24963/ijcai.2018/150
31. Yang X, Yang X, Yang J, Ming Q, Wang W, Tian Q, et al. *Learning High-Precision Bounding Box for Rotated Object Detection Via Kullback-Leibler Divergence*. *Advances in Neural Information Processing Systems*. NeurIPS. (2021).
32. Ma'Sum MA, Jatmiko W, Tawakal MI, Al Afif F. Automatic fetal organs detection and approximation in ultrasound image using boosting classifier and hough transform. In: *International Conference on Advanced Computer Science and Information System*. IEEE (2014).
33. Ronneberger O, Fischer P, Brox T. U-Net: convolutional net-works for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich. (2015).
34. Xiao X, Lian S, Luo Z, Li S. Weighted res-unet for high-quality retina vessel segmentation. In: *International Conference on Information Technology in Medicine and Education (ITME)*. IEEE, Hangzhou. (2018).
35. Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, et al. CE-Net: context encoder network for 2D medical image segmentation. *IEEE Trans Med Imag*. (2019) 39:2281–92. doi: 10.1109/TMI.2019.2903562
36. Qiao D, Zulkernine F. Dilated squeeze-and-excitation U-net for fetal ultrasound image segmentation. In: *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Chiang Mai. (2020).

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yang, Liao, Yang, Guo, Zhang, Yang, Guo, Yin, Liu and Kang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# PET and CT Image Fusion of Lung Cancer With Siamese Pyramid Fusion Network

Ning Xiao<sup>1</sup>, Wanting Yang<sup>1</sup>, Yan Qiang<sup>1\*</sup>, Juanjuan Zhao<sup>1</sup>, Rui Hao<sup>2</sup>, Jianhong Lian<sup>3</sup> and Shuo Li<sup>4</sup>

<sup>1</sup> College of Information and Computer, Taiyuan University of Technology, Taiyuan, China, <sup>2</sup> School of Information Management, Shanxi University of Finance and Economics, Taiyuan, China, <sup>3</sup> Thoracic Surgery, Shanxi Cancer Hospital, Taiyuan, China, <sup>4</sup> Department of Medical Imaging and Medical Biophysics, Western University, London, ON, Canada

## OPEN ACCESS

### Edited by:

Alice Chen,  
National Cancer Institute (NCI),  
United States

### Reviewed by:

Jun Shi,  
Shanghai University, China  
Yan Wang,  
Sichuan University, China

### \*Correspondence:

Yan Qiang  
qiangyan@tyut.edu.cn

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

**Received:** 10 October 2021

**Accepted:** 10 March 2022

**Published:** 31 March 2022

### Citation:

Xiao N, Yang W, Qiang Y, Zhao J,  
Hao R, Lian J and Li S (2022) PET and  
CT Image Fusion of Lung Cancer With  
Siamese Pyramid Fusion Network.  
Front. Med. 9:792390.  
doi: 10.3389/fmed.2022.792390

**Background:** The fusion of PET metabolic images and CT anatomical images can simultaneously display the metabolic activity and anatomical position, which plays an indispensable role in the staging diagnosis and accurate positioning of lung cancer.

**Methods:** In order to improve the information of PET-CT fusion image, this article proposes a PET-CT fusion method via Siamese Pyramid Fusion Network (SPFN). In this method, feature pyramid transformation is introduced to the siamese convolution neural network to extract multi-scale information of the image. In the design of the objective function, this article considers the nature of image fusion problem, utilizes the image structure similarity as the objective function and introduces L1 regularization to improve the quality of the image.

**Results:** The effectiveness of the proposed method is verified by more than 700 pairs of PET-CT images and elaborate experimental design. The visual fidelity after fusion reaches 0.350, the information entropy reaches 0.076.

**Conclusion:** The quantitative and qualitative results proved that the proposed PET-CT fusion method has some advantages. In addition, the results show that PET-CT fusion image can improve the ability of staging diagnosis compared with single modal image.

**Keywords:** PET-CT fusion, image quality, siamese neural network, pyramid transform, structural similarity

## 1. INTRODUCTION

Medical imaging is a technique and process for obtaining images of a certain part of the human body in a non-invasive manner (1–3). With the continuous development of computer imaging technology, medical imaging has derived multi-modal forms. Common medical images can be divided into Computed Tomography (CT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), UltraSound, and so on. Medical images of different modalities can reflect disease information from different angles. The correlation and complementarity of image information from different imaging modality can be used to merge different modalities of medical image. Effective fusion can provide doctors with richer diagnosis and treatment information. For lung cancer, the common imaging screening procedures mainly include CT and PET (4, 5). Among them, CT images have the characteristics of short scanning time and clear images, which can provide clear human bone tissue anatomy and lesion images, and are widely used for screening

diseases such as chest and abdomen; PET uses short-lived radionuclide metabolites (common fluorodeoxyglucose, FDG) to reflect the metabolic activities in the human body to perform imaging. PET has the advantages of high sensitivity and high specificity. The PET-CT fusion image can simultaneously reflect the pathophysiological changes and morphological structure of the lesion. It has important clinical value for the staging of non-small cell lung cancer, the judgment of the recurrence and metastasis of lung cancer, and the formulation of radiation treatment plans.

Medical image fusion technology is an important application area in the field of information fusion technology. The fused image has more information than the source image, which is suitable for human eyes to distinguish, and can further enrich some details of the image, which can provide more practical information for clinical diagnosis. At present, researchers have proposed many PET-CT fusion methods. The mainstream fusion methods mainly include methods based on multi-scale decomposition and methods based on wavelet transformation (6, 7). For example, a multi-modal medical image fusion method in the non-subsampled wavelet transform domain was proposed in (8). This method first performs non-subsampled wavelet transform on the source image, and then uses Pulse Coupled Neural Network(PCNN) and Max selection fusion rule to analyze high frequency sub-band and low frequency sub-band fusion, this strategy simultaneously solves the two problems of energy preservation and detail extraction in image fusion. An image fusion based on guided filtering (9) used average filtering to obtain two-scale base and detail layers, and determined the two-scale weight of the fusion result according to the saliency map of the image. These fusion methods follow certain fusion rules to process the images point by point when fusing images. Therefore, the quality of the fused image is largely affected by the fusion rules, and the noise resistance effect is not high. During multi-scale decomposition and transformation, some original brightness information of the image is often lost.

In recent years, medical image fusion technology based on convolutional neural network (10–12) has been developed rapidly. With the help of symmetric network structure or co-learning method, it automatically learns the direct mapping between the original image and the fused image, which is different from the multi-scale decomposition method and the wavelet transform method, the convolutional neural network learns to extract features from a large number of images autonomously, and can obtain low-level features and high-level semantic features at the same time. In order to obtain a clearer fusion image, the multi-focus convolutional neural network (11) used a high-quality image patch and its blurred patch as input to automatically learn the coding features of the source image and the target image; A novel image fusion framework, IFCNN (10), was designed, which uses convolutional layers to extract salient image features from input images, selects appropriate fusion rules to fuse the extracted features, and finally obtains the fused image through convolutional layer reconstruction. A multi-layer cascaded fusion network was proposed in (12), this end-to-end deep convolutional neural network can automatically perform feature extraction, feature fusion, and image reconstruction on

the fused medical images, and use fast deconvolution to reduce the number of features. The main feature of the convolutional neural network is the invariance of feature translation. It does not require complex fusion rules to obtain high-quality fused images and can also retain the structural information in the original image to the greatest extent. However, the convolutional neural network only takes a single-dimensional picture is used as an input, which lacks the diversity of image scales, and the standard convolution still has the problem of unknowable content (13, 14).

The feature pyramid (15, 16) is a method that can efficiently extract the features of each dimension in the picture. The method of image transformation is used to generate images of various scales. The convolutional neural network model is used to express the characteristics of images of different scales from the low to the top, so as to generate feature maps with stronger content expression ability. In order to consider the multi-scale information of the image, this article introduces the feature pyramid transformation in the traditional convolution, which enhances the content of the original convolution feature on the scale. Specifically, this article proposes a Siamese Pyramid Fusion Network, which implements end-to-end image feature-level fusion by constructing a siamese structure and multi-scale feature modules. The network fusion process includes a multi-scale feature extraction stage and bimodal cross-correlation. The fusion stage and the image reconstruction stage consist of three parts. In particular, the contribution of this article can be summarized in the following three aspects:

- We use a siamese structure composed of a PET encoder and a CT encoder. The two encoders have the same structure and share parameters. This structure can extract the two modal image features of PET and CT separately.
- When extracting image feature maps, we design a multi-scale convolution structure. This structure can not only extract the translation invariance of features in the original image, but also increase the multi-scale features of image.
- When designing the objective function, we propose a novel objective function, which takes the structural similarity between the original image and the fusion image as the backbone, and adds the L1 norm as a regularization term to reduce noise interference.

The rest of the article is organized as follows. The Section 2 describes the proposed algorithm in detail. The Section 3 describes the experiment and results. The discussion is in Section 4 and the conclusion follows in Section 5.

## 2. METHOD

This article uses  $X_{ct}$  and  $X_{pet}$  to represent the original CT image and PET image,  $X_{pc}$  represents the fused PET-CT image,  $X_{pc} = F(X_{ct}, X_{pet})$ ,  $F()$  represents the fusion method. In order to be able to find this fusion method, this article designs the Siamese Pyramid Fusion Network (SPFN). When performing image fusion, SPFN is mainly divided into feature extraction part, fusion part, and image reconstruction part. The feature extraction part is mainly composed of two encoders based on

the Convolutional Layer Coupling Module (CLCM) to extract features from original image. The fusion part is composed of cross-correlation layer to fuse the characteristics of the two modality images. And the image reconstruction part is a decoder composed of three de-convolutional layers, which decodes the fused features to reconstruct the fused PET-CT image. The specific process is shown in the **Figure 1**.

## 2.1. Architecture Design

The main framework of the siamese autoencoder architecture proposed in this article is mainly inspired by the siamese network (17). The siamese network initially takes two samples as input and outputs its embedding high-dimensional space representation to compare the similarity of the two samples. Based on this siamese structure, this article designs an improved siamese autoencoder to fuse the features of two samples. The siamese autoencoder includes two encoders, a fusion layer and a reconstruction decoder.

The two encoders have the same structure, sharing parameters, and respectively accept PET and CT as input. In the encoding stage, the two inputs of the siamese encoder are the same size  $512 \times 512$ , and the images of the two modalities will be input to the two encoders with different parameters at the same time. The encoder designed in this article is mainly composed of CLCM, which specifically includes three parts: convolutional layer, channel coupling module, and spatial pyramid coupling module.

After the features of PET and CT extracted by the encoder, the features of the two modality images will be fused through the cross-correlation layer, which is implemented by inter-correlation operations. It will be described in detail in Section 2.3. Finally, the fused features are fed into the decoder and reconstructed to obtain the final fused image. The decoder is mainly composed of three deconvolutional layers, which aims to reconstruct the final input according to the features in the latent space. The decoder is mainly composed of three de-convolutional layers, which aims to reconstruct the final input according to the features in the latent space.

## 2.2. Convolutional Layer Coupling Module

In the neural network, data will be transmitted along the designed channel. When a module in the network changes, the data will change accordingly, and it can also affect other channels to change accordingly. Data coupling makes the network model more cohesive, while designing different types of modules can allow the model to dynamically focus on specific types of features in the image, which is more conducive to the PET-CT fusion image of tumor staging.

In this article, we designed the Convolutional Layer Coupling Module. This module mainly includes a convolutional layer, a channel coupling module and a spatial pyramid coupling module. The initial convolutional layer is to extract the feature map from original images. The traditional convolutional encoder is composed of stacked convolutional layers (18, 19). The change in the weight of each layer in this stacked convolutional structure will cause the subsequent output to change accordingly. For fusion tasks, the features extracted by the convolutional layer

are different from the heavy classification or segmentation of traditional image processing tasks. There is no need to expand the receptive field to extract features that can distinguish as many categories as possible, but to preserve the details of the image as much as possible. Under the conditions, it is used to characterize the key information in the two modality of PET and CT.

Therefore, after the convolution layer in the encoder, this article does not directly use the stacked convolution operation, but uses the channel coupling module and the spatial pyramid coupling module to extract the multi-scale characterization features of the image. The role of the channel coupling module is to assign different weights for each channel, so that the network can focus on important features and obtain task relevance features. The function of the spatial pyramid coupling module is to transform the spatial information in the original image to another space and generate multi-scale key information (20).

Specifically, this article first uses a set of  $3 \times 3$  size convolution kernels to extract features from the original image. The first layer of convolutional layer can extract some low-level features in the image (21). In the image fusion task, there is no need for the network to process a larger receptive field. In addition, deeper abstract features representing the unknowability of the fusion task will increase the computational complexity of the network. Therefore, after the convolution operation, the stacked convolution and pooling operations are not performed, but the channel coupling module and the spatial pyramid coupling module are used. The two coupling modules perform data coupling. The feature map generated by the convolutional layer determines the input in the channel coupling module. At the same time, the channel coupling module also directly controls the input of the spatial pyramid coupling module.

Given a feature map  $F$  generated by the convolutional layer, after the coupling module, the fusion task relevance feature map and the multi-scale feature map will be sequentially obtained, and the two types of maps obtained are superimposed with the original image to obtain the final The characteristics of the image. The calculation process of the two types of data coupling is as follows:

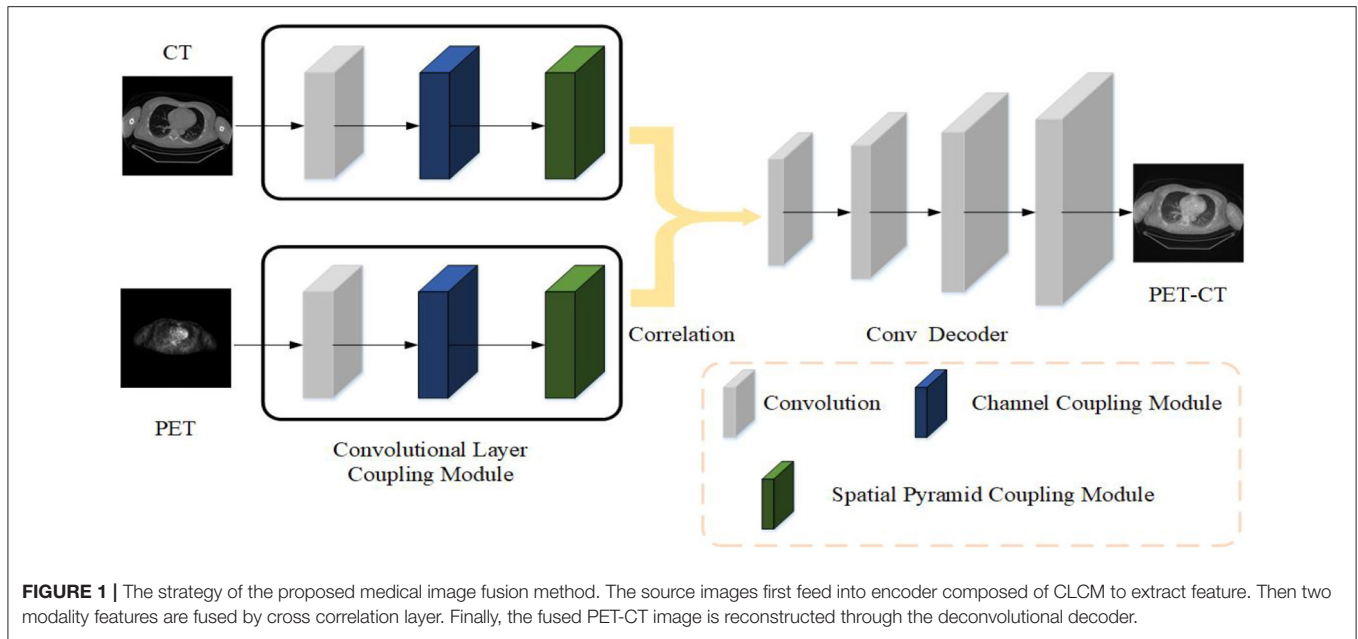
$$F' = Cc(F) \quad (1)$$

$$F'' = Cs(F') \quad (2)$$

$Cc$  and  $Cs$  represent channel coupling operation and spatial pyramid coupling operation, respectively. Under the effect of cross-layer connection, the difficulty of training model parameters is greatly reduced, making it easier to train a coding model with good effect.  $F''$  is the final output. Next, we will introduce the details of the two coupling module in detail.

### 2.2.1. Channel Coupling Module

Channel coupling module mainly use the relationship between feature maps to generate channel coupling features. Each channel in the feature map is regarded as a feature extractor, the features extracted by each feature extractor are different, and the focus of the channel coupling module is to find the most meaningful



features of the input image in these features (22). In order to effectively extract the channel coupling features of the image, this article designs the structure as shown in the **Figure 2**. The channel coupling module will pass the input feature map through global max pooling and global average pooling based on width and height, respectively, and then feed into multi-layer perceptron (MLP). The outputs from MLP are performed element-wise summation operation and sigmoid operation. After a series of operations, the final channel feature maps are obtained. The channel feature maps are used as the input of the spatial coupling module. The calculation process for the features of the channel coupling module is as follows:

$$F^c = Cc(F) = \text{Sig}(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (3)$$

$\text{Sig}()$  indicates that the sigmoid operation is performed on the two sets of results

### 2.2.2. Spatial Pyramid Coupling Module

Spatial pyramid coupling module mainly uses the spatial relationship between pixels in the image to generate spatial feature maps. Since the convolutional layer only keeps the translation of an object in the image invariance, and the scale transformation of the object cannot be processed, so when extracting the spatial feature map, this article adds the spatial pyramid transformation (23) to extract the multi-scale spatial feature map. The spatial pyramid coupling module is equivalent to performing feature convolution from the bottom up on the feature map of the image, and then fusing feature maps of multiple scales.

The spatial pyramid coupling module uses the feature map obtained by the channel coupling module as the input feature map. First, the convolution operation in the spatial pyramid coupling module detaches the input features then uses spatial

pyramid pooling to perform multi-scale transformation of the abstract features to obtain four features of different sizes. And the different scales feature maps are obtained through multi-scale convolution after spatial pyramid coupling module. Finally, the four convolution feature results are performed through the channel concatenation operation to obtain the final spatial pyramid feature map, as shown in **Figure 3**. The calculation process of the spatial pyramid coupling module is as follows:

$$F^s = Cs(F) = \text{Concat}(\text{Conv}(\text{SPP}(\text{Conv}(F^c)))) \quad (4)$$

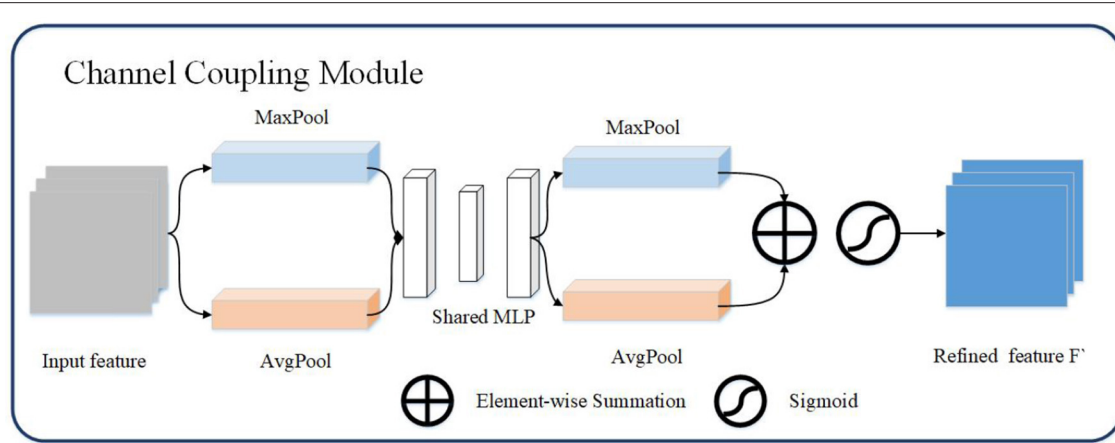
$\text{SPP}()$  represents for spatial pyramid pooling operation. Specifically, the pyramid pooling operation is to divide the original input features using four different pooling scales (16, 4, 2, 1).

### 2.3. Cross Correlation Layer

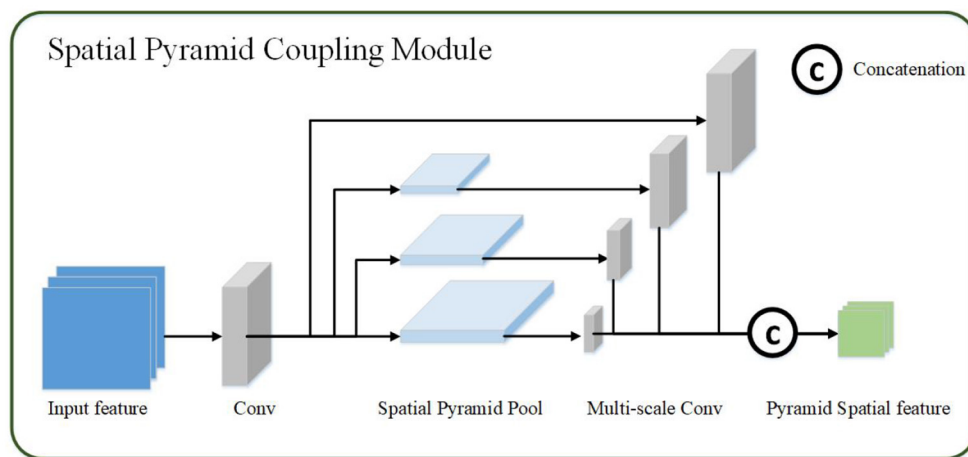
For the fusion of PET features and CT features, the simplest fusion method is to linearly add them. However, this operation ignores the association between adjacent pixels in the area, and lacks the expression of the overall information of the image. In order to better enhance the display ability of the fused image features without losing the original information in PET and CT, this article uses the cross correlation layer (24) to fuse the features extracted by the encoder. Given the image features  $F_{ct}$  and  $F_{pet}$ , their fusion results will be calculated according to the cross-correlation layer. The specific formula is as follows:

$$F_{pc} = \text{Cor}(F_{ct}, F_{pet}) \quad (5)$$

$\text{Cor}()$  represents for cross-correlation layer, and the cross-correlation layer is a special convolutional layer that uses cross-correlation operation. Different from the ordinary convolution function, the cross-correlation function is an operation between



**FIGURE 2 |** The channel coupling module. The channel coupling module utilizes two different pooling operation and feed results to multi-layer perceptron. The output of multi-layer perceptron continue to forward to element-wise summation and sigmoid operation.



**FIGURE 3 |** The spatial pyramid coupling module. The spatial pyramid coupling module utilizes spatial pyramid pooling to get multi-scale feature maps and concatenate them.

two data. When the data is transferred in the network, the weight to be trained is not needed. The calculation equation is as follows

$$Cor(x_1, x_2) = \sum_m \sum_n x_1(m, n) * x_2(m + o, n + o) \quad (6)$$

$x_1, x_2$  represent the feature patch on the two feature maps  $f_{ct}$  and  $f_{pet}$  respectively,  $m, n$  are the size of patch,  $o$  represents the patch stride, this article set  $o$  as 3. Each patch on the feature map  $f_{ct}$  must be cross-correlated with all the patch on the other feature map  $f_{pet}$ . In order to obtain the output of the same size as the original image, the padding pattern selects the “SAME” when performing cross-correlation in this article (25, 26).

## 2.4. Loss Function

For the loss function of the network, this article selects the structural similarity loss (6, 27) as loss function. After obtaining

the fused PET-CT, it will first calculate the structural similarity with the original PET and CT respectively, as in the formula:

$$SSIM(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (7)$$

$x$  and  $y$  are the fused image and the original image, respectively.  $\mu$  represents image mean,  $\sigma_x, \sigma_y$  represent image variance,  $\sigma_{xy}$  represents co-variance of images.  $C_1$  and  $C_2$  are constants (avoid the denominator being 0), the calculation formula is  $C_1 = k_1L^2, C_2 = k_2L^2$ , where  $L$  is the grayscale change of the image. Since this article is performing feature extraction, the image is normalized, so  $L$  is 1.  $K_1$  and  $K_2$  are two constants, the default value is 0.01 and 0.03. The loss of the network can be defined as:

$$L_{ssim}(x, y) = 1 - SSIM(x, y) \quad (8)$$

When using SSIM, there will be a problem with edge noise defects. Therefore, this article introduces an L1 regularization term (28) in the original loss function. The L1 regularization term  $\|\omega\|$  can be used to estimate the difference between the target value  $x$  and the estimated value  $y$ , which can effectively reduce the noise in the image, has a certain degree of robustness, and can also prevent the neural network from overfitting during training. Therefore, the loss of the final fusion network is as follows:

$$L_f(x_{ct}, x_{pet}; y_{pc}|\omega) = 2 - \frac{2\mu_{ct}\mu_{pc} + C_1}{\mu_{ct}^2 + \mu_{pc}^2 + C_1} \cdot \frac{2\sigma_{ct,pc} + C_2}{\sigma_{ct}^2 + \sigma_{pc}^2 + C_2} - \frac{2\mu_{pet}\mu_{pc} + C_1}{\mu_{pet}^2 + \mu_{pc}^2 + C_1} \cdot \frac{2\sigma_{pet,pc} + C_2}{\sigma_{pet}^2 + \sigma_{pc}^2 + C_2} + \|\omega\| \quad (9)$$

### 3. EXPERIMENT AND RESULT

#### 3.1. Dataset

The PET-CT images used in this article are from Soft-tissue-Sarcoma (29). This dataset includes clinical images of 21 patients with lung tumors. All patients underwent FDG-PET and CT screening from November 2004 to November 2011 by McGill University Health Centre (MUHC). The median of intravenous FDG was 420 MBq. The dataset also includes patient information, histopathological type, tumor grade, follow-up information (metastasis, survival rate). In this article, only PET and CT corresponding to the patient's organs including the lungs are used as the experimental dataset, and a total of 840 CTs and corresponding FDG-PET are extracted from the dataset. The PET and CT images of each patient are registered by (30). We adopted a leave-one-out cross-validation strategy to test the effectiveness of the method in this article and divided the data set into a training set and a validation set according to a ratio of 4:1 to 840 pairs of images. In the training verification, 672 imaging data were used as training, and the remaining 168 imaging data were verified, repeated five times. All patient information are de-identified.

#### 3.2. Implementation Details

In the used dataset, the resolution of CT is  $512 \times 512$  pixels, and the resolution of PET is  $128 \times 128$  pixels. Before fusion, this article uses image zoom to up-sampling PET to obtain an image with the same resolution as CT. In addition, the human body's absorption of X-rays recorded in CT images, unit is Hounsfield Unit (HU), and the human body's absorption of isotopes recorded in PET images. Therefore, this article uses the min-max standardization method to normalize the images in the extracted dataset. We implement our fusion algorithm using Tensorflow 1.12.0 on a machine running Ubuntu 16.04 with CUDA 8.0 and CuDNN 5.1. Training is performed on 32 GB NVIDIA GTX 1080 Ti. The parameter initialization in the fusion algorithm uses the Xavier method (31); for the optimization algorithm. This article uses the adaptive optimization algorithm AdaGrad algorithm (32) to optimize the parameters.

#### 3.3. Evaluation Metric

In order to quantitatively evaluate the performance of the proposed fusion algorithm, there are reference image evaluation

indicators and no reference evaluation indicators. This article uses the following seven indicators for evaluation: average value, standard deviation, average gradient, entropy, root mean square error, normalized mutual information, visual fidelity.

The average value  $\bar{x}$  of the image represents the average level of the overall pixels of the image and reflects the brightness of the image. Assuming that the size of the image  $I$  is  $m \times n$ , the average value of the image is averaged for each pixel,

$$\bar{x} = \frac{\sum_{i=1}^m \sum_{j=1}^n x_{ij}}{m \times n} \quad (10)$$

The standard deviation  $\sigma$  of the image represents the degree of dispersion between the pixel value and the average of the image, and reflects the contrast of the image,

$$\sigma = \sqrt{\frac{\sum_i \sum_j (x_{ij} - \bar{x})^2}{m \times n}} \quad (11)$$

The average gradient  $G$  of the image reflects the clarity and texture changes of the image,

$$G = \frac{1}{m \times n} \sum_i \sum_j \sqrt{\frac{(\frac{\partial I}{\partial x})^2 + (\frac{\partial I}{\partial y})^2}{2}} \quad (12)$$

$\frac{\partial I}{\partial x}$  represents the gradient in the horizontal direction,  $\frac{\partial I}{\partial y}$  represents the gradient in the vertical direction.

The entropy  $Ent$  of the image is expressed as the average number of bits in the grayscale set of the image, reflecting the spatial characteristics of the grayscale distribution of the image.

$$Ent = - \sum p_x \ln p_x \quad (13)$$

$p_x$  represents the proportion of pixels in the image that have a grayscale value of  $x$ .

The root mean square error  $RMSE$  of the image is used to measure the difference between the two images. The mean square error is to find the sum of the square of the error for each pixel and find the mean and then square off:

$$RMSE = \sqrt{\frac{1}{m \times n} \sum_i (y_i - \hat{y}_i)^2} \quad (14)$$

$y_i$  and  $\hat{y}_i$  represent the original image and the fused image, respectively. We calculate the RMSE between the fused image and PET CT, respectively, and take the average value as the final result.

The normalized mutual information  $NMI$  (33) of the image reflects the information correlation between the two images,

$$NMI = 2[2 + \frac{Ent(I_1, I_F)}{Ent(I_1) + Ent(I_F)} + \frac{Ent(I_2, I_F)}{Ent(I_2) + Ent(I_F)}] \quad (15)$$

where  $Ent(I_k, I_F)$  is the joint entropy between the input image  $I_k$  and  $I_F$ .

The visual fidelity  $VIF$  (34) of the image was originally an evaluation index using the statistical characteristics of natural scenes. The image information will follow the distortion process, which will result in poor visual quality. Therefore, the image quality can be calculated by calculating the fidelity of the image. The visual fidelity of the image can be defined as:

$$VIF = \frac{\sum_i I(c_i, f_i)}{\sum_i I(c_i, e_i)} \quad (16)$$

Among them,  $I(c, f)$  and  $I(c, e)$  represent the information that the human eye can extract from the original image and the distorted image, and  $c, f, e$  represent image blocks of different scales in the image,  $i$  represents the index of the image block in the image.

In order to verify the quality of the fused images for clinical diagnosis, we also invited four radiologists from the partner hospital to visually evaluate the fused images. Four radiologists evaluated the fusion image from four indicators: Noise Suppression (NS), Artifact Reduction (AR), Detail Information (DI), and Comprehensive Quality (CQ). The evaluation score ranges from 1 to 5 points, with 1 point representing “bad” and 5 points representing “excellent.” Each doctor conducts an independent evaluation.

### 3.4. Comparative Experiment

In order to verify the effectiveness of the proposed method, this article compared five representative methods: Multilayer Concatenation Fusion Network (MCFNET) (12), Guided Filtering (GF) (9), Adaptive Decomposition (AD) (5), Parameter Adaptive Pulse Coupled Neural Network (PAPCNN) (8), Image Fusion Convolutional Neural Network (IFCNN) (10). The implementation of the above methods has corresponding authors to provide source code, all parameters are the default parameters set by the author. This article verifies from qualitative results, quantitative comparisons, and details of fusion results.

This article selects two cases of early lung cancer patients (A and B) from the test dataset for qualitative display, as shown in **Figures 4, 6**. It can be seen from **Figures 4A, 5A** that CT mainly provides detailed edges, contours and other structures of lung lesions and tissues, while PET reflects the accurate location of lung lesions in **Figures 4B, 5B**. The fusion images of different methods have obtained satisfactory results on the texture and edges, which proves the effectiveness of the fusion method for PET-CT. From the perspective of each fusion result, the fusion image of AD contains some noise in **Figures 4C, 5C**. In **Figures 4D, 5D**, the fusion result of GF has low contrast. In **Figures 4E, 5E**, there are slight artifacts near the lung wall. In **Figures 4F, 5F**, the brightness of the fusion result of MCFNET and IFCNN is higher than other results. The contrast and sharpness of the fusion result of the proposed method are relatively high.

For medical images, the brightness of the image is not the main indicator of quality considerations, but mainly the presentation of the lesion area in the image and the display of details. Therefore, this article discusses two details in the original image, as shown in the area marked in **Figures 6A,B, 7A,B**. The details are shown in **Figures 5, 7**. We can clearly observe

two highlighted areas from **Figure 5B**, which means that the nodules in this area have rapid metabolism and the nodules are extremely malignant. In the **Figure 7**, there is an obvious adhesive nodule, however it is not highlighted in PET. This nodule has a slow response rate and low malignancy, which needs follow-up observation. Among these fusion detail results, the fusion results of the proposed algorithm show more excellent results in terms of the contrast between the lesion area and the background and the details of the lesion area.

Next, in order to further verify the effectiveness of the proposed method, this article evaluates the fusion results through different objective indicators and the physician's subjective scores. The specific results are shown in **Tables 1, 2**.

**Table 1** lists the objective results of the entire test dataset for evaluating fusion images of different fusion algorithms. From the results, it can be observed that the standard deviation, average gradient, image entropy, normalized mutual information, and visual fidelity of the original image mentioned in this article have a high level, which proves that the proposed method has a high level of image clarity, contrast and information. The mean value of IFCNN is higher than the method in this article, which reflects that the image fused by IFCNN is better than the method in this article in terms of brightness index, and it also corresponds to the qualitative result.

