Contents lists available at ScienceDirect

Digital Chinese Medicine

journal homepage: http://www.keaipublishing.com/dcmed

MF²ResU-Net: a multi-feature fusion deep learning architecture for retinal blood vessel segmentation

CUI Zhenchao[a, b*], SONG Shujie[a, b], QI Jing[a, b]

a. School of Cyber Security and Computer, Hebei University, Baoding, Hebei 071002, China

b. Machine Vision Engineering Research Center, Hebei University, Baoding, Hebei 071002, China

ARTICLE INFO

Article history
Received 28 October 2022
Accepted 07 December 2022
Available online 25 December 2022

Keywords
Medical image processing
Atrous space pyramid pooling (ASPP)
Residual neural network
Multi-level model
Retinal vessels segmentation

ABSTRACT

Objective For computer-aided Chinese medical diagnosis and aiming at the problem of insufficient segmentation, a novel multi-level method based on the multi-scale fusion residual neural network (MF2ResU-Net) model is proposed.

Methods To obtain refined features of retinal blood vessels, three cascade connected U-Net networks are employed. To deal with the problem of difference between the parts of encoder and decoder, in MF2ResU-Net, shortcut connections are used to combine the encoder and decoder layers in the blocks. To refine the feature of segmentation, atrous spatial pyramid pooling (ASPP) is embedded to achieve multi-scale features for the final segmentation networks.

Results The MF2ResU-Net was superior to the existing methods on the criteria of sensitivity (Sen), specificity (Spe), accuracy (ACC), and area under curve (AUC), the values of which are 0.8013 and 0.8102, 0.9842 and 0.9809, 0.9700 and 0.9776, and 0.9797 and 0.9837, respectively for DRIVE and CHASE DB1. The results of experiments demonstrated the effectiveness and robustness of the model in the segmentation of complex curvature and small blood vessels.

Conclusion Based on residual connections and multi-feature fusion, the proposed method can obtain accurate segmentation of retinal blood vessels by refining the segmentation features, which can provide another diagnosis method for computer-aided Chinese medical diagnosis.

1 Introduction

Automatic segmentation of retinal blood vessels is an essential part of computer-aided diagnosis for diabetic retinopathy and hypertensive retinopathy [1]. Retinal vessel examination is a non-invasive diagnosis method, as a part of the Chinese medicine theory, which is a primary and commonly used examination method for diabetic complications and retinal diseases. With the development of computer vision, retinal vessel examination has become an important diagnosis method for many diseases. Recent research has found that the retinal vessel density of patients with prolactin inhibitory hormone (PIH) is higher than that of healthy pregnant women [2]. The retinal vessels of patients with Alzheimer's disease (AD) in the early stage show an obviously pathological change [3].

At present, the main method for detection of retinal abnormalities is manual examination in practice.

*Corresponding author: CUI Zhenchao, Doctor, E-mail: cuizhenchao@gmail.com.

Peer review under the responsibility of Hunan University of Chinese Medicine.

DOI: 10.1016/j.dcmed.2022.12.008

Citation: CUI ZC, SONG SJ, QI J. MF²ResU-Net: a multi-feature fusion deep learning architecture for retinal blood vessel segmentation. Digital Chinese Medicine, 2022, 5(4): 406-418.

However, the manual examination method has several disadvantages, such as instability in batch processes and subjectivity. Thus, computer-aided systems have became important for retinal disease diagnosis. In this diagnosis system, automatic segmentation of retinal blood vessels is an essential part, and the segmentation affects the accuracy of examination. Thus, the automatic segmentation method of retinal blood vessels plays a cornerstone role in computer-aided diagnosis [4, 5].

The existing methods for retinal vessel segmentation can be classified into machine learning- and deep learning-based methods. The fixed feature detection method in machine learning has been used for retinal blood vessel segmentation. HASHEMZADEH et al. [6] proposed the UNEt TRansformer (UNETR) model for extracting the retinal blood vessels employing a set of effective image features and a combination of supervised and unsupervised machine learning techniques. To solve the problem of high intra-class variance of image features calculated from various vessel pixels, LAI et al. [7] proposed the automatic retinal image analysis (ARIA) methodology, an applied machine-learning technology, to optimize retinal information. BAHADAR et al. [8] proposed a method for extracting retinal blood vessels based on feature classification. The blood vessels were extracted from the color fundus image by applying preprocessing methods and segmentation techniques using matched filters and a modified local entropy thresholding operation. To reduce the time required by ophthalmologists for examining the retinal images, WANG et al. [9] proposed a supervised method for segmenting blood vessels in retinal images based on the extreme learning machine (ELM) classifier. In these machine learning methods, the features used for classification can have a significant impact on the prediction result. To achieve higher efficiency, researchers have proposed high requirements for an automatic and effective feature extractor.

Since their development in computer vision, the deep learning-based methods have become the mainstream for computer vision tasks, including retinal blood vessel segmentation. LISKOWSKI et al. [10] first used a convolutional neural network (CNN) to segment retinal blood vessels. OLIVERIRA et al. [11] proposed a combination of a steady-state wavelet transform and a multi-scale fully convolutional neural network (FCNN) for blood vessel segmentation. WU et al. [12] proposed a novel scale and context sensitive network (SCS-Net) for retinal vessel segmentation, which dynamically adjusts the receptive fields and uses a self-adaptive feature fusion module. To obtain global context, NI et al. [13] proposed the global context attention (GCA) module and the squeeze excitation pyramid pool (SEPP) module for segmentation of retinal blood vessels.

The typical U-shaped network, U-Net, an encoder-decoder model with skip connections, was designed for medical image segmentation [14]. Many medical image segmentation methods were proposed based on the FCNN with U-shaped structure [15, 16]. JIN et al. [5] proposed an automatic segmentation model, DU-Net, to obtain denser feature information and reduce the characteristics of small blood vessels. LAIBACHER et al. [17] proposed the M2U-Net, which has a new encoder–decoder architecture inspired by the U-Net. GU et al. [18] introduced a context encoder between the feature encoder and the decoder module. This context encoder uses a dense atrous convolutional block and multiple kernel pooling to extract more high-level features. Some studies have tried to address this problem by using atrous convolutional layers [19], self-attention mechanisms [20], and image pyramids [21]. However, although the deep learning-based methods have achieved excellent performance in the field of medical image segmentation, they still cannot fully meet the strict requirements of medical applications for tiny vessel segmentation. Because of the subtle contrast between the target vessels and the background of the images, the features of small and tiny vessels cannot be detected by the existing methods. Therefore, the existing methods have a low accuracy rate for small vessels with weak pixel values. Image segmentation is still a challenging task in medical image analysis.

To obtain precise segmentation of retinal vessels, a novel model, the multi-module fusion residual neural network (MF²ResU-Net) model, and a novel loss function are proposed in this paper. In the MF²ResU-Net, three cascade connected U-Nets are used as the backbone. Three shortcuts are used for the blocks of U-Nets to deal with the problem of over-fitting. In each block of U-Nets, to reduce the semantic gap, residual paths are used to connect the same layers of the encoder and decoder instead of shortcuts. To refine the features of retinal vessels and obtain multi-scale features, an atrous space pyramid pooling (ASPP) module is employed in our model. For training our model and solve the problem of imbalance between the foreground and background, we propose a novel loss function based on the dice loss and cross functions, which could reduce the effects of sample imbalance during training.

