MF2ResU-Net: A Multi-Feature Fusion Deep Learning Architecture for Retinal Blood Vessel Segmentation

ZhenChao Cui
Hebei University of Science and Technology
ShuJie Song
Hebei University of Science and Technology
LiPing Chen
Hebei University of Science and Technology
XiangYang Chen
Hebei University of Science and Technology
Jing Qi (qijingalice@163.com)
Beihang University School of Mechanical Engineering and Automation

Research

Keywords: image processing, U-Net, ASPP, residual neural network, multi-module, retinal vessels, image segmentation

Posted Date: June 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-627790/v1

MF²ResU-Net: A multi-feature fusion deep learning architecture for Retinal Blood Vessel Segmentation

Zhenchao Cui [1,2], Shujie Song[1,2], Liping Chen[1,2], Xiangyang Chen[1,2], and Jing Qi[3], *

[1]School of Cyber Security and Computer, Hebei University, Baoding Hebei, 071002, China;

cv_songshujie@sina.com(S.S.), cuizhenchao@gmail.com(Z.C.)

[2]Hebei Machine Vision Engineering Research Center (Hebei University), Baoding Hebei 071002, China;

mc_chenlp@sina.com (L.C.), chenxyhbu@163.com (X.C.)

[3]School of Mechanical engineering and automation, Beihang University, Beijing, 100191, China.

* Correspondence: qijingalice@163.com; Tel.: 13031034959

Abstract: Segmentation of blood vessels becomes an essential step in computer aided diagnosis system for the diseases in several departments of ophthalmology, neurosurgery, oncology, cardiology, and laryngology. Aiming at the problem of insufficient segmentation of small blood vessels by existing methods, a novel method based on multi-module fusion residual neural network model (MF²ResU-Net) was proposed. In the proposed networks, to obtain refined features of vessels, three cascade connected U-Net networks were employed as main networks. To deal with the problem of over-fitting, residual paths were used in main networks. In the blocks of U-Net in MF²ResU-Net, in order to remove the semantic difference in low-level and high-level, shortcut connections were used to combine the encoder layers and decoder layers in the blocks. Furthermore, atrous spatial pyramid pooling was embedded between the encoder and decoder to achieve multi-scale feature of blood vessels. During the training of the networks, to deal with the imbalance between background and foreground, a novel joint loss function was proposed based on the dice and cost- sensitive, which could greatly reduce the impact of unbalance in classes of samples. In experiment section, two retinal datasets, DRIVE and CHASE DB1, were used to test our method, and experiments showed that MF²ResU-Net was superior to 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.

Keyword: image processing; U-Net; ASPP; residual neural network; multi-module; retinal vessels; image segmentation

I. Introstruct

The automatic segmentation for retinal blood vessels is an essential part of computer-aided diagnosis for diabetic retinopathy and hypertensive retinopathy [1]. Retinal vessels examination is a non-invasive diagnosis method, which is a primary and popular examination method for diabetic complications and retinal diseases. At present, the main method for retinal abnormalities is manual examination in practice. However, manual examination method has several disadvantages, such as instability for batch process and subjectivity. And thus, computer-aided system became important for retinal diseases diagnosis. In this diagnosis system, automatic segmentation for retinal blood vessels is an essential part, and the segmentations affect accuracy of examination. Thus, automatic segmentation method for retinal blood vessels played a cornerstone role in computer-aided diagnosis [2-3].

The existing methods for retinal vessels segmentation can be classified into machine learning-based methods and deep learning-based methods. The fixed feature detection method in machine learning was used for the retinal blood vessel segmentation. Hashemzadeh et al. [4] proposed a Unetr: TRansformers (UNETR) for extracting the retinal blood vessels employing a set of effective image features and combination of supervised and unsupervised machine learning techniques. To solve the problem of the intra-class high variance of image features calculated from various vessel pixels, Lai Maria et al. [5] proposed automatic retinal image analysis (ARIA) methodology applied machine-learning technology to optimize retinal information. Bala M. Ponni et al. [6] proposed a method for extracting the retinal blood vessels based on feature classification. The blood vessels were extracted from the color fundus image by applying the preprocessing methods and segmentation techniques using matched filter and modified local entropy thresholding operation. To reduce the ophthalmologists' time for examining the retinal images, Fan Guo et al. [7] proposed a supervised method for segmenting blood vessels in retinal images based on the ELM classifier. For these

machine learning methods, the features used for classification could have a big impact on the result of the prediction. In order to achieve higher efficiency, researchers proposed high requirements for automatic and effective feature extractor.

Since of its development in computer vision, the deep learning-based methods became the mainstream for each task in computer vision, such as retinal blood vessel segmentation. Liskowski P. et al. [8] first used a convolutional neural network (CNN) to segment retinal blood vessels. Oliverira A. et al. [9] proposed a combination of steady-state wavelet transform and multi-scale fully convolutional neural network for blood vessel segmentation; Wu Huisi et al. [10] proposed a novel scale and context sensitive network (SCS−Net) for retinal vessel segmentation, the network dynamically adjusted the receptive fields and used a self-adaptive feature fusion module. To obtain global context, Ni et al. [11] 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 [12], was designed for medical image segmentations, which was an encoder-decoder model with skip connections. Many medical image segmentation methods were proposed based on fully convolutional neural network (FCNN) with U-shaped structure [13-14]. Jiang et al. [3] proposed an automatic segmentation model, DU-Net, to obtain denser feature information and reduce the characteristics of small blood vessels. T. Laibacher et al. [15] proposed the M2U-Net, which had a new encoder-decoder architecture which was inspired by the U-Net. Z. Gu, J. Cheng et al. [16] introduced a context encoder between the feature encoder and decoder module. This context encoder used dense atrous convolutional block and multiple kernel pooling to extract more high-level features. Some studies had tried to address this problem by using atrous convolutional layers [17], self-attention mechanisms [18], and image pyramids [19]. However, although the deep learning-based methods have achieved excellent performance in the field of medical image segmentation, they still could not fully meet the strict requirements of medical applications for tiny vessel segmentation. Since of the subtle contrast between the target vessels and the background of the images, the features of small and tine vessels cannot be detected by the existing methods. This leaded that the existing methods had a low accuracy rate for the small vessels with weak pixel values. Image segmentation was still a challenge task in medical image analysis.

In order to obtain precise segmentation for retinal vessel, a novel model MF^{2}ResU-Net and a novel loss function were proposed in this paper. In the model of MF^{2}ResU-Net, three cascade connected U-nets were used as the backbone. Three shortcuts were used for the blocks of U-Nets to deal with the problem of over-fitting. In each block of

U-Nets, to reduce semantic gap, residual paths were used to connect same layers of encoder and decoder instead of shortcuts. To refine features of retinal vessels and obtain multi-scale features, atrous space pyramid pooling module (ASPP) was employed into our model. For training of our model, and to solve unbalance problem in foreground and background, we proposed a novel loss function-based dice loss function and cross function, which could reduce effects of unbalance of samples during training.