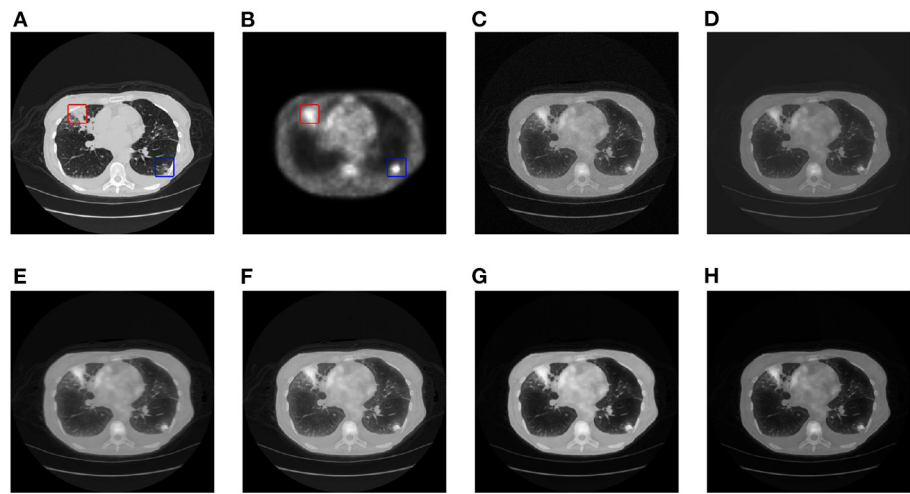
After displaying the objective evaluation label of image quality, this article also lists the doctor's subjective evaluation scores ( $mean \pm std$ ) of the results of different fusion algorithms, as shown in **Table 2**.

The evaluation of image quality is one of the main evaluation indicators of the fusion algorithm. In addition, the computational complexity is also one of the factors considered by the fusion algorithm. Therefore, this article also compares the average time for a pair of PET-CT fusion of several algorithms, as shown in **Table 3**. All fusion methods are run on the same device, and the device parameters are shown in Section 4.2.

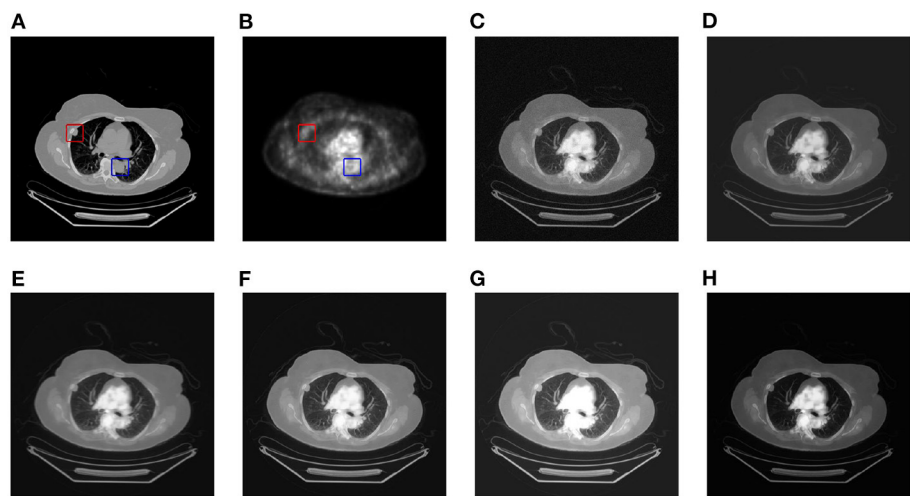
Among these methods, the fusion speed of the method proposed in this article is slightly slower than that of the GF fusion method at 2.12 s, but it is within the acceptable range, which also indicates that the proposed method has considerable potential in clinical applications.

### 3.5. Results for Staging Diagnosis

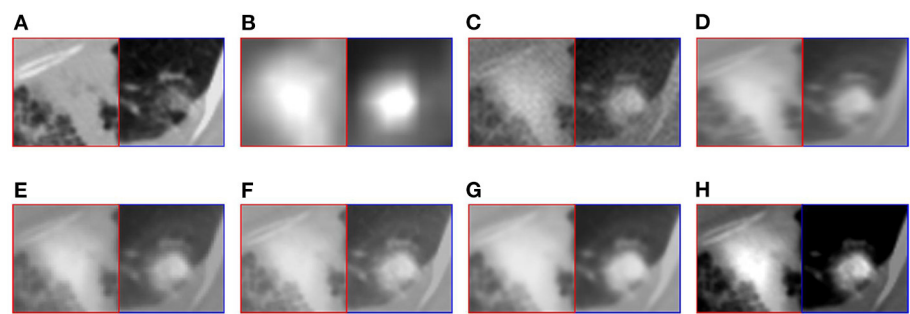
One of the main functions of PET-CT is to determine the staging of tumors in patients with lung cancer. In order to evaluate the fusion of PET-CT for the diagnosis of lung cancer staging, this section uses some simple classification methods to evaluate the collected image data and the corresponding staging information. Specifically, different classification methods are used to train and test PET, CT, PET-CT, including Support Vector Classifier (SVC), Multilayer Perceptron (MLP), K-Nearest Neighbor (KNN), Random Forest (RF), and Naive Bayes Classifier (NB), the comparative performance is shown in the **Table 4**. The kernel function of SVC adopts radial basis function that penalty slack variable is 100 and kernel coefficient is 0.5. The maximum iteration number of MLP is set to 300. The hyper-parameter, neighbor numbers, of KNN is 4. The number of the RF is 100, maximum depth is 2 and the function to measure is gini function.



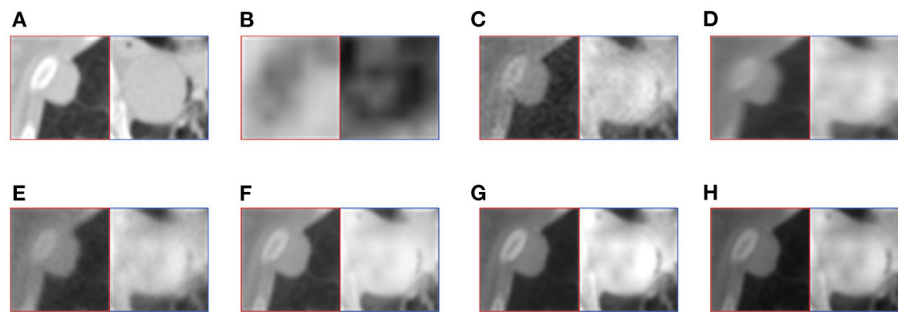
**FIGURE 4 |** The qualitative comparison results of patient A. **(A)** CT; **(B)** PET; **(C)** AD; **(D)** GF; **(E)** PAPCNN; **(F)** MCFNET; **(G)** IFCNN; **(H)** OURS.



**FIGURE 5 |** The qualitative comparison results of patient B. **(A)** CT; **(B)** PET; **(C)** AD; **(D)** GF; **(E)** PAPCNN; **(F)** MCFNET; **(G)** IFCNN; **(H)** OURS.



**FIGURE 6 |** The detail of fusion results of patient A. **(A)** CT; **(B)** PET; **(C)** AD; **(D)** GF; **(E)** PAPCNN; **(F)** MCFNET; **(G)** IFCNN; **(H)** OURS.



**FIGURE 7 |** The detail of fusion results of patient B. (A) CT; (B) PET; (C) AD; (D) GF; (E) PAPCNN; (F) MCFNET; (G) IFCNN; (H) OURS.

**TABLE 1 |** Evaluation metric of different fusion algorithm results.

	Mean	Std	AG	Ent	RMSE	NMI	VIF
AD	0.102	0.026	0.063	0.051	0.030	3.25	0.321
GF	0.134	0.020	0.072	0.049	0.028	3.16	0.075
PAPCNN	0.204	0.021	0.072	0.065	0.030	3.27	0.303
MCFNET	0.260	0.045	0.091	0.062	0.013	3.25	0.292
IFCNN	0.273	0.065	0.082	0.072	0.019	3.21	0.287
OURS	0.157	0.085	0.091	0.076	0.013	3.28	0.350

**TABLE 2 |** Image quality evaluation scores of different algorithm results.

	NS	AR	DI	CQ
AD	2.75 ± 0.43	3.25 ± 0.43	3.75 ± 0.43	3.25 ± 0.43
GF	3.50 ± 0.50	3.00 ± 0.71	3.00 ± 0.71	3.17 ± 0.37
PAPCNN	3.00 ± 0.00	2.50 ± 0.50	3.00 ± 0.00	2.67 ± 0.41
MCFNET	3.75 ± 0.43	3.25 ± 0.43	3.75 ± 0.43	3.58 ± 0.36
IFCNN	4.75 ± 0.43	3.75 ± 0.43	4.75 ± 0.43	4.42 ± 0.28
OURS	4.75 ± 0.43	4.25 ± 0.43	4.50 ± 0.50	4.50 ± 0.37

**TABLE 3 |** Average runtime comparison of different fusion methods.

	AD	GF	PAPCNN	MCFNET	IFCNN	OURS
Mean time (s)	2.73	1.62	3.38	3.36	3.30	2.12
STD	0.04	0.01	0.20	0.14	0.12	0.03

**TABLE 4 |** Classification performance of different modal images for lung cancer staging.

	SVC(%)	MLP(%)	KNN(%)	RF(%)	NB(%)
PET	62.70	56.66	62.78	60.32	61.80
CT	79.37	79.01	74.86	76.17	81.11
PET-CT	82.71	81.49	80.05	82.34	84.01

Naive Bayes classifier uses the multinomial Naive Bayes classifier, the number of sample class is 4.

As shown in **Table 4**, both CT and PET have certain staging diagnostic capabilities for lung cancer. However, the performance in staging diagnosis of lung cancer using PET-CT images is generally higher than other monomodal data. The accuracy of constructing individual classifiers in different sample spaces is different, excluding the easy distinguishing characteristics of PET images. PET-CT can improve generalization ability and stability indeed. This result proves that the PET-CT fusion image has diagnostic performance for the staging of lung cancer, which is helpful to assist doctors in identifying lung tumors.

## 4. DISCUSSION

Taking into account the above comprehensive comparison, it can be concluded that the fusion method proposed in this article has a certain competitive performance on the quality improvement and information retention of PET-CT. Nevertheless, this article still has some shortcomings that need to be studied in the future. Since the purpose of this article is to diagnose lung cancer in stages, the collected image data are all CT and PET images of patients, which are limited by technical problems and lack the use of images such as Angiography (35). By selecting different tracers, it can achieve the best results in disease diagnosis. Angiography can show the tumor vascular characteristics of lung cancer and provide a basis for interventional therapy. If the images of different modalities can be fused, the comprehensive judgment of the tumor is of great significance for the stage diagnosis of the tumor and the formulation of the treatment plan.

In addition, the siamese network proposed in this article requires the same resolution of the input two modal data. However, in clinical practice, the resolution of CT and PET are often different. Therefore, in the preprocessing, this article image zoom PET to expand it to the same resolution as CT. Although PET functional imaging does not display too many structure details compared to CT, image distortion often occurs during the up-sampling process. For the fusion of images with different resolutions, how to avoid distortion caused by image scaling is another new challenge.

The main research of this article is the staging diagnosis of lung cancer through PET-CT after fusion. Although experiments have proved that PET-CT after fusion does have a certain

improvement in the diagnosis of lung cancer, because of the limited space of this article, only some traditional image classification algorithms are used. In the follow-up work, this article will make corresponding research on the staging diagnosis method of lung cancer.

## 5. CONCLUSION

In this article, we propose a novel siamese autoencoder network for the fusion of PET and CT. The CLCM in the proposed siamese autoencoder can extract features from both PET and CT modal images, and while ensuring that the original image information remains feature invariance, it can also increase the multi-scale information of the features. In addition, this article also designs a structural similarity loss function combined with the L1 regularization term as the object of the model solution. We collected 840 pairs of PET-CT images to verify the effectiveness of the proposed fusion method in this article. From the results of quantitative comparison, qualitative comparison and subjective evaluation, the performance of fusion results is relatively outstanding, which proves the effectiveness of the proposed method. In future work, we will conduct research on

the fusion of PET-CT, aiming to propose high accuracy model for staging diagnosis of lung cancer based on PET-CT.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: <https://wiki.cancerimagingarchive.net/display/Public/Soft-tissue-Sarcoma>.

## AUTHOR CONTRIBUTIONS

NX: conceptualization and methodology. WY: visualization. YQ: supervision. JZ and RH: funding acquisition. SL: project administration. JL: data curation and formal analysis. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by National Natural Science Foundation of China (Grant number 61872261) and Natural Science Foundation of Shanxi Province (201901D111319).

## REFERENCES

- Hartman TE, Swensen SJ. CT screening for lung cancer. *Iaslc Thorac Oncol.* (2018) 40:52–8. doi: 10.1016/B978-0-323-52357-8.00007-X
- Decazes P, Hinault P, Veresezan O, Thureau S, Gouel P, Vera P. Trimodality PET/CT/MRI and radiotherapy: a mini-review. *Front Oncol.* (2021) 10:3392. doi: 10.3389/fonc.2020.614008
- Zhou T, Lu H, Hu F, Shi H, Qiu S, Wang H. A new robust adaptive fusion method for double-modality medical image PET/CT. *BioMed Res Int.* (2021) 2021:8824395. doi: 10.1155/2021/8824395
- Luo Y, Wang Y, Zu C, Zhan B, Wu X, Zhou J, et al. 3D transformer-GAN for high-quality PET reconstruction. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Strasbourg: Springer (2021). p. 276–85. doi: 10.1007/978-3-030-87231-1\_27
- Wang J, Li X, Zhang Y, Zhang X. Adaptive decomposition method for multi-modal medical image fusion. *IET Image Process.* (2018) 12:1403–12. doi: 10.1049/iet-ipr.2017.1067
- Kumar N, Hoffmann N, Oelschlägel M, Koch E, Kirsch M, Gumhold S. Structural similarity based anatomical and functional brain imaging fusion. In: *Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy*. Shenzhen: Springer (2019). p. 121–9. doi: 10.1007/978-3-030-33226-6\_14
- Gao R, Du L, Yuen KF. Robust empirical wavelet fuzzy cognitive map for time series forecasting. *Eng Appl Artif Intell.* (2020) 96:103978. doi: 10.1016/j.engappai.2020.103978
- Yin M, Liu X, Liu Y, Chen X. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampling shearlet transform domain. *IEEE Trans Instrum Measure.* (2018) 68:49–64. doi: 10.1109/TIM.2018.2838778
- Li S, Kang X, Hu J. Image fusion with guided filtering. *IEEE Trans Image Process.* (2013) 22:2864–75. doi: 10.1109/TIP.2013.2244222
- Zhang Y, Liu Y, Sun P, Yan H, Zhao X, Zhang L. IFCNN: a general image fusion framework based on convolutional neural network. *Inform Fus.* (2020) 54:99–118. doi: 10.1016/j.inffus.2019.07.011
- Liu Y, Chen X, Peng H, Wang Z. Multi-focus image fusion with a deep convolutional neural network. *Inform Fus.* (2017) 36:191–207. doi: 10.1016/j.inffus.2016.12.001
- Liang X, Hu P, Zhang L, Sun J, Yin G. MCFNet: multi-layer concatenation fusion network for medical images fusion. *IEEE Sensors J.* (2019) 19:7107–19. doi: 10.1109/JSEN.2019.2913281
- Yu W, Sun X, Yang K, Rui Y, Yao H. Hierarchical semantic image matching using CNN feature pyramid. *Comput Vis Image Understand.* (2018) 169:40–51. doi: 10.1016/j.cviu.2018.01.001
- Xia KJ, Yin HS, Wang JQ. A novel improved deep convolutional neural network model for medical image fusion. *Clust Comput.* (2019) 22:1515–27. doi: 10.1007/s10586-018-2026-1
- Prakash O, Kumar A, Khare A. Pixel-level image fusion scheme based on steerable pyramid wavelet transform using absolute maximum selection fusion rule. In: *2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*. Chengdu (2014). p. 765–70. doi: 10.1109/ICICT.2014.6781377
- Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Hawaii: IEEE (2017). p. 2117–25. doi: 10.1109/CVPR.2017.106
- Liu X, Tang X, Chen S. Learning a similarity metric discriminatively with application to ancient character recognition. In: *International Conference on Knowledge Science, Engineering and Management*. Tokyo: Springer (2021). p. 614–26. doi: 10.1007/978-3-030-82136-4\_50
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. (2014).
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV: IEEE (2016). p. 770–8. doi: 10.1109/CVPR.2016.90
- Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT: IEEE (2018). p. 7132–41. doi: 10.1109/CVPR.2018.00745
- Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: *European Conference on Computer Vision*. Zurich: Springer (2014). p. 818–33. doi: 10.1007/978-3-319-10590-1\_53
- Komodakis N, Zagoruyko S. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer. In: *ICLR*. Toulon (2017).

23. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell.* (2015) 37:1904–16. doi: 10.1109/TPAMI.2015.2389824
24. Dosovitskiy A, Fischer P, Ilg E, Hausser P, Hazirbas C, Golkov V, et al. Flownet: learning optical flow with convolutional networks. In: *Proceedings of the IEEE International Conference on Computer Vision*. Santiago (2015). p. 2758–66. doi: 10.1109/ICCV.2015.316
25. Bertinetto L, Valmadre J, Henriques JF, Vedaldi A, Torr PH. Fully-convolutional siamese networks for object tracking. In: *European Conference on Computer Vision*. Amsterdam: Springer (2016). p. 850–65. doi: 10.1007/978-3-319-48881-3\_56
26. Trauth MH. Spectral analysis in Quaternary sciences. *Quater Sci Rev.* (2021) 270:107157. doi: 10.1016/j.quascirev.2021.107157
27. Zhao H, Gallo O, Frosio I, Kautz J. Loss functions for image restoration with neural networks. *IEEE Trans Comput Imaging.* (2016) 3:47–57. doi: 10.1109/TCI.2016.2644865
28. Zhu S, Zanibbi R. A text detection system for natural scenes with convolutional feature learning and cascaded classification. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV: IEEE (2016). p. 625–32. doi: 10.1109/CVPR.2016.74
29. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol.* (2015) 60:5471. doi: 10.1088/0031-9155/60/14/5471
30. Yu H, Zhou X, Jiang H, Kang H, Wang Z, Hara T, et al. Learning 3D non rigid deformation based on an unsupervised deep learning for PET/CT image registration. In: Gimi B, Krol A, editors. *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*. Vol. 10953. Houston, TX: International Society for Optics and Photonics (2019). p. 109531X.
31. Kumar SK. On weight initialization in deep neural networks. *arXiv preprint arXiv:170408863*. (2017).
32. Ruder S. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:160904747*. (2016).
33. Koch-Janusz M, Ringel Z. Mutual information, neural networks and the renormalization group. *Nat Phys.* (2018) 14:578–82. doi: 10.1038/s41567-018-0081-4
34. Sheikh HR, Bovik AC. Image information and visual quality. *IEEE Trans Image Process.* (2006) 15:430–44. doi: 10.1109/TIP.2005.859378
35. Zanon M, Pacini GS, de Souza VVS, Marchiori E, Meirelles GSP, Szarf G, et al. Early detection of lung cancer using ultra-low-dose computed tomography in coronary CT angiography scans among patients with suspected coronary heart disease. *Lung Cancer.* (2017) 114:1–5. doi: 10.1016/j.lungcan.2017.10.004

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xiao, Yang, Qiang, Zhao, Hao, Lian and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Early COPD Risk Decision for Adults Aged From 40 to 79 Years Based on Lung Radiomics Features

Yingjian Yang<sup>1,2</sup>, Wei Li<sup>2\*</sup>, Yingwei Guo<sup>1,2</sup>, Yang Liu<sup>2</sup>, Qiang Li<sup>1,2</sup>, Kai Yang<sup>3</sup>, Shicong Wang<sup>2</sup>, Nanrong Zeng<sup>2</sup>, Wenxin Duan<sup>2</sup>, Ziran Chen<sup>2</sup>, Huai Chen<sup>4\*</sup>, Xian Li<sup>4</sup>, Wei Zhao<sup>5</sup>, Rongchang Chen<sup>3,6,7\*</sup> and Yan Kang<sup>1,2,8\*</sup>

<sup>1</sup> College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China, <sup>2</sup> Medical Device Innovation Center, Shenzhen Technology University, Shenzhen, China, <sup>3</sup> Shenzhen Institute of Respiratory Diseases, Shenzhen People's Hospital, Shenzhen, China, <sup>4</sup> Department of Radiology, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, <sup>5</sup> Medical Engineering, Liaoning Provincial Crops Hospital of Chinese People's Armed Police Forces, Shenyang, China, <sup>6</sup> The Second Clinical Medical College, Jinan University, Guangzhou, China, <sup>7</sup> The First Affiliated Hospital, Southern University of Science and Technology, Shenzhen, China, <sup>8</sup> Engineering Research Centre of Medical Imaging and Intelligent Analysis, Ministry of Education, Shenyang, China

## OPEN ACCESS

### Edited by:

Kuanquan Wang,  
Harbin Institute of Technology, China

### Reviewed by:

Qiao He,  
Swedish Hospital, United States  
Tianzhu Xiang,  
Inception Institute of Artificial  
Intelligence (IIAI), United Arab Emirates

### \*Correspondence:

Wei Li  
liwei2@sztu.edu.cn  
Huai Chen  
chenhuai1977@163.com  
Rongchang Chen  
chenrc@vip.163.com  
Yan Kang  
kangyan@sztu.edu.cn

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

Received: 29 December 2021

Accepted: 04 March 2022

Published: 21 April 2022

### Citation:

Yang Y, Li W, Guo Y, Liu Y, Li Q,  
Yang K, Wang S, Zeng N, Duan W,  
Chen Z, Chen H, Li X, Zhao W, Chen R  
and Kang Y (2022) Early COPD Risk  
Decision for Adults Aged From 40 to  
79 Years Based on Lung Radiomics  
Features. *Front. Med.* 9:845286.  
doi: 10.3389/fmed.2022.845286

**Background:** Chronic obstructive pulmonary disease (COPD), a preventable lung disease, has the highest prevalence in the elderly and deserves special consideration regarding earlier warnings in this fragile population. The impact of age on COPD is well known, but the COPD risk of the aging process in the lungs remains unclear. Therefore, it is necessary to understand the COPD risk of the aging process in the lungs, providing an early COPD risk decision for adults.

**Methods:** COPD risk is evaluated for adults to make an early COPD risk decision from the perspective of lung radiomics features. First, the subjects are divided into four groups according to the COPD stages. Their ages are divided into eight equal age intervals in each group. Second, four survival Cox models are established based on the lung radiomics features to evaluate the risk probability from COPD stage 0 to suffering COPD and COPD stages. Finally, four risk ranks are defined by equally dividing the COPD risk probability from 0 to 1. Subsequently, the COPD risk at different stages is evaluated with varying age intervals to provide an early COPD risk decision.

**Results:** The evaluation metrics area under the curve (AUC)/C index of four survival Cox models are 0.87/0.94, 0.84/0.83, 0.94/0.89, and 0.97/0.86, respectively, showing the effectiveness of the models. The risk rank levels up every 5 years for the subjects who had suffered COPD after 60. For the subjects with COPD stage 0, the risk rank of suffering COPD stage I levels up every 5 years after the age of 65 years, and the risk rank of suffering COPD stages II and III & IV levels up every 5 years after the age of 70 years.

**Conclusion:** Once the age is above 60 years, the patients with COPD need to take action to prevent the progress and deterioration of COPD. Once the age is above 65 years, the patients with COPD stage 0 need to take precautions against COPD.

**Keywords:** COPD risk, aging, COPD stage (GOLD), radiomics, early decision, survival Cox model, Lasso

## INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is characterized by persistent airflow limitation. The gold standard for the diagnosis and evaluation of COPD is the forced expiratory volume in the first second (FEV<sub>1</sub>) and FEV<sub>1</sub>/forced vital capacity (FVC) ratio examined by pulmonary function test (PFT) (1). Previous studies on COPD mainly focus on COPD diagnosis and classification (2), COPD treatment (3, 4), COPD exacerbation prediction (5, 6), and COPD evaluation (7, 8). Many age-associated changes have been confirmed in the respiratory and pulmonary immune systems (9). Age relative risks of COPD mortality increase exponentially in China and the US (10). Age has become one of the factors of the score called Emphysema, Age, Smoking, SIZE (EMPHASIZE) in predicting the presence of clinically significant COPD and future morbidity (11). COPD has the highest prevalence in the elderly and deserves special consideration regarding treatment in this fragile population (9). However, it remains unclear the impact of age on the COPD risk.

The thoracic cavity's size decreases, limiting lung volumes and altering the muscles that aid in respiration with adults aging (9). The PFT result changes with the ages of both healthy people and COPD patients. After birth, the lung tissue will continue developing and growing to maturity. As a result, the alveoli and the small blood vessels in the lung will increase exponentially, and the lung volume will also become more extensive. The median FEV<sub>1</sub> and FVC in the PFT increase with age from 6 to 18 years, which linearly change until the adolescent growth spurt at about 10 years in girls and 12 years in boys (12). The median FEV<sub>1</sub>/FVC, first, decreases, then increases, and finally decreases with age from 6 to 18 years (12). After the lungs mature, the respiratory function of the lungs gradually declines with aging. Previous research (13–18) shows that both the FEV<sub>1</sub> and the FVC decline progressively with age increasing from 20 to 90 years of healthy lifelong non-smokers. The FVC and the FEV<sub>1</sub> in patients with COPD also show a significant decrease during a follow-up period of 4 years (19). The FEV<sub>1</sub> peaks between the age of 20 and 36 years and declines with aging (20). COPD prevalence is 2–3 times higher in people above the age of 60 years (21, 22). The increased burden of COPD seen in the elderly population may be due to age-associated changes in the structure and function of the lung, increasing the pathogenetic susceptibility to COPD (9). These changes, described in elderly lifelong non-smokers, are characterized by airspace dilatation resulting from loss of supporting tissue without alveolar wall destruction, similar to changes seen with COPD (9, 23). Therefore, it is necessary to evaluate COPD risk at different stages with aging for precision medicine.

Compared with PFT, computed tomography (CT) has been regarded as the most effective modality for characterizing and quantifying COPD (24), for example, quantitatively analyzing airway disease and emphysema in patients with COPD. Since the concept of radiomics was formally proposed in 2012 (25), radiomics of the chest CT images has been widely used for the chemotherapy response prediction in non-small-cell lung cancer (26) and pathology invasiveness prediction in patients with solitary pulmonary nodules (27). Recently, radiomics also

has been used in COPD for survival prediction (28, 29), COPD presence prediction (30), and the COPD exacerbations (31). However, radiomics in COPD has not been extensively investigated yet. Currently, there are only potential applications of radiomics features in COPD for the diagnosis, treatment, and follow-up of COPD and future directions (32). In particular, lung radiomics features as an imaging biomarker that reflects the state of lung parenchyma should be applied to COPD risk evaluation for an early COPD risk decision.

In summary, our contributions in this study are briefly described as follows:

- Four survival Cox models are established to evaluate the COPD risk at different COPD stages based on lung radiomics features;
- Earlier COPD risk decisions are made. The start age of the COPD risk rank, which levels up every 5 years, is given for the subjects who had suffered COPD or may suffer COPD at different stages.

## MATERIALS AND METHODS

This section mainly introduces the cohort (materials) and research methods used in this study.

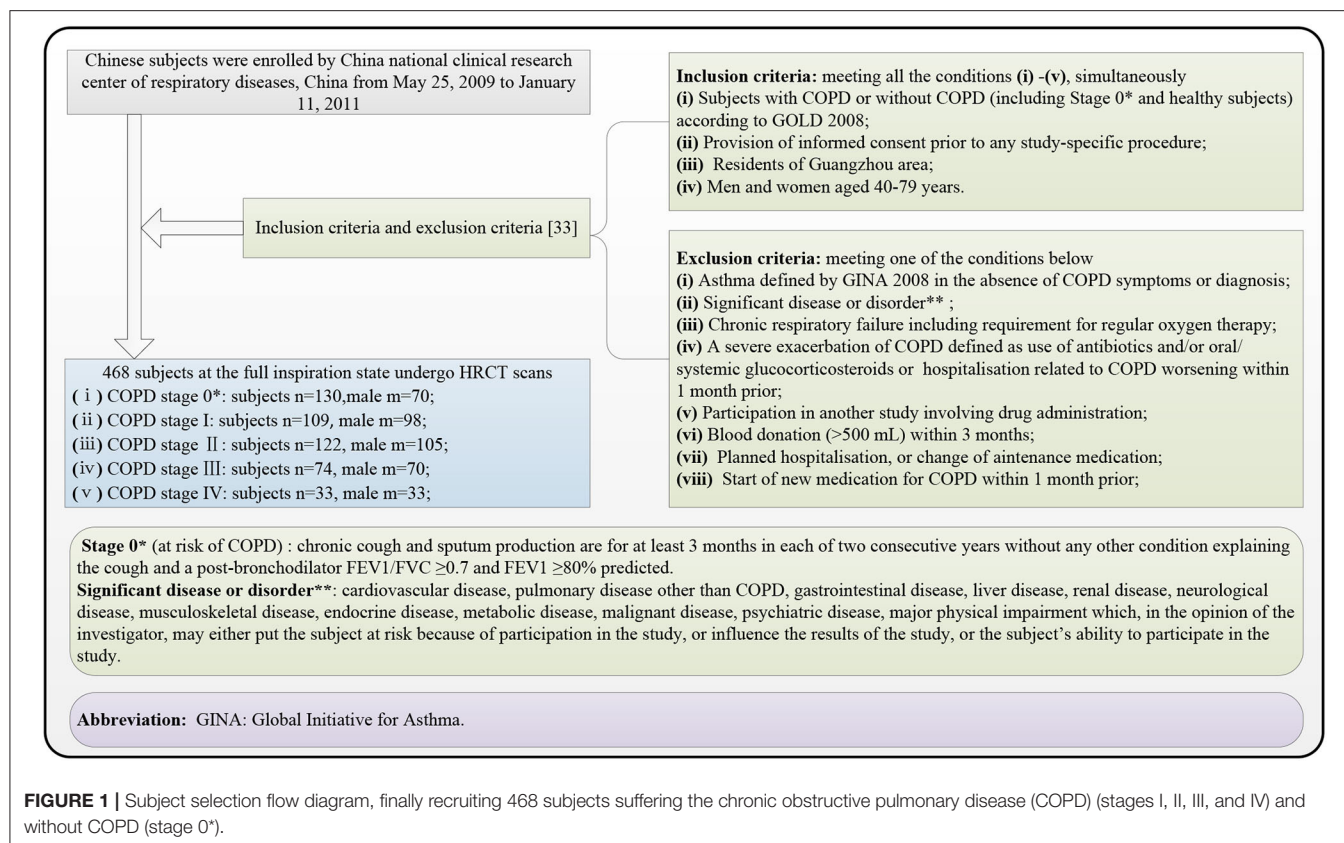
### Materials

The ethics committee had approved this study of the National Clinical Research Center of Respiratory Diseases in Guangzhou Medical University, China. Chinese subjects were enrolled by the China National Clinical Research Center of Respiratory Diseases from May 25, 2009, to January 11, 2011.

**Figure 1** shows the selection flow of the subjects, followed by the inclusion and exclusion criteria (33). The 468 subjects who met the inclusion criteria and the exclusion criteria underwent HRCT scans (manufacturer: TOSHIBA, KVP: 120 kVp, X-ray tube current: 40 mA, slice thickness: 1.0 mm, window center: –600, and window width: 1,250) and PFT after using the bronchodilator. All 468 subjects had been provided written informed consent by the first affiliated hospital of Guangzhou Medical University before chest HRCT scans and PFT. The COPD stage is diagnosed from stages 0 to IV according to (Global Initiative for Chronic Obstructive Lung Disease, GOLD) 2008 criteria accepted by the American Thoracic Society and the European Respiratory Society.

### Methods

A trained deep learning model ResU-Net automatically segments the lung parenchyma images from the chest HRCT images. Then, PyRadiomics automatically calculates the lung radiomics features based on the lung parenchyma images. The 468 subjects are divided into four groups A–D according to the COPD stages, and the ages of 40–79 years are divided into eight equal age intervals in each group. Each group includes the subjects suffering from COPD stage 0 as the 1st subgroup. The 2nd subgroup of the four groups includes the subjects suffering from COPD, stages I, II, and III & IV, respectively. The Lasso model is individually applied to select the lung radiomics features of the four groups. To predict



the COPD risk probability of the four groups, the four survival Cox models are constructed based on the selected lung radiomics features by the Lasso model and their age intervals, generating intuitive radiomics nomograms. The four COPD risk ranks are defined by equally dividing the COPD risk probability from 0 to 1. Subsequently, the COPD risk at different stages is evaluated with varying age intervals.

**Figure 1** shows the overall block diagram of research methods in this study, including the region of interest (ROI) segmentation (refer to the “ROI segmentation” section), lung radiomics feature calculation (refer to the “Lung radiomic features calculation” section), and the COPD risk evaluation for an earlier COPD decision (refer to the “COPD risk evaluation at different stages” section).