The contributions of our work can be elaborated as follows:

(I) To refine the representation features of retinal small vessels, a novel network, MF²ResU-Net, is proposed, which can detect the features of blurry small vessels.

(II) A novel loss function based on the dice loss function and cross entropy, added to a cost-sensitive matrix, is introduced to achieve a more balanced segmentation between the vessel and non-vessel pixels.

2 Methods

Given the difficulty of feature detection for small blurry vessels, it is challenging to obtain satisfactory segmentation results using conventional methods. In this paper, we propose a new model, MF²ResU-Net, to refine small vessels and obtain the segmentation of retinal vessels.

2.1 Fusing residual path U-Net

In the MF²ResU-Net, we use three cascade-connected U-Nets as the backbone network of the module. U-Net is a classic encoder-decoder structure network. A distinctive contribution of the U-Net architecture is the introduction of shortcut connections between the corresponding layers before the max-pooling and after the deconvolution operations. As shown in Figure 1A, the features coming from the encoder are computed in the earlier layers of the network. Conversely, the decoder features go through convolution, down-sampling, and up-sampling, and are supposed to be of much higher level because they are computed at the very deep layers of the network. Thus, there are semantic differences between the same layers of the encoder and decoder, which could affect the results of segmentation.

To remedy the semantic differences between the same layers of the encoder and decoder, SZEGEDY et al. [22] used two 3 × 3 convolutional layers after each pooling layer and before the transposed convolutional layer. These series of two 3 × 3 convolutional operations actually resemble a 5 × 5 convolutional operation. Therefore, according to the approach of the Inception network, the simplest way to augment the U-Net with a multi-resolution analysis capability is to incorporate 3 × 3 and 7 × 7 convolution operations in parallel to the 5 × 5 convolution operation (Figure 1B). Therefore, replacing the convolutional layers with Inception-like blocks should facilitate the U-Net architecture to reconcile the features learnt from the image at different scales. Another possible option is to use stride convolutions. However, in our experiments, although the performance improved, the introduction of additional convolutional layers in parallel increased the memory requirement exaggeratedly. We factorized the bigger, more demanding 5 × 5 and 7 × 7 convolutional layers using a sequence of smaller and lightweight 3 × 3 convolutional blocks (Figure 1C). Our fusing residual path was a cascade-connected structure of blocks, and each block consisted of convolutional layers with 3 × 3 filters and a 1 × 1 filter. This modification greatly reduced the memory requirement. We gradually increased the filters in the layers to prevent the memory requirement of the earlier layers from exceedingly propagating to the deeper part of the network. We also added a residual connection because of their efficacy in biomedical image segmentation along with a 1 × 1 convolutional layer, which may allow us to capture some additional retina spatial information [23]. We named this structure as Res-path. According to the different number of layers, we set different Res-path lengths.

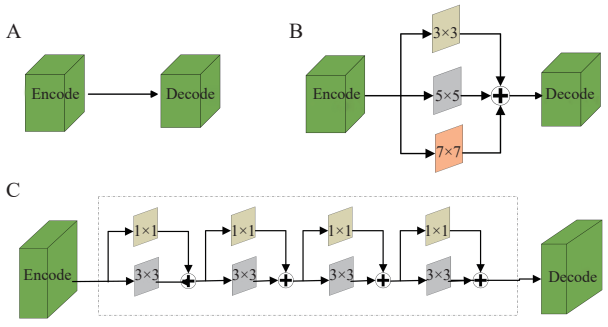

Figure 1 Res-path structure diagram

A, residual connection between decoder and encoder. B, parallel connection. C, fusion connection.

To refine the feature maps of retinal vessels, we used a fusing U-Net, named ResU-Net, for the blocks of our model, which is shown in (Figure 2). To avoid losing features in detection, we used a light U-Net structure with two convolutional layers and two max-pooling layers as down-sampling in the encoder, and two convolutional layers and two deconvolution layers as up-sampling in the decoder. The parameters of the fusing residual path U-Net are presented in Table 1. Moreover, a great challenge in vessel segmentation is the detection of

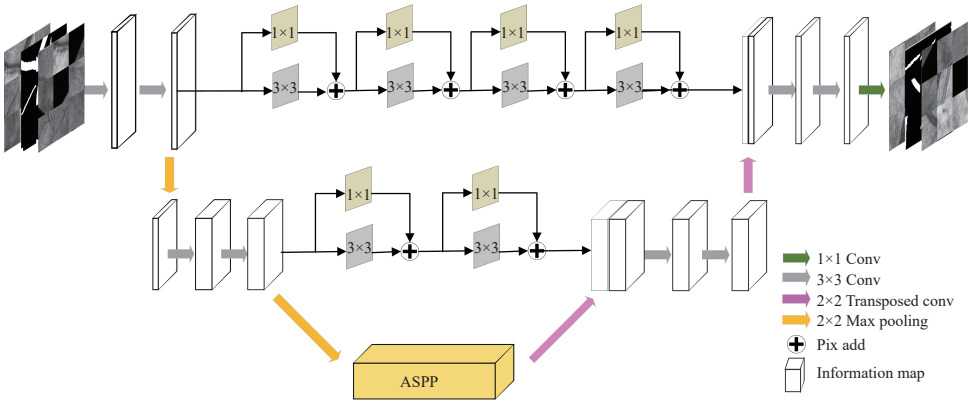

Figure 2 ResU-Net block module

Table 1 Parameters of the ResU-Net block

ResU-Net	Layer (filter size)	Filter
Block1	Conv2D (3 × 3)	32
	MaxPooling (2 × 2)	64
	Conv2D (3 × 3)	64
	Conv2D (3 × 3)	64
	MaxPooling (2 × 2)	128
	ASPP	128
	Conv2D (3 × 3)	64
	Conv2D (3 × 3)	64
	Upsampling (2 × 2)	32
	Conv2D (3 × 3)	32
	Conv2D (3 × 3)	32
	Conv2D (3 × 3)	32
Res-path1	Conv2D (1 × 1)	32
	Conv2D (3 × 3)	32
	Conv2D (1 × 1)	32
	Conv2D (3 × 3)	32
	Conv2D (1 × 1)	32
	Conv2D (3 × 3)	32
	Conv2D (1 × 1)	32
Res-path2	Conv2D (3 × 3)	64
	Conv2D (1 × 1)	64
	Conv2D (3 × 3)	64
	Conv2D (1 × 1)	64
ASPP	Conv2D (3 × 3) Rate: 2	128
	Conv2D (1 × 1)	128
	Conv2D (3 × 3) Rate: 4	128
	Conv2D (1 × 1)	128
	Conv2D (3 × 3) Rate: 8	128
	Conv2D (1 × 1)	128
	Conv2D (3 × 3) Rate: 16	128
	Conv2D (1 × 1)	128

vessels with various shapes and scales. We detected the comprehensive and detailed feature extraction in fundus retinal vascular images by using a parallel model (Figure 3). Between the encoder and decoder, we employed an ASPP module, which can obtain the multi-scale local features of retinal vessels. According to the tubular characteristics of various sizes in vessels, four atrous convolutions with a 3 × 3 convolution size were used for multi-scale feature extraction. The ASPP was inspired by the spatial pyramid pooling method of DeepLab v2 [19], but because that network uses a too large dilation rate, it