The contributions of our work can be elaborated as follows:

1. To refine representation features of retinal small vessels, a novel network, multi-module fusion residual neural network model, MF^{2}ResU-Net, was proposed, by which features of blurry small vessels can be detect.

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

The rest of this paper was organized as follows: Section 2 presented the improved method and fusion; Section 3 introduced experiment data sets and analyzes the experimental results; Section 4 summarized the paper and draws our conclusions.

2 MF^{2}ResU-Net for retinal vessels segmentation

Since of the difficulty on feature detection on small blurry vessels, it was hard to obtain considerable segmentation results by conventional methods. In this paper, we proposed multi-module fusion residual U-Net model, named MF^{2}ResU-Net, to refine the small vessels and to obtain the segmentation of retinal vessels.

2.1 Fusing Residual Paths U-Net

In MF^{2}ResU-Net, we used three cascade-connected U-Net as the backbone network of the module. U-Net was a classic encoder-decoder structure net, A distinctive contribution of the U-Net architecture was the introduction of shortcut connections between the corresponding layers before the max-pooling and after the deconvolution operations. As shown in Fig. 1(a). As the features coming from the encoder were computed in the earlier layers of the network. Opposite, the decoder features of going through convolution, down-sampling and up-sampling were supposed to be of much higher level, because they were computed at the very deep layers of the network. Thus, there were semantic differences between the same layers of encoder and decoder, which could affect the results of segmentation.

In order to remedy semantic differences between the same layers of encoder and decoder, Szegedy et al. [20] used two 3 × 3 convolutional layers after each pooling layer and before transposed convolutional layer, this series of two 3 × 3 convolutional operations actually resemble a 5 × 5 convolutional operations. Therefore, according the approach of the Inception network, the simplest way to augment U-Net with a multi-resolution analysis capability was to incorporate 3 × 3, and 7 × 7 convolution operations in parallel to the 5 × 5 convolution operation, as shown in Fig. 1(b). 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 was to use stride convolutions, but in our experiments, although performance had improved, the introduction of additional convolutional layers in parallel extravagantly increases the memory requirement. We factorized the bigger, more demanding 5 × 5 and 7 × 7 convolutional layers, using a sequence of smaller and lightweight 3 × 3 convolutional blocks, as shown in Fig. 1(c). our fusing residual path was a cascade-connected blocks structure, and each block consists convolutional layers with 3×3 filters and 1×1 filter. This modification greatly reduced the memory requirement, we gradually increased the filters in those, to prevent the memory requirement of the earlier layers from exceedingly propagating to the deeper part of the network. We also add a residual connection because of their efficacy in biomedical image segmentation (Drozdzal et al. [21]) as well as to add 1 × 1 convolutional layer, which may have allowed us to capture some additional retina spatial information. We named this structure as 'Res-path'. And according to the difference number of different layers, set different length Res-path.

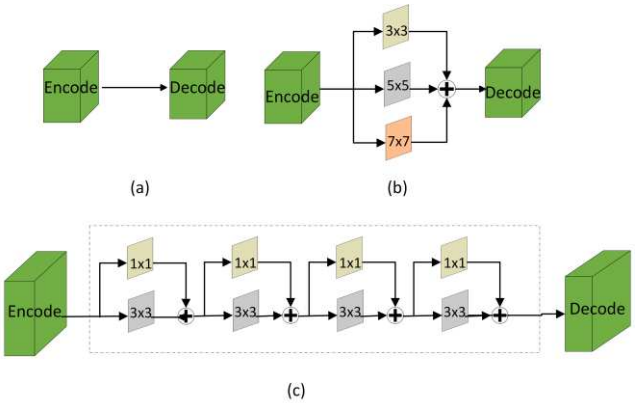


Fig. 1 Res Path structure diagram

In order to refine feature maps of retinal vessels, we used fusing U-Net, named ResU-Net, as the blocks of our model as shown in Fig. 2. To avoid losing feature in detection, we used a light U-Net structure with two convolutional

layers and two max-pooling layers as down-sampling in encoder and two convolutional layers and two deconvolution layers as up-sampling in decoder. The parameters for the fusing residual path U-Net were presented in Table 1.

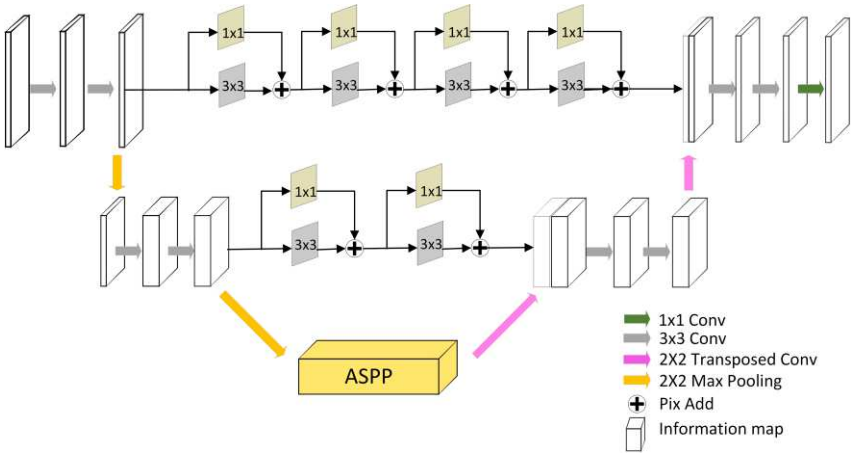


Fig. 2. ResU-Net Block module

In addition, a big challenge in vessel segmentation was to detect the vessels with various shapes and scales. To fundus retinal vascular image to conduct a comprehensive and detailed feature extraction, as shown in Fig. 3, in the middle of the encoder and decoder, we build the module adopted the atrous space pyramid pooling module (ASPP) implemented learning multi-scale characteristics of the local area, ASPP refined feature of retinal vessels in MF^{2}ResU-Net. According to tubular characteristics of various size in vessels, 4 atrous convolutions with the size of 3×3 convolution were used for multi-scale feature extraction. ASPP was inspired by the spatial pyramid pooling method of DeepLab v2 [22], but due to using too large dilation rate, network extract invalid features from blood vessels, too large dilation rate was not suitable for data sets of retinas, so the dilation convolution with a void rate of 24 in ASPP was deleted. The dilation rates in our model were defined as {2,4,8,16}. In order to accelerate computing, 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 pixels and corresponding. The entire parameters in ASPP as shown in Table 1.