### Region of Interest Segmentation

Considering the overall change of lung status with aging, the lung parenchyma, including left and right lungs, is taken as the ROI in this study. The ResU-Net model trained by human chest CT images (34) automatically segment the ROI with red color in **Figure 2A** (size:  $512 \times 512 \times N$ ) from the chest HRCT images (size:  $512 \times 512 \times N$ ), and the detailed architecture of ResU-Net has been described in our previous study (35). All the ROI images had been checked and modified by three experienced radiologists in Shenzhen People's Hospital and the First Affiliated Hospital of Guangzhou Medical University. The trained ResU-Net model

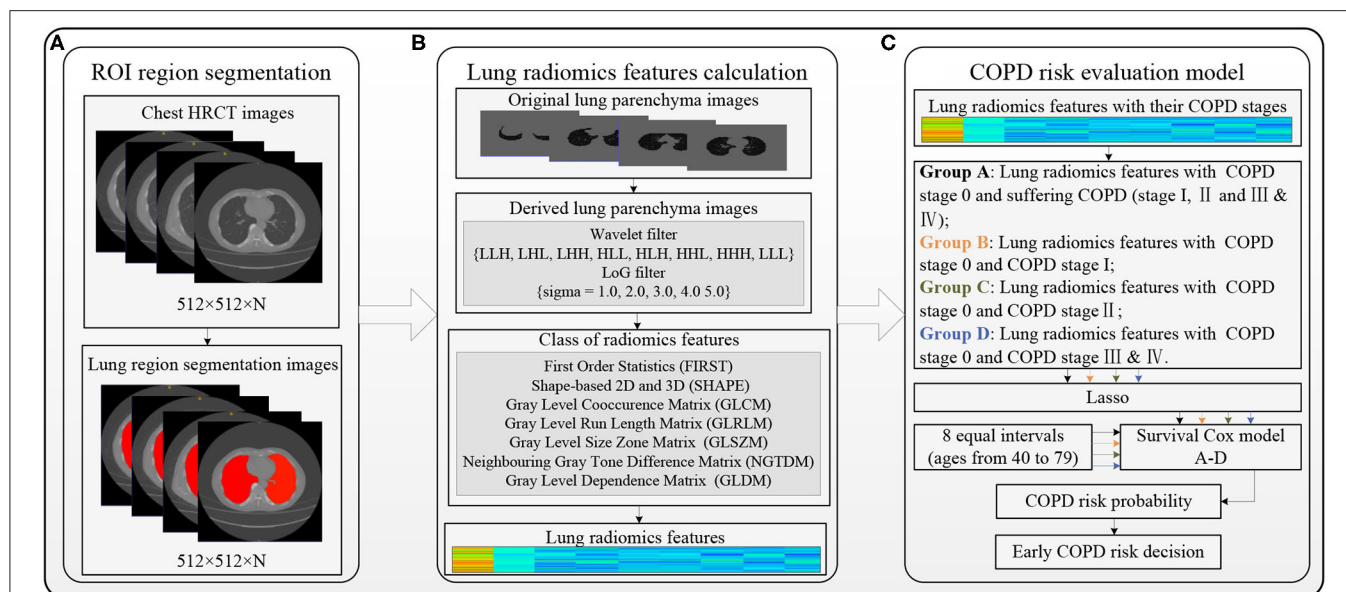
can be downloaded from the website <https://github.com/JoHof/lungmask>.

### Lung Radiomics Feature Calculation

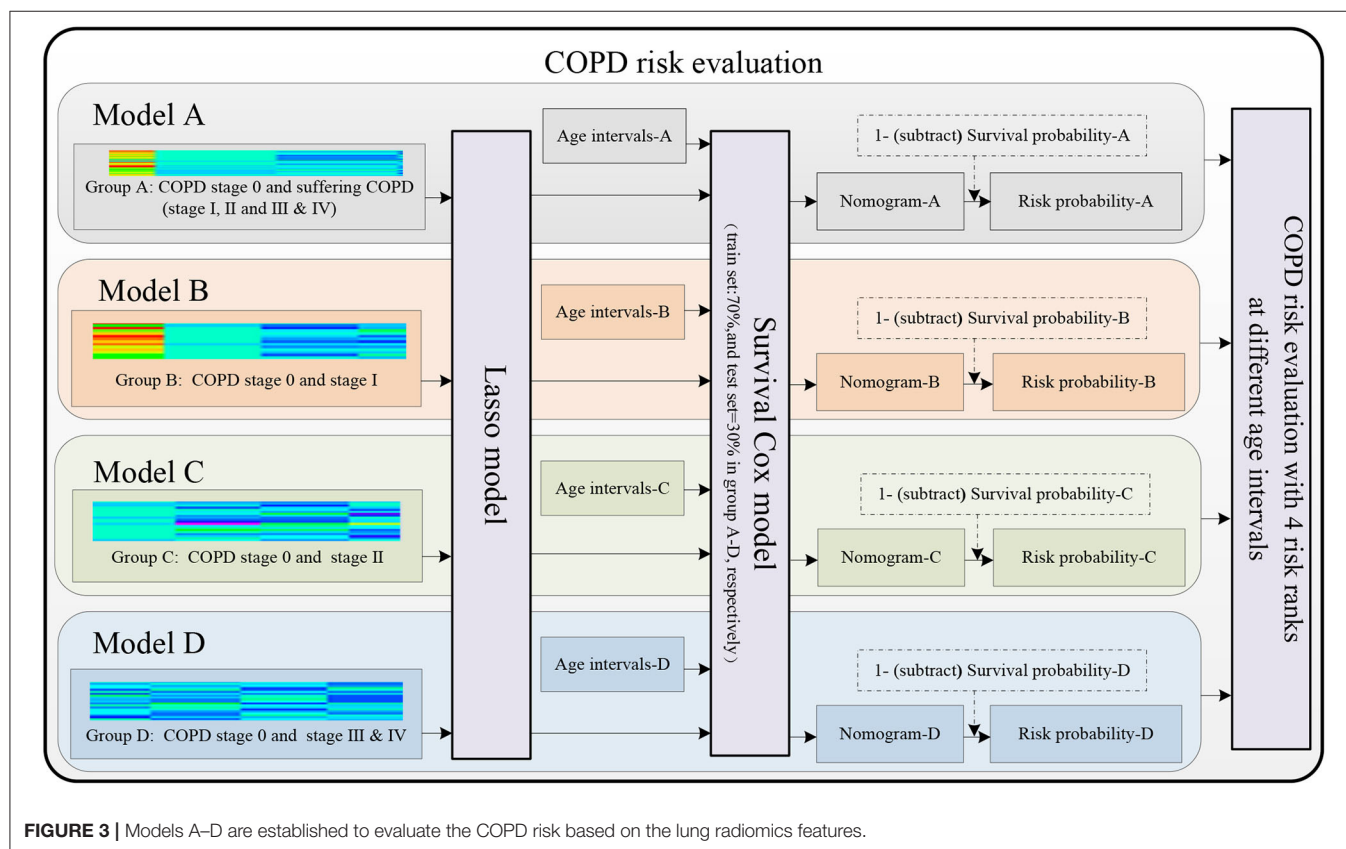
**Figure 2B** shows the lung radiomics feature calculation using PyRadiomics (36). PyRadiomics is available on the website <https://pyradiomics.readthedocs.io/en/latest/index.html>. Before calculating the lung radiomics features from the ROI (lung parenchyma) images, the ROI images with Hounsfield unit (HU) should be extracted from the chest HRCT images by our previous method (37). The original lung parenchyma images are the ROI images with HU. The wavelet filter (38, 39) and Laplacian of Gaussian (LoG) filter (40, 41) are applied to filter the original ROI images, generating two kinds of derived ROI images. The lung radiomics features are calculated based on the original and derived lung parenchyma images by the preset classes shown in **Figure 2B**. Finally, the 1,316 lung radiomics features for each subject are obtained.

### COPD Risk Evaluation at Different Stages

The four groups A–D with the lung radiomics features divide according to their COPD stages, and the eight equal age intervals divide from the age 40–79 years in each group. The four groups A–D include the subjects at COPD stage 0 and suffering from COPD, COPD stages 0 and I, COPD stages 0 and II, and COPD stages 0 and III & IV, respectively. However, the advantages of optimal stability and accuracy of



**FIGURE 2 |** Overall block diagram of the methods in this study. **(A)** Region of interest (ROI) segmentation, **(B)** Lung radiomics feature calculation, and **(C)** COPD risk evaluation model.



**FIGURE 3 |** Models A-D are established to evaluate the COPD risk based on the lung radiomics features.

the least absolute shrinkage and selection operator (Lasso) model have been confirmed (42). It is applied to select the lung radiomics features from 1,316 lung radiomics features.

COPD risk of each group is evaluated by the survival Cox model (43, 44) with the selected lung radiomics features and age intervals.

**Figure 3** shows the four COPD risk evaluation models of the four groups A–D, including the four groups and eight age intervals (refer to the “The four groups and eight age intervals” section), the lung radiomics feature selection by Lasso model (refer to the “Lasso model for lung radiomic features selection” section), and the COPD risk evaluation by the survival Cox model (refer to the “Survival Cox model for COPD risk evaluation” section).

#### The Four Groups and Eight Age Intervals

Four groups A–D are divided for predicting the COPD risk from COPD stage 0 to suffering from COPD (COPD stages I, II, and III & IV), stages I, II, and III & IV using survival Cox models, respectively. **Figure 4A** shows that all the four groups A–D include the subjects who suffer from COPD stage 0 as the 1st subgroup ( $n = 130$ ), and the 2nd subgroup in the four groups A–D are the subjects who suffer from COPD (I, II and III & IV,  $n = 338$ ), I ( $n = 109$ ), II ( $n = 122$ ), and III & IV ( $n = 107$ ), respectively. The reference (19) has given the conclusion that the FVC and FEV<sub>1</sub> in patients with COPD also show a significant decrease during a follow-up period of 4 years. Therefore, the eight equal age intervals are divided from 40 to 79 years in each group every 5 years. **Figure 4B** shows that the ages from 40 to 79 years of the 468 subjects are

equally divided into eight age intervals. **Figure 4C** shows the age distribution map of eight equal age intervals at different COPD stages.

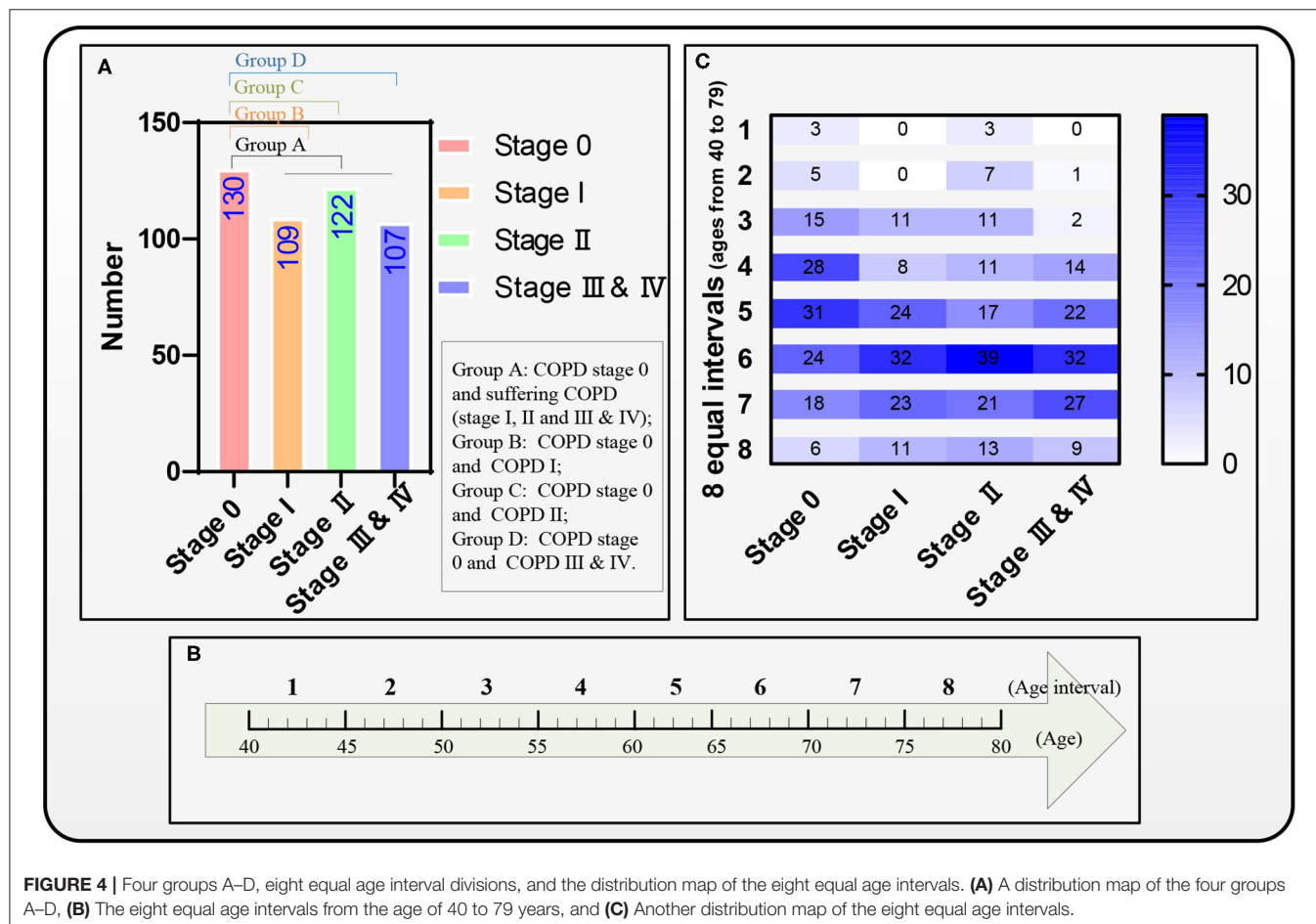
#### Lasso Model for Lung Radiomics Feature Selection

The standard R package lars (Lasso model) is applied separately to select the lung radiomics features affecting the COPD stages from the normalized lung radiomics features of the four groups. A ten-fold cross-validation (a standard R package “cv. Lars” with  $K = 10$ ) is used to ensure the effectiveness of the Lasso model.

The lung radiomics features of the four groups are normalized by the Equation (1).

$$x_{ij}^* = (x_{ij} - \bar{x}_j) / (x_{jmax} - x_{jmin}) \quad (1)$$

where,  $i = 1 \sim 468$  (468 subjects),  $j = 1 \sim 1316$  (1316 lung radiomics features of each subject),  $x_{ij}$  is the  $i$ th row and  $j$ th column of the  $468 \times 1,316$  lung radiomics features,  $\bar{x}_j$ ,  $x_{jmax}$ ,  $x_{jmin}$



**TABLE 1** | The lung radiomics features of the four groups A–D selected by the Lasso model, respectively.

Lung radiomics features selected by Lasso	Type of images	Class	Group A	Group B	Group C	Group D
original_shape_Elongation	Original images	SHAPE features	✓	✓		✓
original_shape_Maximum2DDiameterRow		SHAPE features	✓	✓	✓	✓
original_shape_Maximum2DDiameterSlice		SHAPE features			✓	
original_shape_Sphericity		SHAPE features				✓
original_shape_SurfaceVolumeRatio		SHAPE features	✓	✓		
original_firstorder_10Percentile		FIRST features				✓
original_glszm_GrayLevelNonUniformityNormalized		GLSZM features			✓	
original_glszm_ZoneEntropy		GLSZM features	✓		✓	✓
log.sigma.1.0.mm.3D_firstorder_Maximum		FIRST features			✓	✓
	Derived images generated from LoG filter					
log.sigma.1.0.mm.3D_gldm_ClusterProminence		GLCM features			✓	✓
log.sigma.1.0.mm.3D_gldm_GrayLevelVariance		GLRLM features			✓	
log.sigma.1.0.mm.3D_glszm_SmallAreaEmphasis		GLSZM features			✓	
log.sigma.1.0.mm.3D_glszm_ZoneEntropy		GLSZM features				✓
log.sigma.2.0.mm.3D_firstorder_Maximum		FIRST features				✓
log.sigma.2.0.mm.3D_glszm_SmallAreaLowGrayLevelEmphasis		GLSZM features			✓	
log.sigma.2.0.mm.3D_ngtdm_Contrast		NGTDM features	✓		✓	
log.sigma.2.0.mm.3D_gldm_SmallDependenceLowGrayLevelEmphasis		GLDM features			✓	
log.sigma.2.0.mm.3D_gldm_DependenceVariance	Derived images generated from wavelet filter	GLDM features				✓
log.sigma.3.0.mm.3D_firstorder_10Percentile		FIRST features				✓
log.sigma.5.0.mm.3D_firstorder_10Percentile		FIRST features	✓		✓	✓
log.sigma.5.0.mm.3D_firstorder_TotalEnergy		FIRST features	✓			
log.sigma.5.0.mm.3D_gldm_RunLengthNonUniformity		GLRLM features		✓		
log.sigma.5.0.mm.3D_glszm_SmallAreaEmphasis		GLSZM features			✓	
wavelet.LLH_gldm_ClusterTendency		GLCM features	✓			
wavelet.LLH_glszm_GrayLevelNonUniformityNormalized		GLSZM Features	✓			
wavelet.LLH_glszm_LargeAreaLowGrayLevelEmphasis		GLSZM features		✓		
wavelet.LLH_gldm_GrayLevelNonUniformityNormalized		GLRLM features		✓		
wavelet.LLH_firstorder_Mean		FIRST features			✓	
wavelet.LLH_firstorder_RootMeanSquared		FIRST features			✓	
wavelet.LHL_gldm_SmallDependenceLowGrayLevelEmphasis		GLDM features	✓			
wavelet.LHL_firstorder_Kurtosis		FIRST features		✓	✓	✓
wavelet.HLH_gldm_ShortRunLowGrayLevelEmphasis		GLRLM features				✓
wavelet.LLL_firstorder_10Percentile		FIRST features	✓	✓	✓	
wavelet.LLL_firstorder_Minimum		FIRST features	✓	✓		
wavelet.LLL_firstorder_TotalEnergy		FIRST features		✓		
wavelet.LLL_gldm_Imc2		GLCM features	✓		✓	✓

are the mean, the maximum, the minimum of each radiomics feature  $x_j$ , respectively.

Formula (2) is the Lasso model to select the normalized lung radiomics features of the four groups, respectively.

$$\mathbf{A}_k \leftarrow \arg \min \left\{ \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij}^* \right)^2 + \lambda \sum_{j=0}^p |\beta_j| \right\} \quad (2)$$

where matrix  $\mathbf{A}_k$  denotes the selected lung radiomics features,  $k = 1, 2, 3, 4$  respectively denotes group A–D.  $x_{ij}^*$  denotes each

group's normalized lung radiomics features (the independent variable).  $y_i$  denotes each group's COPD stage (the independent variable). Especially,  $y_i$  in group A denotes the COPD stage 0 and suffering COPD.  $\lambda$  denotes the penalty parameter ( $\lambda \geq 0$ ).  $\beta_j$  denotes the regression coefficient,  $i \in [1, n]$ , and  $j \in [0, p]$ .

#### Survival Cox Model for COPD Risk Evaluation

The survival Cox model (45, 46), the standard R survival package coxph, picks up the final selected lung radiomics features  $x$  of the four groups from the selected lung radiomics features again,

**TABLE 2 |** The final lung radiomics features of the four groups A–D selected from the survival Cox model, respectively.

Group	Definition	Lung radiomics features selected by survival Cox	Coef	HR:exp(coef)/95%CI	Se(coef)	z	p-value
Group A	Radiomics 1	original_shape_SurfaceVolumeRatio	−0.396	0.673/ 0.531–0.853	0.121	−3.275	**
	Radiomics 2	log.sigma.5.0.mm.3D_firstorder_TotalEnergy	0.316	1.372/1.056–1.783	0.134	2.369	*
	Radiomics 3	wavelet.LLL_firstorder_Minimum	−0.248	0.780/ 0.684–0.890	0.067	−3.693	***
Group B	Radiomics 4	wavelet.LLH_glszm_LargeAreaLowGrayLevelEmphasis	0.414	1.512/1.137–2.012	0.146	2.838	**
	Radiomics 5	wavelet.LLL_firstorder_Minimum	−0.384	0.681/0.567–0.818	0.094	−4.106	***
Group C	Radiomics 6	log.sigma.1.0.mm.3D_firstorder_Maximum	0.297	1.346/1.082–1.675	0.112	2.665	**
	Radiomics 7	log.sigma.1.0.mm.3D_glszm_SmallAreaEmphasis	−0.529	0.589/0.399–0.871	0.200	−2.651	**
	Radiomics 8	log.sigma.5.0.mm.3D_firstorder_10Percentile	−0.7350	0.479/0.336–0.685	0.182	−4.045	***
	Radiomics 9	wavelet.LLH_firstorder_RootMeanSquared	−0.230	0.794/0.640–0.985	0.110	−2.096	*
	Radiomics 10	wavelet.LHL_firstorder_Kurtosis	0.529	1.697/1.214–2.373	0.172	3.091	**
	Radiomics 11	wavelet.LLL_firstorder_10Percentile	−1.085	0.338/0.223–0.513	0.213	−5.099	***
Group D	Radiomics 12	original_shape_Maximum2DDiameterRow	−0.374	0.688/0.446–1.062	0.221	−1.689	.
	Radiomics 13	original_firstorder_10Percentile	−0.654	0.520/0.358–0.756	0.191	−3.428	***
	Radiomics 14	log.sigma.1.0.mm.3D_glcM_ClusterProminence	−0.357	0.700/0.4905–0.9978	0.181	−1.972	*
	Radiomics 15	log.sigma.1.0.mm.3D_glszm_ZoneEntropy	0.461	1.585/1.064–2.364	0.204	2.262	*
	Radiomics 16	log.sigma.2.0.mm.3D_firstorder_Maximum	0.186	1.205/1.006–1.443	0.092	2.022	*

by statistically significant hazard ratio (HR). Then, the subjects with the final selected lung radiomics features  $x$  are divided into 70 and 30%. Notably, 70% of the subjects train the survival Cox models A–D by computing estimates of the survival functions, drawing the nomograms A–D, respectively. Then, four survival Cox models adopt the standard R rms package cph with the significance level  $\alpha = 0.1$ . Finally, 30% of the subjects in each group validate the performance of the four trained survival Cox models, respectively. It is noted that the failure event time is the age interval. The event indicator represents the COPD stage 0, and the suffering COPD, stage I, stage II, and stage III & IV of the four groups.

Equation (3) gives the survival probability formula of the survival Cox model.

$$\lambda(t|x) = \lambda_0(t) \cdot e^{\beta^T x} \quad (3)$$

where  $\lambda_0(t)$  denotes the baseline hazard function,  $\beta^T x$  denotes the log-risk function, which is the product of the probability at each event time the event has occurred to the individual.  $\beta$  denotes the weights for optimizing the Cox partial likelihood.

Equation (4) converts the COPD survival probability to the COPD risk probability.

$$Risk_{ijk} = 1 - \lambda_{ijk} \quad (4)$$

where  $k = 1, 2, 3, 4$  respectively denotes the group A–D,  $Risk_{ijk}$  denotes the  $i$ th COPD risk probability in the  $k$ th group, and  $\lambda_{ijk}$  is the  $i$ th COPD survival probability in the  $k$ th group as Equation (3).

Specifically, the COPD risk probability in group A is the COPD stage 0 probability and suffering COPD probability. The

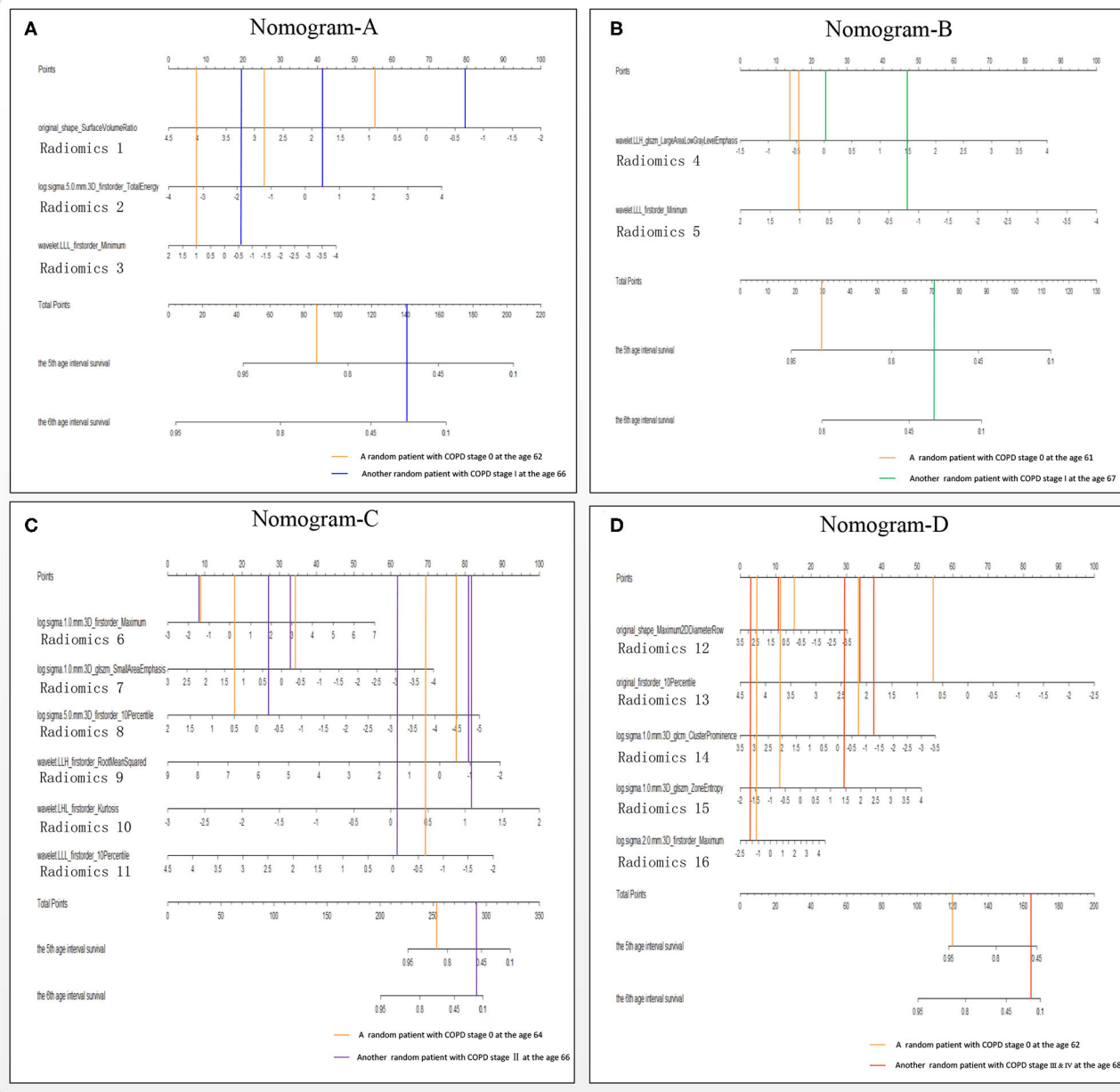
COPD risk probability in groups B–D is the COPD stage 0 probability and suffering COPD stage I probability, stage II probability, and stage III & IV probability, respectively. After calculating the COPD risk probability, each group's COPD risk probability is separated according to the COPD stages. Then, the four COPD risk ranks (mild risk: 0–0.25, moderate risk: 0.25–0.50, severe risk: 0.50–0.75, and extreme risk: 0.75–1) are defined by equally dividing the COPD risk probability from 0 to 1. Subsequently, the COPD risk is evaluated in the four groups at different risk ranks and age intervals.

## RESULTS

This section shows the final selected lung radiomics features, radiomics nomograms, performance evaluation of the four survival Cox models, scatter plots, and curves of COPD risk probability of the four groups, respectively.

### The Selected Lung Radiomics Features in Groups A–D

Tables 1, 2 report the results of the selected lung radiomics features by the Lasso model and the survival Cox model's final selected lung radiomics features of the four groups, respectively. In Table 1, the symbol  $\checkmark$  denotes the lung radiomics features selected by the Lasso model. The weight (coef), HR with 95% confidence interval (CI), and Wald's statistics (z) and significance (p-value) are also reported in Table 2. The symbol in Table 2 “\*\*\*” denotes  $p$ -value  $< 0.001$ , “\*\*” denotes  $p$ -value  $< 0.01$ , “\*” denotes  $p$ -value  $< 0.05$ , and “.” denotes  $p$ -value  $< 0.1$ . For facilitating expression, the final selected lung radiomics features are defined as Radiomics 1–16, respectively. Radiomics 1–16 are used to construct the survival Cox models.



**FIGURE 5 |** Nomograms A–D of the survival Cox models (A–D), taking the 5th and 6th age interval for example.

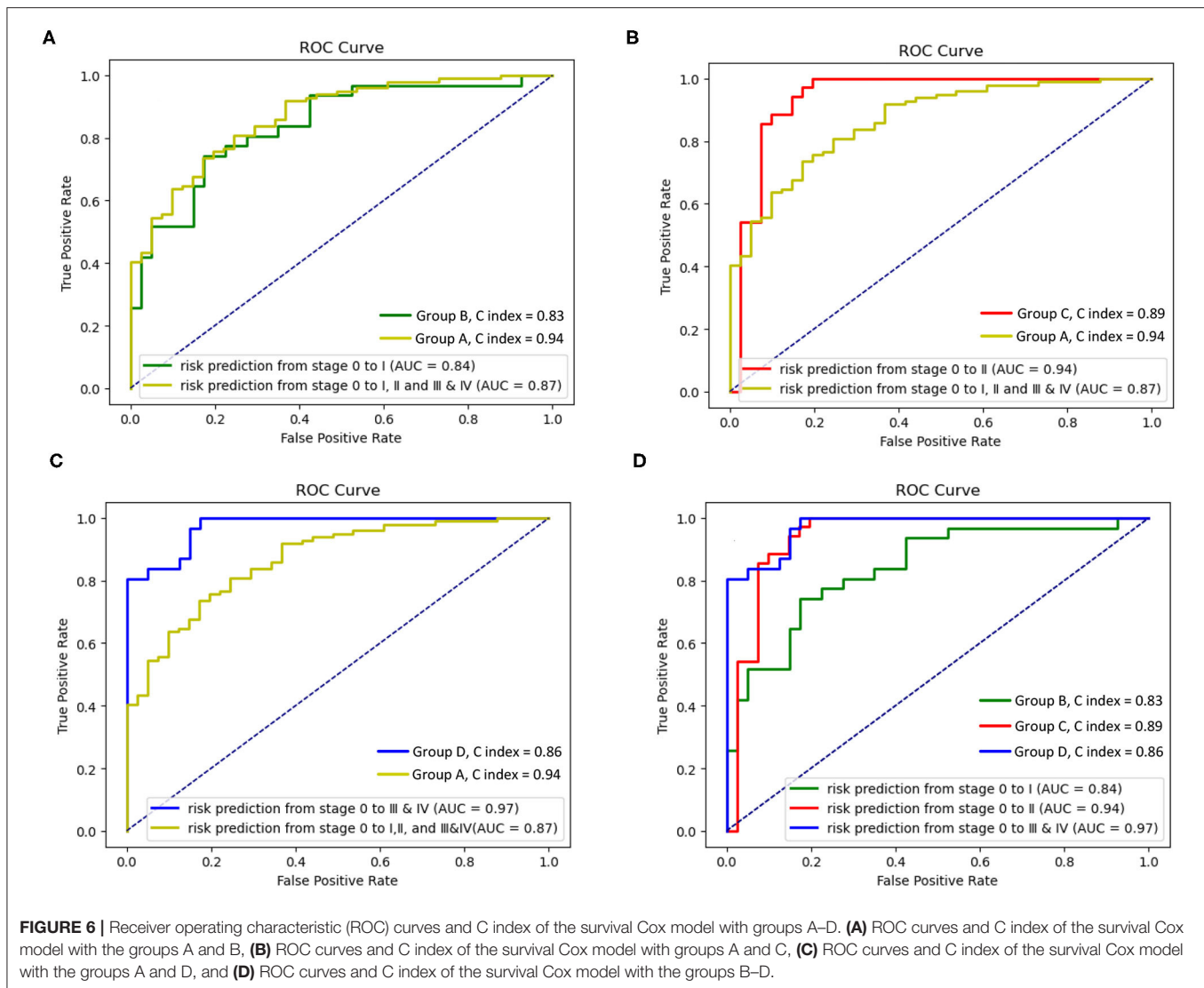
Specifically, Radiomics 1–3 are the factors that affect the COPD risk of the patients who have suffered from COPD with aging. Radiomics 4–5, Radiomics 6–11, and Radiomics 12–16 are the factors that affect the COPD risk of the patients who had suffered the COPD stages I, II, III, and IV with aging, respectively.

## Radiomics Nomograms and Performance Evaluation of Models A–D

Radiomics nomograms are pictorial representations depicting the association between radiomics variables and the probability

of suffering COPD or different COPD stage events, providing an intuitive way to interpret the survival Cox model (44).

**Figure 5** shows four radiomics nomograms of the four models A–D at the 5th and 6th age interval, respectively. The radiomics nomograms of the four groups A–D further indicate the importance of the final selected lung radiomics affecting COPD with aging, respectively. The points of Radiomics  $X$  ( $X = 1, 2, \dots, 16$ ) show the importance of its group. For example, **Figure 5A** indicates that Radiomics 2 is more critical than Radiomics 3 in nomogram-A. **Figure 5B** means that Radiomics 5 is more important than Radiomics 4 in nomogram-B. **Figure 5C** shows



that the order of importance is Radiomics 10, Radiomics 9, Radiomics 11, Radiomics 8, Radiomics 7, and Radiomics 6 in nomogram-C. **Figure 5D** shows that the order of importance is Radiomics 13, Radiomics 14, Radiomics 15, Radiomics 12, and Radiomics 16 in nomogram-D.

