



Segmentation of retinal vessels in fundus images based on U-Net with self-calibrated convolutions and spatial attention modules

YiBiao Rong^{1,2} · Yu Xiong^{1,2} · Chong Li^{1,2} · Ying Chen^{1,2} · Peiwei Wei^{1,2,3} · Chuliang Wei^{1,2} · Zhun Fan^{1,2}

Received: 5 July 2021 / Accepted: 8 February 2023
© International Federation for Medical and Biological Engineering 2023

Abstract

Automated and accurate segmentation of retinal vessels in fundus images is an important step for screening and diagnosing various ophthalmologic diseases. However, many factors, including the variations of vessels in color, shape and size, make this task become an intricate challenge. One kind of the most popular methods for vessel segmentation is U-Net based methods. However, in the U-Net based methods, the size of the convolution kernels is generally fixed. As a result, the receptive field for an individual convolution operation is single, which is not conducive to the segmentation of retinal vessels with various thicknesses. To overcome this problem, in this paper, we employed self-calibrated convolutions to replace the traditional convolutions for the U-Net, which can make the U-Net learn discriminative representations from different receptive fields. Besides, we proposed an improved spatial attention module, instead of using traditional convolutions, to connect the encoding part and decoding part of the U-Net, which can improve the ability of the U-Net to detect thin vessels. The proposed method has been tested on Digital Retinal Images for Vessel Extraction (DRIVE) database and Child Heart and Health Study in England Database (CHASE DB1). The metrics used to evaluate the performance of the proposed method are accuracy (ACC), sensitivity (SE), specificity (SP), F1-score (F1) and the area under the receiver operating characteristic curve (AUC). The ACC, SE, SP, F1 and AUC obtained by the proposed method are 0.9680, 0.8036, 0.9840, 0.8138 and 0.9840 respectively on DRIVE database, and 0.9756, 0.8118, 0.9867, 0.8068 and 0.9888 respectively on CHASE DB1, which are better than those obtained by the traditional U-Net (the ACC, SE, SP, F1 and AUC obtained by U-Net are 0.9646, 0.7895, 0.9814, 0.7963 and 0.9791 respectively on DRIVE database, and 0.9733, 0.7817, 0.9862, 0.7870 and 0.9810 respectively on CHASE DB1). The experimental results indicate that the proposed modifications in the U-Net are effective for vessel segmentation.

Keywords Retinal vessel segmentation · Self-calibrated convolutions · Spatial attention modules

1 Introduction

As the ophthalmic diseases, including diabetic retinopathy, arteriosclerosis and leukemia, can make the retinal vessels change in length, width, angle and vascular proliferation, the structural information of retinal vessels in the fundus

images is important for the diagnosis of ophthalmic diseases [4, 5, 8, 24, 28]. In clinical practice, to obtain the structural information, retinal vessels are manually annotated by ophthalmologists. The disadvantages of manual annotation are obvious, including time-consuming and error-prone. To reduce the workload of ophthalmologists and increase the accuracy, many methods have been proposed for segmenting retinal vessels in fundus images automatically, which can be divided into unsupervised methods and supervised methods roughly.

1.1 Unsupervised methods

The unsupervised methods could be divided into several categories further, which are morphology based methods, vessel tracking methods, model based methods and matched filtering based methods. The morphology based methods [3,

✉ Zhun Fan
zfan@stu.edu.cn

¹ Department of Electronic and Information Engineering, Shantou University, 515063, Guangdong, China

² Key Lab of Digital Signal and Image Processing of Guangdong Province, Shantou University, 515063, Guangdong, China

³ Department of Microbiology and Immunology, Shantou University Medical College, Guangdong 515041, China

[18] are generally designed based on the observation that the intensity of retinal vessels is lower than the background brightness. The vessel tracking methods [16] mainly employ local information to detect vessels with the seed points. The model based methods [10, 32] apply explicit vessel profile to extract the vasculature. The matched filtering based methods [2] employ the filters, which are designed based on the characters of vessels, to enhance the vessels in the images.

1.2 Supervised methods

One of advantages for the unsupervised methods is that they do not require the manual labels to train the models. However, the unsupervised methods are designed based on some strong assumptions, e.g., the intensity of blood vessels is lower than the background brightness [18], once the vessels do not meet the assumptions, the designed methods would fail to detect the vessels.

To avoid making a strong assumption for vessel segmentation, the traditional supervised methods [14, 20, 21, 25] are an alternative way. There are two main steps in the traditional supervised methods, which are feature extraction and classification. Concretely, a series of features are first designed. A classifier is then trained based on the designed features. Once training done, the trained classifier could be used to determine a pixel whether belongs to vessel pixels. In the process, the classifiers are designed based on some common machine learning algorithms, including support vector machine [21], K-nearest neighbor [25], and AdaBoost [14], while the features, e.g., Gabor filter based features [9] and the Gaussian filter based features [20], are generally designed by hand. There are several disadvantages with hand-tuned features, including being time-consuming and difficult to be generalized to other domains.

As the revival of deep learning [11], especially deep convolutional neural networks [12], many researchers have employed them for different tasks, including the task of vessel segmentation. Note that deep convolutional neural networks can learn hierarchical features through multiple levels of abstraction from images automatically, and thus can avoid the process of designing features by hand. The existing methods for vessel segmentation with deep convolutional neural networks are mainly designed based on the structure of U-Net [22], which is a very popular network structure for medical image segmentation. Such as, Alom et al. [1] designed a recurrent convolutional neural network based on the U-Net to segment retinal vessel. Wang et al. [26] proposed a dual encoding method for the U-Net to enhance the capability of the networks in segmentation of vessels. Besides, CSU-Net [27], CS2-Net [19] and Genetic U-Net [29] were also designed for vessel segmentation.

Although the improved U-Nets could achieve good performance, they generally adopted traditional convolution operations, where the size of the convolution kernels is fixed. As a result, the receptive field of an individual convolution operation is single, which does not conducive to the segmentation of retinal vessels with various thicknesses. In addition, the information of thin vessels can be easily filtered by the traditional convolution operations in the U-Net. In this paper, we would like to make the U-Net can learn discriminative representations from different receptive fields, so that the U-Net has a stronger capability in detecting the vessels with different thickness. Besides, we also would like to improve the ability of the U-Net in capturing the thin vessels.