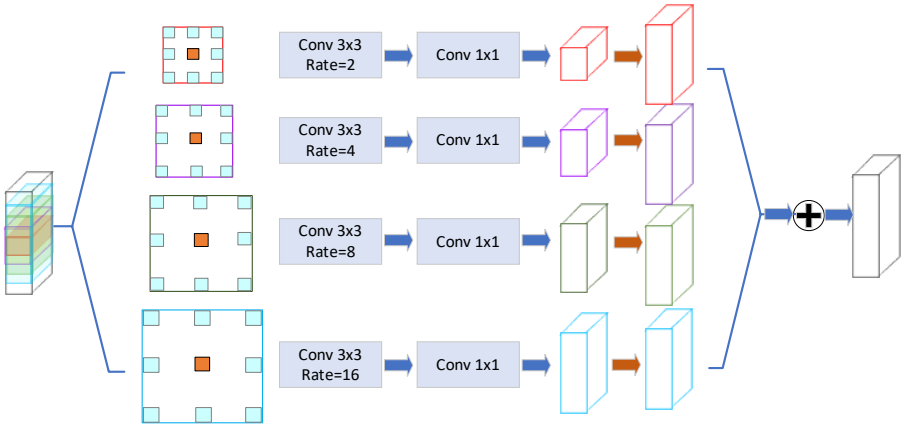


Fig. 3. Atrous spatial pyramid pooling

Table 1. The paramters in ResU-Net block

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

2.2 multi-module residual neural network MF2ResU-Net for retinal vessel segmentation

The aim of this study was to build deep learning models to segment retinal vessels in fundus images, Fig. 4 shows a diagram of the MF^{2}ResU-Net model for retinal vessels segmentation. We cascaded three ResU-Net blocks for feature maps to form our multi-module residual neural network, MF^{2}ResU-Net to remedy the sensitivity problem of the networks for training data, residual links were employed in our model. In order to improved generalization of proposed method, the original images were preprocessed and randomly cropped into small patches to establish training and

validation dataset. The sizes of patches were empirically selected 48 × 48 (pixel × pixel) for training and evaluation. After segmentation by MF^{2}ResU-Net, complete vessel probability map could be obtained by reordering the patches. The connection result of each residual block was represented by Form. (1),

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

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

where a_m represented the ResU-Net input and b_m represented the ResU-Net output, $F(a_m,b_m)$ standed for the residual function. $H(a_m)$ meant the map function of feature. The value of $F(a_m)$ was equivalent to b_m. In Form. (2), $\sigma(W_m)$ represented the activation function. The residual connection result was used as the input of the next module through the ReLU activation function. Finally, we got the output result of each residual connection W_m.

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

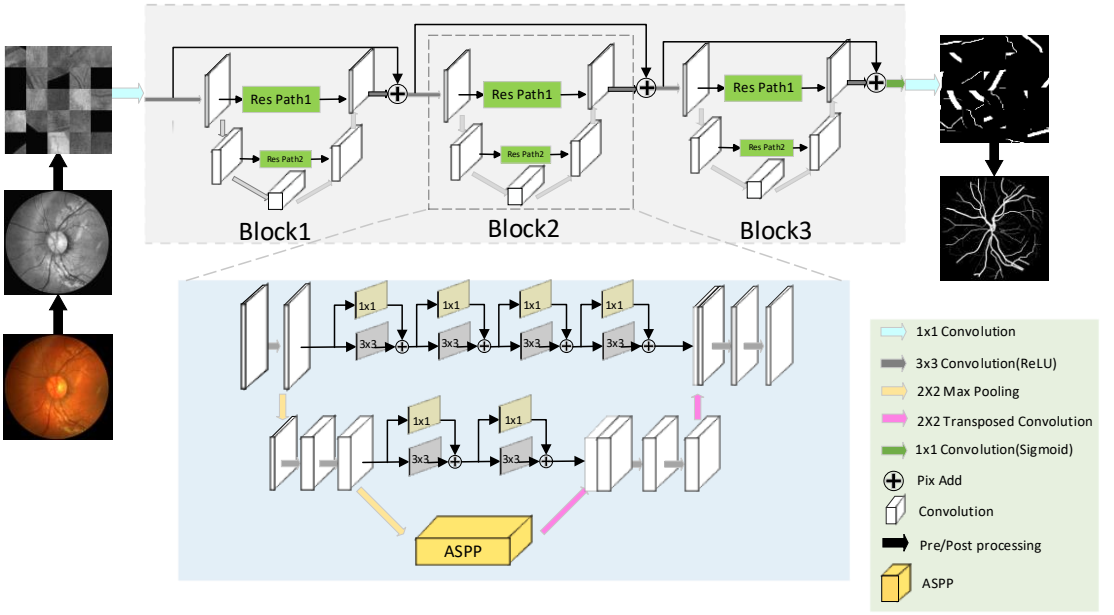


Fig. 4 MF^{2}ResU-Net network structure general map.

2.3 Loss function design

From statistics in a typical retinal blood vessel in data set, we found that the ratio of the pixels of foreground to background was almost 1:12 (The pixel number of foregrounds was 453,800, and the pixel number of backgrounds

was 6,599,200.), which shows that it was an imbalance task for segmentation. And thus, several existing loss functions are not suitable for our task. In order to deal with the imbalance problem, we proposed a novel loss function based on improved binary cross entropy loss function and dice loss function. For binary segmentation problems, the binary cross entropy loss function could be defined as in Form. (3),

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

where N was the number of patch pixels, x_k represented the foreground predicted probability of the input pixel k and y_k was the true label of the pixel k, which was 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 Form.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 $\overline{x_k}$ was an indicator for wrong prediction to x_k, $\overline{x_k} = \begin{cases} 1, & x_k \le 0.5 \\ 0, & x_k > 0.5 \end{cases}$. λ is the penalty parameter for predicting the blood vessel, which was a positive real number. In this loss function. In the loss function of L_{CE}, we could set a large value of λ to enlarge the loss of the wrong prediction in foreground. Since of the ratio of foreground to background, λ was set to 12 in this paper.

To deal with imbalance problem and obtain a perfect criterion on intersection of union, the dice loss function [23] was proposed in segmentation task, which could be presented in Form. (5).

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

where x_k represented the fundus blood vessel region segmented by the algorithm, and y_k denoted the fundus blood vessel region manually segmented by the expert. $|x_k \cap y_k|$ represented the same area of the retinal blood vessel region

segmented by the proposed method and expert. To remedy the numerical problems, we improved the dice loss by introducing a Laplace smooth factor: φ, and the improved dice loss can be presented in Form.(6).

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

To deal with the vanishing gradient problem of dice loss and combine the advantage of two loss functions, a synthetic loss function for the training of MF^{2}ResU-Net model was proposed in this paper,

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

where α was a parameter which controls the contribution of the L_{CE} and L_{dice} loss functions.