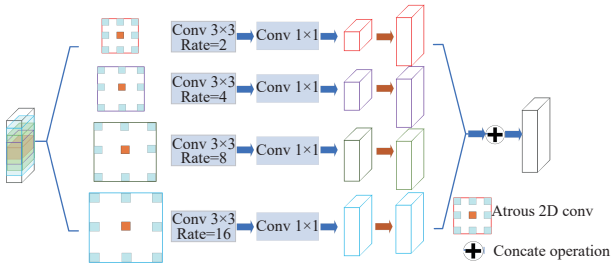

Figure 3 Atrous spatial pyramid pooling

extracts invalid features from blood vessels. An excessively large dilation rate is not suitable for datasets of retinas, and thus the dilation convolution with a void rate of 24 was deleted in the ASPP. The dilation rates in our model were defined as {2, 4, 8, 16}. To accelerate computing, a 1 × 1 convolution was employed after each atrous convolution. In the final feature map, the feature image was up-sampled using the bilinear interpolation method. The resolution of each feature map after atrous convolution was expanded by bilinear interpolation, which made the size of each layer feature map consistent. Finally, the target feature map was formed by four feature maps through the addition of the corresponding pixels. The parameters of the ASPP are listed in Table 1.

2.2 MF²ResU-Net for retinal vessel segmentation

The aim of this study was to build deep learning models to segment retinal vessels in fundus images. Figure 4 shows a diagram of the MF²ResU-Net model for retinal vessel segmentation. We cascaded three ResU-Net blocks for the feature maps to form MF²ResU-Net. To remedy the sensitivity problem of the networks for training data, residual links were employed in our model. To improve the generalization of the proposed method, the original images were preprocessed and randomly cropped into small patches to establish training and validation datasets. The size of the patches for training and evaluation was empirically selected as 48 × 48 (pixel × pixel). After segmentation by the MF²ResU-Net, a complete vessel probability map could be obtained by reordering the patches. The connection result of each residual block was represented by Formula (1).

$$W_m = F(a_m, b_m) + H(a_m) \tag{1}$$

$$a_{m+1} = \sigma(W_m) \tag{2}$$

Where a_m represents the ResU-Net input and b_m represents the ResU-Net output. $F(a_m, b_m)$ stands for the residual function. $H(a_m)$ denotes the map function of features. The value of $F(a_m)$ is equivalent to b_m. In Formula (2), $\sigma(W_m)$ represents the activation function. The residual connection result was used as the input of the next module through the ReLU activation function. Finally, we obtained the output result of each residual connection W_m.

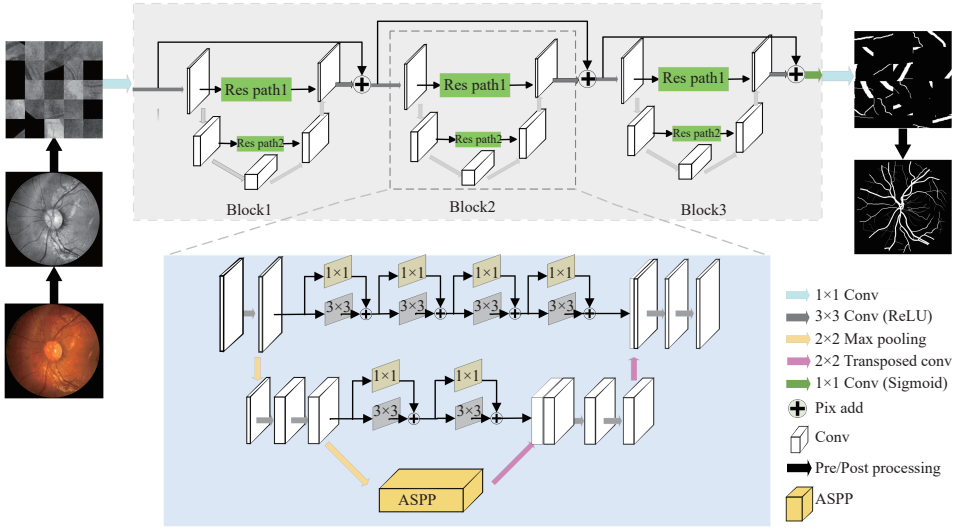

Figure 4 General map of the MF²ResU-Net structure

The segmentation output was obtained by the last block of the fusion ResU-Net and two convolutions, which used a 1 × 1 (pixel × pixel) convolutional kernel and max pooling operation. The ReLU and sigmoid functions were used as activation functions in the convolutions.

2.3 Loss function design

From the statistics of typical retinal blood vessels in the dataset, we found that the ratio of pixels in the foreground to those in the background was almost 1 : 12 (the number of pixels in the foreground was 453 800, whereas that in the background was 6 599 200), which shows that there was an imbalance for the segmentation task. Thus, several existing loss functions are not suitable for our task. To deal with the imbalance problem, we proposed a novel loss function based on an improved binary cross entropy loss function and a dice loss function.

For binary segmentation problems, the binary cross entropy loss function could be defined as in Formula (3).

$$Loss' = -\frac{1}{N}\left(\sum_{k=1}^{N} y_k \log x_k - \sum_{k=1}^{N}(1-y_k)\log(1-x_k)\right) \tag{3}$$

Where N is the number of patch pixels, x_k represents the foreground predicted probability of the input pixel k, and y_k is the true label of pixel k, which is either 1 (foreground) or 0 (background) in this task. For the imbalance problem in our task, we improved the binary cross entropy loss function. As presented in Formula (4), we defined a novel loss function L_{CE}.

$$L_{CE} = -\frac{1}{N}\sum_{k=1}^{N}\begin{bmatrix} y_k & 1-y_k \end{bmatrix}\begin{bmatrix} \log x_k \\ \log(1-x_k) \end{bmatrix}\begin{bmatrix} 1 & \lambda\overline{x_k} \\ 0 & 1 \end{bmatrix} \tag{4}$$

Where $\overline{x_k}$ is an indicator for wrong prediction to x_k, $\overline{x_k} = \begin{cases} 1, & x_k \leqslant 0.5 \\ 0, & x_k > 0.5 \end{cases}$. λ is the penalty parameter for predicting the blood vessel, which is a positive real number. In the loss function of L_{CE}, we could set a large value of λ to enlarge the loss of the wrong prediction in the foreground. Because of the ratio of foreground to background, λ was set to 12 in this study.