To make the U-Net can learn discriminative representations from different receptive fields, we employed self-calibration convolutions (SCC) [17] to replace the traditional convolutions. In a self-calibration convolution, the low-dimensional features are used to calibrate the high-dimensional features, as a result, different receptive fields can be obtained to help the networks learn discriminative features. Concretely, the convolution filters in a specific layer are separated into multiple portions. The low-dimensional feature information is then obtained through the down-sampling operation, which is used to calibrate the convolution transformation of another filter. Benefiting from the heterogeneous convolutions in a self-calibrated convolution, the receptive field of each spatial location can be effectively amplified.

To enhance the ability of the U-Net in capturing the thin vessels, we proposed an improved spatial attention module (ISAM) to connect the encoding part and decoding part of U-Net. As we know, the spatial attention modules (SAM) [30] could find out the areas that need to be attended from the image information. However, the traditional spatial attention modules ignored the spatial information of thin vessels since the information of thin vessels could be easily filtered by convolution operations. In the proposed spatial attention module, the self-calibrated convolutions are included. The convolution operation and spatial attention map generation are performed in a parallel way. As a result, more complete information of vessels can be preserved.

A similar work has been reported in [7], where the traditional U-Net is also improved for retinal vessel segmentation in fundus images via changing the convolutional blocks and employing the spatial attention modules to connect the encoding part and the decoding part in the traditional U-Net. However, there is essential difference between the work presented in [7] and the proposed work. In [7], the authors employed structured dropout convolutional blocks instead of the original convolutional blocks of the U-Net to prevent the network from overfitting. While in this work, we employed self-calibrated convolutions to replace

the original convolutional blocks to make the network can learn discriminative representations from different receptive fields. Besides, although both of the proposed work and the work presented in [7] employed spatial attention module to connect the encoding part and the decoding part of the U-Net, in this work we proposed a modification for the spatial attention module, which is integrating the self-calibrated convolutions into the spatial attention modules.

The contributions of this paper are concluded as follows:

- We employed self-calibrated convolutions to replace the traditional convolutions for the U-Net, which can make the U-Net learn discriminative representations from different receptive fields.
- We proposed an improved spatial attention module, instead of using traditional convolutions, to connect the encoding part and decoding part of U-Net, which can improve the ability of the U-Net to detect thin vessels.
- We verified the proposed modifications for the U-Net via a series of experiments. The experiments showed that the proposed modifications are effective for the U-Net to detect the retinal vessels with various thicknesses.

2 Method

The proposed network structure for vessel segmentation is shown in Fig. 1, which is designed based on the traditional U-Net [22]. There are two differences between the proposed network and the traditional U-Net. The first one is that we employed self-calibrated convolutions to replace the traditional convolutions, shown as black arrows in Fig. 1. The second is that we proposed an improved spatial attention module to connect the encoding part and decoding part of the U-Net, shown as the red block in Fig. 1. Note that, in the traditional U-Net, the encoding part and decoding part are connected via traditional convolutions also.

2.1 Self-calibrated convolutions

Convolution operations are widely used for feature extraction in convolutional neural networks. However, the size of filters for convolutions is generally fixed, which makes it difficult to capture the feature information of different scales. Instead of using the traditional convolution operations, in this paper, we employed self-calibrated convolutions [17] to help the networks learn discriminative representations by augmenting the basic convolution transformation per layer. The work flow of a self-calibrated convolution is illustrated in Fig. 2. In the self-calibrated convolution, the convolution filters in a specific layer are separated into

multiple portions. The low-dimensional feature information is then obtained through the down-sampling operation, which is used to calibrate the convolution transformation of another filter. Benefiting from the heterogeneous convolutions in a self-calibrated convolution, the receptive field of each spatial location can be effectively amplified.

Mathematically, given the input feature $F \in \mathbb{R}^{H \times W \times C}$, which is split into two portions $\{F_1 \in \mathbb{R}^{H \times W \times C/2}, F_2 \in \mathbb{R}^{H \times W \times C/2}\}$ by two 1×1 convolutions, where H , W represent the height and the width of the feature map respectively, C represents the number of channels. The output of the first pathway, as shown in Fig. 2, is $F'_1 \in \mathbb{R}^{H \times W \times C/2}$, which is obtained by performing self-calibrated convolutions on F_1 . This process can be represented by Eq. (1). The output of second pathway is $F'_2 \in \mathbb{R}^{H \times W \times C/2}$, which is obtained by performing traditional convolutions on F_2 , as represented in Eq. (2), and thus the original spatial context is retained in F'_2 .

$$F'_1 = f[\sigma(Up(f(Down(F_1)))) \oplus F_1 \odot f(F_1)] \quad (1)$$

$$F'_2 = f(F_2) \quad (2)$$

In Eq. (1) or Eq. (2), σ denotes the sigmoid function, f represents a convolution operation, $Down$ and Up represent the Down-sampling and Up-sampling respectively. \oplus and \odot represent element-wise addition and multiplication respectively. The final output of a self-calibrated convolution $F' \in \mathbb{R}^{H \times W \times C}$ is the concatenation of the feature map F'_1 and F'_2 , namely,

$$F' = Concatenate \{F'_1; F'_2\} \quad (3)$$

2.2 Spatial attention modules

The spatial attention modules (SAM) [30] can find out the areas that need to be attended from the image information. However, in the task of retinal vessel segmentation, the traditional spatial attention modules ignore the spatial information of thin vessels since the information of thin vessels can be easily filtered by convolution operations. To obtain more complete spatial information, we proposed an improved spatial attention module based on the convolutional attention module [30] to connect the encoding part and decoding part of the traditional U-Net. Figure 3 shows the original spatial attention module and the improved spatial attention module respectively. In the improved spatial attention module, the self-calibrated convolutions are included. The convolution operation and spatial attention map generation are performed in a parallel way. As a result, we can obtain the feature maps with more complete spatial information to improve the segmentation performance.