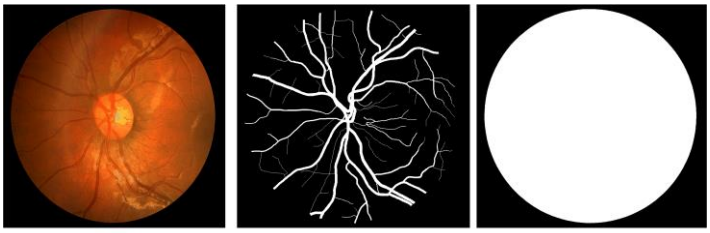
3 Experiment

In this section, we tested our model, MF^{2}ResU-Net, by comparing with several conventional retinal vessel segmentation methods on public datasets: DRIVE, 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, CHASE_DB1 were used to test our model. The DRIVE contains 40 colored retinal vessel images which were obtain from diabetic retinopathy (DR) screening program in the Netherlands [24]. The dataset was randomly divided into two groups, one was the test set and the other was the training set, each group contained twenty images. The size of each image is 565 × 584. CHASE DB1 dataset contained 28 images corresponding to two per patient for 14 children in the program Child Heart and Health Study in England [24]. 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 manually. The result of manually setting the mask for CHASE DB1 was shown in Fig. 5. These two-fundus retinal blood vessel data sets were commonly used retinal blood vessel segmentation algorithm test data sets.

211 The platform of experiment was: 64-bit Win 10, Intel Core i5-4160k, 3.60GHz CPU (Intel), and 6 GB of
212 NVIDA1660Ti GPU (GIGABYTE) for MF^{2}ResU-Net training. The structure of the network was implemented under
213 the open-source deep learning library TensorFlow.



214

215 **Fig. 5. One typical image in CAHSE DB1.** (a) the original retinal vessel image;(b) ground-truth; (c) the manual mask.

216 ***3.2Data pre-processing***

217 The images of the two databases were all colored RGB images, the original retinal blood vessel image had a low
218 contrast. The features of retinal blood vessel were not obvious. In order to improve the performance of the MF^{2}ResU-
219 Net model, the technique that enhances image contrast was used to make the retinal blood vessel features more
220 obvious. Since that gray scale image showed better contrast than the RGB images [25]. We used gray scale image as
221 the input of models. To strength the contrast ratio between vessels and background in retinal images, we used three
222 strategies for image preprocessing, which were normalization, contrast limited adaptive histogram equalization [26]
223 (CLAHE) and gamma correction. Fig. 6 shows preprocessed images of one typical retinal image using the three
224 strategies. The preprocessed retinal image had a high contrast between the blood vessel outline and the background and
225 reduced noise.

226 Data augmentation was widely applied in convolutional neural networks because of its high efficiency and
227 operability. Considering that the DRIVE and CHASE DB1 were small datasets, model will be prone to overfitting and
228 has a poor classification performance. Therefore, it was necessary to augment the dataset for achieving the better
229 results. Four image processing steps were used for augmenting dataset and they are rotating, mirroring, shifting and
230 cropping. To reduce overfitting problem, our models were trained on small patches which were randomly extracted
231 from the images. In order to reduce the calculation complexity and ensure the surrounding local features, small size
232 blocks of 48 × 48 randomly were extracted from the preprocess images were used to train our model. In this paper,
233 190,000 and 200,000 blocks were randomly extracted from DRIVE and CHASE DB1 data sets respectively. And The

90% of blocks were used for training, and the rest blocks are used for testing. Several randomly sampled patches and corresponding labels for experiments were presented in Fig 7, the corresponding label for that patch was decided based on the ground truth images.

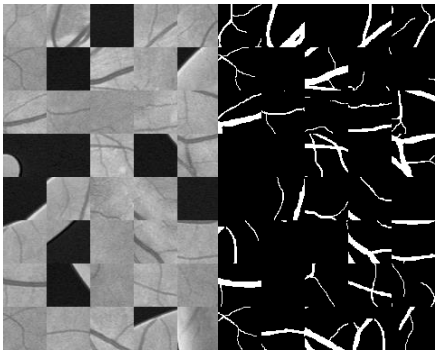


Fig. 7. Local block information map. Typical 48 × 48 patches selected for model training. (a) Patches from the preprocessed image; (b) patches from the corresponding ground truth.

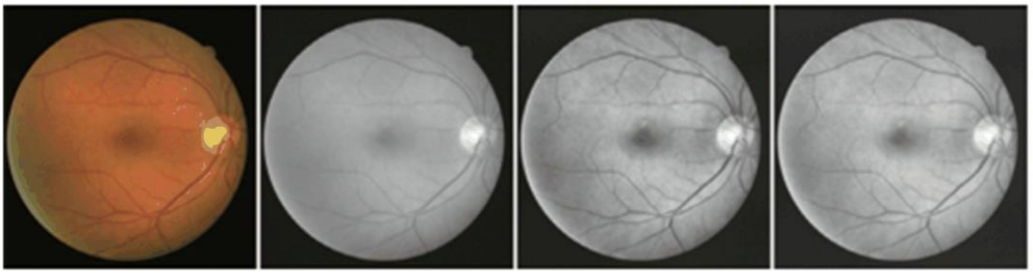


Fig. 6. Retina image preprocessing results. (a) Original image; (b)Normalized image; (c) Image after CLAHE operation; (d) Image after Gamma correction.

3.3 Performance evaluation criteria

The classification result of every pixel could be one of the four types, where the entries were as follows. 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) 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.

In order to quantitatively evaluate segmentation of the proposed algorithm, four evaluation indicators were used in this paper, which were: accuracy (ACC), sensitivity (Sen), specificity (Spe) and F1-Score [28]. In this model, positive refered to blood vessels and negative refered to background. The ACC, Sen, Spe and F1-Score are defined as follows:

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

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

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

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

$$pre = \frac{T_P}{T_P + F_P} \tag{12}$$

$$F_{1-score} = \frac{2 \times pre \times R_{call}}{pre + R_{call}} \tag{13}$$

where ACC was the ratio of the number of correctly detected blood vessels and background pixels to the total number of image pixels; Sen was the ratio of the number of correctly detected retinal blood vessel pixels to the total number of blood vessel pixels; Spe was the ratio of the number of correctly detected non-vessel pixels to the total number of non-vessel pixels; F1-Score was a measure of the similarity between the model and the expert segmentation results. In this model, positive refered to blood vessels and negative refers to background.

The AUC value represents the area under ROC curve. The ROC curve was an important method for measuring the comprehensive performance of image semantic segmentation. Its value ranged from 0 to 1. Condition AUC = 1, a perfect classifier; 0.5 < AUC < 1, better than random classifiers; 0 < AUC < 0.5, worse than the random classifier.