To deal with the imbalance problem and obtain a perfect criterion on intersection of union, the dice loss function [24] was proposed in the segmentation task, which could be expressed as in Formula (5).

$$L'_{dice} = 1 - 2\frac{x_k \cap y_k}{x_k + y_k} \tag{5}$$

Where x_k represents the fundus blood vessel region segmented by the algorithm, and y_k denotes the fundus blood vessel region manually segmented by the expert. $|x_k \cap y_k|$ represents the same area of the retinal blood vessel region segmented by the proposed method and an expert. To remedy the numerical problems, we improved the dice loss by introducing a Laplace smooth factor, φ, and the improved dice loss is expressed in Formula (6).

$$L_{dice} = 1 - 2\frac{\sum_{i=1}^{N} x_k y_k + \varphi}{\sum_{i=1}^{N} x_k^2 + \sum_{i=1}^{N} y_k^2 + \varphi} \tag{6}$$

To deal with the vanishing gradient problem of dice loss and combine the advantages of the two loss functions, a synthetic loss function for the training of the MF²ResU-Net model was proposed in this study.

$$Loss = \alpha L_{CE} + (1-\alpha)L_{dice} \tag{7}$$

Where α is a parameter that controls the contribution of the L_{CE} and L_{dice} loss functions.

3 Results and discussion

We tested our model, MF²ResU-Net, by comparing it with several conventional retinal vessel segmentation methods on the public datasets DRIVE and CHASE DB1. The segmentation performance of the proposed method on the fundus image can be assessed by comparing the segmentation testing results with the image label.

3.1 Datasets and computational platform

The public datasets DRIVE and CHASE DB1 were used to test our model. DRIVE contains 40 colored retinal vessel images, which were obtained from a diabetic retinopathy (DR) screening program in the Netherlands [25]. The dataset was randomly divided into two groups, the test and training sets, and each group contained 20 images. The size of each image was 565 × 584. The CHASE DB1 dataset contained 28 images of 14 children, corresponding to two images per patient, from the Child Heart and Health Study in England [25]. The size of the image in CHASE DB1 was 999 × 960. In our experiment, 14 images were used for training, and the other 14 images were used for testing. Each picture in the CHASE DB1 was segmented by a professional physician to obtain a manual result image, but there was no corresponding mask, and we needed to set it manually. The result of manually setting the mask for CHASE DB1 is shown in Figure 5. These two fundus retinal blood vessel datasets are commonly used as test datasets for retinal blood vessel segmentation algorithms.

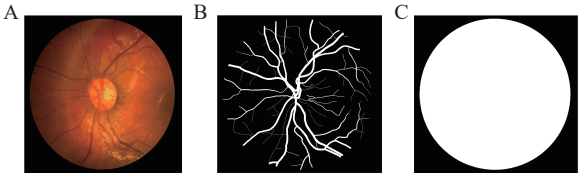

Figure 5 Typical image in CHASE DB1
A, original retinal vessel image. B, ground-truth. C, manual mask.

The platform of the experiment for MF²ResU-Net training was a 64-bit Win 10 system with an Intel Core i5-4160k, 3.60 GHz CPU (Intel) and a 6 GB NVIDA1660Ti GPU . The structure of the network was implemented under the open-source deep learning library TensorFlow.

3.2 Data pre-processing

The images of the two databases were all colored RGB images, and the original retinal blood vessel images had a low contrast. The features of retinal blood vessels were not obvious. To improve the performance of the MF²ResU-Net model, a technique that enhances the image contrast was used to make the retinal blood vessel features more obvious. As the gray scale images showed better contrast than the RGB images [26], we used gray scale images as the input of models. To strengthen the contrast ratio between vessels and background in retinal images, we applied three strategies for image preprocessing, which were normalization, contrast limited adaptive histogram equalization (CLAHE) [27], and gamma correction. Figure 6 shows preprocessed images of one typical retinal image using these three strategies. The preprocessed retinal image had a higher contrast between the blood vessel outline and the background and reduced noise.

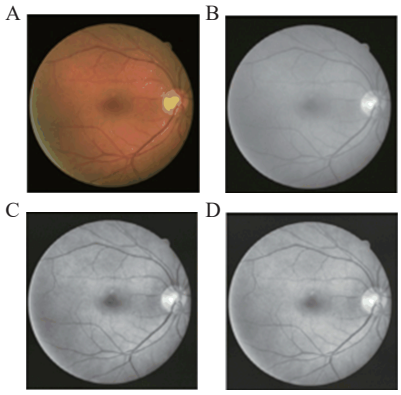

Figure 6 Retina image preprocessing results
A, original image. B, normalized image. C, image after CLAHE operation. D, image after Gamma correction.

Data augmentation is widely applied in convolutional neural networks because of its high efficiency and operability. Considering that DRIVE and CHASE DB1 are small datasets, the model will be prone to overfitting and poor classification performance. Therefore, it was necessary to augment the dataset for achieving better results. Four image processing steps were used for augmenting the dataset: rotating, mirroring, shifting, and cropping. To reduce the overfitting problem, our models were trained on small patches, which were randomly extracted from the images. Further, to reduce the calculation complexity and ensure the surrounding local features, small size blocks of 48 × 48 were randomly extracted from the preprocessed images and used to train our model. In this study, 190 000 and 200 000 blocks were randomly extracted from the DRIVE and CHASE DB1 datasets, respectively; 90% of blocks were used for training, and the remaining blocks were used for testing. Several randomly sampled patches and the corresponding labels for the experiments are presented in Figure 7. The corresponding labels of the patches were decided based on the ground truth images.

3.3 Performance evaluation criteria

The classification result of every pixel could be one of four types with the following entries. True positive (TP) indicated a vessel pixel classified correctly as a vessel pixel. False positive (FP) indicated a non-vessel pixel classified wrongly as a vessel pixel. True negative (TN)

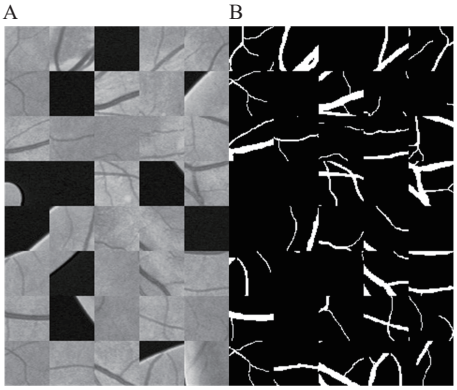

Figure 7 Local block information map

A, patches from the preprocessed image. B, patches from the corresponding ground truth. Typical 48 × 48 patches selected for model training.

indicated a non-vessel pixel classified correctly as a non-vessel pixel. False negative (FN) indicated a vessel pixel classified wrongly as a non-vessel pixel.