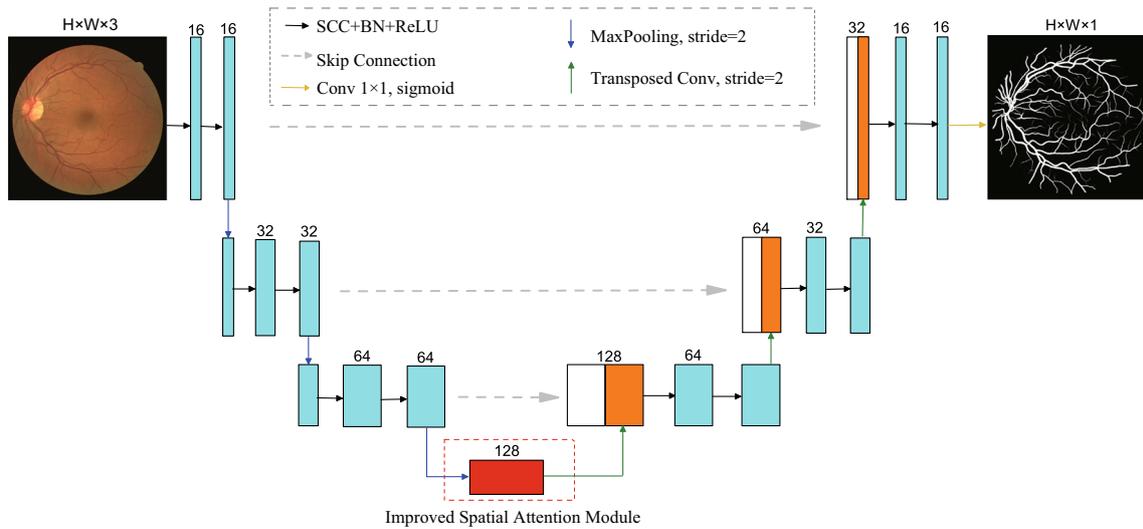


Fig. 1 The structure of the proposed network. There are two differences between the proposed network and the traditional U-Net. The first one is that we employed self-calibrated convolutions to replace the traditional convolutions, shown as black arrows in the figure. The

second is that we proposed an improved spatial attention module to connect the encoding part and decoding part of the U-Net, shown as the red block in the figure. Note that, in the traditional U-Net, the encoding part and decoding part are connected via traditional convolutions also

To compute the spatial attention map $M_s \in \mathbb{R}^{H \times W \times 1}$, given the input features $F_s \in \mathbb{R}^{H \times W \times C}$, the average-pooling and max-pooling operations are applied along the channel axis to get outputs $F_{MP} \in \mathbb{R}^{H \times W \times 1}$ and $F_{AP} \in \mathbb{R}^{H \times W \times 1}$ respectively. The output feature $F'_s \in \mathbb{R}^{H \times W \times C}$ of the improved spatial attention module is calculated as:

$$\begin{aligned}
 F'_s &= SCC(SCC(F_s)) \otimes M_s(F_s) \\
 &= SCC(SCC(F_s)) \otimes \sigma(f([MaxPool(F_s); AvgPool(F_s)])) \\
 &= SCC(SCC(F_s)) \otimes \sigma(f([F_{MP}; F_{AP}]))
 \end{aligned}
 \tag{4}$$

where $SCC()$ denotes the self-calibrated convolutions, the σ is sigmoid function and f represents a convolution operation.

3 Experiments

3.1 Databases

The proposed method has been verified on two public databases for vessel segmentation, which are DRIVE (Digital Retinal Images for Vessel Extraction) [25] and CHASE DB1 (Child Heart and Health Study in England Database) [6]. DRIVE database consists of 40 color fundus photographs obtained from a diabetic retinopathy screening program, in which 20 samples were used for training and the rest were used for testing. Each fundus image in DRIVE is composed of 565×584 pixels and the corresponding ground truth was annotated by human observers. CHASE DB1 contains 28 retinal images and the size of each image