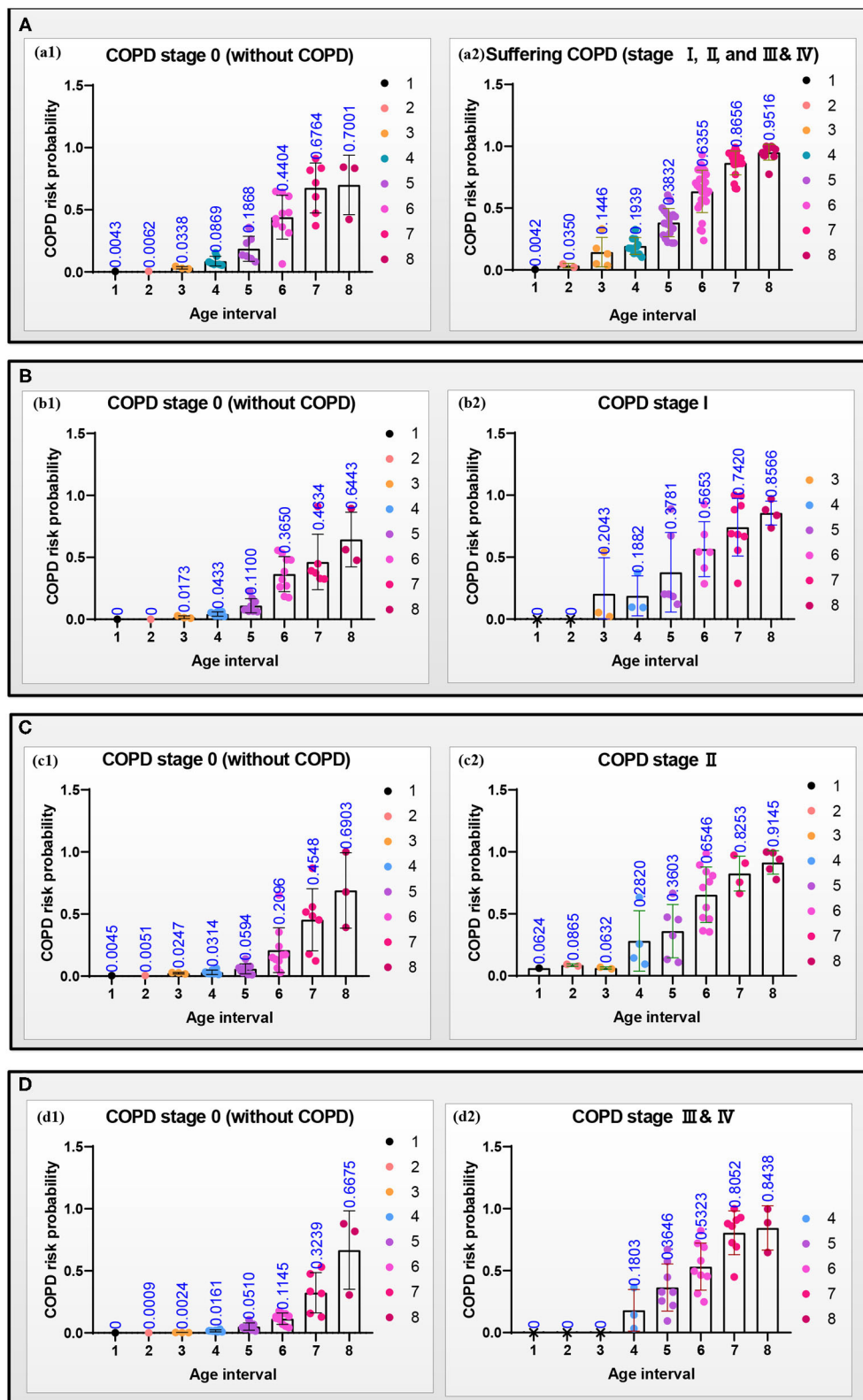
**Figure 6** reports the models A–D's performances to illustrate the effectiveness of the models. After verifying the effectiveness of the models, **Figure 7** intuitively reports the results of the COPD risk probability at different age intervals predicted by models A–D. **Figure 8** further compares the COPD risk probability of different COPD stages at different age intervals to illustrate the impact of suffering COPD or the COPD stage on the COPD risk. Finally, **Figure 9** summarizes the results in **Figure 8** to make an earlier COPD risk decision for adults.

The four receiver operating characteristic (ROC) curves are drawn to evaluate the models A–D's performances. **Figures 6A–C** show that the area under the curve (AUC, performance measurement for classification) of model A (AUC = 0.87) is

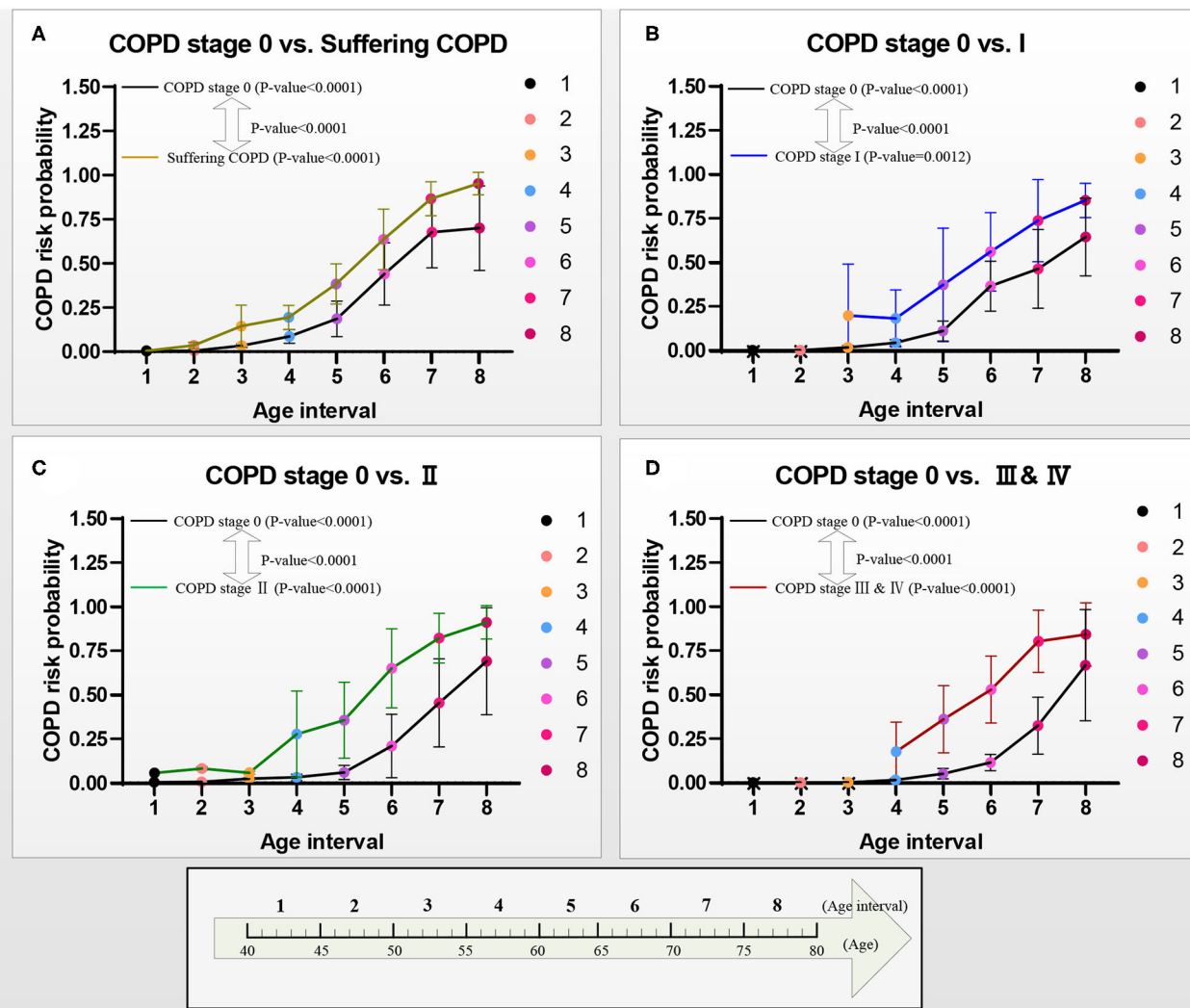
higher than that of model B (AUC = 0.84), and the AUC of the model C and the model D (model C: AUC = 0.94; model D: AUC = 0.97) is higher than that of model A (AUC = 0.87). **Figure 6D** shows that the AUC of the model D (model D: AUC = 0.97) is higher than that of model B and model C (model B: AUC = 0.84; model C: AUC = 0.89), and the AUC of the model C is higher than that of model B. The concordance index (C index, a standard performance metric for survival Cox analysis) of model A (C index = 0.94) is higher than other models (model B: C index = 0.83; model C: C index = 0.89; model D: C index = 0.86).

## COPD Risk Probability

**Figure 7** shows the scattering plots of each group's separated COPD risk probability according to the COPD stages. Models A–D predict the COPD risk probability at different age intervals with 30% of the final selected lung radiomics features  $x$ . **Figure 7a1–d1** show that the COPD risk probability of the



**FIGURE 7 |** Scattering plots with bar (mean with SD) of the COPD risk probability in the four groups (A–D) at different age intervals, respectively. (a1–d1) The COPD risk probability of the subjects at the COPD stage 0, and (a2–d2) the COPD risk probability of the patients who had suffered the COPD, COPD stages I, II, and III & IV, respectively.



**FIGURE 8 |** COPD risk probability curves with age increasing from COPD stages 0 to I, II, and III & IV, respectively. **(A)** The COPD risk probability in group A, **(B)** the COPD risk probability in group B, **(C)** the COPD risk probability in group C, and **(D)** the COPD risk probability in group D.

COPD stage 0 in the four groups significantly increases with aging. Similarly, **Figure 7a2–d2** also show that the COPD risk probability of suffering COPD, stage I, stage II, and stage III & IV in the four groups significantly increases with aging.

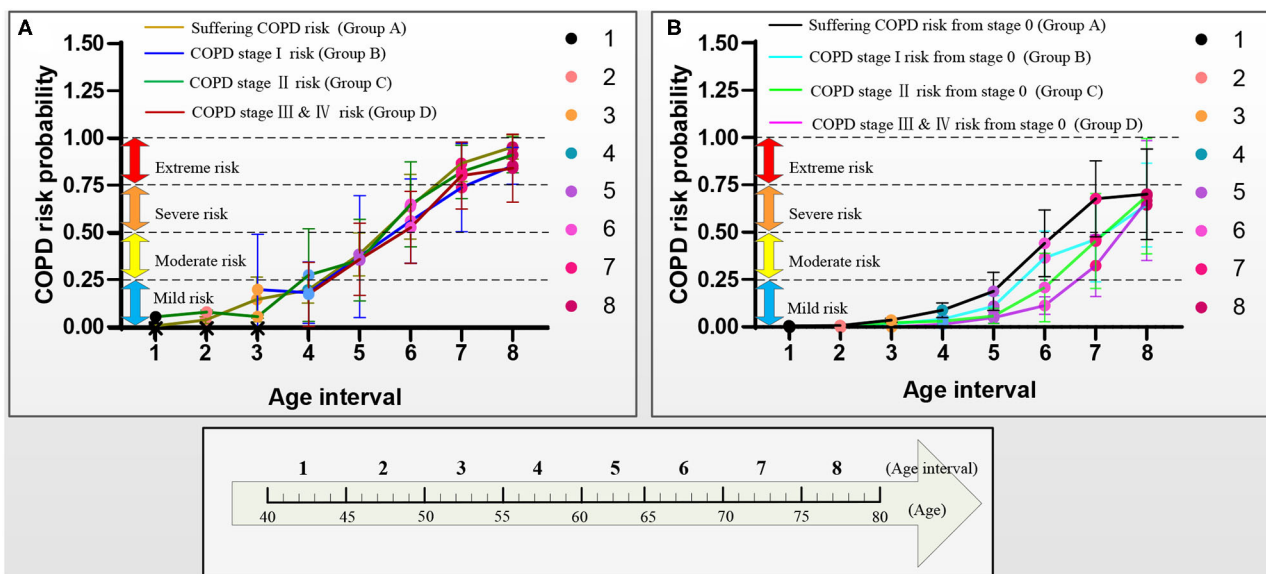
**Figure 8** shows the curves of the COPD risk probability of the four models A–D at different age intervals. The COPD risk probability (mean with SD) of all the COPD stages 0, I, II, and III & IV in the four groups significantly increases with aging. Overall, the COPD risk probability of the patients who had suffered the COPD, COPD stages I, II, and III & IV is higher than that of subjects at COPD stage 0 in the four groups, respectively.

**Figure 9** further shows the COPD risk probability curves of the patients who had suffered the COPD and the subjects at the COPD stage 0, respectively.

**Figure 9A** shows that the COPD risk of the patients who had suffered the COPD stages I, II, and III & IV increases with aging. Specifically, the 1st–4th age intervals of the patients who had

suffered from COPD, the COPD stages I, II, and III & IV are basically at mild risk. In contrast, the 5th age interval of those patients is at moderate risk, the 6th age interval of those patients is at severe risk, and the 7th and 8th age intervals of those patients are at extreme risk.

**Figure 9B** shows that the COPD risk of subjects at COPD stage 0 in the four groups increases with aging. Overall, in the same age interval, the COPD risk of the subjects at the COPD stage 0 reduces in order of suffering the COPD stages I, II, and III & IV. The risk of those subjects who may suffer from COPD is greater than that of COPD stages I, II, and III & IV with aging. Similarly, the risk of those subjects who may suffer from the COPD stage I is greater than that of the COPD stages II and III & IV with aging, and the risk of those subjects who may suffer from the COPD stage II is greater than that of the COPD stage III & IV with aging. Specifically, for the subjects with COPD stage 0, the COPD risk of the 1st–5th age intervals in groups A and B



**FIGURE 9 |** COPD risk probability curves with aging in the four groups A–D, respectively. **(A)** COPD risk probability curves of the patients who had suffered COPD, COPD stages I, II, and III & IV, and **(B)** COPD risk probability curves of subjects at COPD stage 0.

and 1st–6th age intervals in groups C and D are at mild risk. The COPD risk of the 6th age interval in group A, the 6th and 7th age intervals in group B, and the 7th age interval in groups C and D are at moderate risk. The COPD risk of the 7th and 8th age intervals in groups A and B and the 8th age interval in groups C and D are at severe risk. The COPD risk of the subjects who may suffer COPD stage I (10 years: the 6th and 7th age intervals) develops slower than that of the subjects who may suffer the COPD stages II and III & IV (5 years: the 7th age interval) in the moderate risk rank. However, until the age of 79 years, the COPD risk will not develop to the extreme risk rank.

## DISCUSSION

In this study, four survival Cox models of the four groups A–D are developed based on the lung radiomics features to evaluate COPD risk at different stages for adults aged from 40 to 79 years, providing a COPD risk decision for adults. All four survival Cox models' effectiveness has been evaluated. They are not only effective in evaluating the risk of suffering from COPD (COPD stages I, II, and III & IV) but also effective in more detailed differentiation of COPD stages (COPD stages 0 and I, COPD stages 0 and II, COPD stages 0 and III & IV).

The lung radiomics features, as an imaging biomarker reflecting the lung structure (SHAPE radiomics features) and lung tissue (the other radiomics features except for SHAPE), are first used to evaluate COPD risk at different stages with aging. Finally, independent features Radiomics 1–16, affecting the COPD evolution from stage 0 to suffering COPD, COPD stages I, II, and III & IV, are determined. Radiomics 1, 5, 10, and 13 are the most direct risk features for COPD evolution from stage 0 to suffering COPD, COPD stages I, II, and III &

IV, respectively. Our study discovers that different lung radiomics features affect the COPD risk at different stages with aging. The trends of COPD risk probability, which increase with aging, are in line with the change law of the aging process in the lung (9, 13).

For the subjects who have suffered the COPD, COPD stages I, II, and III & IV, our study discovers that although the COPD risk increases with aging, each COPD risk rank of the patients who had suffered from the COPD, COPD stages I, II, and III & IV, basically has the same age intervals. Therefore, age is not only a point to distinguish the “young COPD” (young patients with COPD aged <50 years) (47, 48) from the non-young COPD but also it can distinguish different COPD risk ranks. No matter which is COPD stage at, it is regarded as a relatively safe COPD risk rank (the mild risk) before the age of 60 years. The law of COPD risk ranks to level up with age has also been revealed for the subjects after 60 years. The risk rank of these subjects levels up every 5 years. Therefore, once the patients with COPD are above 60 years, they need to pay close attention to prevent the progress and deterioration of COPD. If necessary, manual intervention should be considered, including increasing appropriate practical exercise (49) and COPD care treatment (50) in the hospital.

For the subjects at the COPD stage 0, the COPD risk which may suffer the COPD, COPD stages I, II, and III & IV also increases with aging. Our study discovers that the age intervals of each COPD risk rank of the subjects who may suffer the COPD, COPD stages I, II, and III & IV, from the COPD stage 0 are inconsistent. Age also can distinguish different COPD risk ranks. It is a relatively safe COPD risk rank (the mild risk) of the subjects who may suffer the COPD and the COPD stage I before the age of 65 years. It is also a relatively safe COPD risk rank of subjects who may suffer the COPD stages II and III & IV before the age of 70 years. For the subjects at COPD stage 0, the risk rank of

suffering COPD stage I levels up every 5 years after the age of 65 years, and the risk rank of suffering COPD stages II and III & IV levels up every 5 years after the age of 70 years. Therefore, the age of 65 years is the start age of the increased risk of the subjects who may suffer the COPD stage I, and the age of 70 years is the start age of the increased risk of the subjects who may suffer the COPD stages II and III & IV. Although the COPD risk cannot reach the extreme risk rank, once the subjects with COPD stage 0 are above 65 or 70 years, they also need to take precautions against COPD.

There are some limitations of the methods and materials in this study. First, the survival Cox model considers both events and time, but it can only analyze two opposite events. At the same time, the model itself has high requirements for the collinearity of input data. Although we used the Lasso model to remove the collinearity of the lung radiomics features, some valuable features may be omitted. Then, there is a lack of subjects aged 40–50 years in the COPD stage I and subjects aged 40–55 years in the COPD stages III & IV. That is because relatively few people aged 40–50 years who suffer from COPD stage I will go to the hospital for treatment and undergo the CT scan, and few people suffer the COPD stages III & IV at the age of 40–55 years. Finally, although lung radiomics features have met the needs of COPD risk evaluation, an improved deep learning survival Cox model (45), with a deep feed-forward neural network, has also been used to improve the model's performance further. Regrettably, the deep learning survival Cox model's AUC and C index are not enhanced. We believe that lung radiomics features with quantitative CT parameters and/or clinical text data will improve the model's performance.

## CONCLUSION

Four effective models are established to evaluate COPD risk from COPD stage 0 to suffered COPD, COPD stages I, II, and III & IV, respectively. The early COPD risk decision is made based on the COPD risk results. The start age of the COPD risk rank, which levels up every 5 years, is given for the subjects who had suffered COPD or may suffer COPD at different stages. It concludes that once the age is above 60 years, the patients with COPD need to be paid close attention to prevent the progress and deterioration of COPD, and once the age is above 65 years, the patients with COPD stage 0 need to take precautions against COPD.

## REFERENCES

1. GOLD 2021. *Global Initiative for Chronic Obstructive Lung Disease* (2021).
2. Virani A, Baltaji S, Young M, Dumont T. Chronic obstructive pulmonary disease: diagnosis and GOLD Classification. *Crit Care Nurs Q*. (2021) 44:335. doi: 10.1097/CNQ.0000000000000335
3. Bender BG. Nonadherence to COPD Treatment: what have we learned and what do we do next? *Copd J Chronic Obstruct Pulmon Dis*. (2012) 12:880. doi: 10.3109/15412555.2012.679880
4. Baha A, Kokturk N. Physician's attitude against COPD guidelines and the choice of first-line treatment for COPD *Respir Med*. (2021) 176:106273. doi: 10.1016/j.rmed.2020.106273
5. Kronborg T, Hangaard S, Cichosz SL, Hejlesen O. A two-layer probabilistic model to predict COPD exacerbations for patients in telehealth. *Comput Biol Med*. (2021) 128:104108. doi: 10.1016/j.compbimed.2020.104108

## DATA AVAILABILITY STATEMENT

The lung radiomics features supporting the conclusions of this article will be available with a reasonable request.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by National Clinical Research Centre of Respiratory Diseases in Guangzhou Medical University, China. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

WL, HC, RC, and YK: conceptualization. YY, WL, YG, YL, and QL: methodology. YY, KY, SW, NZ, and WD: software. WL, YG, YL, and ZC: validation. YY, WL, HC, and XL: formal analysis. YY, WZ, and XL: investigation. KY, HC, and XL: resources. YY, WL, YG, YL, and ZC: data curation. YY: writing—original draft. WL, YL, and YK: writing—review and editing. YY, YG, QL, and SW: visualization. HC, RC, and YK: supervision. WL and YK: project administration. WL, HC, RC, and YK: funding acquisition. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Natural Science Foundation of Guangdong Province, China (2019A1515011382), the Stable Support Plan for Colleges and Universities in Shenzhen, China (SZWD2021010), the Scientific Research Fund of Liaoning Province, China (JL201919), and the National Natural Science Foundation of China (62071311).

## ACKNOWLEDGMENTS

The authors thank the Department of Radiology, the First Affiliated Hospital of Guangzhou Medical University for providing the dataset.

6. Crooks MG, den Brinker AC, Thackray-Nocera S, van Dinther R, Wright CE, Morice AH. Domiciliary cough monitoring for the prediction of COPD Exacerbations. *Lung*. (2021) 199:9859. doi: 10.1007/s00408-021-00435-9
7. Hassett DJ, Borchers MT, Panos RJ. Chronic obstructive pulmonary disease (COPD): Evaluation from clinical, immunological and bacterial pathogenesis perspectives. *J Microbiol*. (2014) 52:211–26. doi: 10.1007/s12275-014-4068-2
8. Strassmann A, Guler M, Steurer-Stey C, Dalla Lana K, Carron T, Braun J, et al. Nationwide implementation of the self-management program "Living well with COPD": process and effectiveness evaluation using a mixed-methods approach - ScienceDirect. *Patient Educ Couns*. (2021) 6:18.
9. Lowery EM, Brubaker AL, Kuhlmann E, Kovacs EJ. The aging lung. *Clin Interv Aging*. (2013) 8:1489. doi: 10.2147/CIA.S51152
10. Wen H, Xie C, Wang L, Wang F, Wang Y, Liu X, et al. Difference in long-term trends in COPD mortality between China and the US, 1992–2017:

- an age-period-cohort analysis. *Int J Environ Res Public Health*. (2019) 16:1529. doi: 10.3390/ijerph16091529
11. Han MK, Tayob N, Kim V, Murray S, Barr G, Rennard SI, et al. COPD: What IS. NEW IN. IMAGING?: CT. emphysema score (EMPHYsema, age, smoking, SIZE – EMPHASIZE) identifies clinically significant COPD. *ATS*. (2014) 14:315–21. doi: 10.1164/ajrccm-conference.2014.189.1\_MeetingAbstracts.A4336
  12. Wang X, Dockery DW, Wypij D, Fay ME, Ferris BG. Pulmonary function between 6 and 18 years of age. *Pediatr Pulmonol*. (2010) 15:75–88. doi: 10.1002/ppul.1950150204
  13. Tan WC, Bourbeau J, Hernandez P, Chapman K, Cowie R, FitzGerald MJ, et al. Canadian prediction equations of spirometric lung function for Caucasian adults 20 to 90 years of age: results from the Canadian Obstructive Lung Disease (COLD) study and the lung health canadian environment (LHCE) study. *Canad Respirat J Canad Thorac Soc*. (2016) 18:321. doi: 10.1155/2011/540396
  14. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med*. (1999) 159:179–87. doi: 10.1164/ajrccm.159.1.9712108
  15. Gutierrez C, Ghezzi RH, Abboud RT, Cosio MG, Dill JR, Martin RR, et al. Reference values of pulmonary function tests for Canadian. *Caucasians Can Respir J*. (2004) 11:414–24. doi: 10.1155/2004/857476
  16. Tan WC, Bourbeau J, FitzGerald JM, Cowie R, Chapman K, Hernandez P, et al. Can age and gender explain the variation in COPD rates across large urban cities? a population-study in Canada. *Int J Tubercul Lung Dis*. (2011) 2:211. doi: 10.5588/ijtld.11.0211
  17. Morris JE, Koski A, Johnson LC. Spirometric standards for healthy nonsmoking adults. *Am Rev Respir Dis*. (1971) 103:57–67.
  18. Quanjer PH, Enright PL, Miller MR, Stocks J, Ruppel G, Swanney MP, et al. The need to change the method for defining mild airway obstruction. *Eur Respir J*. (2011) 37:720–2. doi: 10.1183/09031936.00135110
  19. Cukic V, Lovre V, Ustamujic A. The changes of pulmonary function in copd during four-year period. *Materia Socio Medica*. (2013) 25:88–92. doi: 10.5455/msm.2013.25.88-92
  20. Rossi A, Ganassini A, Tantucci C, Grassi V. Aging and the respiratory system. *Med Clin North Am*. (1983) 67:419–31.
  21. Buist AS, McBurnie MA, Vollmer WM, Gillespie S, Burney P, Mannino DM, et al. BOLD Collaborative Research Group. international variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet*. (2007) 370:741–50. doi: 10.1016/S0140-6736(07)61377-4
  22. Fukuchi Y, Nishimura M, Ichinose M, Adachi M, Nagai A, Kuriyama T, et al. COPD in Japan: the Nippon COPD epidemiology study. *Respirology*. (2004) 9:458–65. doi: 10.1111/j.1440-1843.2004.00637.x
  23. Sharma G, Hanania NA, Shim YM. The aging immune system and its relationship to the development of chronic obstructive pulmonary disease. *Proc Am Thorac Soc*. (2009) 6:573–80. doi: 10.1513/pats.200904-022RM
  24. Lynch D. Progress in Imaging COPD, (2004). - 2014. *Chronic Obstruct Pulmon Dis J Copd Found*. (2014) 1:73. doi: 10.15326/jcopdf.1.1.2014.0125
  25. Estépar RS, Jr., Kinney GL, Black-Shinn JL, Bowler RP, Kindlmann GL, Ross JC, et al. Computed tomographic measures of pulmonary vascular morphology in smokers and their clinical implications. *Am J Respirat Critic Care Med*. (2013) 188. doi: 10.1164/rccm.201301-0162OC
  26. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RG, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Euro. J. Cancer: Offic. J. Euro. Organiz. Res. Treat. Cancer*. (2012) 12:36. doi: 10.1016/j.ejca.2011.11.036
  27. Chang R, Qi S, Yue Y, Zhang X, Song J, Qian W. Predictive radiomic models for the chemotherapy response in non-small-cell lung cancer based on computerized-tomography images. *Front Oncol*. (2021) 2021:2548. doi: 10.3389/fonc.2021.646190
  28. Huang L, Lin W, Xie D, Yu Y, Cao H, Liao G, et al. Development and validation of a preoperative CT-based radiomic nomogram to predict pathology invasiveness in patients with a solitary pulmonary nodule: a machine learning approach, multicenter, diagnostic study. *Euro Radiol*. (2022) 2022:1983–96. doi: 10.1007/s00330-021-08268-z
  29. Au RC, Tan WC, Bourbeau J, Hogg JC, Kirby M. Impact of image pre-processing methods on computed tomography radiomics features in chronic obstructive pulmonary disease. *Phy Med Biol*. (2021) 66:245015. doi: 10.1088/1361-6560/ac3eac
  30. Yun J, Cho Y H, Lee S M, et al. Deep radiomics-based survival prediction in patients with chronic obstructive pulmonary disease. *Sci Rep*. (2021) 11:1–9. doi: 10.1038/s41598-021-94535-4
  31. Au RC, Tan WC, Bourbeau J, Hogg JC, Kirby M. Radiomics Analysis to Predict Presence of Chronic Obstructive Pulmonary Disease Symptoms Using Machine Learning[M]//TP121. TP121 COPD: FROM CELLS TO THE CLINIC. *American Thoracic Soc*. (2021) 2021:A4568–A4568. doi: 10.1164/ajrccm-conference.2021.203.1\_MeetingAbstracts.A4568
  32. Liang C, Xu J, Wang F, Chen H, Tang J, Chen D, et al. Development of a Radiomics Model for predicting COPD Exacerbations Based on Complementary Visual Information[M]//TP41. TP041 DIAGNOSIS AND RISK ASSESSMENT IN COPD. *American Thoracic Soc*. (2021) 2021:A2296–A2296. doi: 10.1164/ajrccm-conference.2021.203.1\_MeetingAbstracts.A2296
  33. Wu G, Ibrahim A, Halilaj I, Leijenaar RT, Rogers W, Gietema HA, et al. The Emerging Role of Radiomics in COPD and Lung Cancer. *Respiration*. (2020) 99:1–9. doi: 10.1159/000505429
  34. Hofmanninger J, Prayer F, Pan J, Röhrich S, Prosch H, Langs G. Automatic lung segmentation in routine imaging is a data diversity problem, not a methodology problem. *Euro Radiol Experiment*. (2020) 4:1–13. doi: 10.1186/s41747-020-00173-2
  35. Yang Y, Li Q, Guo Y, Liu Y, Li X, Guo J, et al. Lung parenchyma parameters measure of rats from pulmonary window computed tomography images based on ResU-Net model for medical respiratory researches. *Mathematic Biosci Eng*. (2021) 18:4193–4211. doi: 10.3934/mbe.2021210
  36. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
  37. Yang Y, Guo Y, Guo J, Gao Y, Kang Y. “A method of abstracting single pulmonary lobe from computed tomography pulmonary images for locating COPD,” In *Proceedings of the Fourth International Conference on Biological Information and Biomedical Engineering* (2020), pp. 1–6.
  38. Yang GZ, Hansell DM, CT. image enhancement with wavelet analysis for the detection of small airways disease. *Transact Med Imaging*. (1997) 16:953–61. doi: 10.1109/42.650893
  39. Birring SS, Peake MD. Symptoms and the early diagnosis of lung cancer. *Thorax*. (2005) 60:268–9. doi: 10.1136/thx.2004.032698
  40. Neyenssac F. Contrast Enhancement Using the Laplacian-of-a-Gaussian Filter. *Cvgip Graphicmodelimage Process*. (1993) 55:447–63. doi: 10.1006/cgip.1993.1034
  41. Shi Z, Bai J, He L, Nakamura T, Yao Q, Itoh H. “A method for enhancing lung nodules in chest radiographs by use of LoG Filter[C]// (2009),” In *2nd International Congress on Image and Signal Processing* (New York, NY: IEEE) (2009).
  42. Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Statistic Soc Series B (Methodological)*. (2020) 58:267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x
  43. Lin H, Zelterman D. Modeling Survival Data: Extending the Cox Model. New York: Springer; (2000).
  44. Liang W, Yao J, Chen A, Lv Q, Zanin M, Liu J, et al. Early triage of critically ill COVID-19 patients using deep learning. *Nature Commun*. (2020) 11:1–7. doi: 10.1038/s41467-020-17280-8
  45. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol*. (2018) 18:1–12. doi: 10.1186/s12874-018-0482-1
  46. Cox DR. Regression models and life-tables. *J Roy Statistic Soc Series B (Methodological)*. (1972) 34:187–202. doi: 10.1111/j.2517-6161.1972.tb00899.x
  47. Morice AH, Celli B, Kesten S, Lystig T, Tashkin D, Decramer M. COPD in young patients: A pre-specified analysis of the four-year trial of tiotropium (UPLIFT). *Respiratory Med*. (2010) 104:1659–67. doi: 10.1016/j.rmed.2010.07.016
  48. Sanchez-Salcedo P, Divo M, Casanova C, Pinto-Plata V, de-Torres JP, Cote C, et al. Disease progression in young patients with COPD: rethinking the Fletcher and Peto model. *Euro Respirat J*. (2014) 44:613. doi: 10.1183/09031936.00208613

49. Gloeckl R, Marinov B, Pitta F. Practical recommendations for exercise training in patients with COPD. *Euro Respirat Rev.* (2013) 22:178–86. doi: 10.1183/09059180.00000513
50. López-Campos JL, Gallego EQ, Hernández LC. Status of and strategies for improving adherence to COPD treatment. *Int J Chronic Obstruct Pulmonary Dis.* (2019) 14:1503. doi: 10.2147/COPD.S170848

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of

the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yang, Li, Guo, Liu, Li, Yang, Wang, Zeng, Duan, Chen, Chen, Li, Zhao, Chen and Kang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Brain Tumor Image Segmentation Method Based on Quantum Entanglement and Wormhole Behaved Particle Swarm Optimization

Tianchi Zhang<sup>1</sup>, Jing Zhang<sup>2,3\*</sup>, Teng Xue<sup>2,3</sup> and Mohammad Hasanur Rashid<sup>2,3</sup>

<sup>1</sup> School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing, China, <sup>2</sup> School of Information Science and Engineering, University of Jinan, Jinan, China, <sup>3</sup> Shandong Provincial Key Laboratory of Network-Based Intelligent Computing, Jinan, China

## OPEN ACCESS

### Edited by:

Yong Xu,  
Harbin Institute of Technology, China

### Reviewed by:

R. Shantha Selva Kumari,  
Mepco Schlenk Engineering  
College, India  
Zhifan Gao,  
Sun Yat-sen University, China

### \*Correspondence:

Jing Zhang  
ise\_zhangjing@ujn.edu.cn

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

**Received:** 13 October 2021

**Accepted:** 17 March 2022

**Published:** 10 May 2022

### Citation:

Zhang T, Zhang J, Xue T and  
Rashid MH (2022) A Brain Tumor  
Image Segmentation Method Based  
on Quantum Entanglement and  
Wormhole Behaved Particle Swarm  
Optimization. *Front. Med.* 9:794126.  
doi: 10.3389/fmed.2022.794126

**Purpose:** Although classical techniques for image segmentation may work well for some images, they may perform poorly or not work at all for others. It often depends on the properties of the particular image segmentation task under study. The reliable segmentation of brain tumors in medical images represents a particularly challenging and essential task. For example, some brain tumors may exhibit complex so-called “bottle-neck” shapes which are essentially circles with long indistinct tapering tails, known as a “dual tail.” Such challenging conditions may not be readily segmented, particularly in the extended tail region or around the so-called “bottle-neck” area. In those cases, existing image segmentation techniques often fail to work well.

**Methods:** Existing research on image segmentation using wormhole and entangle theory is first analyzed. Next, a random positioning search method that uses a quantum-behaved particle swarm optimization (QPSO) approach is improved by using a hyperbolic wormhole path measure for seeding and linking particles. Finally, our novel quantum and wormhole-behaved particle swarm optimization (QWPSO) is proposed.