To evaluate the segmentation of the proposed algorithm quantitatively, four evaluation indicators were used, which were accuracy (ACC), sensitivity (Sen), specificity (Spe), and F1-Score [27]. In this model, positive referred to blood vessels and negative referred to background. The ACC, Sen, Spe, and F1-Score are defined as follows.

$$\mathrm{ACC} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{8}$$

$$\mathrm{Sen} = \frac{T_P}{T_P + T_N} \tag{9}$$

$$\mathrm{Spe} = \frac{T_N}{T_N + F_P} \tag{10}$$

$$R_{call} = \frac{T_P}{T_P + F_N} \tag{11}$$

$$\mathrm{Pre} = \frac{T_P}{T_P + F_P} \tag{12}$$

$$\text{F1-Score} = \frac{2 \times \mathrm{Pre} \times R_{call}}{\mathrm{Pre} + R_{call}} \tag{13}$$

Where ACC is the ratio of the number of correctly detected blood vessel and background pixels to the total number of image pixels; Sen is the ratio of the number of correctly detected retinal blood vessel pixels to the total number of blood vessel pixels; Spe is the ratio of the number of correctly detected non-vessel pixels to the total number of non-vessel pixels; and F1-Score is a measure of the similarity between the model and the expert segmentation results.

The AUC value represents the area under the ROC curve. The ROC curve is an important method for measuring the comprehensive performance of image semantic segmentation. Its value ranges from 0 to 1. The condition AUC = 1 indicates a perfect classifier; 0.5 < AUC < 1 means that better than random classifiers; and 0 < AUC < 0.5 means that worse than random classifiers.

3.4 Experimental results

We tested the proposed method by comparing it with several existing retinal vessel segmentation methods. The experiment was dived into two parts: ablation experiments and comparative experiments.

3.4.1 Ablation and basic network comparison experiments Figure 8 shows the main processing of ablation experiments. After image preprocessing, one retinal image could be segmented in many image blocks with size of 48 × 48. By using the vessel detection model, we obtained the segmentation result composed of the segmented image. As the network deepened, the segmentation results were better. From the detail map, in the third row and the fourth column, it can be observed that the three-fusion network had the best detail segmentation and complex curvature segmentation functions.

Table 2 and 3 present the results of ablation experiments. In the ablation experiments, we compared the proposed model with U-Net, single residual U-Net (Block1), double residual U-Net (Block2), and four residual U-Net (Block4) on the DRIVE and CHASE DB1 datasets. As presented in Table 2, on the DRIVE dataset, the Sen, F1, AUC, and ACC of MF^{2}ResU-Net were 0.801 3, 0.822 3, 0.979 7, and 0.970 0, respectively, which were the highest values among the compared methods. This means that the results of MF^{2}ResU-Net on DRIVE were better than those of other methods. As listed in Table 3, the Sen, F1, AUC, and ACC of MF^{2}ResU-Net were 0.810 2, 0.814 2, 0.983 7, and 0.977 6, respectively, on the CHASE DB1 dataset. These were the highest values among all results, which means that the generalization ability of MF^{2}ResU-Net was better than that of other block methods. The analysis of the experimental results of Block4, as presented in the discounted indicators in Figure 8, Table 3 and 4, indicates that Block4 achieved higher segmentation indicators than those of Block1 and Block2, but they were slightly lower than or the same as those of MF^{2}ResU-Net, and the training time of Block4 was longer. Therefore, the three-module residual network, MF^{2}ResU-Net, was superior to the other four networks in terms of segmentation statistics.

We compared our proposed model with two state-of-the-art networks. One was the customized implementation of U-Net, which we introduced above. The other was DeepLab v2. DeepLab v2 is an advanced segmentation network, which combines deep convolutional nets, atrous convolution, and fully connected conditional random fields (CRFs). In MF^{2}ResU-Net, we combined atrous

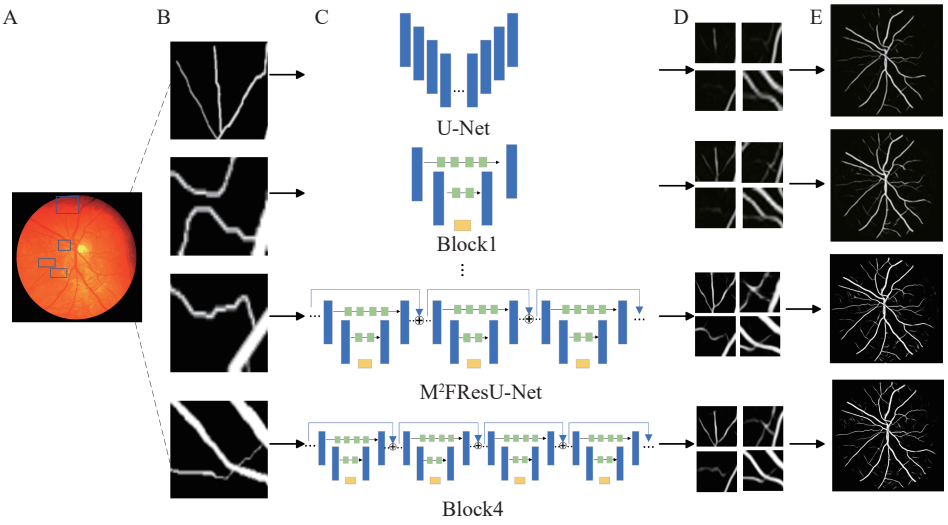

Figure 8 Comparison diagram of detail segmentation of three network structures

A, original image. B, original image samples. C, snapshots of the proposed MF²ResU-Net, Block1, Block2, and U-Net. D, samples of network segmentation results. E, re-composition of segmentation results.

Table 2 Statistical results of different modules on the DRIVE dataset

Model	DRIVE				
	Sen	F1-Score	Spe	AUC	ACC
Block1	0.7491	0.8066	0.9842	0.9768	0.9542
Block2	0.7559	0.8093	0.9843	0.9785	0.9591
MF²ResU-Net	0.8013	0.8223	0.9842	0.9797	0.9700
Block4	0.7741	0.8159	0.9820	0.9782	0.9647

Table 3 Statistical results of different modules on the CHASE DB1 dataset

Model	CHASE DB1				
	Sen	F1-Score	Spe	AUC	ACC
Block1	0.7538	0.7848	0.9814	0.9747	0.9585
Block2	0.7736	0.7958	0.9810	0.9796	0.9602
MF²ResU-Net	0.8102	0.8142	0.9809	0.9837	0.9776
Block4	0.7942	0.8119	0.9820	0.9824	0.9722

convolution and U-Net. In order to highlight the excellence of our model, we compared the three models, DeepLab v2, U-Net, and MF²ResU-Net based on the DRIVE and CHASE DB1 datasets. We evaluated the model using the test data. Sen, Spe, ACC, F1, and AUC were compared and the results are presented in Table 4 and 5. It can be observed from the tables that the MF²ResU-Net achieves the highest values for most of the metrics. The global accuracy for DeepLab v2, U-Net, and MF²ResU-Net were 0.9636, 0.9527, and 0.9700 on DRIVE, and 0.9624, 0.9549, and 0.9776 on CHASE, respectively. Moreover, we evaluated the models using ROC curves (Figure 9). The closer the ROC curve to the top-left border in the ROC

Table 4 Performance of the three models tested on DRIVE

Model	DRIVE				
	Sen	F1-Score	Spe	AUC	ACC
U-Net	0.7376	0.7987	0.9840	0.9640	0.9527
DeepLab v2	0.7631	0.8083	0.9837	0.9749	0.9636
MF²ResU-net	0.8013	0.8223	0.9842	0.9797	0.9700

Table 5 Performance of the three models tested on CHASE DB1

Model	CHASE DB1				
	Sen	F1-Score	Spe	AUC	ACC
U-Net	0.7020	0.7574	0.9831	0.9709	0.9549
DeepLab v2	0.7619	0.8001	0.9848	0.9801	0.9624
MF²ResU-net	0.8102	0.8142	0.9809	0.9837	0.9776

coordinates, the more accurate is a model. These results showed that the curves of the MF²ResU-Net were the most top-left among the three models, whereas the U-Net curves were the lowest of the three. The MF²ResU-Net also had the largest area under the ROC curve (AUC).