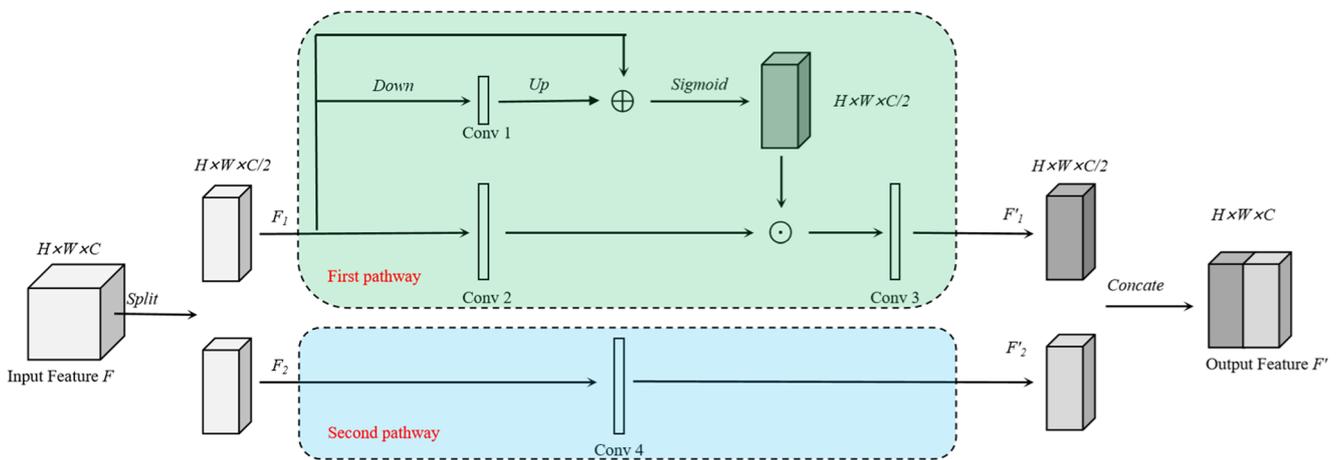


Fig. 2 The work flow of a self-calibrated convolution

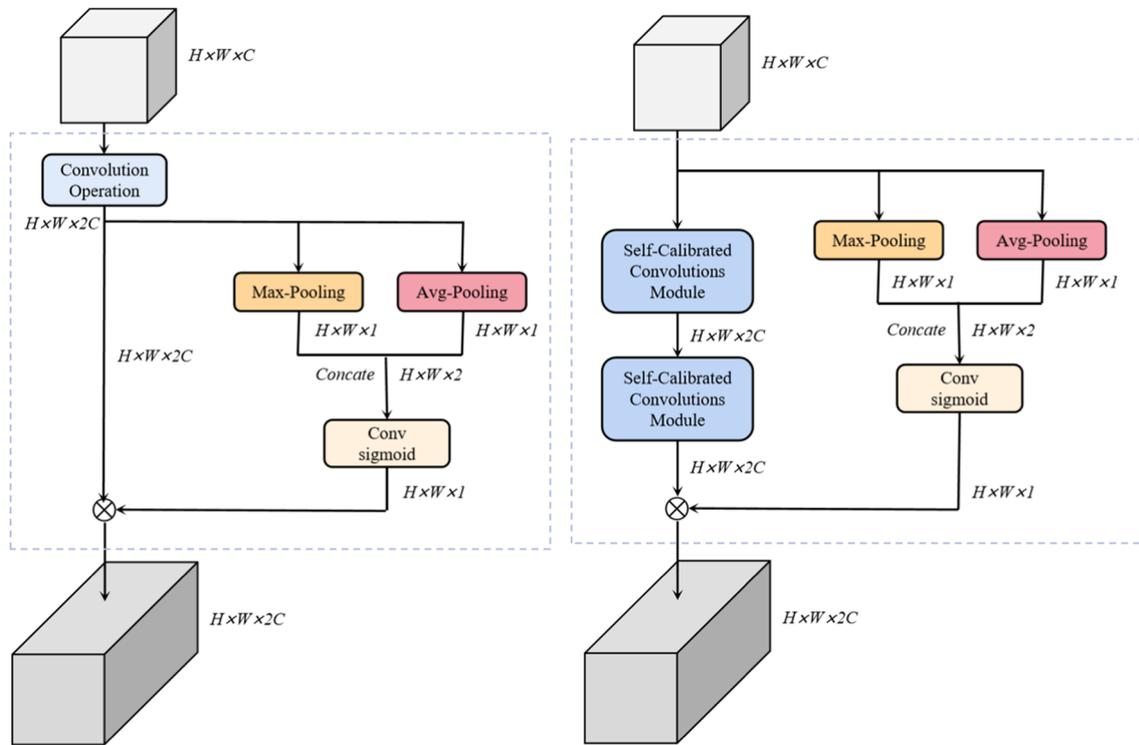


Fig. 3 The original spatial attention module (left) and the proposed spatial attention module (right)

is 999×960 pixels. The images in CHASE DB1 were collected from both left and right eyes of 14 children. To train and test the proposed method, the images in CHASE DB1 were divided into training set and test set also, in which the first 20 images were used for training and the others for testing.

3.2 Implementation details

3.2.1 Loss function

The task of retinal vessel segmentation can be regarded as a binary classification problem in pixel level. Since cross entropy is widely used as the loss function in deep learning networks to deal with binary classification problems, in this work, the cross entropy is used as the loss function to train the networks:

$$Loss = - \sum_{i=1} \hat{y}_i \log y_i + (1 - \hat{y}_i) \log(1 - y_i) \quad (5)$$

where \hat{y}_i represents the ground truth and y_i the segmented results obtained by the proposed method.

3.3 Metrics

The metrics used to evaluate the performance of the proposed method are accuracy (ACC), sensitivity (SE), specificity (SP), F1-score (F1) and the area under the

receiver operating characteristic curve (AUC). Table 1 summarizes the definition of each metric, where TP is short for true positive, FP false positive, TN true negative and FN false negative.

3.4 Results

3.4.1 Empirical study of parameter setting

A computer equipped with NVIDIA TITAN Xp GPU with 12GB memory was used to train the networks based on the Keras and TensorFlow frameworks. During training, the stochastic gradient descent was used for minimizing the loss function. The hyperparameters for the proposed method included learning rate, the number of epochs and the batch size, which might influence the performance of the proposed method. We employed the DRIVE database

Table 1 The metrics used to evaluate the performance of the proposed method

Metrics	Description
$ACC(accuracy)$	$ACC=(TP+TN)/(TP+FN+TN+FP)$
$SE(sensitivity)$	$SE=TP/(TP+FN)$
$SP(specificity)$	$SP=TN/(TN+FP)$
$F1(F1-score)$	$F1=(2 \times TP)/(2 \times TP+FP+FN)$
AUC	Area Under the ROC curve

to study the influence of the hyperparameters on the performance of the proposed method. We set a default value for each parameter. Then we allowed one change while the others were equal to the default values. As a result, we can explore how varying the hyperparameter may affect the performance of the proposed method. The default values of learning rate, the number of epochs and the batch size are 0.001, 40 and 4 respectively.

Figure 4(a), (b) and (c) summarize how varying the learning rate, the number of epochs and the batch size affected the performance respectively. It can be observed that a larger learning rate (e.g., 0.1) or a smaller learning rate (e.g., 0.0001) could lead to a deteriorated performance, which indicates that the proposed method got stuck in a worse local minima during training. The performance of the proposed method is improved with the increase of the number of epochs until it reaches 40. When the number of the epochs is greater than 40, the performance of the algorithm tends to be stagnant. The influence of the batch size on the performance of the proposed is slight. However, if the batch size is set too large, more memory is required during training. According to the above analysis, we set the learning rate, the number of epochs and the batch size equal to 0.001, 40 and 4 for the following experiments.

3.4.2 Ablation studies

Figure 5 shows some segmentation examples obtained by the proposed method. In Fig. 5, the pixels are represented with different color, in which green color means that the pixels are segmented correctly by the proposed method, namely belongs to TP category, red FP, black TN, and blue FN.

In order to verify the performance of the proposed method, ablation experiments were performed on DRIVE and CHASE DB1 respectively. Tables 2 and 3 summarize the results obtained by U-Net, U-Net+SCC, U-Net+SAM,

U-Net+ISAM and U-Net+SCC+ISAM respectively. Note that U-Net+SCC means that the convolution operations in the traditional U-Net are replaced by self-calibrated convolutions. U-Net+SAM means that the traditional spatial attention module is used to replace the traditional convolutions to connect the encoding part and decoding part of the U-Net. U-Net+ISAM means that the improved spatial attention module is used to replace the traditional convolution operations to connect the encoding part and decoding part of the U-Net. U-Net+SCC+ISAM means that not only the convolution operations in the traditional U-Net are replaced by self-calibrated convolutions, but also the traditional convolution operations used to connect the encoding part and decoding part of U-Net are replaced by the improved spatial attention module. The last column in Tables 2 and 3 summarize the number of parameters of each model.