**Results:** Experimental results show that our QWPSO algorithm can better cluster complex “dual tail” regions into groupings with greater adaptability than conventional QPSO. Experimental work also improves operational efficiency and segmentation accuracy compared with current competing reference methods.

**Conclusion:** Our QWPSO method appears extremely promising for isolating smeared/indistinct regions of complex shape typical of medical image segmentation tasks. The technique is especially advantageous for segmentation in the so-called “bottle-neck” and “dual tail”-shaped regions appearing in brain tumor images.

**Keywords:** image segmentation, quantum entanglement, wormhole behavior, QPSO, QWPSO

## INTRODUCTION

The accurate analysis of medical images, especially brain tumors, is essential in reducing clinical mortality rates. Brain tumors grow quickly and often appear as highly irregular and “complex shaped” in medical images. This characteristic tumor appearance is called a “dual tail sign” or “bottle-neck.” Usually, it occurs close to a meningioma, and the dual tail feature appears due to thickening, enhancement, and double distal tapering of the tumor in this area. Existing medical image segmentation methods often wholly ignore the smeared region or require long processing periods to obtain more accurate segmentation. However, precise medical image segmentation is essential in helping to better recognize and diagnose tumors. Thus, there is a pressing need for improved methods to help solve challenging tumor image segmentation problems. Many researchers believe that quantum theory offers a mysterious key that may help us interpret our future world (1–4). Significantly, the practical image segmentation method combines quantum theory with artificial algorithms (5–8), such as Quantum-behaved particle swarm optimization (QPSO). QPSO has been shown to perform well in clustering and image segmentation tasks involving complex object shapes (9). However, prior work with QPSO has not considered highly complex and irregular forms or indistinct smearing problems that are apparent in difficult medical image segmentation tasks.

Can we analyze the cause of the complex shape of brain tumors from the microscopic process and mechanism formation of brain tumor cells? What is the relationship between the complex shape of a brain tumor and the internal microscopic structure between the tumor cells?

In 2017, Maldacena and Susskind (10) reported that the fimbriae (finger-like threads appearing on bacteria) are visible as crooked tentacles, dragging DNA into the bacteria in a way that was somewhat analogous to the action of a wormhole between black holes. Particularly, it has been shown that a wormhole-like process exists in the synthesis of cells (11). Our previous research patent for invention (ZL200810209785.8) on the protein folding process proved that  $f(x) = \cos(nx) + \sin(nx)$  could represent the oscillation of protein folding in a cell (12). The  $f(x) = \cos(nx) + \sin(nx)$  is also a representation of a sin curve when proteins are in the folding process of forming a cell. It prompts the exciting question of whether there could be a wormhole effect amongst tumor cells.

Moreover, we might apply wormhole theory to improve QPSO for solving indistinct or highly complex ‘bottle-neck,’ smeared, or irregular shaped segmentation problems in medical images. To date, most wormhole physics has been applied in computing parallel connection problems or network attack prevention tasks and, to a limited extent, in ortholog prediction algorithms and gene clustering (13). However, the contribution and highlight of our research objective is to validate the application of wormhole theory to QPSO by proposing a novel method of quantum and wormhole-behaved particle swarm optimization (QWPSO) for complex medical image segmentation.

The rest of the study is organized as follows. In section Method, we first discuss the possibility of inducing wormhole

behavior to achieve the complex shape in image segmentation. Then, we present the theory of wormhole path measurement and analyze the difference between wormhole path measurement and the Delta potential well measurement in the QPSO method. Finally, we put forward a novel segmentation method that we call QWPSO for complex shapes of brain tumors based on the wormhole path measurement. In section Results and Discussion, we apply the QWPSO algorithm to segment medical images, especially the complex shaped brain tumor images, and implement comparative experiments. Finally, some conclusions are given in the last section.

## METHOD

This section first discusses the possibility of similarity between wormhole behavior and the complex shape of brain tumor segmentation to primarily determine the tumor contour of “bottle-neck” and “dual-tailed.” Secondly, it analyzes QPSO algorithm and finally proposes the QWPSO algorithm to improve QPSO.

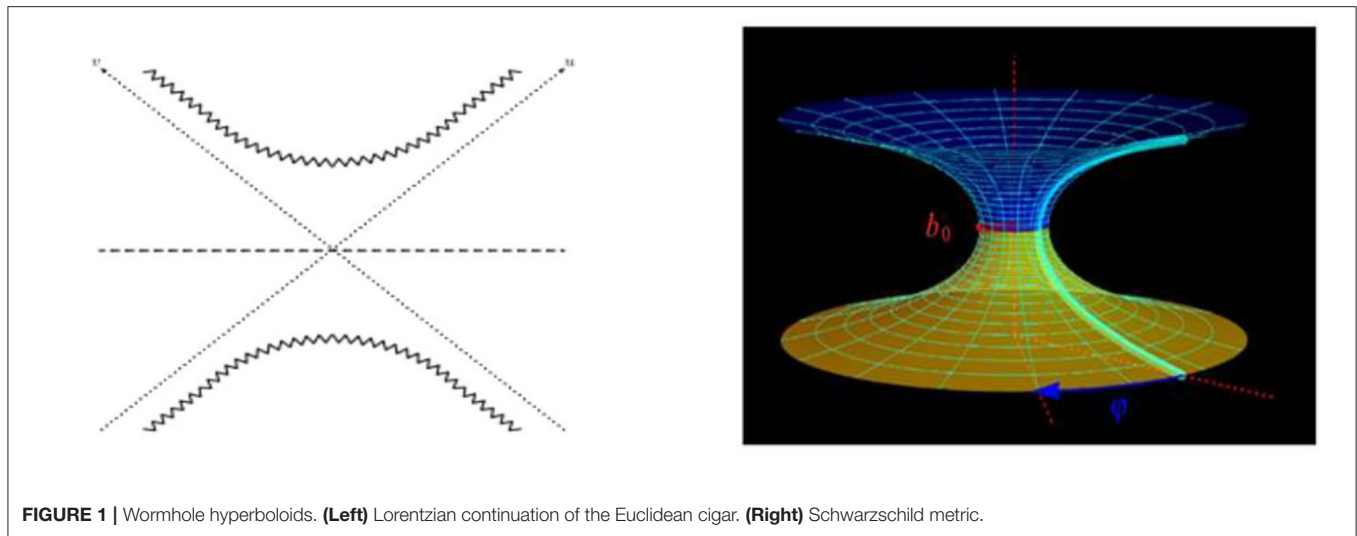
### The Possibility of Inducing From Wormhole Behavior to Brain Tumor Contour

The concept of “wormhole” was first proposed by Austrian physicist Ludwig Frum (14) in 1916 and was perfected by Einstein and Nathan Rosen (15) in 1935. Therefore, “wormhole” is also known as “Einstein-Rosen Bridge” (16). Worm-holes, commonly known as wormholes in space-time, is thought to be possibly curved shortcuts in the universe that allow objects to instantly travel through space and time. Figuratively speaking, a wormhole is a space tunnel connecting two distant spaces and times like a whirlpool in an ocean, ubiquitous but fleeting (17). These space-time vortices are caused by a combination of star rotation and gravity. Just as a whirl, it can make a part of a body of water closer to the bottom or make two parts of space that are relatively far apart become very close in an instant.

New research found that a wormhole’s super strong magnetic field can keep it open by relying on a Phantom matter (18). Scientists believe that instead of a positive case, which produces energy, it also has a negative mass, sucking up all the energy around it. Because exotic matter has both positive energy and negative mass, it can create repulsive effects to prevent the wormhole from closing, thus stabilizing the “wormhole” energy field. In 2013, two distinguished theoretical physicists, Maldesina and Sarskander, explored the behavior of quantum entanglement in the macroscopic area. In their study, they boldly proposed the following:  $EPR = ER$ . EPR refers to quantum entanglement (19), and ER is short for wormhole (20). This puzzling formula links microscopic and macroscopic phenomena, pointing out that the exotic matter that stabilizes the wormhole energy field is quantum entanglement.

Wormhole features according to Maldesina and Sarskander (2013): physical space is by a space of two identical sheets, a particle being represented by a “bridge” connecting these sheets. The details are:

1. Wormholes are fragile and tiny (21).



**FIGURE 1 |** Wormhole hyperboloids. **(Left)** Lorentzian continuation of the Euclidean cigar. **(Right)** Schwarzschild metric.

2. The wormhole formation and wormhole stabilization process depends on a unique effect of exotic matter, which is the entangled state of quantum entanglement (22).
3. Changes in the magnetic field cause wormholes (23).
4. The shape of the wormhole is derived from the rotation of a baseline, and the baseline is hyperbolic. The embedded curved space is a hyperboloid (24).

Worm-holes can be described as the Lorentzian continuation of the Euclidean cigar. The Schwarzschild metric, shown in **Figure 1**, is the most famous wormhole model (25). It is a two-sided eternal black hole. The horizons are the diagonal dotted lines. The past and future singularities are the zigzag hyperbolas at the bottom and top (24).

### Schwarzschild Metric Wormhole Model Equation

$$ds^2 = -c^2 dt^2 + \frac{dr^2}{1 - \frac{b_0^2}{r^2}} + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \quad (1)$$

Where  $c$  is the speed of light,  $r$  is the radius of the throat part of the wormhole,  $\theta$  is the zenith angle between positive  $z$ -axis, and  $\varphi$  is the azimuth angle between the positive  $X$ -axis in the spherical coordinate system. In the two-dimensional static spherically symmetric solution of a plane, Equation (1) can be simplified as:

$$ds^2 = \frac{dr^2}{1 - \frac{b_0^2}{r^2}} + r^2 d\varphi^2 \quad (2)$$

Meanwhile, the equation of the embedded surface is:

$$z(r) = \pm b_0 \ln \left[ \frac{r}{b_0} + \sqrt{\left( \frac{r}{b_0} \right)^2 - 1} \right] \quad (3)$$

where  $b_0 = 2GM$  (26),  $M$  is the object's mass,  $G$  is the universal gravitational constant, and  $r$  is the radius of the throat of the

hyperbolic neck. Specifically,  $r$  is the distance of the curve represented by a radius line. At the same time, the wormhole's hyperbolic Equation (3) describes the spatial shape of the entire hyperboloid obtained by rotating numerous radius lines.

Roman Konoplya (27), a research associate at the People's Friendship University of Russia (RUDN) Institute for Gravity and Cosmology, proposed that the shape and mass of the wormhole can be calculated from the displacement value and the range of high-frequency gravitational waves. He first mathematically described the shape of a symmetrical wormhole based on its range of fluctuations. Then, using a quantum mechanical approximated the wormhole, we therefore simplified Equation (3) into Equation (4):

$$z(r) = \pm b_0 \ln(a) \quad (4)$$

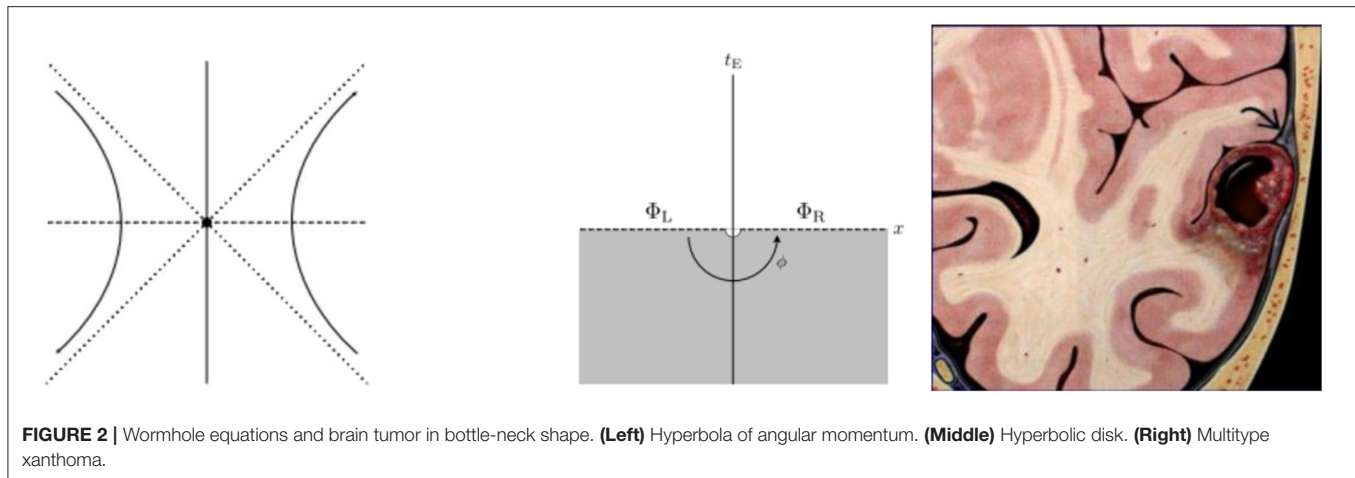
We used a hyperbolic disk to detail the equation for a hyperbola with angular momentum in all directions (4). The coefficient of  $a$  can be replaced by  $\Delta\theta/2$  and  $b_0$  can be replaced by  $2/\zeta$ . Hence, Equation (4) is written in detail as the following Equation (5):

$$x(r) = r + r' + (2/\zeta) \ln(\Delta\theta/2) \quad (5)$$

Some shapes of brain tumors look like 'bottle-necks' as hyperbolic shapes, such as the Multitype xanthoma shown in **Figure 2**. Is there any relationship between the shape of brain tumors and wormholes? Let us analyze in detail below.

### The Cause of Brain Tumor

Mounting evidence specifically from long-term mobile phone use (cumulative exposure) shows that it can cause brain tumors, including glioma and acoustic neuroma, and appreciable long-term deficits in learning abilities and memory functions. Thus, it raises public concern and compels investigation (27–32). In Morgan's view (33), many results and several epidemiology studies are consistent with radiofrequency fields from which states that mobile phones can cause brain cancer. There are many causes to increase the risk of brain cancer, such as cumulative



**FIGURE 2 |** Wormhole equations and brain tumor in bottle-neck shape. **(Left)** Hyperbola of angular momentum. **(Middle)** Hyperbolic disk. **(Right)** Multitype xanthoma.

hours of use, long-time use, and radiated power. Saikhedkar's findings (34) indicate that extensive neurodegeneration on radio waves increased the unstable production of reactive oxygen. It is caused by the exhaustion of enzymatic and non-enzymatic antioxidants and increased lipid peroxidation. It indicates that radio waves increase the unstable production of reactive oxygen, causing extensive neurodegeneration in selective areas of CA1 (cornu ammonis 1), CA3 (cornu ammonis 3), DG (dentate gyrus), and the cerebral cortex. This extensive neuronal damage results in alterations in behavior related to memory and learning. Pronounced effects of electromagnetic fields may interfere with the results of laboratory tests on murine experimental models in veterinary or biomedical research (35). Electromagnetic radiations may result in chromosomal aberrations by either illegitimate recombination events or reduction of functionality of nonhomologous end-joining (36). An association with high-dose ionizing radiation and brain tumors has been observed in A-bomb studies, nuclear-test fallout data, therapeutic radiation for cancer and benign conditions, and occupational and environmental studies (37). Information is somewhat limited regarding the specific histologic type of tumor, particularly for increasing brain tumor risk. In 2015, the Swedish team and 13 other countries reported significant risks associated with gliomas from exposure to electromagnetic radiation, which was reclassified by the International Agency for Research on Cancer (IARC) from group 2B (probable) to Group 2A (highest level) (38, 39). Researchers found that mobile phone users had an eight-fold increased risk of brain tumors among people exposed to electromagnetic radiation in cities (40, 41).

### The Similarities Between Wormholes and Brain Tumors

1. Both wormholes and brain tumors are caused by magnetic fields regardless of the super-strong magnetic field or electromagnetic radiation.
2. Their origin is the same because wormhole formation depends on quantum entanglement, while brain tumors are microscopic cells formed from mutated particles.

3. The formation processes of wormhole and brain tumors are all unstable as both are caused by exposure to magnetic fields.
4. Some brain tumors are the "bottle-neck" shapes that are the same as the hyperbolic shape of a wormhole

As for the cause, the initial formation process and the shape of brain tumors and wormholes are all similar. Hence, in the next section, we propose a wormhole behavior method to segment the "bottle-neck" shaped brain tumor.

### QPSO

The particle swarm optimization (PSO) method originally comes from a simulation of the social behavior of birds while flocking. However, PSO is not a global convergence-guaranteed algorithm. This is because at each iteration, the particles are restricted to a finite search space. Alternatively, the QPSO approach is one in which individual particles are assumed to have quantum behavior. QPSO is based on the quantum theory of a Delta potential well which offers a robust global searching ability (42, 43). Furthermore, the particles in QPSO can appear anywhere during the iterations, thus enhancing the population diversity.

In a Delta potential well, QPSO particles, in the process of optimization, move around the center area of the Delta potential toward the best position  $P$  for which the quantum potential  $V(x_{id})$  is expressed as  $V(x_{id}) = -\lambda\delta(x_{id} - p_{id})$ , where  $\lambda$  is weight,  $\delta(x_{id} - p_{id})$  is the Dirac delta function, and  $y_{id} = x_{id} - p_{id}$ . For the calculation of the particle's fitness values, we must know the exact particle position of  $x_{id}$ . However, we only know the probability density function of  $Q(y_{id})$  from the quantum state of each particle  $y_{id}$  as shown below:

$$Q(y_{id}) = |\psi(y_{id})|^2 = \frac{1}{L} e^{\frac{-2|y_{id}|}{L}} \quad (6)$$

where  $L = h^2/m\gamma$ ,  $\gamma$  is the intensity of the potential well,  $m$  is the particle mass, and  $h$  is Planck's constant. As a given particle moves toward the potential well's center,  $\psi(y_{id})$  is the spin field operator, while the quantum state function  $Q(y_{id})$  represents the location of a particle probabilistically. To make the wave collapse to an actual state for each particle, we must use a method to

estimate the position of the particles (44, 45). Employing the Monte Carlo random simulation (46), it is assumed that  $s$  is a lucky number within the range of  $(0, 1/L)$ , that is:

$$s = \frac{1}{L} \text{rand}(0, 1) = \frac{1}{L}u, \text{ and } u = \text{rand}(0, 1) \quad (7)$$

Take Equation (6) into the random number Equation (7),  $s = \frac{1}{L}e^{-\frac{2|y_{id}|}{L}}$ , and  $u = e^{-\frac{2|y_{id}|}{L}}$ . Consequently,  $y_{id} = \pm \frac{L}{2} \ln(1/u)$  and  $y_{id} = x_{id} - p_{id}$ . Therefore, the estimated position of the particle  $x_{id}$  can be obtained by the following prototype:

$$X_{id} = P_{id} \pm \frac{L}{2} \ln\left(\frac{1}{\mu}\right) \quad (8)$$

where  $L$  is the characteristic length of the potential well and  $\mu$  indicates the random value between 0 and 1 that represents the arbitrary distance between particles in the quantum potential well.  $P_{id}$  is the best position of the particle.

Suppose  $P = (P_1, P_2, \dots, P_M)$ , then the particles coordinates of  $P$  is given by:

$$P = (\varphi_1 \times P_{id} + \varphi_2 \times P_{gd}) / (\varphi_1 + \varphi_2) \quad (9)$$

$$\text{Mbest} = \frac{\sum_{i=1}^M P(t)}{M} \quad (10)$$

where  $\varphi_1 = \text{rand}(0, 1)$ ,  $\varphi_2 = \text{rand}(0, 1)$ ,  $P_{id}$  represents the  $i$ th components of the personal best position of the particle, and  $P_{gd}$  represents the global best position of the population. Mbest is the mean best position. The following iterative step is defined as the local best position of all particles on average and is calculated as follows:

If the random digital  $\mu > 0.5$ ,

$$x(t+1) = P - \alpha \cdot |\text{Mbest} - x(t)| \cdot \ln(1/\mu) \quad (11)$$

If the random digital  $\mu \leq 0.5$ ,

$$x(t+1) = P + \alpha \cdot |\text{Mbest} - x(t)| \cdot \ln(1/\mu) \quad (12)$$

Where  $\alpha$  is the expansion coefficient of the speed in controlling convergence, and it represents the maximum number of iterations.

## The QWPSO Method

Because particles in QPSO move around the central area of the Delta potential well, the existing QPSO approach, when applied to long-range searches such as when two regions are far apart, fails to segment well (47). However, there exists the notion of a wormhole in quantum theory. It offers an unusual correlation between particles, wherein actions performed on one particle immediately affect another reverse particle no matter how far apart the two particles are. We therefore propose a new quantum and QWPSO method, the details of which now follow. All nodes exist in a metric space, where distance abstracts to node similarities (48, 49). Hence, more similar nodes are closer in the area, and more

alike or close nodes are more likely to be connected. Thus, particle optimization consists of links with the probability that decreases with the hidden distance. It gives two metric spaces between each pair of nodes: observable and hidden. Visible teams are joined with neighborhood nodes by entanglement (50, 51), while remote pairs can be expressed as a kind of wormhole.

Hence, we conclude that the features of the wormhole metric are as follows:

- All nodes exist in a metric space.
- The separation distance in this space represents one way of describing the similarity of the node. The more similar the nodes, the closer in the area they appear. Worm-holes link the other measure of similarity between nodes.
- The network consists of wormhole links. These exist with the probability that decreases with the hidden distance. Thus, more similar/close nodes are more likely to be connected.
- Worm-holes link long-distance nodes as a consequence of their negative curvature.
- A node forwards information to its neighbor closest to the destination in the wormhole space.
- Clustering is a consequence of the metric property of the wormhole spaces.

Worm-holes in Schwarzschild's solution form naturally in the cosmos, as it contains no matter and is merely full of curved space-time (52). Therefore, wormhole paths are asymptotically the shortest. However, many wormhole paths are successful depending on the image space geometry (51–59). Consequently, we put forward the measure that the wormhole is Hyperbolic in shape.

## The Novel Wormhole Measure of Hyperbolic Path

Assuming a wormhole is a hyperbolic disc, we present the novel hyperbolic wormhole equation as of  $N = ce^{R/2}$ , where  $R$  is the radius,  $N$  is the number of nodes in the network, and  $c$  controls its average degree. The node distribution of uniform angular density is  $\rho_\theta(\theta) = 1/(2\pi)$ , where the range of  $\theta$  is from 0 to  $57.32$ , namely,  $\theta \leq 360/2\pi$ . The node degree at a distance  $r$  from the disc center in an exponential radial density is  $\rho(r) = \sinh r / (\cosh R - 1) \approx e^{r-R}$ , and a simple approximation,  $\rho(r) \approx (4c/\pi)e^{(R-r)/2} \approx e^{-\zeta r/2}$ , connects each pair of nodes located at  $(r, q)$  and  $(r', q')$ , for which the connection probability is:  $p = e^{\zeta(x-R)/2}$ .

The wormhole measure of hyperbolic path  $x$  is as follows:

$$x = r + r' + (2/\zeta)\ln(\Delta\theta/2) \quad (13)$$

where the range of  $\Delta\theta$  is:  $0 < \Delta\theta < 57.32$  and  $\zeta$  indicates the distance coefficient. When there is a wormhole between nodes, we modify the measure of QPSO as a wormhole path measure in a hyperbolic path of QWPSO.

## QWPSO Method

The node probability distribution of the wormhole path measure  $\rho(r)$  is:

$$\rho(r) \approx e^{-\zeta r/2} \quad (14)$$

The position of a particle in the wormhole path measure is:

$$X_{iw} = P_{id} \pm (2/\zeta) \ln(\Delta\theta/2) \quad (15)$$

$P_{id}$  is the best position of the wormhole particle, and  $x(t+1)$  represents the next step for the iteration variable wormhole particle which is defined as the local best position of all particles on average.

If the angle between nodes  $\Delta\theta > 2$ , then

$$x(t+1) = P(t) - (2/\zeta) \cdot |M_{best} - x(t)| \cdot \ln(\Delta\theta/2) \quad (16)$$

On the other hand, if the angle between nodes  $\Delta\theta \leq 2$ , then

$$x(t+1) = P(t) + (2/\zeta) \cdot |M_{best} - x(t)| \cdot \ln(\Delta\theta/2). \quad (17)$$

where  $M_{best}$  is at the mean best position described as  $M_{best} = \frac{\sum_{i=1}^M P(t)}{M}$ ,  $P(t)$  represents the position of the particle  $P_{id}$  at time  $t$ , and  $M$  represents the number of particles.

## The Difference Between QPSO and QWPSO

Our proposed QWPSO method is based on a measure of entanglement and wormhole theory. Using clustering, we firstly analyze and determine the connection type, i.e., is it entanglement or wormhole? If the connection is by trap, we find particles by a random link and cluster. If there is a wormhole connection between nodes, we employ our proposed wormhole measure, Equation (13), to find the particles and then cluster them. The main difference between QPSO and QWPSO is the coefficient  $\alpha$  in equations (11), (12), and  $\zeta$  in equations (16) and (17).  $\zeta$  is related to distance, while  $\alpha$  is related to speed. It means that while every step in QWPSO has a definite path, we know where to find a random process that finds the next particle in QPSO. Therefore, the efficiency in QWPSO is higher than QPSO due to the characteristic of a definite path exiting the wormhole. This is because  $\ln(\Delta\theta/2)$  in equations (16), (17) in QWPSO, and the range of  $\Delta\theta$  is  $0 < \Delta\theta < 57.32$ . Otherwise, in QPSO, the integer random value of  $\mu$  in the function of  $\ln(1/\mu)$  is from 0 to 32767 depending on the computing power of a computer. Corresponding to the angular coordinate, the value range of  $\mu$  in QPSO is from 0 to 360. Hence, 57.32 in 360 equals to 15.9%, the running time of QWPSO is only 15.9 % of the QPSO, and the efficiency of QWPSO is higher than that of QPSO. In addition, the critical difference between the existing QPSO approach and our QWPSO method is the definition of wormhole limitations. We conclude the three definitions for an existing wormhole as follows.

The three definitions for an existing wormhole:

(These limitations are more specific to image segmentation)

- (1) The number of nodes clustered by the wormhole is not less than two, i.e., there are at least two nodes as particles;

- (2) Node positions are not in the neighborhood, but their gray values are similar;
- (3) The similarity matches the wormhole measure.

If the cluster nodes meet the three limitations, the segmentation can be done by our proposed QWPSO method.

## The Framework of the QWPSO Algorithm

As mentioned in previous sections, two distinguished theoretical physicists, Maldesina and Sarskander, explored the behavior of quantum entanglement in the macroscopic field. They boldly proposed the equation of “EPR=ER,” where EPR refers to quantum entanglement and E.R. is short for wormhole (6, 50, 60–64). The puzzling formula links microscopic and macroscopic phenomena and points out that the wormhole is caused by quantum entanglement. Inspired by this, this study presents the novel concept of seed and pixel particles. The seed particle is in quantum entanglement which exists a wormhole between each seed particle. In contrast, the pixel particle is opposite the seed particle, and there is no quantum entanglement and wormhole between the particles. Therefore, our proposed method of QWPSO consists of two sections. First, we cluster particles into seed and pixel particles. Secondly, we determine a wormhole between two seed particles by wormhole Equation (13) and segment the image using the QWPSO algorithm. Otherwise, if there are no seed particles, and therefore no wormholes between particles, the image segmentation is performed by QPSO. The detail of the QWPSO framework is shown in **Figure 3**.

1. Cluster particles into seed particles and pixel particles:

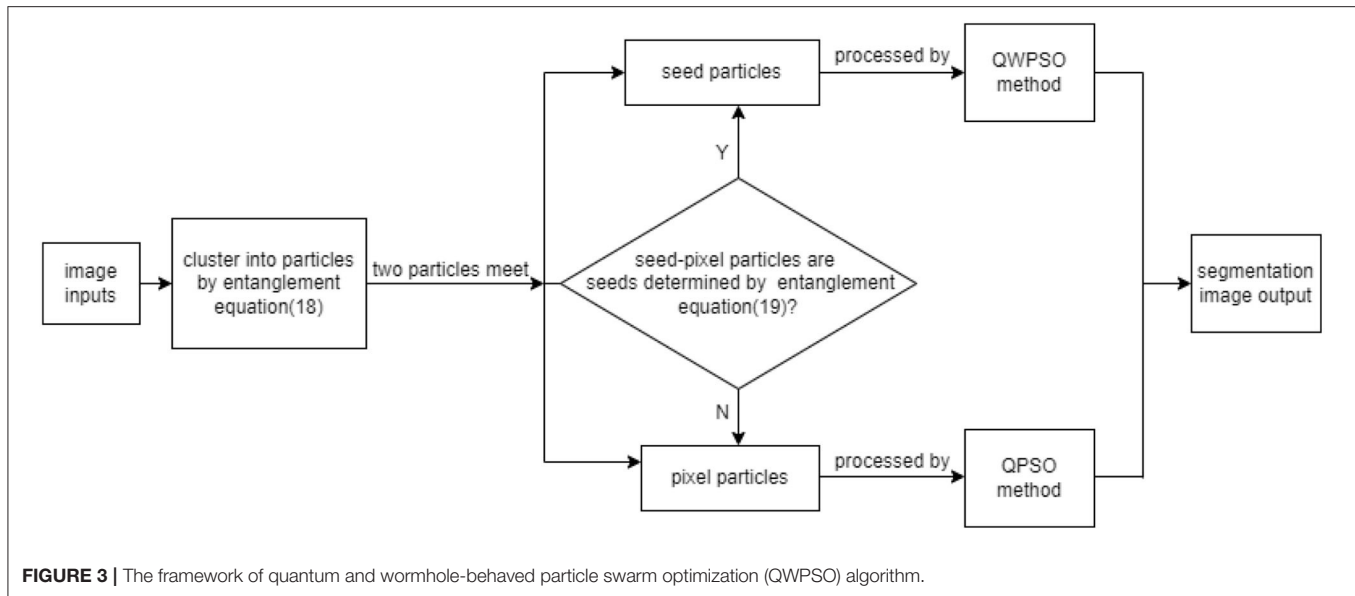
Two particles are found by a random process to determine if one particle is within the neighborhood range of the other particle. This is done by comparing the characteristics of the gray pixel value and position information between them. Assuming the two-particle positions,  $x(i, j)$  and  $x(k, l)$ , their gray pixel values are  $f_{ij}$  and  $f_{kl}$ .  $TH_o$  and  $TH_f$  are the threshold values of position variance and pixel gray value difference, where  $f$  represents the gray value difference of the two particles and  $\Delta d$  represents the root mean square difference of the particles' position. Only when two particles satisfy the entanglement Equation (18) are they considered within one cluster as seed particles. Otherwise, they are pixel particles.

$$\Delta f = |f_{ij} - f_{kl}| \leq TH_f \text{ and } \Delta d = \sqrt{(i-k)^2 + (j-l)^2} \leq TH_o \quad (18)$$

2. If a pixel particle encounters a seed particle:

If a pixel particle meets a seed particle, the gray seed value  $f_{kl}$  is replaced by the average gray value of the particles in the seed area, represented by  $\bar{f}$ . Only when two particles satisfy the entanglement Equation (19) are the two particles entangled together. They are then considered to be within one cluster as a new seed particle. The entangle equation is:

$$\Delta f = |f_{ij} - \bar{f}| \leq TH_f \text{ and } \Delta d = \sqrt{(i-k)^2 + (j-l)^2} \leq TH_o \quad (19)$$



### 3. If two seeds meet:

Find a seed particle by Equation (13). If the two seed particles meet, and there exists a wormhole between them, then the entanglement equations (8), (11), and (12) are replaced by the measure of the wormhole equations (13), (16), and (17) respectively.

### The QWPSO Algorithm

The process and flow chart of the QWPSO algorithm is shown in **Figure 4**. There are two sections in the algorithm. The left one is the seeds particle with wormhole path and the right one is pixel particles that have no wormhole path between them. According to the different paths, it will be processed with different equations.

The steps in conducting the QWPSO algorithm is listed below:

Step 1: Input the image and initialize the position vector for each particle.

Step 2: Cluster particles into seeds and pixels. In cases where two-pixel particles meet, check if one particle is within the neighborhood range of another particle by Equation (18), and then group them otherwise, go to step 5.

Step 3: In cases where pixel quantum particles meet a seed quantum particle, check whether any particle is within the neighborhood range of the seed particle using Equation (19). Then, group them. Otherwise, go to step 5.

Step 4: In cases where two seed particles meet, calculate their distance by Equation (13) and go to step 6.

Step 5: If the random digital  $\mu > 0.5$ , then calculate according to Equation (11). Otherwise, calculate using Equation (12), and then cluster particles into foreground and background regions.

Step 6: If the angle between nodes  $\Delta\theta > 2$ , then calculate according to Equation (16). Otherwise, calculate using

Equation (17), and then cluster particles into foreground and background regions.