3.4.2 Comparison against existing methods To test our model for retinal vessel segmentation, we compared it with several state of art models. In Figure 10, we present the results and details of several typical retinal vessel images using several methods. The first column in Figure 10 shows the original images from DRIVE and CHASE DB1. The second column is the ground-truth of the segmentation. The third, fourth, and fifth columns, respectively, present the results of U-Net, DU-Net [5], and MF²ResU-Net. The second, fourth, and sixth lines, respectively,

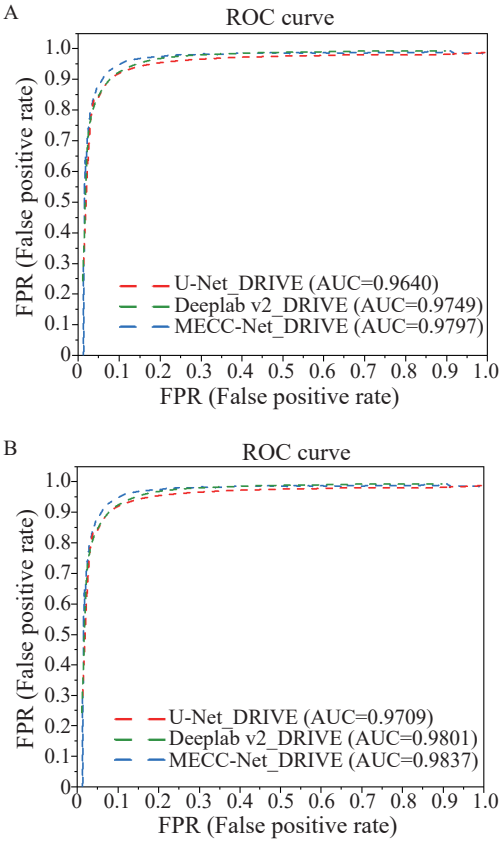

Figure 9 ROC curves of different models

A, U-Net, DeepLab v2, and MF²ResU-Net on DRIVE. B, U-Net, DeepLab v2, and MF²ResU-Net on CHASE DE1.

present enlarged versions of the red and green blocks in the corresponding original images obtained using U-Net, DU-Net, and MF²ResU-Net.

As can be observed from the entire images in Figure 10, the results of MF²ResU-Net were closer to the ground-truth than those of other methods, which means that our model is superior. From the enlarged version of segment blocks, MF²ResU-Net was able to detect small vessels that were disturbed by the complex background, but the small vessels in the results of the compared methods were blurred or disappeared. MF²ResU-Net also produced a more obvious expression of vascular features, and could segment subtle blood vessels that were not obvious. Figure 11 shows the segmentation details of the two datasets. In Figure 11, the first line shows local details and the third line shows an overlap area of main blood vessels adjacent to each other. The network must accurately segment this blood vessel, but in LadderNet [28], because the blood vessel was unclear and there was a lot of noise interference, the blood vessel was broken when it was segmented. Obviously, our method had a better segmentation effect in this case, as shown in the second and fourth lines of Figure 11. LadderNet only has the U-Net's common convolutional layer feature extraction function. With the increase in depth of the network, the constant pooling operation led to loss of a large number of local details.

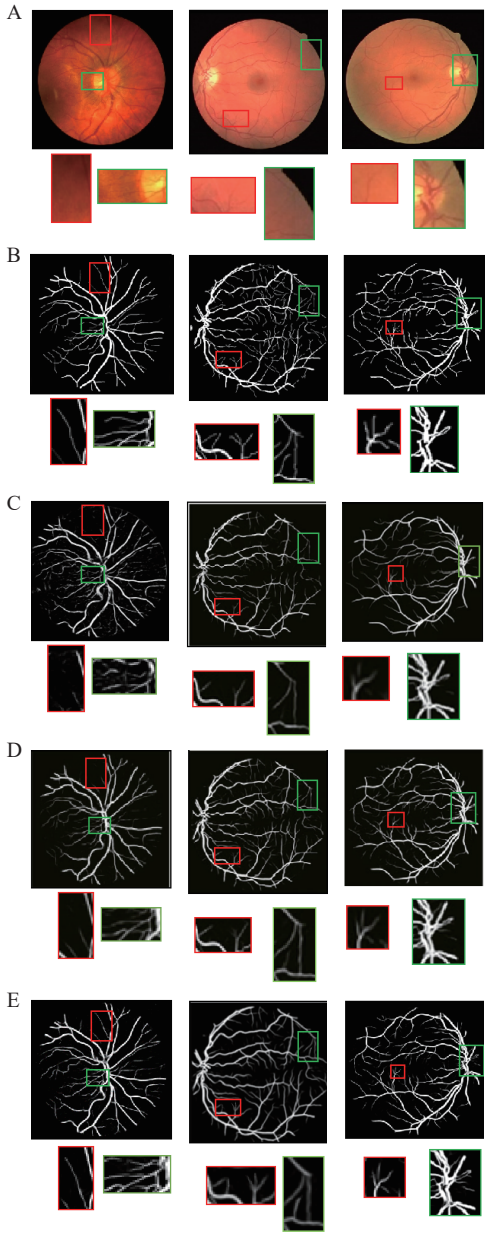

Figure 10 Results and enlarged versions of segmentation of different algorithms on the DRIVE and CHASE DB1 datasets

A, original image. B, ground-truth. C, segmentation results of U-Net. D, DU-Net. E, MF²ResU-Net.

In contrast, in our method, the Res-paths were integrated into the neural network to better capture small-shaped blood vessels in the retina. The multi-module fusion operation not only deepened the depth of the convolutional network, but also made up for the loss of blood vessels in the deep learning network with the help of residual connection. MF²ResU-Net extracted more details of blood vessels in some connection areas than LadderNet and achieved more ideal segmentation results in the segmentation of small blood vessels. Therefore, the multi-module fusion extraction convolutional neural network based on the initial module could make the features of small retinal vessels more discriminative.