It can be observed that the modified U-Net is more complex than the traditional U-Net. The number of parameters in the traditional U-Net is 6.38M. While the number of parameters in U-Net+SCC reaches to 8.68M. Although the modified U-Net is more complex than the traditional U-Net, the performance of the proposed method is increased. As can be observed from Tables 2 and 3, the network with the self-calibrated convolutions performs better than the traditional U-Net, which indicates that self-calibrated convolutions are effective for retinal vessel segmentation compared with the traditional convolution operations. In addition, the performance of the U-Net with the improved spatial attention module is better than the one with traditional spatial attention module, which verifies the effectiveness of the improved spatial attention module. Besides, the results obtained by U-Net+SCC+ISAM are best on most of metrics, which indicates that the proposed modifications on the U-Net are effective for vessel segmentation.

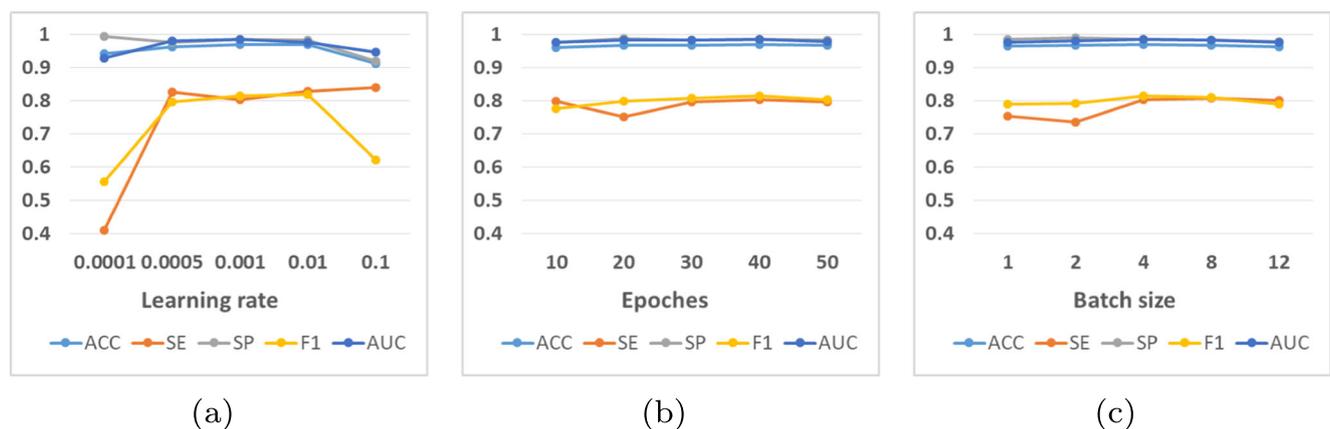
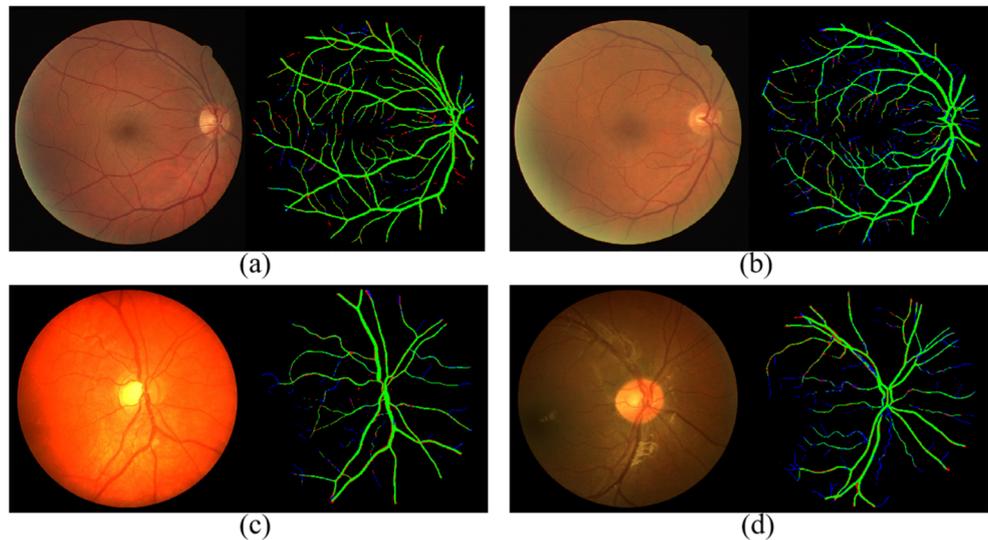


Fig. 4 (a) The performance with different learning rate. (b) The performance with different number of epochs. (c) The performance with different batch size

Fig. 5 Detected examples on DRIVE (first row) and CHASE DB1 (second row) databases. (a) $F1 = 0.8497$, $ACC = 0.9730$, $SE = 0.9191$, $SP = 0.9779$, $AUC = 0.9912$; (b) $F1 = 0.7888$, $ACC = 0.9611$, $SE = 0.7288$, $SP = 0.9868$, $AUC = 0.9786$; (c) $F1 = 0.8228$, $ACC = 0.9802$, $SE = 0.8638$, $SP = 0.9876$, $AUC = 0.9928$; (d) $F1 = 0.7658$, $ACC = 0.9716$, $SE = 0.7334$, $SP = 0.9877$, $AUC = 0.9839$



To demonstrate the effectiveness of the proposed modifications on the U-Net visually, an example is given in Fig. 6, which contains a fundus image, ground truth and the segmentation results obtained by U-Net, U-Net+SCC, U-Net+SAM, U-Net+ISAM and U-Net+SCC+ISAM respectively. It is observed that U-Net+SCC achieves a better segmentation result for vessels with various thicknesses compared with the U-Net. Besides, although the ability of U-Net+SAM in inhibition to noise is obvious, many thin vessels can not be detected by U-Net+SAM. Compared with the U-Net+SAM, U-Net+ISAM can preserve more detailed information of thin vessels, which proves the power of the proposed spatial attention module in detecting thin vessels. When we combined the proposed modifications for the U-Net together, we could obtain more correct and continuous blood vessels, which verifies the effectiveness of the proposed modifications for the U-Net further.