3.4 Experimental result

In this section, 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

Fig. 8 shown the main processing of ablation experiments. After the image preprocessing, one retinal image could be segmented many image blocks with size of 48x48, as shown in Fig. 8(b). By using the vessel detection model, we

obtained the segmentation result composed of the segmented image. As the network deepens, the segmentation results were better. From the detail map in the third row and (d) column, it could be analyzed that the three-fusion network had the best detail segmentation and complex curvature segmentation functions.

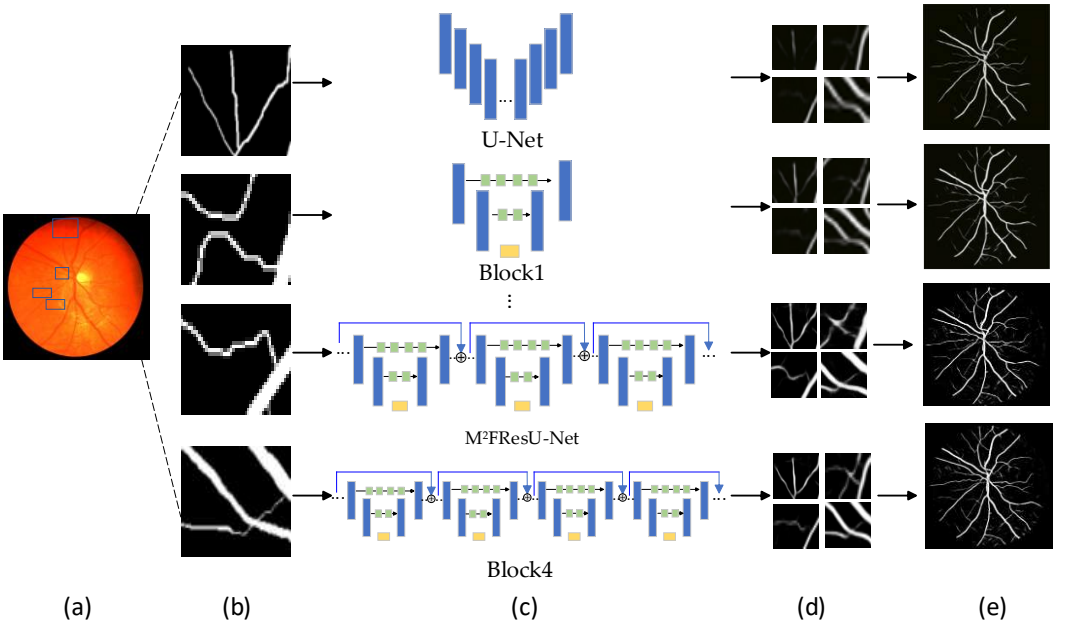


Fig. 8. Detail comparison diagram of segmentation of three network structures. (a) Original image; (b) original image samples; (c) Snapshots of proposed MF^{2}ResU-Net, Block1, Block2 and U-Net. (d) Corresponding to the sample of network segmentation results (e) Re-composition of segmentation results.

Table 2 and 3 present the results of ablation experiments. In 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, we can see that on the dataset of DRIVE, the Sen, F1, AUC, and ACC of MF^{2}ResU-Net were 0.8013, 0.8223, 0.9797 and 0.9700 respectively, which were the highest values in comparing methods. This meant that the results of MF^{2}ResU-Net on DRIVE were better than other methods. As shown in Table 3, we can see that the Sen, F1, AUC, and ACC of MF^{2}ResU-Net were 0.8102, 0.8142, 0.9837, 0.9776 respectively on the dataset of CHASE DB1. The values were the highest values in all results, which means that the generalization ability of MF^{2}ResU-Net was better than the results of other blocks methods. However, compared with the analysis of the experimental results of Block4, as shown in the discounted indicators in Fig. 8, Table 3 and Table 4, Block4 had higher indicators than Block1 and Block2, but slightly lower or the same segmentation indicators

than MF^{2}ResU-Net, and the training time of Block4 was longer. The three-module residual network, MF^{2}ResU-Net, was superior to the other four networks in terms of segmentation statistic.

Table 2. The statistical results of different modules on the dataset of DRIVE.

Models	DRIVE				
	Sen	F1	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
MF^{2}ResU-Net	**0.8013**	**0.8223**	0.9842	**0.9797**	**0.9700**
Block4	0.7741	0.8159	0.9820	0.9782	0.9647

Table 3. The statistical results of different modules on the dataset of CHASE DB1.

Models	CHASE DB1				
	Sen	F1	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
MF^{2}ResU-Net	**0.8102**	**0.8142**	0.9809	**0.9837**	**0.9776**
Block4	0.7942	0.8119	0.9820	0.9824	0.9722

We compared our proposed model with two state-of-the-art networks. One was the customized implementation of U-Net, which we had introduced above; The other was the DeepLab v2. DeepLab v2 was an advanced segmentation network, which combines deep convolutional nets, atrous convolution, and fully connected CRFs. In MF^{2}ResU-Net, we combined atrous convolution and U-Net. In order to highlight the excellence of our works. We compared the three models, DeepLab v2, U-Net and MF^{2}ResU-Net based on the DRIVE and CHASE DB1datasets.We evaluated the model using the test data. Sen, Spe, ACC, F1 and AUC were compared and shown in Tables 4 and 5. It shows from the tables that the MF^{2}ResU-Net achieves the highest values for most of the metrics. The global accuracy for DeepLab v2, U-Net, and MF^{2}ResU-Net were 0.9636, 0.9527, 0.9700 on DRIVE, 0.9624, 0.9549, 0.9776 on CHASE, respectively. What was more, we evaluated the models using ROC curves, which was shown in Fig. 9. The closer the ROC curve to the top-left border was in the ROC coordinates, the more accurate a model. These results shown that the curves of MF^{2}ResU-Net were the most top-left one among the three models while the U-Net curve was the lowest one of the three. Besides, the figures also shown that the MF^{2}ResU-Net obtains the largest area under the ROC curve (AUC).