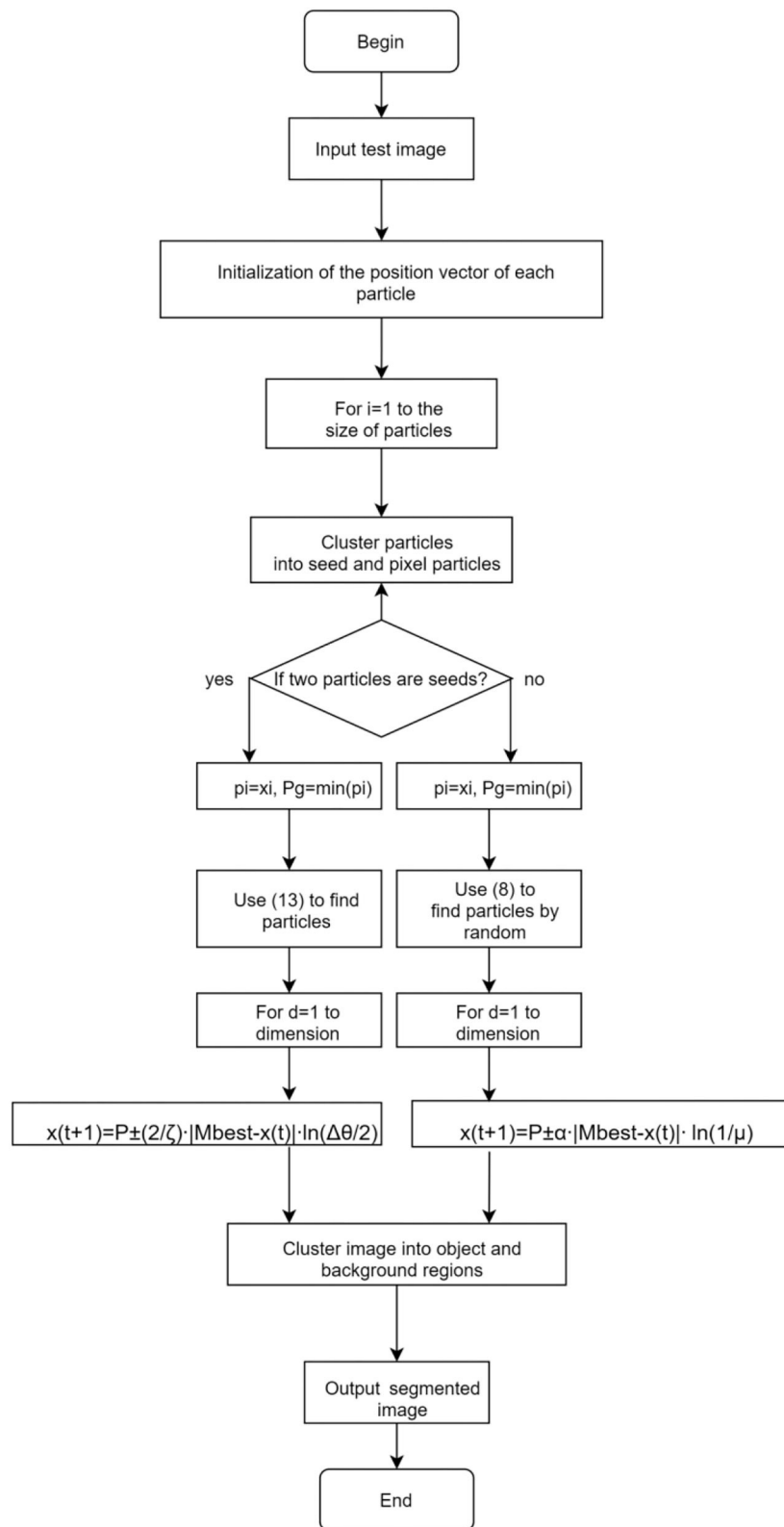
Step 7: if all particles are clustered, output the segmented image and then exit. Otherwise, return to step 2.

### The Contributions and Highlights of the QWPSO Algorithm

As for the complex so-called “bottle-neck” shapes in brain tumor image segmentation, to essentially solve the problem of ‘dual tail’ shape segmentation, we propose a novel method of QWPSO algorithm. The novelty, challenges, advantages, and limitations are as follows.

The novelty and challenge of the QWPSO algorithm:

1. The study of wormhole behavior comes from the microscopic process of DNA dragged into the bacteria. Between the tumor cells, we have sensed that fimbriae appeared as crooked tentacles to drag DNA into the bacteria, which is somewhat analogous to the action of a wormhole between black holes.
2. The study comes from the research of synthesis of cells as well. They prove that a wormhole-like process exists in the synthesis of cells.
3. The study origins from our previous research patent for invention (ZL2008 1 0209785.8) on the protein folding process proved that  $f(x) = \cos(nx) + \sin(nx)$  could represent the oscillation of protein folding in a cell. The  $f(x) = \cos(nx) + \sin(nx)$  is also a representation of a curve that looks like a worm-hole.
4. Based on the novel research of the microscopic structure of DNA into the bacteria, the wormhole-like process that existed in the synthesis of cells, and the protein curve folding process, we prompt an exciting and challenging research to discuss the relationship between the shape of the tumor and the shape of the wormhole.



**FIGURE 4 |** Flow chart of the QWPSO algorithm.

5. This study's main contributions and highlights introduce a wormhole behavior method to improve QPSO into QWPSO. First, we proposed the novel concept of seeds and pixel particles. Thus, the QWPSO consists of two sections. Then, we present all the wormhole behavior equations, frames, and algorithms for QWPSO.
6. The difference or superiority of the proposed QWPSO method compared with the existing brain image segmentation methods mainly aims to solve the segmentation problem of special-shaped tumors, especially the shape of "bottle-neck" and "dual tail" based on the similar shape between the special shaped tumors and the shape of wormhole behavior.

The advantages of the QWPSO algorithm:

1. We propose the novel wormhole measure equation applied to the method of QWPSO. The wormhole measure is represented by the hyperbolic path, with angles describing the wormhole in all directions.
2. We propose the novel framework of the QWPSO algorithm with two sections. Firstly, the coarse clustering aims to achieve two groups of particles: seed particles have wormholes, and pixel particles do not. Secondly, and key to our method, is the refined clustering by quantum entanglement and the wormhole measure equation with seed particles.
3. The wormhole theory of a hyperbolic path in QWPSO is proposed instead of a random path as in QPSO. The running efficiency of QWPSO is higher than that of QPSO.
4. The QWPSO algorithm enables more accurately a segment in complex 'bottle-neck' and indistinct shapes, typical of trailing brain tumor images in cases where other segmentation algorithms often fail.

The limitations of the QWPSO algorithm:

1. The proposed QWPSO algorithm is designed for a brain tumor with a unique shape, but in a human tumor, there are various tissues and parts with such curved shapes, such as lung, liver, spleen, etc. Next, our study extends from brain tumors to image segmentation of other organs with curved shapes.
2. The proposed method of the QWPSO algorithm should be extended to the image segmentation of particular curved shape targets in other fields besides medical images. Therefore, we will study the application of this method and expand into more research areas in future studies.

## RESULTS AND DISCUSSION

Magnetic resonance imaging and CT images are typically used to analyze medical brain images. In this section, we consider three tests for the two types of brain images. Test 1 included ten MRI brain images, including tumors of complex shape with long tails or bottle-neck contours. We wished to investigate whether our method is feasible and valuable in segmenting this challenging brain image, and determine the distance coefficient value of  $\Delta\theta$  representing the angle between nodes. Test 2 aimed to test another important CT medical image beside the image of MRI for the comparative test to examine whether


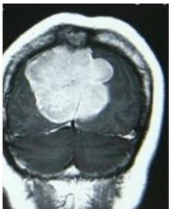
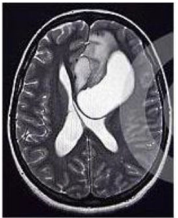
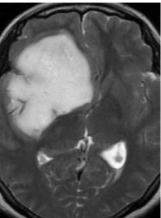
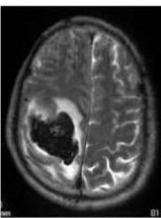










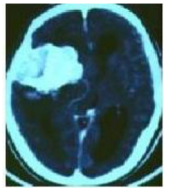
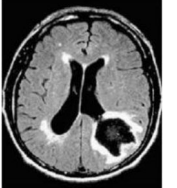
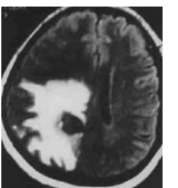
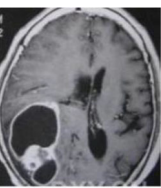
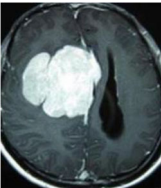










our proposed way is better than existing. There are four CT brain images for which we compare results with five current related reference methods. They are QPSO (65, 66), J. Sun cooperation quantum-behaved particle swarm optimization (SunCQPSO) (67), the Dynamic-context cooperation quantum-behaved particle swarm optimization algorithm (CCQPSO) (59), partitioned and cooperative quantum-behaved particle swarm optimization (SCQPSO) (45), and the improved quantum particle swarm optimization-intelligent fuzzy level set (IQPFLS) (8). We aimed to prove if our proposed QWPSO algorithm had better adaptability for object region shape, operational efficiency, and segmentation accuracy than QPSO and other typical competing reference methods. Test 3 is specifically for demonstrating the advantage of the proposed QWPSO method, the compared tests were implemented in 10 studies listed in references (8, 68–77), of which publication years were from 2018 to 2021.

### Test 1: MRI Brain Image Segmentation

Test 1 aimed to determine the distance coefficient value of  $\Delta\theta$  in the proposed QWPSO approach. It consisted of 10 images, and they are from the benchmark datasets of brain tumor segmentation (BRATS) (78). Their names and tumor types are Glioma 1 and 2, Occipital, Ependymoma 1 and 2, Edema, Meningioma 1 and 2, Hematoma, and Tuberculoma. We aimed to test if our proposed method, QWPSO, could segment the tumor with complex object shapes called neck and tail features. Images of Ependymoma 1, Hematoma, Tuberculoma, and Ependymoma 2 include neck features, while the others all have tails features either long or short. From observation of the segmented images, shown in **Table 1**, it can be seen that the pictures with neck features, especially Ependymoma 1, Hematoma, and Tuberculoma as segmented by our QWPSO method, have a better and more accurate contour line than those segmented by using the QPSO method. The other images with tail features segmented by QWPSO also perform better than those segmented using the QPSO method.

The quantitative evaluation parameters of the image segmentation process were Time (Running time), P (Precision), R (Recall), and F (F-measure). These were used to evaluate whether the method can achieve good results in image segmentation. The time parameters represented the running time to assess the algorithm's efficiency. P, R, and F were used to access and compare consistency, accuracy, and sensitivity, respectively. P is the fraction of retrieved relevant instances. It determines how beneficial the results are. The recall rate R is derived from our original sample, and it tells us how many positive examples in the sample were predicted to be correct. Finally, it was compared with the prediction. Therefore, P and R indicators are sometimes contradictory, so they need to be considered comprehensively by F. Specifically, P tells the accuracy, and F demonstrates the final and mixed evaluated results. The two parameters, P and F, are more critical among the P, R, and F parameters. The detailed evaluation parameters obtained from test 1 are shown in **Table 2**. The range of  $\Delta\theta$  in the 10 brain images was 0.06 to 40.00. It is within the range we

**TABLE 1 |** MRI brain image segmented by quantum-behaved particle swarm optimization (QPSO) and quantum and wormhole-behaved particle swarm optimization (QWPSO) methods.

Method	Glioma 1	Occipital	Ependymoma 1	Glioma 2	Edema
Original Image					
QPSO method					
QWPSO method (Proposed in this paper)					
	Meningioma 1	Hematoma	Tuberculoma	Ependymoma 2	Meningioma 2
Original Image					
QPSO method					
QWPSO method (Proposed in this paper)					

**TABLE 2** | Evaluation of parameters in test 1.

Image	QWPSO evaluate parameters					QPSO evaluate parameters			
	$\Delta\theta$ / rad/s	Time/s	P/%	R/%	F/%	Time/s	P/%	R/%	F/%
Glioma 1	2.60	0.849	1.0000	0.9097	0.9527	0.863	0.9401	0.9090	0.9491
Occipital	2.00	0.865	1.0000	0.9903	0.9951	0.882	0.9996	0.9877	0.9936
Ependymoma 1	4.00	0.820	1.0000	0.9941	0.9970	0.840	0.9929	0.9988	0.9958
Glioma 2	40.00	0.829	0.9966	0.9868	0.9917	0.875	0.9741	1.0000	0.9869
Edema	1.60	0.823	0.9836	0.9941	0.9888	0.885	0.8344	1.0000	0.9097
Meningiom1	0.06	0.801	1.0000	0.9973	0.9986	0.950	0.9997	0.9888	0.9983
Hematoma	2.00	0.842	0.9995	1.0000	0.9967	0.976	0.9649	1.0000	0.9821
Tuberculoma	0.60	0.882	1.0000	0.9891	0.9945	0.922	0.9891	1.0000	0.9945
Ependymoma 2	2.00	1.144	0.9936	0.9256	0.9584	1.211	0.9048	1.0000	0.9500
Meningiom2	2.00	0.888	1.0000	0.9942	0.9971	1.000	1.000	0.9929	0.9964

predicted and is less than  $360/2\pi$ , which is within the scope of 0 to 57.32.

The running time (Time) of our proposed method QWPSO was also less than the QPSO method. This is because our approach ran an angle  $\Delta\theta$  within a changing range from in each step, while the QPSO ran a random value in every step. This meant that the changing rise in QWPSO had higher efficiency than the haphazard approach used in the QPSO method. Test 1 has also shown that the run time of QWPSO was less in the range of 84 to 98% than that of the QPSO. The evaluation parameters of P, R, and F, especially parameter P, for our QWPSO policy were optimal, i.e., they were all greater than the value obtained by QPSO. Moreover, the parameter R for QWPSO was better than that of QPSO. Furthermore, the parameter F was obtained from our proposed method, QWPSO, which was better than the method QPSO, except for the value 0.9945 for Tuberculoma, which is equal to both QPSO and QWPSO. In summary, for the essential evaluation parameters of Time, P, and F, our QWPSO method outperforms that which is obtained when using the QPSO method. This means that our process of QWPSO offers higher efficiency and greater accuracy than the QPSO method in the ten complex tumor-shaped medical image segmentation tasks.

## Test 2: CT Brain Image Segmentation

Test 2 aimed to test another important C.T. medical image beside the image from MRI. It explored the results compared with the latest improved QPSO methods: SunCQPSO, CCQPSO, SCQPSO, and the IQPFLS method. We considered four complex-shaped tumors that typically appear in CT medical tumor images, all of which included shapes with a long tail known as a 'dual tail' and the so-called 'bottle-neck' feature. The test CT images comprised CT201.86, CT201.136, CT201.29, and CT200.2, which were all taken from reference (78). The segmented images are shown in Table 3.


In Table 3, it may be observed that the images segmented by our proposed method of QWPSO have a precise contour, especially at the region with the shape of long-tail known as the 'dual tail' and around the so-called 'bottle-neck' region. For

example, in image CT201.86, there is a 'dual tail' in the object region. Thus, the trail is clear and distinct when segmented by our proposed method. However, the trail is comparatively fuzzy and unclear when segmented by competing methods like QPSO, SunCQPSO, CCQPSO, SCQPSO, and the IQPFLS method. Similar results were obtained for images CT201.29 and CT200.2. In the case of image CT201.136, there was a bottle-neck shape within the object region, for which our proposed method has achieved a distinct perfect bottle-neck curve in the object contour. On the other hand, neither a distinct 'bottle-neck' nor angle was segmented by the other methods.

The evaluated parameters of P, R, and F are shown in Table 4. We use  $\Delta\theta$  of 0.6 in our QWPSO method to segment the images. From Table 4, we can see that all the evaluation parameters (P, R, and F) show an improved performance for our QWPSO method compared with that obtained using QPSO, SunCQPSO, CCQPSO, SCQPSO, and IQPFLS. As for the mean values for R in our approach, QWPSO is all in the range of 0.7423 to 0.9990. Moreover, it is greater than the range of 0.3116 to 0.8876 obtained using the other five methods. The mean value R increased from 1.12 to 2.382 times. As for the value P, our QWPSO ways are all better than the compared methods except for our QWPSO and IQPFLS process in Image CT200.2 which had the same value of 0.7546. This is because there was no distinct 'bottle-neck' shape within the object region in Image CT200.2. The value of F is the combination of precision P and recall rate R. This reflects the total score of image segmentation. Our method QWPSO in the range of 0.7484 to 0.9995 is greater than the range of 0.3988 to 0.8171 obtained by the other five methods. The mean value F has increased 1.876 to 1.223 times. Namely, our proposed method QWPSO has significant advantages, especially in distinct 'bottle-neck' shape images. Furthermore, our approach's running time ranges from 0.810 to 0.900/s, which is less than needed for any of the other four methods.

In summary, based on the two tests, we conclude that our proposed method of QWPSO offers an advantage when applied to typical MRI and CT medical image segmentation tasks, especially for segmenting complex indistinct tumor shapes. Compared with the existing methods of QPSO, SunCQPSO,

**TABLE 3 |** Comparison segmentation test using CT brain images.

Method	Image CT201.86	Image CT201.136	Image CT201.29	Image CT200.2
Original image				
QPSO				
SunCQPSO				
CCQPSO				
SCQPSO				
IQPFLS				
QWPSO (Our proposed method)				

**TABLE 4 |** Evaluate parameters in test 2.

Image name	Method name	Time/s	P/	R/%	F/%
CT201.86	QPSO	0.932	0.6769	0.7985	0.7327
	SunCQPSO	1.000	0.6323	0.8876	0.7385
	CCQPSO	1.022	0.6822	0.8168	0.7435
	SCQPSO	1.120	0.6823	0.8164	0.7434
	IQPFLS	1.108	0.9476	0.7144	0.8146
	QWPSO	0.900	0.9996	0.9990	0.9993
CT201.136	QPSO	0.900	0.6078	0.5548	0.5801
	SunCQPSO	0.933	0.5670	0.6594	0.6097
	CCQPSO	0.912	0.5508	0.6220	0.5843
	SCQPSO	0.990	0.5734	0.6597	0.6135
	IQPFLS	1.102	0.9208	0.7344	0.8171
	QWPSO	0.856	0.9998	0.9993	0.9995
CT201.29	QPSO	0.912	0.6439	0.7887	0.7090
	SunCQPSO	1.000	0.5536	0.3116	0.3988
	CCQPSO	1.021	0.5657	0.3909	0.4624
	SCQPSO	1.020	0.6056	0.3355	0.4318
	IQPFLS	1.099	0.7571	0.6675	0.7094
	QWPSO	0.874	0.7944	0.7951	0.7948
CT200.2	QPSO	0.850	0.7216	0.6062	0.6589
	SunCQPSO	0.912	0.7186	0.6192	0.6652
	CCQPSO	0.923	0.7464	0.6583	0.6996
	SCQPSO	0.931	0.7654	0.6622	0.7101
	IQPFLS	1.111	0.7858	0.7097	0.7458
	QWPSO	0.810	0.7546	0.7423	0.7484

CCQPSO SCQPSO, and IQPFLS, our approach offers improved performance in terms of operational efficiency. In addition, we reduced the running time to shorter and higher segmentation accuracy, both under manual observational inspection and in quantitative analysis using established evaluation parameters.

### Test 3: Compared With Existing Methods

To demonstrate the advantage of the proposed QWPSO method, test 3 implements a comparison between ten studies listed from (8, 68–77) in references whose publication years were from 2018 to 2021. **Table 5** shows the original image, reference segmentation results, suggested QWPSO method, and ground truth in the 1st to 4th columns. The ground truth is obtained by an evaluation program. The detail is that we first input the original image into the evaluation program, and then input our segmented image. The program gave a standard segmentation result (red) based on the original image. Our segmentation result was blue, red, and blue overlap, indicating a good segmentation result. As seen in **Table 5**, despite the great challenge of these images due to the low contrast and high-intensity inhomogeneities, the QWPSO segmentation results are pretty consistent with the ground truth, and it successfully recovers the contours of the tumor substructures, especially in the region with bottle-neck. Although, for example, the green circle regions shown in the image of (8) illustrate the main differences between the segmentation result and the ground truth, their results lead to the fuzzy and missing of the bottle-neck

parts. Still, our QWPSO segmentation results can enforce spatial consistency. Consequently, the contours of different reference images are well segmented by our QWPSO method.




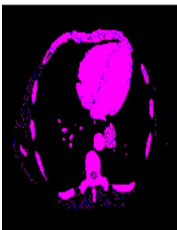
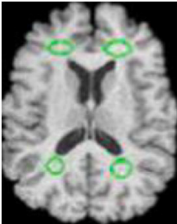


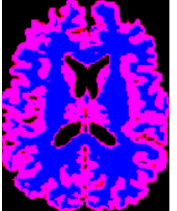
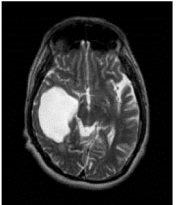
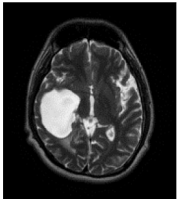

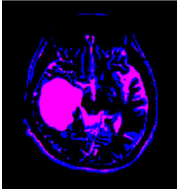
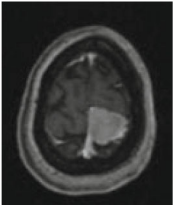

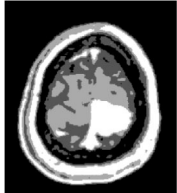

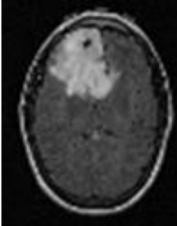
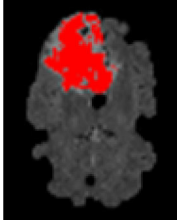
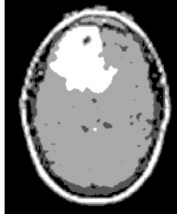

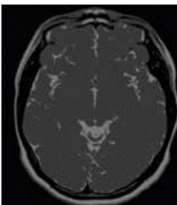
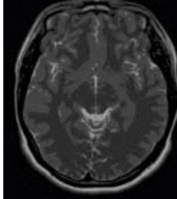
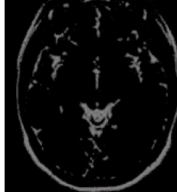
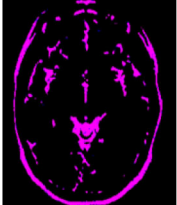
In addition to the P, R, and F parameters, the receiver operating curve (ROC) and Hausdorff evaluation parameters were added in this part in order to better compare the methods proposed and the experimental results of the reference studies. The ROC curve reflects the relationship between sensitivity and specificity, while the value of Hausdorff can measure the distance between proper subsets in a metric space. The smaller the Hausdorff value is, the higher the edge matching. ROC curve is shown in **Figure 5**, while **Tables 6, 7** demonstrate the performance-evaluated parameters of segmentation Precision, Recall, F-measure, and Hausdorff (H, the abbreviation for Hausdorff in this study). All values of P, R, F, and H of the reference studies and our segmentation results are in each row.

**Table 6** demonstrates the reference studies (8, 68–71) published from 2021 to 2020. These studies used newly proposed methods within two years of the methods' publication in higher-level international journals from 2018 to 2019. **Table 7** includes references (72–77) published from 2019 to 2018. In **Table 6**, for example, the first row P of Allouui et al. (68) is 0.9176, and the value of 0.9785 is the P results of our proposed QWPSO method. The corresponding segmentation results in rows P, R, and F show that our proposed methods are higher than their reference studies.

The first row in **Table 7** shows that the mean values of P compared to (72) are from 0.7623 to 0.7911. Our value of 0.9789 is the average P. As for the last row of Hausdorff, our H values are all less than the compared values, with the least one being three. Despite this, the H value compared with (73) is 22.0907, and the corresponding Ground Truth (GT) image in **Table 5** is in pink without blue, indicating that the blue is completely overlapped. In addition, the segmented image and the standard segmentation image are almost identical and cannot be distinguished by the naked eye. Thus, regardless of whether the studies were published in the periods 2018–2019 or 2020–2021, the segmentation results of our proposed QWPSO method are all higher than theirs. This further illustrates that the segmented images with prominent bottle-neck regions have even better performance, i.e., reference images of (70, 71, 75, 76), of which our evaluated parameter P are all 1.0000.

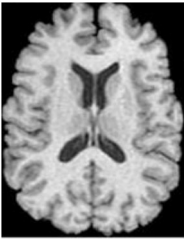

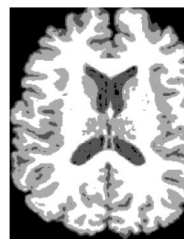

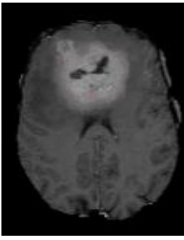
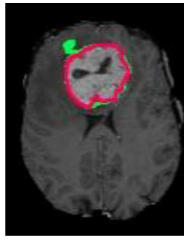
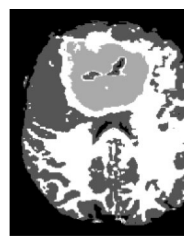
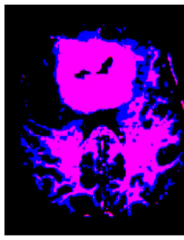
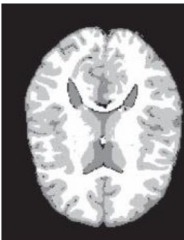
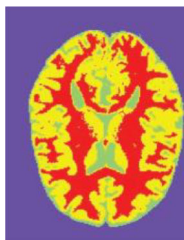
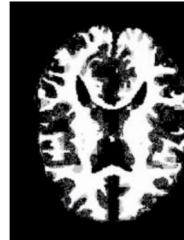
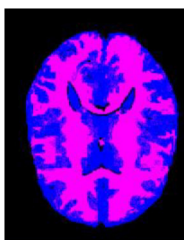



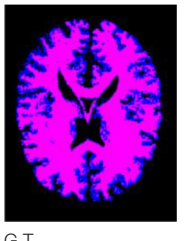
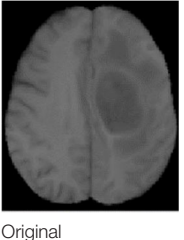
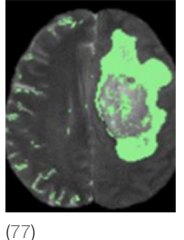
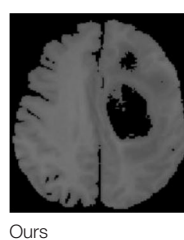

The ROC curve, also known as the “subject operating characteristic curve” or sensitivity curve, is mainly used for the prediction accuracy from X to Y. The ROC curve reflects the relationship between sensitivity and specificity. The X-axis is 1–specificity, also known as false-positive rate. The closer the X-axis is to zero, the higher the accuracy. The Y-axis is called sensitivity, also known as true positive rate (sensitivity). The higher the Y-axis, the better the accuracy. According to the position of the Curve, the whole graph is divided into two parts. The Area under the Curve (AUC) indicates the accuracy of prediction. The higher the AUC value is, the higher the accuracy of prediction. The closer the curve is to the top left corner. Hence, the smaller the X, the larger the Y, and the higher the prediction accuracy. The ROC curve of the segmentation results of our proposed method and reference studies are shown in **Figure 5**.

TABLE 5 | Comparison segmentation test.

			
Original	(68)	Ours	Ground Truth(G.T.)
			
Original	(8)	Ours	G.T.
			
Original	(69)	Ours	G.T.
			
Original	(70)	Ours	G.T.
			
Original	(71)	Ours	G.T.
			
Original	(72)	Ours	G.T.

(Continued)

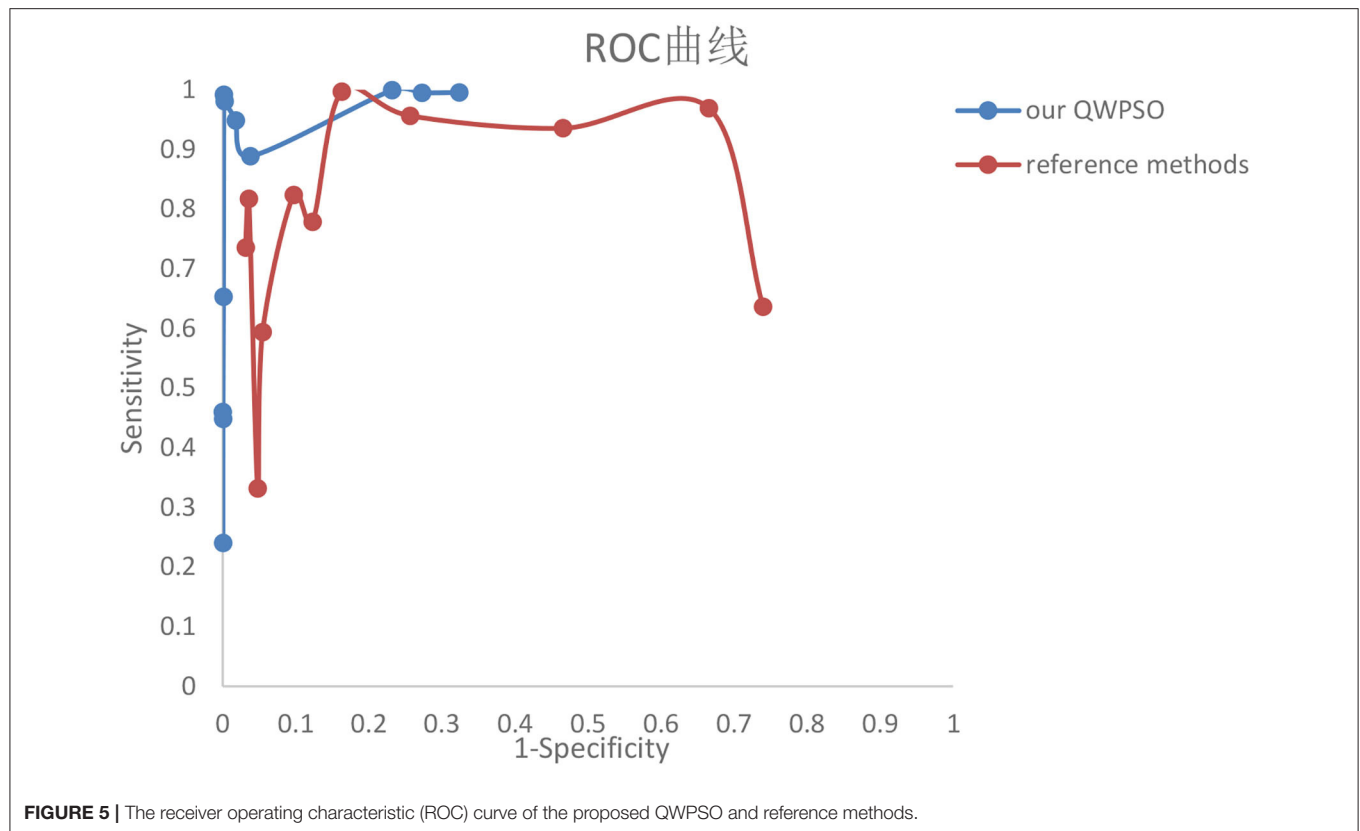
TABLE 5 | Continued

 Original	 (73)	 Ours	 G.T.
 Original	 (74)	 Ours	 G.T.
 Original	 (75)	 Ours	 G.T.
 Original	 (76)	 Ours	 G.T.
 Original	 (77)	 Ours	 G.T.

In comparing the two curves of our proposed QWPSO method and the reference methods in **Figure 5**, we can state that no matter the AUC, our proposed QWPSO method is better than the reference studies. Therefore, we conclude that our proposed QWPSO method has a higher segmentation accuracy than the reference methods.

DISCUSSION

The current related works with the proposed QWPSO are the PSO, QPSO, and their improving algorithms, such as SunCQPSO, CCQPSO SCQPSO, and IQPFLS. The limitations of all these related works are that they solve the general shape



**FIGURE 5 |** The receiver operating characteristic (ROC) curve of the proposed QWPSO and reference methods.

**TABLE 6 |** Evaluate parameters in test 3-part one.

	Allioui et al. (68), Ours	Radha et al. (8), Ours	Mahesa et al. (69), Ours	Vijh et al. (70), Ours	Sharif et al. (71), Ours
P	0.9176, 0.9785	0.8875, 0.9429	-, 0.9877	0.9800, 1.0000	0.994-0.998, 1.0000,
R	0.9248, 0.9541	-,0.5034	-,0.7088	-,0.9784	-,0.9265
F	0.9212,0.9661	-,0.6563	-,0.8253	-,0.9891	-, 0.9619
H	35.4683, 21.3073	24.5967, 18.1384	47.2017, 33.1813	93.1933, 10.9545	35.0571, 17.1172,

**TABLE 7 |** Evaluate parameters in test 3-part two.

	Khairuzzaman et al. (72), Ours	Ibungomacha Singh et al. (73), Ours	Hasan et al. (74), Ours	Guerrouet et al. (75), Ours	Pham et al. (76), Ours	Ma et al. (77), Ours
P	<b>0.7623–0.7911, 0.9789,</b>	<b>0.8650, 0.9993</b>	0.9220, 0.9656	0.855–0.991, 1.0000,	0.9126–0.9835, 1.0000	0.900, 0.9811
R	-,0.9479	<b>-,0.9920</b>	-,0.9544	-,0.5605	-,0.7218	0.850,0.8747
F	-, 0.9632	<b>-,0.9956</b>	-,0.9599	-, 0.7184	-,0.8384	0.870,0.9248
H	70.3847,17.5784	<b>22.0907, 3</b>	44.2945,28.7054	15.9374,14.6969	24.7790,10.1980	50.2905,17.8045

of the image segmentation problem rather than specifically for the special curved 'bottle-neck' shape of the image segmentation problem. However, in terms of technical improvement, there exists an interesting evolutionary relationship between the PSO, QPSO, and our QWPSO segmentation algorithm. While our proposed QWPSO algorithm is based on the existing QPSO technique, the QPSO approach is an improved version of the PSO algorithm. It is perhaps helpful to consider these algorithms'

unique properties and functions. Particles are dynamically represented in the PSO approach. Particles adjust their speed according to the flight experience of both individuals and groups. The PSO algorithm works well for some image segmentation tasks but works less for others, particularly those involving complex and indistinct object shapes. The QPSO algorithm represents the latest intelligent optimization algorithm. Each particle moves well according to quantum behavior based on

a Delta potential where they are centered during the various iteration steps. In this way, the QPSO algorithm can usefully enhance the population diversity and has a more robust global searching ability than the earlier PSO algorithm. Furthermore, this also means that the QPSO algorithm is better for more complex object shapes than in segmentation tasks. However, in the case of our QWPSO algorithm, all nodes are additionally considered to potentially have existing similar nodes which are more likely to be connected by a wormhole. This characteristic of the QWPSO algorithm provides a unique, advantageous, and powerful global searching ability for defining complex and unique object contour shapes in challenging image segmentation tasks. This unique ability to connect long-distance particles is a significant contribution of our QWPSO algorithm.