3.4.3 Comparison with state-of-the-art methods

Several state-of-the-art methods for vessel segmentation, including R2U-Net [1], DEU-Net [26] and CSU-Net [27], were selected for comparison. Tables 4 and 5 summarize the results obtained by different methods on DRIVE and CHASE DB1 respectively. The ACC, SE, SP, F1 and AUC obtained by the proposed method are 0.9680, 0.8036,

0.9840, 0.8138 and 0.9840 respectively on DRIVE, and 0.9756, 0.8118, 0.9867, 0.8068 and 0.9888 respectively on CHASE DB1, which are very competitive with the state-of-the-art methods, or even better than some of the state-of-the-art methods, including the DEU-Net [26] and CSU-Net [27]. The comparative experiments demonstrate that the proposed modifications for the U-Net are effective for retinal vessel segmentation. In addition, it could be observed that the Genetic U-Net [29], which is a kind of neural architecture search based method, shows impressive performance in vessel segmentation. As will be pointed out in the discussion section, we would employ the proposed modules to augment the search space for the neural architecture search based methods as a future work, so that the neural architecture search based methods can search a more powerful network for vessel segmentation.

4 Discussion and conclusion

In this paper, we proposed two modifications for the traditional U-Net for retinal vessel segmentation in fundus images. Concretely, we employed self-calibrated convolution operations to replace the traditional convolution operations in the U-Net. As a result, the networks can learn discriminative representations from different receptive fields.

Table 2 Ablation studies on DRIVE dataset

Methods	ACC	SE	SP	F1	AUC	Params
U-Net	0.9646	0.7895	0.9814	0.7963	0.9791	6.38M
U-Net+SCC	0.9665	0.8067	0.9818	0.8084	0.9810	8.68M
U-Net+SAM	0.9659	0.7795	0.9838	0.8001	0.9780	6.39M
U-Net+ISAM	0.9663	0.8066	0.9820	0.8089	0.9815	6.39M
U-Net+SCC+ISAM	0.9680	0.8036	0.9840	0.8138	0.9840	8.69M

Table 3 Ablation studies on CHASE DB1 dataset

Methods	ACC	SE	SP	F1	AUC	Params
U-Net	0.9733	0.7817	0.9862	0.7870	0.9810	6.38M
U-Net+SCC	0.9744	0.8158	0.9851	0.8006	0.9872	8.68M
U-Net+SAM	0.9740	0.7933	0.9861	0.7934	0.9859	6.39M
U-Net+ISAM	0.9753	0.8114	0.9863	0.8053	0.9862	6.39M
U-Net+SCC+ISAM	0.9756	0.8118	0.9867	0.8068	0.9888	8.69M

In addition, we designed an improved spatial attention module, instead of using traditional convolution operations, to connect the encoding part and decoding part in the U-net, which can improve the ability of U-Net in detecting thin vessels.

To verify the effectiveness of the self-calibrated convolution operations, a comparative experiment was conducted. Concretely, the traditional convolution operations in the

U-Net were replaced by the self-calibrated convolution operations to obtain the modified U-net. The modified U-Net and the traditional U-net were then compared in the same databases. We found that the modified U-Net outperforms the traditional U-Net, which indicated that the self-calibrated convolution operations are more effective than the traditional convolution operations in capturing the vessel information. Similarly, We also conducted a comparative