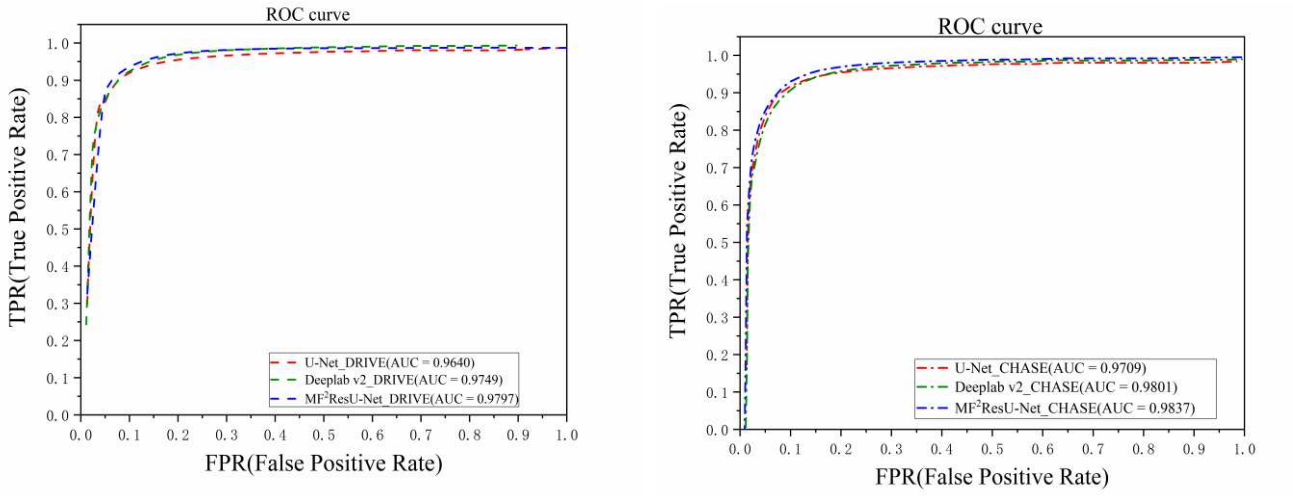
Table 4. Performance of the three models tested on DRIVE

Models	DRIVE

	Sen	F1	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
MF^{2}ResU-net	0.8013	0.8223	0.9842	0.9797	0.9700

Table 5. Performance of the three models tested on CHASE DB1

Models	CHASE DB1				
	Sen	F1	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
MF^{2}ResU-net	0.8102	0.8142	0.9809	0.9837	0.9776



(a) (b)

Fig. 9. ROC curves of different models. (a)U-Net, DeepLab v2 and MF^{2}ResU-Net of ROC in DRIVE. (B) U-Net, DeepLab v2 and MF^{2}ResU-Net of ROC in CHASE DE1.

3.4.2Comparison against existing methods

To test our model for retinal vessel segmentation, we compared it with several states of art. We presented the results and details of several typical retinal vessel images by using several methods. The first column in Figure 10 was the original images from DRIVE and CHASE DB1; the second column was ground-truth of the segmentation; the third column, fourth column, the fifth column, respectively present the results of U-Net, DU-Net [3] and MF^{2}ResU-Net; and

the second line, the fourth line, the sixth line respectively presented the enlarged versions of red and green blocks in corresponding original images by using U-Net, DU-Net and MF^{2}ResU-Net.

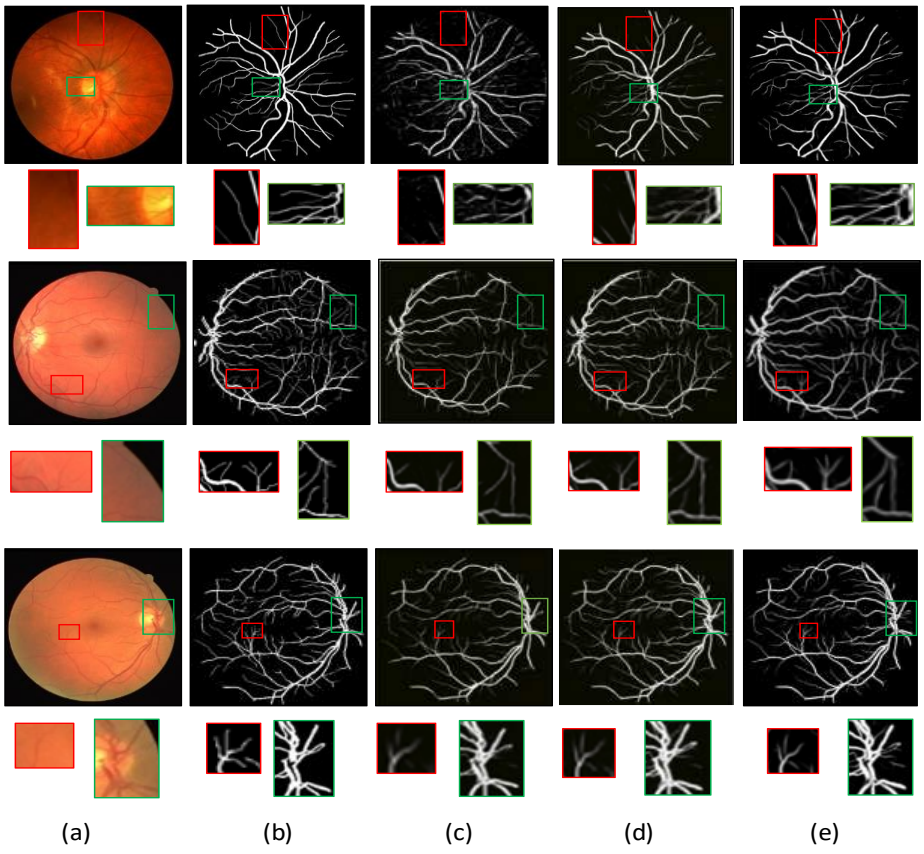


Fig. 10. The results and enlarged versions of segmentation of different algorithms on DRIVE and CHASE DB1datasets. (a) Original image; (b) ground-truth;(c) U-Net segmentation results;(d) DU-Net segmentation results;(e) MF^{2}ResU-Net segmentation results.

From the whole images in Fig. 10, the results of MF^{2}ResU-Net were closer to the ground-truth than other methods, which meaned that our model was superior to the comparing methods. From the enlarged version of segment blocks, MF^{2}ResU-Net could detect small vessels which were disturbed by complex background, but the small vessels in the results of comparing methods were blurred or disappeared. MF^{2}ResU-Net also had a more obvious expression of vascular features, and could segment the subtle blood vessels that were not obvious. Fig. 11 shows the segmentation details of the two datasets. Fig. 11 the first line of the local details and the third line was show, the main blood vessels overlap area and adjacent to each other, the network must accurately segment this blood vessel, but in LadderNet [27],

when the blood vessel was unclear and there was a lot of noise interference, the blood vessel breaks when it was segmented. Obviously, our method had a better segmentation effect in this case, by as shown in Fig. 11 second and fourth lines. LadderNet only has U-Net's common convolutional layer feature extraction function. With the increase of depth of network constantly pooling operation led to loss of a large number of local details. In contrast, in our method, 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 deepen 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. MF^{2}ResU-Net extracted more details of blood vessels in some connection areas than that in 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.