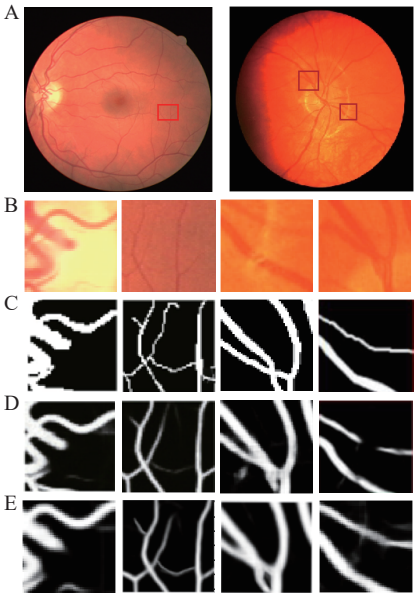

Figure 11 Comparison of detail segmentation

A, original image. B, ground-truth. C, U-Net segmentation. D, LadderNet. E, MF²ResU-Net.

We also compared our method with several recently proposed state-of-the-art approaches. In Table 6 and 7, we summarize the type of algorithm, year of publication, and performance on the DRIVE and CHASE DB1 datasets. As can be observed, our model achieved the highest values on Sen, Spe, ACC, and AUC on the DRIVE dataset, which were 0.8013, 0.9842, 0.9700, and 0.9797, respectively. Table 5 shows the statistical results of the methods on the CHASE dataset. Our proposed method achieved the highest values on Sen, F1, ACC, and AUC, which were 0.8102, 0.8142, 0.9776, and 0.9837, respectively. Thus, the segmentation results obtained by MF²ResU-Net were more accurate than those of the compared methods. In general, the proposed model was superior compared to the other retinal vascular segmentation algorithms. Concretely, LI et al. [29] extracted an image patch with a size of 16 × 16, which might fall in the flat area, and thus it was not a good choice for a high-resolution dataset. ORLANDO et al. [30] used a fully-connected CRF model for blood vessel segmentation, but labeling all

Table 6 Performance indicators of different algorithms in the DRIVE dataset

Model	DRIVE					
	Year	Sen	Spe	F1-Score	ACC	AUC
LI et al. [29]	2016	0.7569	0.9816	NA	0.9527	0.9738
ORLANDO et al. [30]	2017	0.7897	0.9684	0.7857	0.9454	0.9506
R2U-Net [31]	2018	0.7799	0.9810	0.8171	0.9556	0.9784
MS-NFN [32]	2018	0.7844	0.9819	NA	0.9567	0.9807
DEU-Net [33]	2019	0.7940	0.9816	0.8270	0.9567	0.9772
MOU et al. [34]	2020	0.8010	0.9800	NA	0.9667	0.9700
MSFFU-Net [35]	2020	0.7762	0.9835	NA	0.9694	0.9790
MF²ResU-Net	2023	0.8013	0.9842	0.8223	0.9700	0.9797

NA represents not available by their authors.

Table 7 Performance indicators of different algorithms on the CHASE DB1 dataset

Model	CHASE DB1					
	Year	Sen	Spe	F1-Score	ACC	AUC
LI et al. [29]	2016	0.7507	0.9793	NA	0.9581	0.9716
ORLANDO et al. [30]	2017	0.7277	0.9712	0.7332	0.9458	0.9524
R2U-Net [31]	2018	0.7756	0.9820	0.7928	0.9634	0.9815
MS-NFN [32]	2018	0.7538	0.9847	NA	0.9637	0.9823
DEU-Net [33]	2019	0.8074	0.9821	0.8037	0.9661	0.9812
TAMIM et al. [36]	2020	0.7585	0.9846	0.7580	0.9620	NA
Sine-Net [37]	2020	0.7856	0.9845	NA	0.9694	0.9824
MF²ResU-Net	2023	0.8102	0.9809	0.8142	0.9776	0.9837

NA represents not available by their authors.

blood vessels was a challenge. In the methods proposed in recent years [31-37], the performance was satisfactory but the training procedure was complicated. Although these methods reached new state-of-the-art performance in some metrics, they are not practical for real applications. The modified MF²ResU-Net model acts as a soft and rapid automation system, which provides the ophthalmologist with primary information and knowledge about the blood vessels, such as size, tortuosity, crossing, lesions structure, and hard and soft exudates. This model cannot replace the role of ophthalmology doctors, but it can help the doctor in early diagnosis, increasing the result accuracy. We plan to develop our MF²ResU-Net to other types of vessel segmentation, aiming to obtain more accurate results in medical image analysis tasks. This would be a fruitful area for further work.

4 Conclusion

In this paper, to help computer-aided diagnosis on eye diseases and related conditions and Chinese and Western diagnosis learning, we present a novel residual neural network based on U-Net for retinal vessel segmentation. The experimental results proved that the method succeeded both in absolute terms and in comparison with nine other state-of-the-art similar methods using two well-known publicly available datasets. The proposed method encompasses many elements that contribute to its success. To refine the segmentation features of retinal vessels, we used residual paths to connect the encoder and decoder of the U-Net, and an ASPP was used between the encoder and decoder to obtain global features. To improve the segmentation, a multi-model was employed based on ResU-Net blocks. To test the performance of our model, ablation and comparison experiments were conducted based on Sen, Spe, F1, ACC, and AUC criteria on the DRIVE and CHASE datasets. The experimental results demonstrated that the proposed algorithm was able to obtain the best values on most of the criteria, and the segmentation results showed that our model was more robust than the compared methods on a complex background. This means that MF²ResU-Net was superior to the compared methods in retinal vessel segmentation on the DRIVE and CHASE DB1 datasets. Therefore, our model provides the supplement and development of TCM theory in the treatment of retina-related diseases, and then provides a reference for the participation scheme of TCM in the treatment of retina-related diseases.

Fundings

Key R&D Projects in Hebei Province (22370301D), Scientific Research Foundation of Hebei University for Distinguished Young Scholars (521100221081), and Scientific Research Foundation of Colleges and Universities in Hebei Province (QN2022107).

Competing interests

The authors declare no conflict of interest.

References

[1] IRSHAD S, AKRAM M. Classification of retinal vessels into arteries and veins for detection of hypertensive retinopathy. 2014 Biomedical Engineering Conference (CIBEC), 2014: 133–136.

[2] SUN P, ZHAO X, TANG M, et al. Changes of retinal microcirculation in postnatal patients with pregmancy-induced hypertension based on OCTA. International Journal of Engineering Science, 2022, 22(11): 1861–1866.

[3] ZHANG Q, WANG W, DU Y. Retinal optical coherence tomography and optical coherence tomography angiography in Alzheimer's disease. Chinese Journal of Behavioral Medicine and Brain Science, 2022, 31(10): 955–960.

[4] CHENG CY, ZHENG Y. Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors. Ophthalmology, 2011, 118(5): 812–818.

[5] JIN Q, MENG Z, PHAM TD, et al. DUNet: a deformable network for retinal vessel segmentation. Knowledge-Based Systems, 2019, 178: 149–162.

[6] HATAMIZADEH A, YANG D, ROTH H, et al. UNETR: transformers for 3D medical image segmentation. Proceedings of IEEE/CVF Winter Conference on Application of Computer Vision, 2022: 574–584.