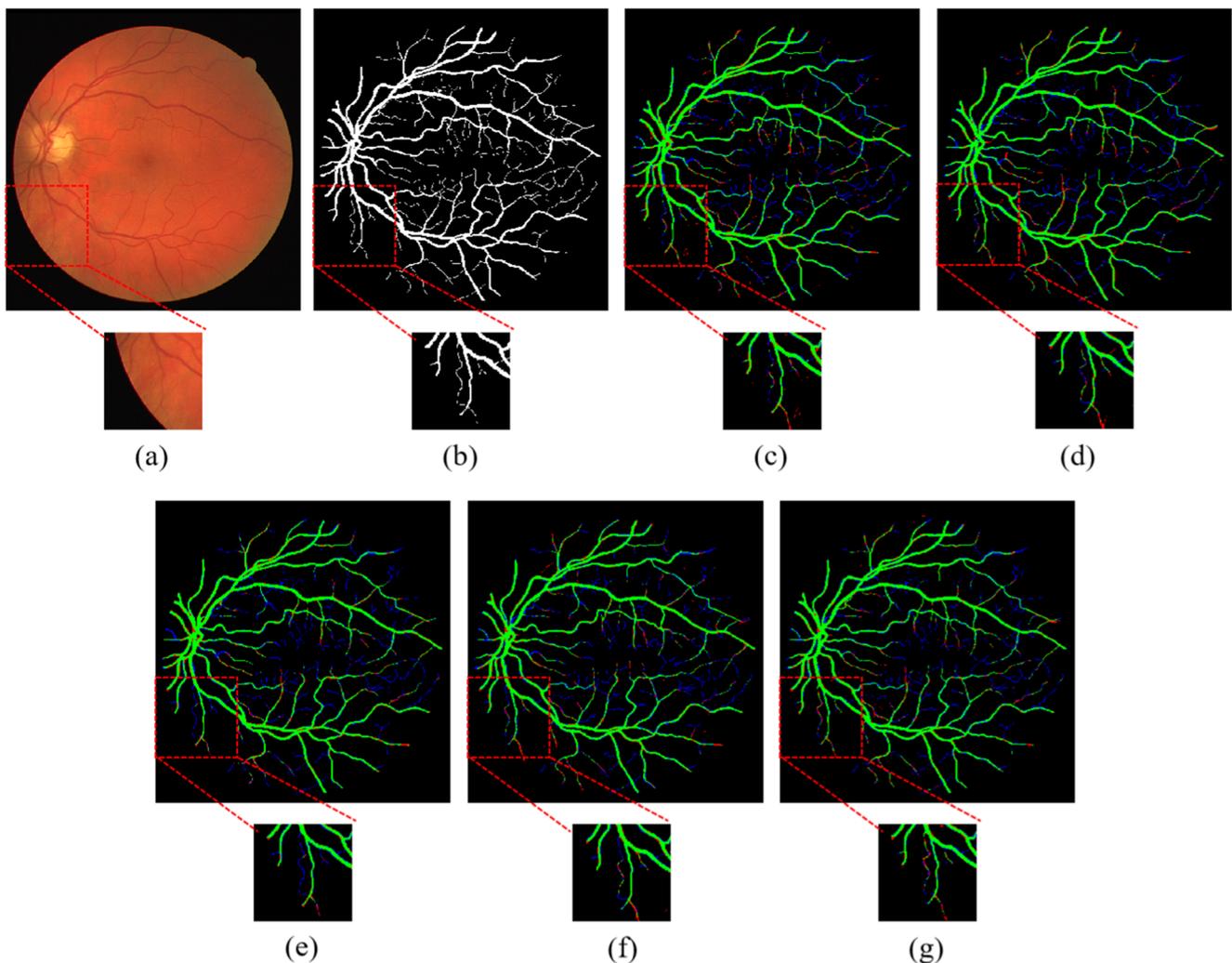

Fig. 6 (a) Fundus image; (b) Ground truth; (c) U-Net; (d) U-Net+SCC; (e) U-Net+SAM; (f) U-Net+ISAM; (g) U-Net+SCC+ISAM

Table 4 Comparison with different methods on DRIVE

Methods	Year	ACC	SE	SP	F1	AUC
Azzopardi et al. [2]	2015	0.9442	0.7655	0.9704	–	0.9614
Roychowdhury et al. [23]	2016	0.9520	0.7250	0.9830	–	0.9620
Liskowsk and Krawiec. [15]	2016	0.9495	0.7763	0.9768	–	0.9720
Li et al. [13]	2016	0.9527	0.7569	0.9816	–	0.9738
MS–NFN [31]	2018	0.9567	0.7844	0.9819	–	0.9807
R2U–Net [1]	2019	0.9556	0.7792	0.9813	0.8171	0.9784
DEU–Net [26]	2019	0.9567	0.7940	0.9816	0.8270	0.9772
CSU–Net [27]	2020	0.9565	0.8071	0.9782	0.8251	0.9801
CS2–Net [19]	2021	0.9553	0.8154	0.9757	0.8228	0.9784
Genetic U–Net [29]	2022	0.9707	0.8300	0.9843	0.8314	0.9885
U–Net+SCC+ISAM (Proposed)	–	0.9680	0.8036	0.9840	0.8138	0.9840

experiment to verify the effectiveness of the improved spatial attention module. As the results shown in Tables 2 and 3, we found that the improved spatial attention module is more effective in capturing the thin vessels. Moreover, many state-of-the-art methods for vessel segmentation, including CSU-Net [27], CS2-Net [19] and Genetic U-Net [29], were selected for comparison. The comparative experiments showed that the proposed method is very competitive with the state-of-the-art methods.

From Tables 2 and 3, we can also observe that the modified U-Net is more complex than the traditional U-Net. Even so, we think the increased complex in the modified U-Net is acceptable for improving the performance. As we know, detecting the vessels with various thickness automatically and accurately plays a very important role for the precision medicine, meanwhile is a very challenging task. In this paper, we proposed two modifications to make the U-Net can learn discriminative representations from different receptive fields and to improve the ability of the U-Net in detecting the thin vessels. We verified the effectiveness of the proposed modifications for the U-Net

via a series of experiments. Figure 6 is an example to demonstrate the superiority of the proposed method, where Fig. 6(c) is the results obtained by the traditional U-Net and Fig. 6(g) the proposed method. As can be observed, the proposed method can obtain a more correct result. In addition, more thin vessels are preserved.

Note that it does not mean that the more complex the model, the better the performance. As the work presented in [29], the architecture obtained by the neural architecture search based methods offered a superior performance with less parameters. On the other hand, the proposed method might fail to capture the vessels when the contrast between the vessels and non-vessels is low. To address the limitation of the proposed method, we think a promising way is employing the proposed modules to augment the search space for the neural architecture search based methods, so that the neural architecture search based methods can search a more powerful network for vessel segmentation. This opinion, namely employing the proposed modules to augment the search space for the neural architecture search based methods, would be considered in our future work.

Table 5 Comparison with different methods on CHASE DB1

Methods	Year	ACC	SE	SP	F1	AUC
Azzopardi et al. [2]	2015	0.9442	0.7655	0.9704	–	0.9614
Roychowdhury et al. [23]	2016	0.9530	0.7201	0.9824	–	0.9532
Li et al. [13]	2016	0.9581	0.7507	0.9793	–	0.9793
MS–NFN [31]	2018	0.9637	0.7538	0.9847	–	0.9825
R2U–Net [1]	2019	0.9634	0.7756	0.9820	0.7928	0.9815
DEU–Net [26]	2019	0.9661	0.8074	0.9821	0.8037	0.9812
CSU–Net [27]	2020	0.9706	0.8427	0.9836	0.8105	0.9824
CS2–Net [19]	2021	0.9651	0.8329	0.9784	0.8141	0.9851
Genetic U–Net [29]	2022	0.9769	0.8463	0.9857	0.8223	0.9914
U–Net+SCC+ISAM (Proposed)	–	0.9756	0.8118	0.9867	0.8068	0.9888

Funding This work was supported in part by the National Key R&D Program of China under grant 2021ZD0111502, in part by the Science and Technology Planning Project of Guangdong Province of China under grant 180917144960530, in part by the Project of Educational Commission of Guangdong Province of China under grant 2017KZDXM032, in part by Science Research Startup Foundation of Shantou University NTF20021, in part by the State Key Lab of Digital Manufacturing Equipment and Technology under grant DMETKF2019020, in part by the Project of Robot Automatic Design Platform combining Multi-Objective Evolutionary Computation and Deep Neural Network under grant 2019A050519008, in part by the Guangdong Natural Science Foundation under grant 2022A1515011396.

Declarations

Competing interests The authors declare no competing interests.