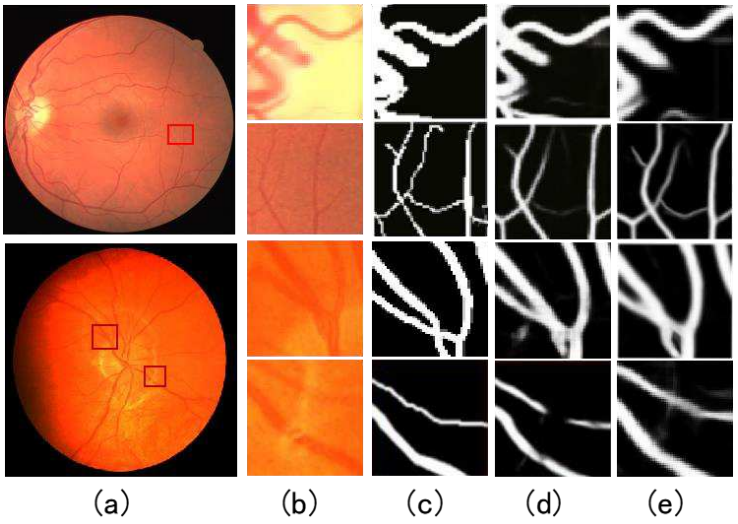


Fig. 11. Detail comparison of segmentation results. (a) Original image; (b) ground-truth; (c) U-Net segmentation; (d) LadderNet; (e) MF^{2}ResU-Net.

We also compared our method with several published state-of-the-art approaches which were proposed recently. From Table 6 and 7, we summarize the type of algorithm, year of publication, and performance on DRIVE and CHASE DB1 dataset. We could see that our model achieved the highest values on Sen, Spe, ACC and AUC for the dataset of DRIVE, which were 0.8013, 0.9842, 0.9700 and 0.9797 respectively. Table 5 shows the statistical results of the methods on dataset of CHASE. Thus, proposed method achieved the highest values on Sen, F1, ACC and AUC,

which were 0.8102, 0.8142, 0.9776 and 0.9837 respectively. These meant that the segmented results by MF^{2}ResU-Net were more accurate than the results of comparing methods.

Table 6. Performance indicators of different algorithms in the DRIVE dataset. N.A: not available by their authors.

Models	DRIVE					
	Years	Sen	Spe	F1	ACC	AUC
Li et al. [28]	2016	0.7569	0.9816	N.A	0.9527	0.9738
Orlando et al. [29]	2017	0.7897	0.9684	0.7857	0.9454	0.9506
R2U-Net [30]	2018	0.7799	0.9810	0.8171	0.9556	0.9784
MS-NFN [31]	2018	0.7844	0.9819	N.A	0.9567	**0.9807**
DEU-Net [32]	2019	0.7940	0.9816	0.8270	0.9567	0.9772
Mou et al. [33]	2020	0.8010	0.9800	N. A	0.9667	0.9700
MSFFU-Net [34]	2020	0.7762	0.9835	N.A	0.9694	0.9790
MF^{2}ResU-Net	2021	**0.8013**	**0.9842**	**0.8223**	**0.9700**	0.9797

In general, the proposed model was superior to the comparing retinal vascular segmentation algorithms. Concretely, Li et al. [28] extracted an image patch with a size of 16 × 16, which might fall into the flat area, thus it was not a good choice for a high-resolution dataset. Orlando et al. [29] used fully connected conditional random field model for blood vessel segmentation, but labeling all blood vessels was a challenge. The methods proposed in recent years, such as literature [30-36]. Although the performance was desirable, the training procedure was complicated, although these methods reached new state-of-the-art performance in some metrics, which was not practical for real application. A modified MF^{2}ResU-net model acts as a soft and rapid automation system to provide the ophthalmology with primary information and knowledge about the blood vessels, such as size, tortuosity, crossing, lesions structure, hard and soft exudates. This model could not uproot the role of the ophthalmology doctors, but it could help doctor's early diagnosis, and increased the result accuracy. We also planned to extend our MF^{2}ResU-net to other vessel segmentation, aiming to obtain more accurate results in medical image analysis tasks. This would be a fruitful area for further work.

Table 7. Performance indicators of different algorithms in the CHASE DB1 dataset.

Models	CHASE DB1					
	Years	Sen	Spe	F1	ACC	AUC
Li et al. [28]	2016	0.7507	0.9793	N.A	0.9581	0.9716
Orlando et al. [29]	2017	0.7277	0.9712	0.7332	0.9458	0.9524
R2U-Net [30]	2018	0.7756	0.9820	0.7928	0.9634	0.9815
MS-NFN [31]	2018	0.7538	**0.9847**	N.A	0.9637	0.9823
DEU-Net [32]	2019	0.8074	0.9821	0.8037	0.9661	0.9812

Tamim N et al [35]	2020	0.7585	0.9846	0.7580	0.9620	N.A
Sine-Net [36]	2020	0.7856	0.9845	N.A	0.9694	0.9824
MF^{2}ResU-Net	2021	**0.8102**	0.9809	**0.8142**	**0.9776**	**0.9837**

4.Conclusions

In this paper, we presented a novel residual neural network based on U-Net for retinal vessel segmentation. The experimental results proved that the method succeeded both absolutely 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 wholly contribute to its success. To refine segmentation feature of retinal vessel, we used residual path to connect encoder and decoder of U-Net, and ASPP was used between encoder and decoder to obtain global feature. To improve the segmentation, multi-model was employed based on ResU-Net blocks. To test the performance of our model, ablation and comparison experiments were conduct in experimental section with criteria of Sen, Spe, F1, ACC and AUC on the datasets of DRIVE and CHASE. The experimental results demonstrated that the proposed algorithm could obtain best values on the most criteria, and the segmentation results shown that our model was more robust than the comparing methods on the complex background, which means that MF^{2}ResU-Net was superior to the comparing methods in retinal vessel segmentation on DRIVE and CHASE DB1.

Declaration:

Ethics approval and consent to participate：Not applicable

Consent for publication：Not applicable

Competing interests: The authors declare that we have no competing interests

Funding: This work is partially supported by Natural Science Foundation of Heibei Province (F2015201033.F2017201069).

Authors' contributions：Conceptualization, Song Shujie; methodology, Cui Zhenchao; software, Song Shujie and ChenLiping; validation, ChenXiangyang, Song Shujie and Qijing; formal analysis, ChenLiping. and ChenXiangyang; writing—original draft preparation, Song Shujie; writing—review and editing, Cui Zhenchao; visualization, Qijing; funding acquisition, Cui Zhenchao.

Acknowledgements: Natural Science Foundation of Heibei Province (F2015201033.F2017201069).

Availability of data and material：Not applicable

References

[1] Irshad S. Akram M. Classification of retinal vessels into arteries and veins for detection of hypertensive retinopathy. Biomedical Engineering Conference, CIBEC, 2014 Cairo International, IEEE, 2014; pp. 133–136.

[2] Cheng C.Y. Zheng Y. Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors. Ophthalmology 118 (5) (2011) 812–818.

[3] Qiangguo Jin, Zhaopeng Meng; Tuan D. Pham. DUNet: A deformable network for retinal vessel segmentation[J]. Knowledge-Based Systems, 2019.