[7] LAI M, LEE J, CHIU S, et al. A machine learning approach for retinal images analysis as an objective screening method for children with autism spectrum disorder. eClinical Medicine, 2020, 28: 100588.

[8] BAHADAR KK, AKHALIQ A, SHAHID M. Correction: a morphological hessian based approach for retinal blood vessels segmentation and denoising using region based Otsu thresholding. PLoS One, 2016, 11(9): e0162581.

[9] WANG SN, ZHU CZ, ZOU BJ, et al. Retinal blood vessel segmentation using extreme learning machine. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2017, 21(7): 1280–1290.

[10] LISKOWSKI P, KRAWIEC K. Segmenting retinal blood vessels with deep neural networks. IEEE Transactions on Medical Imaging, 2016, 35(11): 2369–2380.

[11] OLIVEIRA A, PEREIRA S, SILVA CA. Retinal vessel segmentation based on fully convolutional neural networks. Expert Systems with Applications, 2018, 112(12): 229–242.

[12] WU H, WANG W, ZHONG J. SCS-Net: a scale and context sensitive network for retinal vessel segmentation. Medical Image Analysis, 2021, 70(10): 102025.

[13] NI J, WU J, TONG J, et al. GC-Net: global context network for medical image segmentation, Computer Methods and Programs in Biomedicine, 2020, 190: 105121.

[14] RONNEBERGER O, FISCHERB P, BROX T. U-Net: convolutional networks for biomedical image segmentation. International Conference on Medical Computing and

Computer-assisted Intervention, 2015: 234–241.
[15] ISENSEE F, JAEGER PF, KOHL SAA, et al. nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 2021, 18(2): 203-211.
[16] JIN Q, MENG Z, SUN C, et al. RA-UNet: a hybrid deep attention-aware network to extract liver and tumor in CT scans. Frontiers in Bioengineering and Biotechnology, 2020, 1471(8): 1–15.
[17] LAIBACHER T, WEYDE T, JALALI S. M2U-Net: effective and efficient retinal vessel segmentation for resource-constrained environments. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, 2019: 1–10.
[18] GU ZW, GU J, CHENG J, et al. CE-Net: context encoder network for 2D medical image segmentation. IEEE Transactions on Medical Imaging, 2019, 38(10): 2281–2292.
[19] CHEN L, PAPANDREOU G, KOKKINOS I, et al. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834–848.
[20] SCHLEMPER J, OKTAY O, SCHAAP M, et al. Attention gated networks: learning to leverage salient regions in medical images. Medical Image Analysis, 2019, 53: 197–207.
[21] ZHAO H, SHI J, QI X, et al, Pyramid scene parsing network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6230–6239.
[22] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning. AAAI Conference on Artificial Intelligence, 2017, (4): 1–12.
[23] DROZDZAL M, VORONTSOV E, CHARTRAND G, et al. The importance of skip connections in biomedical image segmentation. Deep Learning and Data Labeling for Medical Applications, 2016: 179–187.
[24] LI D, RAHARDJA S. BSEResU-Net: an Attention-based before-activation residual U-Net for retinal vessel segmentation. Computer Methods and Programs in Biomedicine, 2021, 205: 106070.
[25] STAL J, ABRAMOFF MD. Ridge-based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging, 2004, 23(4): 501–509.
[26] ZHAO Y, LIU Y, WU X, et al. Correction: retinal vessel segmentation: an efficient graph cut approach with retinex and local phase. PLoS One, 2015, 10(4): e0127486.
[27] LI Y, CAI Y, GAO H. Retinal vessel segmentation algorithm based on hybrid phase feature. Journal of Computer Applications, 2018, 38(7): 2083–2088.
[28] ZHUANG JT. LadderNet: multi-path networks based on U-Net for medical image segmentation. ArXiv Preprint, 2018. doi: 1810.0781.
[29] LI Q, FENG B, XIE L. A cross-modality learning approach for vessel segmentation in retinal images. IEEE Transactions on Medical Imaging, 2016, 35(1): 109–118.
[30] ORLANDO J, PROKOFYEVA E, BLASCHKO M. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. IEEE Transactions on Biomedical Engineering, 2017, 64(1): 16–27.
[31] ALOM M, HASAN M, YAKOPCIC C, et al. Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation. ArXiv Preprint. 2018. doi: 1802.06955.
[32] WU YC, XIA Y, SONG Y, et al. Multiscale network followed network model for retinal vessel segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2018, (11071): 119–126.
[33] WANG B, QIU S, HE H. Dual encoding U-Net for retinal vessel segmentation. International Conference on Medical Image Computing and Computer Assisted Intervention, 2019: 84–92.
[34] MOU L, CHEN J, CHENG Z, et al. Dense dilated network with probability regularized walk for vessel detection. IEEE Transactions on Medical Imaging, 2019, 39(5): 1392–1403.
[35] YANG D, LIU G, REN M. A multi-scale feature fusion method based on U-Net for retinal vessel segmentation. Entropy, 2020, 22(8): 1–21.
[36] TAMIN N, ELSHRKAWEY M, AZIM GA, et al. Retinal blood vessel segmentation using hybrid features and multi-layer perceptron neural networks. Symmetry, 2020, 12(6): 894.
[37] İBRAHIM A, OSMAN S. Sine-Net: a fully convolutional deep learning architecture for retinal blood vessel segmentation. Engineering Science and Technology, 2021, 24(2): 271–283.

MF²ResU-Net: 一种面向视网膜血管分割的多特征融合深度网络构架

崔振超[a, b*], 宋姝洁[a, b], 齐静[a, b]

a. 河北大学网络空间安全与计算机学院, 河北 保定 071002, 中国
b. 河北大学机器视觉工程研究中心, 河北 保定 071002, 中国

【摘要】目的 为解决计算机辅助中医诊断中图像分割精确问题,提出一种基于多尺度的融合冗余连接的多层结构模型的新型神经网络模型(MF2ResU-Net)。**方法** 为获得精细化的血管特征,提出将三个U-Net模型进行级联。为解决U-Net模型中编解码部分语义连接差异性问题,MF2ResU-Net提出利用shortcut连接编解码部分。为了更加精细化分割特征,算法将空洞空间金字塔池化模型嵌入编解码结构中,最终形成所提网络。**结果** MF2ResU-Net算法在视网膜血管数据集CHASE DB1和DRIVE中,利用指标敏感度(Sen)、特异性(Spe)、精确性(ACC)和线下面积(AUC)分别取得了0.8013和0.8102、0.9842和0.9809、0.9700和0.9776, 以及0.9797和0.9837。这些结果表明本方法优于对比方法,并且对于视网膜血管分割,本方法具有较高的有效性及鲁棒性。**结论** 本文所提的基于冗余连接的多特征融合深度学习网络模型可通过获得更加精细化特征获取的方式获得更加准确的视网膜血管分割结果,可为计算机辅助中医诊断提供一种诊断方法提取途径。

【关键词】医学图像处理;空洞空间金字塔池化;冗余连接;多级模型;视网膜血管分割