References

- Alom MZ, Yakopcic C, Hasan M, Taha TM, Asari VK (2019) Recurrent residual U-Net for medical image segmentation. *J Med Imaging* 6(1):014,006
- Azzopardi G, Strisciuglio N, Vento M, Petkov N (2015) Trainable cosfire filters for vessel delineation with application to retinal images. *Med Image Anal* 19(1):46–57
- Fan Z, Lu J, Wei C, Huang H, Cai X, Chen X (2018) A hierarchical image matting model for blood vessel segmentation in fundus images. *IEEE Trans Image Process* 28(5):2367–2377
- Fazlali HR, Karimi N, Soroushmehr SR, Shirani S, Nallamothu BK, Ward KR, Samavi S, Najarian K (2018) Vessel segmentation and catheter detection in x-ray angiograms using superpixels. *Med Biol Eng Comput* 56(9):1515–1530
- Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, Barman SA (2012) Blood vessel segmentation methodologies in retinal images—a survey. *Comput Methods Programs Biomed* 108(1):407–433
- Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, Barman SA (2012) An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans Biomed Eng* 59(9):2538–2548
- Guo C, Szemenyei M, Yi Y, Wang W, Chen B, Fan C (2020) SA-UNet: Spatial attention U-Net for retinal vessel segmentation. [arXiv:2004.03696](https://arxiv.org/abs/2004.03696)
- Guo X, Xiao R, Zhang T, Chen C, Wang J, Wang Z (2020) A novel method to model hepatic vascular network using vessel segmentation, thinning, and completion. *Med Biol Eng Comput* 1–16
- Hamamoto Y, Uchimura S, Watanabe M, Yasuda T, Mitani Y, Tomita S (1998) A Gabor filter-based method for recognizing handwritten numerals. *Pattern Recogn* 31(4):395–400
- Lam BS, Gao Y, Liew AWC (2010) General retinal vessel segmentation using regularization-based multiconcavity modeling. *IEEE Trans Med Imaging* 29(7):1369–1381
- Lecun Y, Bengio Y, Hinton GE (2015) Deep learning. *Nature* 521(7553):436–444
- Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
- Li Q, Feng B, Xie L, Liang P, Zhang H, Wang T (2015) A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Trans Med Imaging* 35(1):109–118
- Li X, Wang L, Sung E (2008) AdaBoost with SVM-based component classifiers. *Eng Appl Artif Intel* 21(5):785–795
- Liskowski P, Krawiec K (2016) Segmenting retinal blood vessels with deep neural networks. *IEEE Trans Med Imaging* 35(11):2369–2380
- Liu I, Sun Y (1993) Recursive tracking of vascular networks in angiograms based on the detection-deletion scheme. *IEEE Trans Med Imaging* 12(2):334–341
- Liu JJ, Hou Q, Cheng MM, Wang C, Feng J (2020) Improving convolutional networks with self-calibrated convolutions. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 10,096–10,105
- Mendonca AM, Campilho A (2006) Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Trans Med Imaging* 25(9):1200–1213
- Mou L, Zhao Y, Fu H, Liu Y, Cheng J, Zheng Y, Su P, Yang J, Chen L, Frangi AF et al (2021) CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging. *Med Image Anal* 67:101,874
- Nguyen V, Blumenstein M (2011) An application of the 2D gaussian filter for enhancing feature extraction in off-line signature verification. In: *2011 international conference on document analysis and recognition*. IEEE, pp 339–343
- Ricci E, Perfetti R (2007) Retinal blood vessel segmentation using line operators and support vector classification. *IEEE Trans Med Imaging* 26(10):1357–1365
- Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp 234–241
- Roychowdhury S, Koozekanani DD, Parhi KK (2014) Blood vessel segmentation of fundus images by major vessel extraction and subimage classification. *IEEE J Biomed Health Inf* 19(3):1118–1128
- Salem SA, Salem NM, Nandi AK (2007) Segmentation of retinal blood vessels using a novel clustering algorithm (RACAL) with a partial supervision strategy. *Med Biol Eng Comput* 45(3):261–273
- Staal J, Abramoff MD, Niemeijer M, Viergever MA, Van Ginneken B (2004) Ridge-based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging* 23(4):501–509
- Wang B, Qiu S, He H (2019) Dual encoding U-Net for retinal vessel segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Springer, pp 84–92
- Wang B, Wang S, Qiu S, Wei W, Wang H, He H (2020) CSU-Net: a context spatial u-net for accurate blood vessel segmentation in fundus images. *IEEE J Biomed Health Inform* 25(4):1128–1138
- Wang W, Wang W, Hu Z (2019) Segmenting retinal vessels with revised top-bottom-hat transformation and flattening of minimum circumscribed ellipse. *Med Biol Eng Comput* 57(7):1481–1496
- Wei J, Zhu G, Fan Z, Liu J, Rong Y, Mo J, Li W, Chen X (2022) Genetic U-Net: automatically designed deep networks for retinal vessel segmentation using a genetic algorithm. *IEEE Trans Med Imaging* 41(2):292–307
- Woo S, Park J, Lee JY, Kweon IS (2018) CBAM: Convolutional block attention module. In: *Proceedings of the European conference on computer vision (ECCV)*, pp 3–19
- Wu Y, Xia Y, Song Y, Zhang Y, Cai W (2018) Multiscale network followed network model for retinal vessel segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Springer, pp 119–126
- Zhao Y, Rada L, Chen K, Harding SP, Zheng Y (2015) Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images. *IEEE Trans Med Imaging* 34(9):1797–1807

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Yibiao Rong received the Ph.D. degree from Soochow University, Suzhou, China. He is currently a lecturer with Shantou University, Shantou, China. His research interests include medical image processing, pattern recognition and machine learning



Yu Xiong received his M.S. degree from Shantou University, Shantou, China. His research interests include medical image processing and deep learning.



Chong Li is currently pursuing the Ph.D. degree in Shantou University, Shantou, China. His current research interests include medical image analysis, image processing and machine learning.



Ying Chen is currently pursuing the M.S. degree in Shantou University, Shantou, China. Her current research interests include medical image analysis, image processing, and deep learning.



Peiwei Wei is currently pursuing a M.S. degree in Shantou University, Shantou, China. His research interests include medical image processing, machine learning and deep learning.



Chuliang Wei received the Ph.D. degrees from the University of Liverpool, UK. He is currently a Professor with Shantou University, Shantou, China. His research interests include intelligent control, smart medical systems, image processing.



Zhun Fan received the Ph.D. degree from Michigan State University, USA. He is currently a Professor with Shantou University, Shantou, China. His research interests include intelligent control, machine learning and image processing.