The above experimental results indicate that the QWPSO algorithm has good application for complex and specialized object contour segmentation, particularly for object regions typically encountered in medical tumor images that possess 'dual tails' and 'bottle-neck' feature shapes. For example, in **Table 2**, for the segmentation of MRI brain images, we present evaluation parameters that include Time, P, R, and F. We obtained the best values from the proposed QWPSO algorithm, namely, 0.801, 1.000, 1.000, and 0.9986 for Time, P, R, and F, respectively. Furthermore, the average running time of the QWPSO algorithm is 0.8743 s, which compares favorably with 0.9404 s for the QPSO algorithm, giving a decrease of 0.0961/s or  $\sim 12\%$  reduction in average running time. Together these results indicate that our QWPSO algorithm has high efficiency in segmenting complex and specially shaped objects. As for the evaluation parameters of P, R, and F, our QWPSO algorithm attains optimal values of almost 1 for all. A parameter  $p$ -value of 0.9995 was obtained. While parameter R was slightly better for the QPSO algorithm, parameter F, which is the product of P and R, was better in all cases for our proposed QWPSO approach. It indicates that the accuracy of our proposed QWPSO algorithm is better than the QPSO algorithm for MRI brain image segmentation. **Table 4** presents results obtained for the segmentation of CT brain images. All parameters, Time, P, R, and F, indicate improved performance over the improved QPSO algorithms, such as SunCQPSO, CCQPSO, and SCQPSO. Therefore, in general, the QWPSO algorithm offers greater adaptability to object region shape, together with better operational efficiency and segmentation accuracy over QPSO or improved QPSO algorithms for typically challenging MRI and CT brain image segmentation tasks. Furthermore, the evaluated parameters P, R, F, and H of the proposed QWPSO algorithm are shown in **Tables 6, 7**. All are better than the compared typical references within the last three years, especially achieving the highest value 1 for P among the images with distinct bottle-neck regions, such as images from references studies (70, 71, 75, 76). Lastly, the AUC area in the ROC curve of our proposed method in **Figure 5** shows higher accuracy than that of all the reference studies.

## CONCLUSION

This study has presented a QWPSO algorithm for challenging image segmentation tasks. We applied wormhole-inspired theory

to our method and put forward a hyperbolic wormhole path measure equation that seeds and links particles to improve the performance of the existing QPSO segmentation method. The QPSO method uses random positioning in the search space even if there are long distances between particles. Our QWPSO algorithm can cluster long-distance regions into groupings and has better adaptability than the existing QPSO algorithm and the current improved QPSO algorithms. Experimental results, both from MRI and CT, have demonstrated enhanced performance in segmenting rare brain tumors with tailing and bottle-neck regions. In addition, our QWPSO method improved operational efficiency and segmentation accuracy compared with current competing reference methods. Because there are many image segmentation that consists of similar curved targets, in the future, we are committed to extending the proposed QWPSO algorithm in this study to the segmentation of curved tumors in other organs' medical images, such as lung, liver, and/or spleen tumors. In addition to medical images, the proposed QWPSO algorithm should also be extended to other research areas of curving or bending target image segmentations.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://mp.weixin.qq.com/s/d1H2j5vrVo-RHRAVI8ygQ>; <https://www.med.upenn.edu/cbica/brats2020/data.html>; <https://www.med.upenn.edu/cbica/brats2019/data.html>; <https://www.med.upenn.edu/sbia/brats2018/data.html>.

## AUTHOR CONTRIBUTIONS

TZ designed the method, conducted the experiments, and wrote the manuscript. JZ helped to improve the results and modified the manuscript. TX handled this project. MR revised the manuscript and figures and enhanced their overall quality. All authors read this revised manuscript and gave final approval for the final submission.

## FUNDING

This research was supported by: (1) 2021–2023 National Natural Science Foundation of China under Grand (Youth) No. 52001039. (2) 2022–2025 National Natural Science Foundation of China under Grant No. 52171310. (3) 2020–2022 Shandong Natural Science Foundation Funding in China No. ZR2019LZH005. (4) 2022–2023 Research fund from Science and Technology on Underwater Vehicle Technology Laboratory under Grant 2021CJQ-SYSJJ-LB06903.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2022.794126/full#supplementary-material>

## REFERENCES

- Moskowitz C. Tangled up in spacetime. *Sci. Am.* (2017) 316:32–7. doi: 10.1038/scientificamerican0117-32
- Hoof GT. Explicit construction of local hidden variables for any quantum theory up to any desired accuracy. *arXiv.* (2021) 210304335. doi: 10.48550/arXiv.2103.04335
- Sabeti M, Karimi L, Honarvar N, Taghavi M, Boostani R. Quantumized genetic algorithm for segmentation and optimization. *Biomed Eng-Appl Basis Commun.* (2020) 32:2050022. doi: 10.4015/S1016237220500222
- Vaze R, Deshmukh N, Kumar R, Saxena A. Development and application of quantum entanglement inspired particle swarm optimization. *Knowl Based Syst.* (2021) 219:106859. doi: 10.1016/j.knsys.2021.106859
- Benatti F, Floreanini R, Franchini F, Marzolino U. Remarks on entanglement and identical particles. *Open Syst Inf Dyn.* (2017) 24:1740004. doi: 10.1142/S1230161217400042
- Shabanifard M, Amirani MC. A modified quantum-behaved particle swarm optimization algorithm for image segmentation. In: *2011 19th Iranian Conference on Electrical Engineering*. Tehran: IEEE. (2011) p. 1–6.
- Singh A, Sinha U. Entanglement protection in higher-dimensional systems. *arXiv.* (2020) 200107604. doi: 10.48550/arXiv.2001.07604
- Radha R, Gopalakrishnan R, A. medical analytical system using intelligent fuzzy level set brain image segmentation based on improved quantum particle swarm optimization. *Microprocess Microsyst.* (2020) 79:103283. doi: 10.1016/j.micpro.2020.103283
- Caraiman S, Manta VI. Histogram-based segmentation of quantum images. *Theor Comput Sci.* (2014) 529:46–60. doi: 10.1016/j.tcs.2013.08.005
- Maldacena J, Susskind L. Cool horizons for entangled black holes. *Fortschritte der Physik.* (2013) 61:781–811. doi: 10.1002/prop.201300020
- Ellison CK, Kan J, Dillard RS, Kysela DT, Ducret A, Berne C. et al. Obstruction of pilus retraction stimulates bacterial surface sensing. *Science.* (2017) 358:535–8. doi: 10.1126/science.aan5706
- Zhang J, Guo M, Li S, Peng W. A New Method for Simulating Protein Folding Process–Snake Algorithm. In: *2008 International Conference on Internet Computing in Science and Engineering*. Harbin: IEEE. (2008). p. 39–42. doi: 10.1109/ICICSE.2008.65
- Flamm L. Beiträge zur Einsteinschen Gravitationstheorie. *Phys. Z.* (1916) 17:448–53.
- Flamm L. Beiträge zur Einsteinschen gravitationstheorie. *Hirzel.* (1916).
- Einstein A, Rosen N. The particle problem in the general theory of relativity. *Phys. Rev.* (1935) 48:73. doi: 10.1103/PhysRev.48.73
- Lindley D. The birth of wormholes. *Physics.* (2005) 15:11. doi: 10.1103/PhysRevFocus.15.11
- Marolf D, Polchinski J, Marolf D, Polchinski J, Jacobson T, Maldacena J, et al. *Black Hole Firewall Problem.* (2013).
- Einstein A, Podolsky B, Rosen N. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* (1935) 47:777. doi: 10.1103/PhysRev.47.777
- Einstein A, Rosen N. The particle problem in the general theory of relativity. *Phys. Rev.* (1935) 48:73–7.
- Iqbal N, Ross S. Towards traversable wormholes from force-free plasmas. *Sci Post Phys.* (2022) 12:086. doi: 10.21468/SciPostPhys.12.3.086
- Gao P, Jafferis DL, Wall AC. Traversable wormholes via a double trace deformation. *J High Energy Phys.* (2017) 2017:1–25. doi: 10.1007/JHEP12(2017)151
- Maldacena J, Milekhin A, Popov F. Traversable wormholes in four dimensions. *arXiv.* (2018) 180704726. doi: 10.48550/arXiv.1807.04726
- Kim SW, Lee H. Exact solutions of a charged wormhole. *Phys Rev D.* (2001) 63:064014. doi: 10.1103/PhysRevD.63.064014
- Jafferis DL, Schneider E. Stringy ER=EPR. *arXiv.* (2021) 210407233. doi: 10.48550/arXiv.2104.07233
- Heinicke C, Hehl FW. Schwarzschild and Kerr solutions of Einstein's field equation: an Introduction. *Int J Modern Phys D.* (2015) 24:1530006. doi: 10.1142/S0218271815300062
- Philipp D, Perlick V. Schwarzschild radial perturbations in Eddington–Finkelstein and Painlevé–Gullstrand coordinates. *Int J Modern Phys D.* (2015) 24:1542006. doi: 10.1142/S0218271815420067
- Konoplya R, Zhidenko A. Detection of gravitational waves from black holes: is there a window for alternative theories? *Physics Lett B.* (2016) 756:350–3. doi: 10.1016/j.physletb.2016.03.044
- Samoi. *Dynamics of neurological symptoms in patients with brain tumors during combined therapy (operation plus betatron electromagnetic radiation).* Samoilov V I. Dynamics of neurological symptoms in patients with brain tumors during combined therapy (operation plus betatron electromagnetic radiation). *Vopr. Neurohir.* 1967, 31(2): 42–46.
- Stipić D, Beroš V, Gnjidić Ž. Does mobile phone use cause brain tumors? *Medicina Fluminensis: Medicina Fluminensis.* (2011) 47:200–5.
- Jargin SV. Electromagnetic radiofrequency radiation with special reference to otorhinolaryngology and brain tumors. *Braz J Otorhinolaryngol.* (2019) 85:129. doi: 10.1016/j.bjorl.2018.09.003
- Gogineni S. *Computational Study of Electromagnetic Wave Induced by Mobilephones on Brain Tissues and its Biological Implications.* San Antonio: The University of Texas at San Antonio. (2010).
- Berg G, Spallek J, Schütz J, Schlehofer B, Böhler E, Schläfer K. et al. Occupational exposure to radio frequency/microwave radiation and the risk of brain tumors: interphone Study Group, Germany. *Am J Epidemiol.* (2006) 164:538–48. doi: 10.1093/aje/kwj247
- Morgan LL, Miller AB, Sasco A, Davis DL. Mobile phone radiation causes brain tumors and should be classified as a probable human carcinogen (2A). *Int J Oncol.* (2015) 46:1865–71. doi: 10.3892/ijo.2015.2908
- Saikhedkar N, Bhatnagar M, Jain A, Sukhwil P, Sharma C, Jaiswal N. Effects of mobile phone radiation (900 MHz radiofrequency) on structure and functions of rat brain. *Neurological Res.* (2014) 36:1072–9. doi: 10.1179/1743132814Y.0000000392
- Zymantiene J, Juozaitiene V, Zelvyte R, Oberauskas V, Spancerniene U, Sederevicius A. et al. Effect of electromagnetic field exposure on mouse brain morphological and histopathological profiling. *J Vet Res.* (2020) 64:319–24. doi: 10.2478/jvetres-2020-0030
- Bhargav H, Srinivasan T, Varambally S, Gangadhar B, Koka P. Effect of mobile phone-induced electromagnetic field on brain hemodynamics and human stem cell functioning: Possible mechanistic link to cancer risk and early diagnostic value of electronphoton imaging. *J Stem Cells.* (2015) 10:287.
- Ostrom QT, Bauchet L, Davis FG, Deltour I, Fisher JL, Langer CE. et al. The epidemiology of glioma in adults: a state of the science review. *Neuro-oncol.* (2014) 16:896–913. doi: 10.1093/neuonc/nou087
- Choi YJ, Moskowitz JM, Myung SK, Lee YR, Hong YC. Cellular phone use and risk of tumors: Systematic review and meta-analysis. *Int J Environ Res Public Health.* (2020) 17:8079. doi: 10.3390/ijerph17218079
- Castano-Vinyals G, Sadetzki S, Vermeulen R, Momoli F, Kundi M, Merletti F. et al. Wireless phone use in childhood and adolescence and neuroepithelial brain tumours: Results from the international MOBI-Kids study. *Environ Int.* (2022) 160:107069. doi: 10.1016/j.envint.2021.107069
- Carlberg M, Hardell L. Evaluation of mobile phone and cordless phone use and glioma risk using the Bradford Hill viewpoints from 1965 on association or causation. *Biomed Res Int.* (2017) 2017:9218486. doi: 10.1155/2017/9218486
- Prasad M, Kathuria P, Nair P, Kumar A, Prasad K. Mobile phone use and risk of brain tumours: a systematic review of association between study quality, source of funding, and research outcomes. *Neurological Sciences.* (2017) 38:797–810. doi: 10.1007/s10072-017-2850-8
- Zhao F, Liu Y, Huo K, Zhang Z. Radar target classification using an evolutionary extreme learning machine based on improved quantum-behaved particle swarm optimization. *Math Probl Eng.* (2017) 2017:7273061. doi: 10.1155/2017/7273061
- Venkatesan A, Parthiban L. Medical image segmentation with fuzzy c-means and kernelized fuzzy C-means hybridized on PSO and QPSO. *Int Arab J Inf Technol.* (2017) 14:53–9.
- Zhang C, Xie Y, Liu D, Wang L. Fast threshold image segmentation based on 2D fuzzy fisher and random local optimized QPSO. *IEEE Trans Image Process.* (2016) 26:1355–62. doi: 10.1109/TIP.2016.2621670

45. Li Y, Bai X, Jiao L, Xue Y. Partitioned-cooperative quantum-behaved particle swarm optimization based on multilevel thresholding applied to medical image segmentation. *Appl Soft Comput.* (2017) 56:345–56. doi: 10.1016/j.asoc.2017.03.018
46. Sun J, Feng B, Xu W. Particle swarm optimization with particles having quantum behavior. In: *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No. 04TH8753)*. Portland, OR: IEEE. (2004). p. 325–331.
47. Sutphin GL, Mahoney JM, Sheppard K, Walton DO, Korstanje R, WORMHOLE. novel least diverged ortholog prediction through machine learning. *PLoS Comput Biol.* (2016) 12:e1005182. doi: 10.1371/journal.pcbi.1005182
48. Lobo FS. From the Flamm–Einstein–Rosen bridge to the modern renaissance of traversable wormholes. *Int J Modern Phys D.* (2016) 25:1630017. doi: 10.1142/S0218271816300172
49. Venegas-Andraca SE, Ball J. Processing images in entangled quantum systems. *Quantum Inf Process.* (2010) 9:1–11. doi: 10.1007/s11128-009-0123-z
50. Papadopoulos F, Kitsak M, Serrano M, Boguná M, Krioukov D. Popularity versus similarity in growing networks. *Nature.* (2012) 489:537–40. doi: 10.1038/nature11459
51. Krioukov D, Papadopoulos F, Kitsak M, Vahdat A, Boguná M. Hyperbolic geometry of complex networks. *Physical Review E.* (2010) 82:036106. doi: 10.1103/PhysRevE.82.036106
52. Maldacena J. Black holes, wormholes and the secrets of quantum spacetime. *Sci Am.* (2016) 315:26–31. doi: 10.1038/scientificamerican1116-26
53. Nielsen MA, Chuang I. Quantum computation and quantum information. *Am Assoc Phys Teachers.* (2002) 558–9. doi: 10.1119/1.1463744
54. Bhattacharjee D. *Extremal Kerr Black Holes, Naked Singularity & Wormholes. Preprint.* (2020) doi: 10.35543/osf.io/syx73
55. Ye Z. 地球上有没有虫洞吗 地球是否出现过虫洞. Available online at: <http://www.lieqiba.com/tansuofaxian/86304.html>.
56. Bagrodia R, Chen Ya, Gerla M, Kwan B, Martin J, Palnati P, et al. Parallel simulation of a high-speed wormhole routing network. In: *Proceedings of the Tenth Workshop on Parallel and Distributed Simulation*. Atlanta, GA (1996). p. 47–56. doi: 10.1145/238793.238813
57. Tun Z, Maw AH. Wormhole attack detection in wireless sensor networks. *Int J Electron Commun.* (2008) 2:2184–9. doi: 10.1109/SAI.2016.7556151
58. Khurana E, Fu Y, Chakravarty D, Demicheli F, Rubin MA, Gerstein M. Role of non-coding sequence variants in cancer. *Nat Rev Genet.* (2016) 17:93–108. doi: 10.1038/nrg.2015.17
59. Li Y, Jiao L, Shang R, Stolkin R. Dynamic-context cooperative quantum-behaved particle swarm optimization based on multilevel thresholding applied to medical image segmentation. *Inf Sci.* (2015) 294:408–22. doi: 10.1016/j.ins.2014.10.005
60. Zhou D, Sun J, Xu WB. Quantum-behaved particle swarm optimization algorithm with cooperative approach. *Control and Decision.* (2011) 26:582–6. doi: 10.1109/IWACI.2010.5585123
61. Pant M, Thangaraj R, Abraham A. A new quantum behaved particle swarm optimization. In: *Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation*. Atlanta, GA (2008). p. 87–94. doi: 10.1145/1389095.1389108
62. Bonahon F. Surfaces to hyperbolic knots: From Euclidean surfaces to hyperbolic knots. *Am. Math. Soc.* (2009) 49:384.
63. Chowdhury BD. Cool horizons lead to information loss. *J High Ener Phys.* (2013) 2013:1–24. doi: 10.1007/JHEP10(2013)034
64. Morris MS, Thorne KS, Yurtsever U. Wormholes, time machines, and the weak energy condition. *Phys Rev Lett.* (1988) 61:1446. doi: 10.1103/PhysRevLett.61.1446
65. Xu S, Mu X, Ma J. Discrete quantum-behaved particle swarm optimization for 2-D maximum entropic multilevel thresholding image segmentation. In: *Chinese Automation Congress (CAC)*. Wuhan: IEEE. (2015). p. 651–656.
66. Zhao J, Wang X, Zhang H, Hu J, Jian X, Side. Scan sonar image segmentation based on neutrosophic set and quantum-behaved particle swarm optimization algorithm. *Marine Geophysical Res.* (2016) 37:229–41. doi: 10.1007/s11001-016-9276-1
67. Gao H, Xu W, Sun J, Tang Y. Multilevel thresholding for image segmentation through an improved quantum-behaved particle swarm algorithm. *IEEE Trans Instrum Meas.* (2009) 59:934–46. doi: 10.1109/TIM.2009.2030931
68. Alloui H, Sadgal M, Elfazziki A. Optimized control for medical image segmentation: improved multi-agent systems agreements using Particle Swarm Optimization. *J Ambient Intell Humaniz Comput.S.* (2021) 12:8867–85. doi: 10.1007/s12652-020-02682-9
69. Mahesa R, Wibowo E. Optimization of fuzzy c-means clustering using particle swarm optimization in brain tumor image segmentation. *J Theor Appl Inf Technol.* (2020) 98:19.
70. Vijh S, Sharma S, Gaurav P. Brain tumor segmentation using OTSU embedded adaptive particle swarm optimization method and convolutional neural network. In: *Data Visualization and Knowledge Engineering*. Springer. (2020). p. 171–94. doi: 10.1007/978-3-030-25797-2\_8
71. Sharif M, Amin J, Raza M, Yasmin M, Satapathy SC. An integrated design of particle swarm optimization (PSO) with fusion of features for detection of brain tumor. *Pattern Recognit Lett.* (2020) 129:150–7. doi: 10.1016/j.patrec.2019.11.017
72. Khairuzzaman AKM, Chaudhury S. Brain MR image multilevel thresholding by using particle swarm optimization, Otsu method and anisotropic diffusion. *IJAMC.* (2019) 10:91–106. doi: 10.4018/IJAMC.2019070105
73. Ibungomacha Singh T, Laishram R, Roy S. Comparative study of combination of swarm intelligence and fuzzy C means clustering for medical image segmentation. In: *Smart Computational Strategies: Theoretical and Practical Aspects*. Springer. (2019). p. 69–80. doi: 10.1007/978-981-13-6295-8\_7
74. Hasan AM, A. hybrid approach of using particle swarm optimization and volumetric active contour without edge for segmenting brain tumors in MRI scan. *IJEEI.* (2018) 6:292–300. doi: 10.52549/ijeei.v6i3.592
75. Guerrou EH, Ait-Aoudia S, Michelucci D, Mahiou R. Hidden Markov random field model and Broyden–Fletcher–Goldfarb–Shanno algorithm for brain image segmentation. *J Exp Theor Artif Intell.* (2018) 30:415–27. doi: 10.1080/0952813X.2017.1409280
76. Pham TX, Siarry P, Oulhadj H. Integrating fuzzy entropy clustering with an improved PSO for MRI brain image segmentation. *Appl Soft Comput.* (2018) 65:230–42. doi: 10.1016/j.asoc.2018.01.003
77. Ma C, Luo G, Wang K. Concatenated and connected random forests with multiscale patch driven active contour model for automated brain tumor segmentation of MR images. *IEEE Trans Med Imaging.* (2018) 37:1943–54. doi: 10.1109/TMI.2018.2805821
78. Multimodal Brain Tumor Segmentation. (2019) *Acropolis Convention Center - Nice, France*. Available online at: <https://www.cbica.upenn.edu/BraTS19>.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Zhang, Xue and Rashid. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Knee Bone and Cartilage Segmentation Based on a 3D Deep Neural Network Using Adversarial Loss for Prior Shape Constraint

Hao Chen<sup>1</sup>, Na Zhao<sup>2\*</sup>, Tao Tan<sup>3\*</sup>, Yan Kang<sup>4</sup>, Chuanqi Sun<sup>5</sup>, Guoxi Xie<sup>5</sup>, Nico Verdonchot<sup>6</sup> and André Sprengers<sup>7</sup>

<sup>1</sup> Department of Biomechanical Engineering, University of Twente, Enschede, Netherlands, <sup>2</sup> School of Instrument Science and Engineering, Southeast University, Nanjing, China, <sup>3</sup> Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, Netherlands, <sup>4</sup> College of Health Science and Environmental Engineering, Shenzhen Technology University, Shenzhen, China, <sup>5</sup> Department of Biomedical Engineering, The Sixth Affiliated Hospital, Guangzhou Medical University, Guangzhou, China, <sup>6</sup> Orthopaedic Research Laboratory, Radboud University Medical Center, Nijmegen, Netherlands, <sup>7</sup> Department of Biomedical Engineering and Physics, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands

## OPEN ACCESS

### Edited by:

Alice Chen,  
National Cancer Institute (NCI),  
United States

### Reviewed by:

Chunpeng Wu,  
Duke University, United States  
Xiaowei Han,  
Nanjing Drum Tower Hospital, China

### \*Correspondence:

Na Zhao  
zhaona@seu.edu.cn  
Tao Tan  
taotans@seu.edu.cn

### Specialty section:

This article was submitted to  
Precision Medicine,  
a section of the journal  
Frontiers in Medicine

Received: 11 October 2021

Accepted: 14 April 2022

Published: 20 May 2022

### Citation:

Chen H, Zhao N, Tan T, Kang Y,  
Sun C, Xie G, Verdonchot N and  
Sprengers A (2022) Knee Bone  
and Cartilage Segmentation Based on  
a 3D Deep Neural Network Using  
Adversarial Loss for Prior Shape  
Constraint. *Front. Med.* 9:792900.  
doi: 10.3389/fmed.2022.792900

Fast and accurate segmentation of knee bone and cartilage on MRI images is becoming increasingly important in the orthopaedic area, as the segmentation is an essential prerequisite step to a patient-specific diagnosis, optimising implant design and preoperative and intraoperative planning. However, manual segmentation is time-intensive and subjected to inter- and intra-observer variations. Hence, in this study, a three-dimensional (3D) deep neural network using adversarial loss was proposed to automatically segment the knee bone in a resampled image volume in order to enlarge the contextual information and incorporate prior shape constraints. A restoration network was proposed to further improve the bone segmentation accuracy by restoring the bone segmentation back to the original resolution. A conventional U-Net-like network was used to segment the cartilage. The ultimate results were the combination of the bone and cartilage outcomes through post-processing. The quality of the proposed method was thoroughly assessed using various measures for the dataset from the Grand Challenge Segmentation of Knee Images 2010 (SKI10), together with a comparison with a baseline network U-Net. A fine-tuned U-Net-like network can achieve state-of-the-art results without any post-processing operations. This method achieved a total score higher than 76 in terms of the SKI10 validation dataset. This method showed to be robust to extract bone and cartilage masks from the MRI dataset, even for the pathological case.

**Keywords:** cartilage segmentation, bone segmentation, MRI, deep learning, CNN

## INTRODUCTION

Quantitative analysis of knee joint structure is a topic of increasing interest as its applications continue to broaden from direct diagnostic purposes to the implant design and preoperative and intraoperative planning. Due to the non-invasive nature and capability to discriminate cartilage from adjacent tissues, magnetic resonance imaging (MRI) is the most effective imaging device to

perform knee joint analysis. However, due to the low contrast among different tissues (similar longitudinal and transverse relaxation time), image artefacts, and intensity of inhomogeneity problems in MRI (1), the accurate segmentation of the knee joint is still an open problem, especially in the knee with a degenerative disease (2).

To obtain an accurate mask for knee bone and cartilage, fully manual and semi-automatic segmentation approaches were often applied to clinical studies (3–5). Nonetheless, they were time-consuming and the reproducibility highly depends on the knowledge of experts. Hence, an automated method to segment the knee joint structure was of great interest in the past decade (6, 7). The popular methods for this aim can be divided into model-based (8–10), atlas-based (11, 12), and classification-based (1, 2, 13) methods. Although these three types of methods showed promising results to automate the knee structure segmentation, they might perform poorly in the case of high subject variability (2).

Recently, deep convolutional neural network (CNN)-based methods have achieved enormous success in biomedical imaging problems, such as classification (14) and segmentation (15–18). Regarding knee joint structure segmentation, Prasoon et al. (19) first applied the two-dimensional (2D) tri-planar CNNs (axial, coronal, and sagittal plane) to classify a pixel label (background or tibial cartilage) by providing local image patches around that pixel. Nevertheless, Ronneberger et al. (18) pointed out that there were two drawbacks to the above architecture, large redundancy and a trade-off between localisation accuracy and the use of context, and proposed a dense prediction network with skip connection, U-Net. This kind of architecture considered both the low-level and high-level features for voxel classification and was applied to the knee joint segmentation by Liu et al. (2), Zhao et al. (20), and Ambellan et al. (21). In general, the pixel-wise or voxel-wise loss, e.g., cross-entropy loss and dice loss, was utilized as the loss function for U-Net. However, there was no guarantee of the spatial consistency of the final output (22); thereafter, a further optimisation step was always required to refine the segmentation result such as deformable model (2), conditional random field (CRF) (20) and statistical shape model (SSM) (21). Although the deformable model and CRF considered the relevant spatial information to refine the segmentation, it might cause serious boundary leakage in the low-contrast regions (22). Ambellan et al. (21) proposed to utilize SSM to refine segmentation using the anatomical prior knowledge and achieved the state-of-the-art result. Nevertheless, the introduction of SSM resulted in a lot of extra calculations and the regulation was limited to the variability of the training dataset. Overall, although deep learning-based methods have been demonstrated as the state-of-the-art methods in knee joint segmentation, there is still much room for improvement.

In this study, we aim to further study a three-dimensional (3D) CNN-based method to perform knee bone and cartilage segmentation. The contributions in this article are: (i) Different neural networks are proposed for bone and cartilage segmentation based on their features and a post-processing step is designed to generate the final segmentation result; (ii) the adversarial loss and a restoration network are proposed

to optimize the neural network for bone segmentation and (iii) the performance of proposed method is tested on a public dataset from the Medical Image Computing and Computer-Assisted Intervention (MICCAI) Segmentation of Knee Images 2010 (SKI10) grand challenge and is fully compared with the performance of the various CNN models (3D U-Net, V-Net, nnU-Net and cascade nnU-Net) and some traditional methods.

## MATERIALS AND METHODS

### Data Description

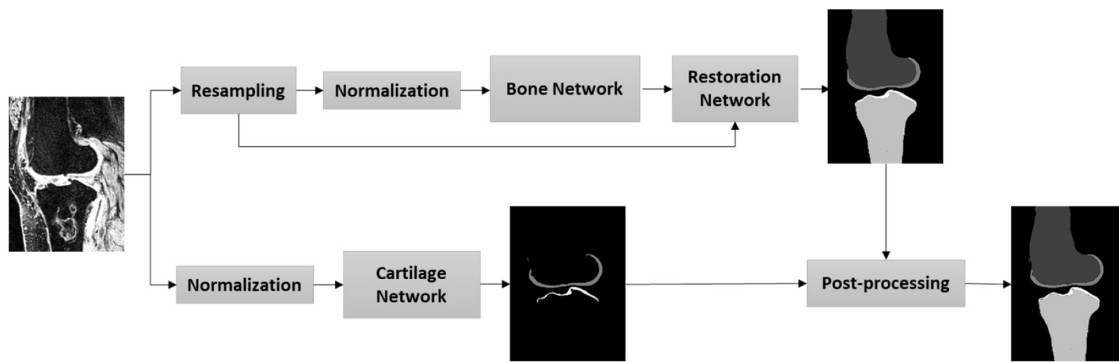
The data used in this study were from the SKI10 competition, which was focused on the knee bone and cartilage segmentation (6). The image datasets were acquired in a sagittal manner with a pixel spacing of 0.4 mm × 0.4 mm and a slice thickness of 1 mm. The total number of the knee images used in this study was 100 (60 for training and 40 for testing), and the cases of left and right knees were approximately equally distributed. Among the scans, 90% of the data were acquired at 1.5 T and the rest of the data were acquired at 3 and 1 T. The majority of data used T1 weighting and the rest of them were acquired with T2 weighting. All the images were acquired for surgery planning of partial or complete knee replacement and, therefore, a high degree of pathological deformations of the knee was included in the dataset.

### Automatic Workflow for Knee Bone and Cartilage Segmentation

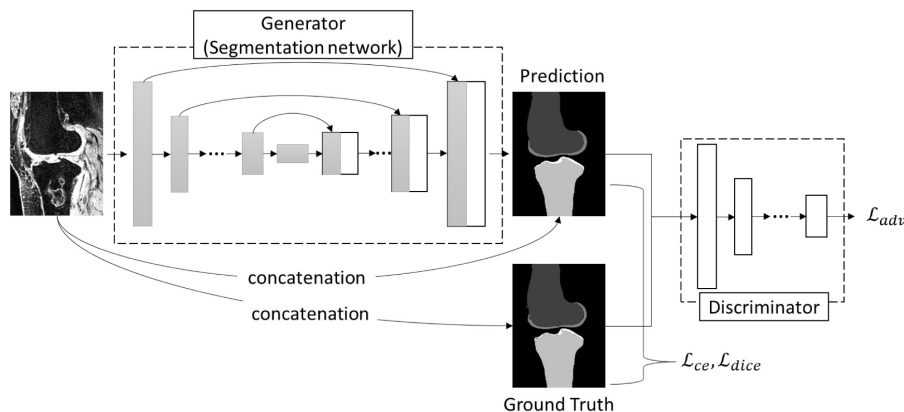
In this study, we aimed to establish a fully automatic workflow to extract knee joint structure (bone and cartilage) with highly accurate and robust segmentation, including the pathological data. **Figure 1** depicts the steps of the proposed workflow. First, MRI images were resampled to enlarge the field of view by the networks; second, an image normalization method standardizes the image to a similar intensity range; third, the bone and cartilage were segmented by the bone network (**Supplementary Figures 1–3** and **Supplementary Table 1**) in a resampled resolution; fourth, the segmented bone and cartilage masks from the bone network were restored to the original resolution through a restoration network (**Supplementary Figure 4**); fifth, the cartilage was segmented through a cartilage network in original resolution; last, the outputs of the cartilage network and the restoration network were post-processed for the final results.

### Pre-processing

Our pre-processing included pixel size normalization and intensity normalization. The first pre-processing step in this study was volume resampling. One of the main challenges in medical image segmentation using deep learning is the volume size, as it is too large to feed into the networks due to the lack of the graphics processing unit (GPU) memory. A patch-wise strategy was an option to solve this issue by breaking down the volume into multiple patches (overlapping or random patches) to fit the GPU memory requirement (23). Yet, this strategy may result in a higher variance among the patches and lose the contextual information (24), especially for the large target. For the bone segmentation, we downsampled the image volume



**FIGURE 1 |** Proposed workflow for the knee bone and cartilage segmentation.



**FIGURE 2 |** The architecture of the bone network.

by a factor of 2 resulting in a new spacing by  $0.8 \times 0.8 \times 2$ . With the resampling step, the input patch can cover more contextual information for bone segmentation. In contrast, the cartilage segmentation based on CNN is relatively sensitive to the resampling due to its small volume size. Hence, for the cartilage segmentation, we input the neural network of the image with the original size.

The second step of pre-processing was the intensity normalization. The imaging noise from the reconstruction of MRI volume, such as DC spike, results in the extreme intensity of some voxels (25). A robust intensity cut-off was selected to prevent the long intensity tail effect for both the bone and cartilage segmentation (25). In this study, the minimum and maximum cut-offs were selected as the threshold with the first and last 2% cumulative intensity histogram. Then, a following z-score strategy was adopted to normalize the intensity by subtracting the mean and dividing by the standard deviation (SD).