[4] Hatamizadeh A., Yang D., Roth H, "UNETR: Transformers for 3d medical image segmentation," 2021.

[5] Lai Maria . A machine learning approach for retinal images analysis as an objective screening method for children with autism spectrum disorder[J]. E Clinical Medicine. 2020: 100588-.

[6] Bahadar Khan K, A Khaliq A, Shahid M (2016) Correction: A Morphological Hessian Based Approach for Retinal Blood Vessels Segmentation and Denoising Using Region Based Otsu Thresholding. PLOS ONE 11(9): e0162581.

[7] Fan Guo. Retinal Blood Vessel Segmentation Using Extreme Learning Machine[J]. Journal of Advanced Computational Intelligence and Intelligent Informatics. 2017, 21(7): 1280-1290.

[8] Liskowski P.; Krawiec K. Segmenting retinal blood vessels with deep neural networks[J]. IEEE Transactions on Medical Imaging. 2016,35(11):2369-2380.

[9] Oliveira A.; Pereira S.; Silva C A. Retinal vessel segmentation based on fully convolutional neural networks[J]. Expert Systems with Applications. 2018,112:229-242.

[10] Wu Huisi. SCS-Net: A Scale and Context Sensitive Network for Retinal Vessel Segmentation[J]. Medical Image Analysis. 2021, 70(prepublish): 102025-.

[11] Jiajia Ni; Jianhuang Wu; Jing Tong; Zhengming Chen; Junping Zhao. GC-Net: Global context network for medical image segmentation[J]. Computer Methods and Programs in Biomedicine,2020,190.

[12] Ronneberger; Olaf; Fischer. (2015). U-net: Convolutional networks for biomedical image segmentation. In International conference on Springer.

[13] K. S. P. J. M.-H. K. Isensee F, Jaeger PF, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," Nat Methods, vol. 18(2):203-211, 2021.

[14] Jin Q., Meng Z., Sun C., Cui H., and Su R., "Ra-unet: A hybrid deep attention-aware network to extract liver and tumor in ct scans," Frontiers in Bioengineering and Biotechnology, vol. 8, p. 1471, 2020.

[15] Laibacher T.; Weyde T.; Jalali S. M2U-Net: Effective and efficient retinal vessel segmentation for resource-constrained environments. ArXiv preprint. 2018, arXiv:1811.07738.

[16] Z. Gu; J. Cheng; H. Fu; K. Zhou; H. Hao; Y. Zhao; T. Zhang; S. Gao; J. Liu. CE-Net: context encoder network for 2D medical image segmentation, IEEE Trans. Med. Imaging 38 (2019) 2281–2292, doi: 10.1 109/TMI.2019. 2903562.

[17] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab:Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834–848, 2018.

[18] Schlemper J., Oktay O., Schaap M., M. Heinrich, Kainz B., B. Glocker, and D. Rueckert, “Attention gated networks: Learning to leverage salient regions in medical images,” Medical Image Analysis, vol. 53, pp. 197–207, 2019.

[19] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,”in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2017, pp. 6230–6239.

[20] Szegedy, Christian, Ioffe, Sergey, Vanhoucke, Vincent, & Alemi, Alexander A (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In AAAI, Vol. 4 (p. 12).

[21] Drozdzal, Michal, Vorontsov, Eugene, Chartrand, Gabriel, Kadoury, Samuel, & Pal, Chris (2016). The importance of skip connections in biomedical image segmentation. In Deep learning and data labeling for medical applications (pp. 179–187). Springer.

[22] Chen Liang-Chieh, Papandreou George, Kokkinos Iasonas, Murphy Kevin, Yuille Alan L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. [J]. IEEE transactions on pattern analysis and machine intelligence, 2018,40(4).

[23] Li Di, Rahardja Susanto. BSEResU-Net: An Attention-based Before-activation Residual U-Net for Retinal Vessel Segmentation[J]. Computer Methods and Programs in Biomedicine,2021(prepublish).

[24] Stal J.; Abramoff; M.D. Ridge-Based Vessel Segmentation in Color Images of the Retina. IEEE Trans. Med. Imaging. 2004, 23, 501–509.

[25] Zhao Y, Liu Y, Wu X, Harding SP, Zheng Y (2015) Correction: Retinal Vessel Segmentation: An Efficient Graph Cut Approach with Retinex and Local Phase. PLOS ONE 10(4): e0127486.

[26] LIYuan Yuan; CAI Yi Huan. Retinal vessel segmentation algorithm based on hybrid phase feature. Journal of Computer Applications. 2018, 38(7): 2083-2088.

[27] Zhuang J. LadderNet: Multi-path networks based on u-net for medical image segmentation[J]. ArXiv Preprint. 2018, arXiv:1810.0781.

[28] Li Q. L.; Feng B. W.; Xie L. P. A cross-modality learning approach for vessel segmentation in retinal images[J]. IEEE Transactions on Medical Imaging. 2016.35(1):109-118.

[29] Orlando, J.I.; Prokofyeva, E.; Blaschko, M.B. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. IEEE Trans. Biomed. Eng. 64(1), 16-27 (2017).

[30] Alom, M.Z.; Hasan, M. Recurrent residual convolutional neural network based on U-Net (r2u-net) for medical image segmentation. ArXiv Preprint. 2018, arXiv:1802.06955.

[31] Wu, Y. Multiscale network followed network model for retinal vessel segmentation. Frangi A.; Schnabel J.; Davatzikos C. AlberolaLopez C.; Fichtinger, G.(eds.) MICCAI 2018. LNCS, vol. 11071, pp. 119-126. Springer, Heidelberg (2018).

[32] Wang B.; Qiu S.; He H. Dual Encoding U-Net for Retinal Vessel Segmentation. Shen D. et al. (eds) Medical Image Computing and Computer Assisted Intervention-MICCAI 2019. MICCAI 2019. Lecture Notes in Computer Science, vol 1176.

[33] L. Mou, L. Chen, J. Cheng, Z. Gu, Y. Zhao, J. Liu, Dense dilated network with probability regularized walk for vessel detection, IEEE Trans. Med. Imaging 39 (2020) 1392–1403, doi: 10.1 109/TMI.2019.2950051.

[34] Dan Yang; Guoru Liu; Mengcheng Ren. A Multi-Scale Feature Fusion Method Based on U-Net for Retinal Vessel Segmentation. 2020, 22(8).

[35] Tamim N, Elshrkawey M, Azim G A, et al. Retinal Blood Vessel Segmentation Using Hybrid Features and Multi-Layer Perceptron Neural Networks[J]. Symmetry, 2020, 12(6):894.

[36] İbrahim Atli; Osman Serdar Gedik. Sine-Net: A fully convolutional deep learning architecture for retinal blood vessel segmentation[J]. Engineering Science and Technology, an International Journal.2020(prepublish).