## Deep Neural Network for Bone and Cartilage Segmentation

### Architecture of the Networks

Since the advent of U-Net (18), many architecture modifications have been proposed to further improve the performance of the

segmentation task. However, Isensee et al. (26) demonstrated that not all of them were effective and pointed out that a typical U-Net architecture can achieve state-of-the-art results with a thorough design of adaptive pre-processing, training scheme, and inference strategy. In this study, we extended the idea of nnU-Net (26) by adding the adversarial loss to refine the segmentation and used nnU-Net as a baseline for the segmentation performance comparison. The architecture of the bone network was similar to pix2pix network (27) (Figure 2), which consisted of a generator trained for mask prediction and a discriminator trained to discriminate the produced masks ('fake') from ground truth labels ('real') (Figure 2). The framework of the generator in this study consisted of an encoding path to encode the valid features and a decoding path to perform a voxel-based classification. The encoding path contained the repeated layers of two convolutions, followed by an instance normalization, a leaky rectified linear unit, and a max pooling operation with stride 2 for downsampling. The upsampling path also contained the repeated layers of convolution, but a skip connection was adopted by a concatenation of the correspondingly cropped feature from the contraction path and the output of the up convolutions from the last layer. At the final layer, a final  $1 \times 1 \times 1$  convolution was used to map each component feature vector to the desired number of classes, and a Softmax calculation was followed at last

**TABLE 1** | Comparison of automatic segmentation methods based on the Segmentation of Knee Images 2010 (SKI10) validation data.

Team (reference)	Total score	Femur bone		Tibia bone		Femur cartilage		Tibia cartilage	
		AvgD (mm)	RMSD (mm)	AvgD (mm)	RMSD (mm)	VOE (%)	VD (%)	VOE (%)	VD (%)
Vincent et al. (10)	52.3 ± 8.6	0.88 ± 0.24	1.49 ± 0.44	0.74 ± 0.21	1.21 ± 0.34	36.3 ± 5.3	−25.2 ± 10.1	34.6 ± 7.9	74.0 ± 7.7
Seim et al. (9)	54.4 ± 8.8	1.02 ± 0.22	1.54 ± 0.30	0.84 ± 0.19	1.24 ± 0.28	34.0 ± 12.7	7.7 ± 19.2	29.2 ± 8.6	−2.7 ± 18.2
Shan et al. (12)	40.0 ± 7.7	—	—	—	—	—	—	—	—
*Liu et al. (2)	64.1 ± 9.5	0.56 ± 0.12	1.08 ± 0.21	0.50 ± 0.14	1.09 ± 0.28	28.4 ± 6.9	8.1 ± 12.3	33.1 ± 7.1	−1.2 ± 17.4
Dam et al. (30)	67.1 ± 8.0	0.68 ± 0.22	1.25 ± 0.41	0.50 ± 0.18	0.91 ± 0.35	26.9 ± 6.0	0.8 ± 13.5	25.1 ± 6.7	0.41 ± 13.4
*Ambellan et al. (21)	74.0 ± 7.7	0.43 ± 0.13	0.74 ± 0.27	0.35 ± 0.07	0.59 ± 0.19	20.99 ± 5.08	7.18 ± 10.51	19.06 ± 5.18	4.29 ± 12.34
3D U-net (17)	48.1 ± 12.3	1.77 ± 1.85	5.24 ± 3.99	2.60 ± 2.59	7.50 ± 5.29	23.80 ± 7.25	−5.45 ± 8.37	20.60 ± 6.40	5.48 ± 15.11
V-Net (28)	55.7 ± 10.7	0.88 ± 0.61	3.36 ± 2.46	1.04 ± 0.95	4.23 ± 3.53	21.91 ± 4.48	1.17 ± 9.14	20.08 ± 5.62	6.12 ± 16.57
Cascade nnU-Net (26)	75.4 ± 8.1	0.37 ± 0.12	0.63 ± 0.29	0.32 ± 0.15	0.57 ± 0.39	22.71 ± 4.88	1.76 ± 10.03	21.21 ± 5.83	7.05 ± 13.66
*nnU-Net 2D (26)	73.4 ± 10.7	0.37 ± 0.15	0.69 ± 0.35	0.38 ± 0.27	0.80 ± 0.77	21.34 ± 5.59	4.49 ± 11.46	21.43 ± 5.67	5.74 ± 13.41
*nnU-Net 3D full res (26)	72.5 ± 14.2	0.56 ± 1.00	1.67 ± 2.96	0.44 ± 0.57	1.34 ± 2.46	19.45 ± 5.06	6.79 ± 10.29	18.09 ± 5.09	8.32 ± 11.31
*nnU-Net 3D low res (26)	75.3 ± 9.3	0.35 ± 0.12	0.65 ± 0.30	0.34 ± 0.23	0.75 ± 1.19	21.72 ± 4.70	3.66 ± 12.14	21.78 ± 5.39	6.58 ± 12.11
*Proposed method	76.2 ± 7.6	0.38 ± 0.15	0.69 ± 0.37	0.29 ± 0.07	0.52 ± 0.12	19.45 ± 5.06	6.78 ± 10.29	18.09 ± 5.09	8.32 ± 11.31

\* indicates the deep learning-related method; 'res' indicates resolution.

to output a probability for each class. Both the U-Net-like (17) and V-Net-like (28) architectures were used for the generator in this study, which might result in some slight variations compared to the above description, and the detail of all the used networks in this study is summarized in the **Supplementary Material**.

The architecture of the discriminator of the bone network was a convolutional 'PatchGAN' classifier that uses the module form of convolution-batch normalization-ReLu (27). The input of the discriminator was the combination of the image patch and the corresponding segmentation patch. The detail of the architecture is provided in the **Supplementary Material**.

The input of the restoration network was the concatenation of the resampled image and the segmented mask from the bone network. The architecture of the restoration network consisted of two convolutional layers, followed by an upsampled deconvolutional layer, and then finally another two convolutional layers to convert the feature maps into the desired number of classes.

The architecture of the cartilage network was nnU-Net 3D at full resolution (26). The input of the cartilage network was in the original resolution, with a patch size of  $160 \times 192 \times 64$ .

The details of both the cartilage network and restoration network are described in the **Supplementary Material**.

### Loss Function

As **Figure 2** and Equation (1) illustrate, to test the optimal loss options for a robust knee bone segmentation, the loss function,  $\mathcal{L}_{gen}$ , used in the generator (bone network) consisted of three parts: category cross-entropy loss ( $\mathcal{L}_{cce}$ ), dice loss ( $\mathcal{L}_{dice}$ ), and adversarial loss ( $\mathcal{L}_{adv}$ ).  $\mathcal{L}_{cce}$  and  $\mathcal{L}_{dice}$  concern the low-level pixel-wise prediction, while the  $\mathcal{L}_{adv}$  preserves the higher-level consistency conditioned on the input.

$$\mathcal{L}_{gen}(x, y; \theta_{gen}, \theta_{disc}) = \lambda_{cce} \mathcal{L}_{cce}(G(x; \theta_{gen}), y) + \lambda_{dice} \mathcal{L}_{dice}(G(x; \theta_{gen}), y) + \lambda_{adv} \mathcal{L}_{adv}(G(x; \theta_{gen}), x; \theta_{disc}), \quad (1)$$

where  $x$  and  $y$  are the input image volume and the corresponding label.  $\lambda_{cce}$ ,  $\lambda_{dice}$ , and  $\lambda_{adv}$  are the weights for the corresponding losses and the loss is ignored if the corresponding weight sets to 0.  $\theta_{gen}$  and  $\theta_{disc}$  are the parameters of the networks of the generator and discriminator, respectively. The pixel-wise category cross-entropy loss is formulated as  $\mathcal{L}_{cce}(\hat{y}, y) = \frac{1}{whd} \sum_i^{whd} \sum_j^c y_{i,j} \ln(\hat{y}_{i,j})$ , where  $c$  represents the number of target classes and  $w$ ,  $h$ , and  $d$  indicate the width, height, and depth of the volume patch. The pixel-wise dice loss is formulated as:

$$\mathcal{L}_{dice}(\hat{y}, y) = - \sum_i^c \frac{2 \sum_j^{whd} y_{i,j} \ln(\hat{y}_{i,j})}{\sum_j^{whd} y_{i,j}^2 + \sum_j^{whd} \ln(\hat{y}_{i,j})^2}.$$

For the adversarial loss, we chose the adversarial loss of the Least Squares Generative Adversarial Network (LSGAN) (29) in this study and, therefore, is formulated as:

$$\mathcal{L}_{adv}(x; \theta_{gen}, \theta_{disc}) = \mathcal{L}_{MSE}(D(G(x; \theta_{gen}); \theta_{disc}), 1), \quad (2)$$

where  $\mathcal{L}_{MSE}(\hat{z}, z) = (\hat{z} - z)^2$ , and  $x$  indicates the input patch. The discriminator attempts to learn the differences between the label and prediction distributions by minimising the loss function as:

$$\mathcal{L}_{disc}(G(x; \theta_{gen}), y) = \mathcal{L}_{MSE}(D(G(x; \theta_{gen}); \theta_{disc}), 0) + \mathcal{L}_{MSE}(y, 1), \quad (3)$$

where  $x$  and  $y$  indicate the input patch and the corresponding annotation, respectively.

For the cartilage network, the loss function is formulated as:

$$\mathcal{L}_{cart}(\hat{y}, y; \theta_{cart}) = \lambda_{cce} \mathcal{L}_{cce}(\hat{y}, y) + \lambda_{dice} \mathcal{L}_{dice}(\hat{y}, y), \quad (4)$$

where  $y$  and  $\hat{y}$  indicate ground truth and the prediction result of the cartilage network, respectively, and  $\theta_{cart}$  indicates the parameters of the cartilage network.

**TABLE 2 |** Segmentation accuracy for the SK10 validation dataset between baseline networks and the proposed methods.

	Femur bone			Tibia bone			Femur cartilage			Tibia cartilage		
	DSC	Sens	Spec	DSC	Sens	Spec	DSC	Sens	Spec	DSC	Sens	Spec
2D	0.98 ± 0.01	0.98 ± 0.02	1.00 ± 0.00	0.98 ± 0.02	0.98 ± 0.03	1.00 ± 0.00	0.88 ± 0.04	0.90 ± 0.04	1.00 ± 0.00	0.86 ± 0.04	0.89 ± 0.06	1.00 ± 0.00
3D F	0.98 ± 0.01	0.98 ± 0.01	1.00 ± 0.00	0.98 ± 0.02	0.98 ± 0.03	1.00 ± 0.00	0.89 ± 0.03	0.92 ± 0.04	1.00 ± 0.00	0.88 ± 0.03	0.92 ± 0.04	1.00 ± 0.00
3D L	0.98 ± 0.01	0.98 ± 0.01	1.00 ± 0.00	0.98 ± 0.01	0.98 ± 0.02	1.00 ± 0.00	0.88 ± 0.03	0.89 ± 0.05	1.00 ± 0.00	0.86 ± 0.04	0.89 ± 0.05	1.00 ± 0.00
Proposed	0.98 ± 0.01	0.98 ± 0.01	1.00 ± 0.00	0.98 ± 0.01	0.98 ± 0.01	1.00 ± 0.00	0.89 ± 0.03	0.92 ± 0.04	1.00 ± 0.00	0.88 ± 0.03	0.92 ± 0.04	1.00 ± 0.00

Two-dimensional (2D), nnU-Net 2D; three-dimensional (3D) F, nnU-Net 3D full resolution; 3D L, nnU-Net 3D low resolution; DSC, dice similarity coefficient; Sens, sensitivity; Spec, specificity.

For the restoration network, the loss function was formulated as:

$$\mathcal{L}_{restore}(\hat{y}, y; \theta_{restore}) = \mathcal{L}_{cce}(\hat{y}, y), \quad (5)$$

where  $y$  and  $\hat{y}$  indicate ground truth and the prediction result of the restoration network, respectively, and  $\theta_{restore}$  indicates the parameters of the cartilage network.

### Training Procedure

One common challenge in deep learning training is limited training data. Data augmentation is one of the options to be taken to prevent overfitting and has been generally accepted as an add-in in the deep learning method. The data augmentation adopted in this study was random scaling (0.85–1.15), random elastic deformations, gamma correction augmentation, and random mirroring along the frontal axis (simulating the left or right knee joint).

In order to implement a fair comparison among the different architectures, the training strategy similar to a previous study (26) was adopted. There are 6,000 training batches in an epoch. The Adam optimizer with an initial learning rate of  $1 \times 10^{-3}$  was utilized for both the generator and the discriminator in this study, and the learning rate was reduced by a factor of 5 if the loss was not improved in the last 5 epochs and the training was stopped if the loss was not improved in the last 20 epochs. The maximum epoch was limited to 500. The proposed deep CNNs were implemented in Python 3.7 using PyTorch with a 3.7-GHz Intel (R) i7 Xeon (R) E5-1620 V2 CPU and a GTX 1080 Ti graphics card with 11 GB GPU memory.

### Inference

In the inference phase, the new input image volume was split into many sub-volume patches and input to the networks. Then, the class of each voxel was determined by the largest probability of the output probability maps from the neural network. At last, we needed to combine all the sub-volume patches back to form a full volume.

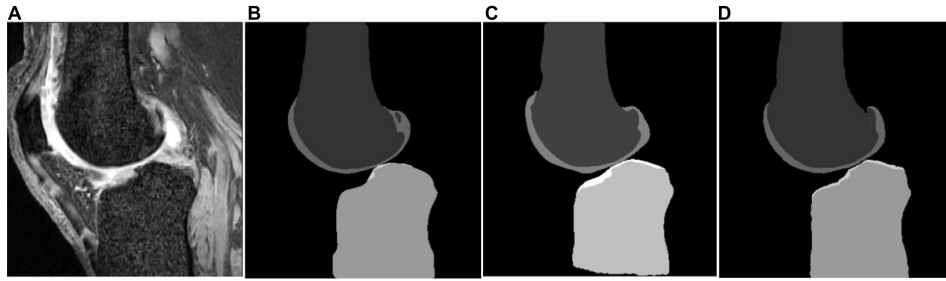
### Post-processing

The main purpose of the post-processing is to combine the advantages of the bone network and the cartilage network in order to generate final bone and cartilage masks. Compared to the cartilage mask from the cartilage network, the bone network could provide less mis-segmented results due to the large contextual information, however, less accurate due to lower resolution. Therefore, the output of the cartilage mask from the bone network after the restoration network was dilated by a  $7 \times 7 \times 7$  kernel, which was later used to filter the cartilage mask from the cartilage network. Finally, the ultimate output of the proposed workflow was the combination of the bone mask from the restoration network and the filtered cartilage mask of the cartilage network.

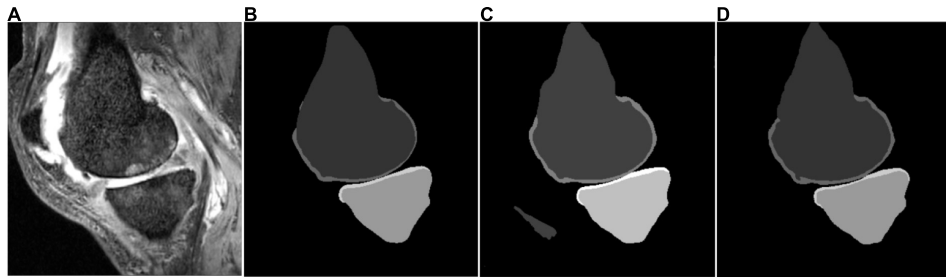
## Evaluation Design

### Methods Designed by the Segmentation of Knee Images 2010

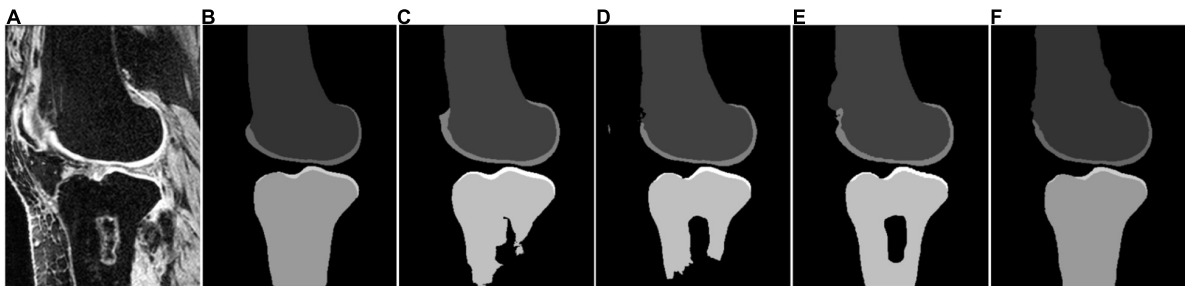
The evaluation method for knee bone and cartilage was different. Regarding bone segmentation, average surface distance (AvgD)



**FIGURE 3** | Segmentation results based on different schemes: **(A)** sagittal slice of the image; **(B)** ground truth; **(C)** nnU-Net two-dimensional (2D); and **(D)** proposed method.



**FIGURE 4** | Segmentation results based on different schemes: **(A)** sagittal slice of the image; **(B)** ground truth; **(C)** nnU-Net three-dimensional (3D) full resolution; and **(D)** proposed method.



**FIGURE 5** | Segmentation results based on different schemes: **(A)** sagittal slice of the image; **(B)** ground truth; **(C)** nnU-Net 2D; **(D)** nnU-Net 3D full; **(E)** nnU-Net 3D low; and **(F)** proposed method.

and root mean square symmetric surface distance (RMSD) were proposed (6, 30).

$$\text{AvgD} = \frac{1}{N_S + N_R} \left( \sum_{i=1}^{N_S} \min_{r \in \partial R} \|s_i - r\|_2 + \sum_{i=1}^{N_R} \min_{s \in \partial S} \|r_j - s\|_2 \right), \quad (6)$$

$$\text{RMSD} = \sqrt{\frac{1}{N_S + N_R} \left( \sum_{i=1}^{N_S} \min_{r \in \partial R} \|s_i - r\|_2 + \sum_{i=1}^{N_R} \min_{s \in \partial S} \|r_j - s\|_2 \right)}, \quad (7)$$

where  $\partial R$  and  $\partial S$  are the boundary of the automatic segmentation and reference segmentation, respectively, and  $N_S$  and  $N_R$  are the number of boundaries, respectively.

For the cartilage segmentation, volume difference (VD) and volume overlap error (VOE) were proposed (6, 21).

$$\text{VD} = 100 \cdot \frac{|S| - |R|}{R}, \quad (8)$$

$$\text{VOE} = 1 - \frac{|S \cap R|}{|S \cup R|}, \quad (9)$$

where  $S$  and  $R$  indicate automatic segmentation and reference segmentation, respectively. As indicated by Heimann et al. (6), the cartilage boundaries to the sides were not always accurate; regions of interest (ROIs) for cartilage mask comparison were used in the above calculation.

**TABLE 3** | Results of the different loss functions based on the proposed network.

Loss	Total score	Femur bone		Tibia bone		Femur cartilage		Tibia cartilage	
		AvgD (mm)	RMSD (mm)	AvgD (mm)	RMSD (mm)	VOE (%)	VD (%)	VOE (%)	VD (%)
CE loss	73.85 ± 9.37	0.43 ± 0.36	1.17 ± 1.88	0.48 ± 0.79	1.25 ± 2.64	21.46 ± 5.17	4.44 ± 9.83	18.44 ± 5.07	6.11 ± 13.49
SD loss	67.54 ± 14.78	0.92 ± 1.50	2.57 ± 4.15	0.82 ± 1.85	2.08 ± 4.57	19.89 ± 5.70	7.65 ± 10.01	18.64 ± 6.49	13.08 ± 13.04
CE loss + SD loss	74.38 ± 10.39	0.38 ± 0.23	1.08 ± 1.33	0.31 ± 0.30	0.58 ± 0.71	20.00 ± 5.63	6.60 ± 9.98	18.62 ± 5.95	10.75 ± 13.41
Proposed loss	76.2 ± 7.6	0.38 ± 0.15	0.69 ± 0.37	0.29 ± 0.07	0.52 ± 0.12	19.45 ± 5.06	6.78 ± 10.29	18.09 ± 5.09	8.32 ± 11.31

## Dice Similarity Coefficient

The Dice similarity coefficient (DSC) score is defined as:

$$\text{DSC} = \frac{2T_P}{2T_P + F_P + F_N}, \quad (10)$$

$$\text{Sensitivity} = \frac{T_P}{T_P + F_N}, \quad (11)$$

$$\text{Specificity} = \frac{T_N}{T_N + F_P}, \quad (12)$$

where  $T_P$  is true positive,  $T_N$  is false negative,  $F_P$  is false positive, and  $F_N$  is false negative. The thickness difference is calculated by the thickness difference from each vertex along the normal vector between automated and manual segmentation masks.

## RESULTS

**Table 1** summarizes the results of previous studies (2, 9, 10, 12, 21, 30), baseline networks [nnU-Net (26, 31), including the 2D version, 3D full-resolution version, and 3D low-resolution version], and the proposed methods for the SKI10 validation dataset in terms of the SKI10 metrics (6). The bone and cartilage segmentation results with proposed networks reached a total score of  $76.2 \pm 7.6$ , which was for the first time higher than 75 using the validation dataset [the second rater's score was 75 in a previous study (6)]. Overall, the results of deep learning-based methods outperformed the traditional methods [atlas based (12) and statistical shape-based methods (9, 10, 30)]. The new baseline (nnU-Net) could achieve state-of-the-art results without any post-processing. Still, the proposed method outperformed the baseline.

Moreover, **Table 2** shows the accuracy evaluation for the SKI10 dataset between the baseline networks and the proposed methods in terms of the DSC, sensitivity, and specificity. For the cartilage result, the DSC is only calculated in the defined ROI according to a previous study (6). The DSC scores of the proposed method are  $0.98 \pm 0.01$ ,  $0.98 \pm 0.01$ ,  $0.89 \pm 0.03$ , and  $0.88 \pm 0.03$  for femur bone, tibia bone, femur cartilage, and tibia cartilage, respectively. Overall, the performance of the proposed methods achieved the highest score.

Some segmentation results on the SKI10 validation set are shown in **Figures 3–5**, which compared the baseline networks with the proposed method. The results of nnU-Net

2D might mis-segment the low-contrast region (bottom of **Figure 3C**), while the result of nnU-Net 3D full resolution might mis-segment some of the unrelated regions (left bottom of **Figure 4C**). A segmentation result of knee joint image with specific pathological tissue is given in **Figure 5**. All the baseline networks failed to segment it successfully and the proposed method with the adversarial loss showed a robust result (**Figure 5F**).

In addition, an ablation study about the loss function selection is shown in **Table 3**. The proposed loss function is capable of improving the segmentation performance.

Computation time for the whole segmentation pipeline for one subject is measured as around 1 min on a consumer-grade workstation (CPU: Intel Xeon E5 2.3 GHz; GPU: GeForce GTX 1080 Ti).

## DISCUSSION

In this study, we presented an end-to-end deep learning-based workflow for knee bone and cartilage segmentation and evaluated the workflow thoroughly on a published dataset, the SKI10 (6). It was the first time that a total score greater than 76 was achieved on the SKI10 validation dataset, which was comparable to the inter-observer variability of two expert readers (6).

The attempt of applying deep learning-based methods to the knee bone and cartilage segmentation was not new and has achieved a lot of state-of-the-art results (2, 21). Nevertheless, most of the previous attempts added a post-processing step [deformable model (2), conditional random field (CRF) (20) and statistical shape model (SSM) (21)] to refine the outcome of the deep learning methods on the area of false segmentation. The main reason behind this is that the information of highly patient-specific areas might not be derived from the training dataset (21). To confirm the necessity of the post-processing, a generic U-Net architecture with fine-tuned hyper-parameter (31) was tested in this study as the baseline. State-of-the-art results can be achieved using the simple nnU-Net architectures (see **Table 1**). Nonetheless, due to the loss of Z-axis information and contextual information, the performance of bone segmentation of generic 2D nnU-Net and 3D nnU-Net full resolution might perform poorly in the low-contrast region (**Figure 3C**), bone- or cartilage-like region (**Figure 4C**), and region with pathological case (**Figures 5C,D**). **Table 1** has shown a good bone segmentation result using the 3D nnU-Net low-resolution

version but not in cartilage segmentation. This is because the target volume of cartilage is relatively small, which resulted in the loss of the cartilage information, especially in the pathological area. In this sense, whether resampling the image volume is necessary to improve the segmentation performance should be considered carefully based on the size of the target and the memory of the GPU.

Moreover, **Figures 5C–E** have shown that all the nnU-Net architectures fail to segment the bone with the specific pathological feature, which demonstrates the necessity of post-processing from previous studies. In this study, we introduced the adversarial loss to serve as a shape regulation penalty to improve bone segmentation. Although the adversarial loss (32) has been proposed to improve the segmentation performance previously, to the best of our knowledge, it was the first time to serve as a shape consistency term to apply to knee bone MRI image segmentation. **Figure 5F** has shown that the introduction of adversarial loss results in state-of-the-art results for bone segmentation despite the pathological case. In addition, a possible alternative method to improve the segmentation performance for the pathological case is to increase the training set size, especially for the pathological case.

This study has a number of limitations. First of all, due to the limited memory of Nvidia 1080 Ti, the number of feature channels of the first layer in nnU-Net experiments is 20 rather than 30 as stated in previous research (31). Further experiments with a better GPU should be implemented to investigate the performance influence of the number of feature channels. An additional limitation is that there are still a lot of ablation studies, which can be implemented to discuss the segmentation performance based on different choices of hyper-parameters. Nevertheless, we believe that the experiment results are enough to share with the community to help the development of fully automatic segmentation of the knee joint. Moreover, the bone segmentation was segmented in a relatively lower resolution in order to enlarge context information. Isensee et al. (26) proposed a cascaded mode to further improve the low-resolution segmentations. However, the training data for these two networks should be different; otherwise, it will easily result in an over-fitted network. As Isensee (33) stated that the cascaded mode was not so much better than the 3d\_lowres and 3d\_full\_res mode in most cases, we believe that the results of 3d\_lowres and 3d\_full\_res are sufficient to be a baseline and we will add the comparison with a cascaded mode in the future when a more annotated dataset is available.

## REFERENCES

1. Zhang K, Lu W, Marziliano P. Automatic knee cartilage segmentation from multi-contrast MR images using support vector machine classification with spatial dependencies. *Magn Reson Imaging*. (2013) 31:1731–43. doi: 10.1016/j.mri.2013.06.005
2. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R. Deep convolutional neural network and 3D deformable approach for tissue segmentation in

## CONCLUSION

To conclude, we presented a robust pipeline to segment the knee bone and cartilage. The result of the proposed method is the first time achieved more than 76 in a well-known dataset, the SKI10 validation, to the best of our knowledge. The lower-resolution strategy and the introduction of adversarial loss improve the shape consistency of the bone segmentation, while a fine-tuned V-Net network was further boosted to achieve a promising result for the cartilage segmentation. Future studies will include segmentation for more knee joint structures such as ligaments and menisci.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

HC, NZ, TT, and AS conceived the study. HC, NZ, and TT designed the experiments. HC implemented the model and experiments. HC and AS wrote the manuscript. YK, NV, and AS helped to supervise the study. CS and GX helped to test the segmentation performance based on the previous CNN methods (3D U-Net, V-Net, etc.). All authors discussed the results and contributed to the final version of the manuscript.

## FUNDING

This study leading to these results has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement number 323091 awarded to NV. This study was also supported by the China Exchange Programme (CEP) from Koninklijke Nederlandse Akademie Van Wetenschappen (KNAW).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2022.792900/full#supplementary-material>

musculoskeletal magnetic resonance imaging. *Magn Reson Med*. (2018) 79:2379–91. doi: 10.1002/mrm.26841

3. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage*. (2006) 31:1116–28. doi: 10.1016/j.neuroimage.2006.01.015
4. Shim H, Chang S, Tao C, Wang JH, Kwok CK, Bae KT. Knee cartilage: efficient and reproducible segmentation on high-spatial-resolution MR images with the

- semiautomated graph-cut algorithm method. *Radiology*. (2009) 251:548–56. doi: 10.1148/radiol.2512081332
5. McWalter EJ, Wirth W, Siebert M, von Eisenhart-Rothe RM, Hudelmaier M, Wilson DR, et al. Use of novel interactive input devices for segmentation of articular cartilage from magnetic resonance images. *Osteoarthritis Cartilage*. (2005) 13:48–53. doi: 10.1016/j.joca.2004.09.008
  6. Heimann T, Morrison BJ, Styner MA, Niethammer M, Warfield S. Segmentation of knee images: a grand challenge [Conference presentation]. In *Proceedings of the MICCAI Workshop on Medical Image Analysis for the Clinic, Beijing, China*. Beijing (2010).
  7. Heimann T, Meinzer HP. Statistical shape models for 3D medical image segmentation: a review. *Med Image Anal*. (2009) 13:543–63. doi: 10.1016/j.media.2009.05.004
  8. Frappi P, Crozier S, Warfield SK, Ourselin S. Automatic segmentation and quantitative analysis of the articular cartilages from magnetic resonance images of the knee. *IEEE Trans Med Imaging*. (2010) 29:55–64. doi: 10.1109/TMI.2009.2024743
  9. Seim H, Kainmueller D, Lamecker H, Bindernagel M, Malinowski J, Zachow S. Model-based auto-segmentation of knee bones and cartilage in MRI data. In *Proceedings MICCAI Workshop on Medical Image Analysis for the Clinic, Beijing, China*. Beijing (2010).
  10. Vincent G, Wolstenholme C, Scott I, Bowes M. Fully automatic segmentation of the knee joint using active appearance models. *Medical Image Analysis for the Clinic: A Grand Challenge*. Vol. 1. (2010). p. 224.
  11. Lee JG, Gumus S, Moon CH, Kwok CK, Bae KT. Fully automated segmentation of cartilage from the MR images of knee using a multi-atlas and local structural analysis method. *Med Phys*. (2014) 41:092303. doi: 10.1118/1.4893533
  12. Shan L, Zach C, Charles C, Niethammer M. Automatic atlas-based three-label cartilage segmentation from MR knee images. *Med Image Anal*. (2014) 18:1233–46. doi: 10.1016/j.media.2014.05.008
  13. Folkesson J, Dam EB, Olsen OF, Pettersen PC, Christiansen C. Segmenting articular cartilage automatically using a voxel classification approach. *IEEE Trans Med Imaging*. (2007) 26:106–15. doi: 10.1109/TMI.2006.886808
  14. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging*. (2016) 35:1299–312. doi: 10.1109/TMI.2016.2535302
  15. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal*. (2017) 36:61–78. doi: 10.1016/j.media.2016.10.004
  16. Vigneault DM, Xie W, Ho CY, Bluemke DA, Noble JA.  $\Omega$ -Net (Omega-Net): fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks. *Med Image Anal*. (2018) 48:95–106. doi: 10.1016/j.media.2018.05.008
  17. Çiçek O, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation [Conference presentation]. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Athens, Greece*. Athens (2016).
  18. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [Conference presentation]. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Munich, Germany*. Munich (2015).
  19. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network [Conference presentation]. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013, Nagoya, Japan*. Nagoya (2013). doi: 10.1007/978-3-642-40763-5\_31
  20. Zhou Z, Zhao G, Kijowski R, Liu F. Deep convolutional neural network for segmentation of knee joint anatomy. *Magn Reson Med*. (2018) 80:2759–70. doi: 10.1002/mrm.27229
  21. Ambellan F, Tack A, Ehlke M, Zachow S. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: data from the osteoarthritis initiative. *Med Image Anal*. (2019) 52:109–18. doi: 10.1016/j.media.2018.11.009
  22. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. *Med Image Anal*. (2019) 58:101552. doi: 10.1016/j.media.2019.10.1552
  23. Kamnitsas K, Chen L, Ledig C, Rueckert D, Glocker B. Multi-scale 3D convolutional neural networks for lesion segmentation in brain MRI [Conference presentation]. In *Proceedings of MICCAI-ISLES 2015, Munich, Germany*. Munich (2015).
  24. Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. *J Digit Imaging*. (2019) 32:582–96. doi: 10.1007/s10278-019-00227-x
  25. Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp*. (2002) 17:143–55. doi: 10.1002/hbm.10062
  26. Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, et al. *nnU-net: Self-Adapting Framework for U-Net-Based Medical Image Segmentation*. *arXiv [preprint]*. (2018). Available online at: <https://arxiv.org/pdf/1809.10486.pdf> (accessed November 1, 2019).
  27. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks [Conference presentation]. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA*. Honolulu, HI (2017).
  28. Milletari F, Navab N, Ahmadi SA. V-net: fully convolutional neural networks for volumetric medical image segmentation [Conference presentation]. In *Proceedings of the 2016 fourth international conference on 3D vision (3DV), Stanford, CA, USA*. Stanford, CA (2016).
  29. Mao X, Li Q, Xie H, Lau RY, Wang Z, Smolley SP. Least squares generative adversarial networks [Conference presentation]. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy*. Piscataway, NJ: IEEE (2017).
  30. Dam EB, Lillholm M, Marques J, Nielsen M. Automatic segmentation of high- and low-field knee MRIs using knee image quantification with data from the osteoarthritis initiative. *J Med Imaging (Bellingham)*. (2015) 2:024001. doi: 10.1117/1.JMI.2.2.024001
  31. Isensee F, Petersen J, Kohl SA, Jäger PF, Maier-Hein KH. *nnU-net: Breaking the Spell on Successful Medical Image Segmentation*. *arXiv [preprint]*. (2019). Available online at: <https://www.arxiv-vanity.com/papers/1904.08128/>
  32. Yang D, Xu D, Zhou SK, Georgescu B, Chen M, Grbic S, et al. Automatic liver segmentation using an adversarial image-to-image network [Conference presentation]. In *Proceedings of the Medical Image Computing and Computer Assisted Intervention – MICCAI 2017, Quebec City, QC, Canada*. Quebec (2017).
  33. Isensee F. *Problem with Cascade Training*. (2019). <https://github.com/MIC-DKFZ/nnUNet/issues/33> (accessed September 1, 2019).

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chen, Zhao, Tan, Kang, Sun, Xie, Verdonschot and Sprengers. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

**Visit us:** [www.frontiersin.org](http://www.frontiersin.org)

**Contact us:** [frontiersin.org/about/contact](http://frontiersin.org/about/contact)



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership