Computers in Biology and Medicine

LHR-Net:A lightweight high-resolution network for retinal vessel segmentation

--Manuscript Draft--

Manuscript Number:	CIBM-D-25-01595
Article Type:	Full Length Article
Section/Category:	Medical Imaging: Image Processing and Reconstruction
Keywords:	retinal vessel segmentation, high-resolution, spatial information, parallel channel attention mechanism, multi-scale features
Abstract:	Retinal vessel segmentation takes a crucial part in diagnosis of ophthalmic diseases. However, the low contrast of images and the complex morphological structure of blood vessels pose challenges to automatic segmentation of retinal vessels.To address this issues, we propose a lightweight high resolution network (LHR-Net) which is based on HR-Net. LHR-Net consists of one high-resolution primary path, two low-resolution branches and one multi-scale feature extraction branch, in which before-activation residual blocks (BRB) are introduced to enhance the feature extraction capability of the network. In order to generate rich semantic and spatial information in the high-resolution main path and promote better prediction of microvessels, the parallel channel attention mechanism (PCAM) is embedded in it. Furthermore, we adopt atrous convolution with different dilatation rates to capture the multi-scale features of blood vessels. To evaluate the segmentation performance of LHR-Net, experiments were performed on the fundus datasets DRIVE and STARE, with accuracy of 96.96% and 97.40%, sensitivity of 84.70% and 84.60%, and AUC of 98.69% and 98.76%, respectively. Compared with other state-of-the-art methods, our network has better segmentation results, while its more lightweight, with only 1.0M parameters.

LHR-Net:A lightweight high-resolution network for retinal vessel segmentation

Jun Feng[a], Haibin He[a], Jinmei Guo[a], Huazhen Zhang[a], Jingjing Xiao[a], Liming Liang[b]

[a]School of Computer Science and Engineering, Jiangxi College of Applied Technology, Ganzhou 341000, Jiangxi, China
[b]School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, Jiangxi, China

e-mail of corresponding author Jun Feng: 15770830668@163.com

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 51365017, 61463018), the Natural Science Foundation of Jiangxi Province (No. 20192BAB205084), the Science and Technology Research Youth Project of Jiangxi Provincial Department of Education (No. GJJ2200848), the Science and technology project of Jiangxi Provincial Department of Education(No. GJJ2405103).

Abstract: Retinal vessel segmentation takes a crucial part in diagnosis of ophthalmic diseases. However, the low contrast of images and the complex morphological structure of blood vessels pose challenges to automatic segmentation of retinal vessels.To address this issues, we propose a lightweight high resolution network (LHR-Net) which is based on HR-Net. LHR-Net consists of one high-resolution primary path, two low-resolution branches and one multi-scale feature extraction branch, in which before-activation residual blocks (BRB) are introduced to enhance the feature extraction capability of the network. In order to generate rich semantic and spatial information in the high-resolution main path and promote better prediction of microvessels, the parallel channel attention mechanism (PCAM) is embedded in it. Furthermore, we adopt atrous convolution with different dilatation rates to capture the multi-scale features of blood vessels. To evaluate the segmentation performance of LHR-Net, experiments were performed on the fundus datasets DRIVE and STARE, with accuracy of 96.96% and 97.40%, sensitivity of 84.70% and 84.60%, and AUC of 98.69% and 98.76%, respectively. Compared with other state-of-the-art methods, our network has better segmentation results, while its more lightweight, with only 1.0M parameters.
Keywords: retinal vessel segmentation; high-resolution; spatial information; parallel channel attention mechanism; multi-scale features

1 Introduction

The retinal vessels are the only vascular system in the human body that can be visualized in depth using non-invasive means. In general, diseases such as diabetes, hypertension, high myopia, and glaucoma can lead to changes in retinal vessel structure. If untreated, retinal vessel structural alterations can eventually cause vasculopathy, including narrowing of blood vessels, capillary atherosclerosis, and microvascular aneurysms [1], which can result in inconvenience to people's daily lives. In fact, detecting changes in vascular morphology and structure can help diagnose these diseases, which makes retinal vessel segmentation of great clinical importance. In the clinical diagnosis of ophthalmic diseases, retinal vessel segmentation is usually performed manually by doctors, which is not only time-consuming and laborious, but also difficult to achieve accurate segmentation due to the high variability of the segmentation results [2]. Therefore, it is necessary to develop an accurate and automatic retinal segmentation method to reduce the workload and improve the efficiency of clinical diagnosis.

At present, retinal vessel segmentation methods are divided into two main categories: unsupervised segmentation methods and supervised segmentation methods [3]. The unsupervised segmentation methods are based on the inherent properties of the blood vessels for learning and do not need to refer to manually labeled samples. Many unsupervised segmentation methods have been proposed: Zhang et al. [4] proposed a matched-filter segmentation algorithm that detects blood vessels by thresholding the response of the retinal image to the matched filter, which significantly reduces the blood vessel misdetection produced by the original matched filter. Tian et al. [5] presented an improved retinal blood vessel segmentation algorithm based on the traditional Frangi filtering and mathematical morphology methods, which has high segmentation accuracy and low complexity. Oscar et al. [6] used optimized top-hat and homomorphic filtering to achieve effective retinal vessel segmentation with better specificity and accuracy than most unsupervised methods. In [7], an unsupervised retinal vessel segmentation algorithm based on principal component analysis and local ternary pattern was proposed. A new vessel tracking algorithm was

adopted in [8], which considered only three types of vessel structures, including normal, branching, and crossing vessels. Experiments show that the algorithm can track major blood vessels, but still suffers from missing tiny branches in some cases. Zhao et al. [9] extended the classical active contour model into the IPACHI model, which is more suitable for detecting blood vessels after the introduction of infinite perimeter regularization, but still has insufficient microvessel segmentation and missegmentation of the optic disk. Although unsupervised methods achieve segmentation of a large number of unlabeled fundus retinal images, the low segmentation accuracy and poor generalization ability cannot achieve the ideal results.Thus, it is very necessary to find a more optimal method for automatic retinal vessel segmentation.

In contrast to unsupervised segmentation methods, supervised segmentation methods necessitate the incorporation of manually labeled images as training data labels for constructing the corresponding algorithmic models, thereby yielding superior segmentation performance and enhanced robustness. In recent years, the rapid advancement of computer technology and high-performance hardware has led to an increasing number of scholars and experts dedicating their efforts to the study of deep learning, which has demonstrated state-of-the-art performance in various computer vision tasks such as semantic segmentation, object detection, and image classification [10-12]. With the introduction of networks such as FCN, U-Net, SegNet, DeepLab [13-16], many supervised retinal vessel segmentation algorithms based on deep learning have been proposed: In [17], to fully utilize multi-scale information and encoder features, Shi et al. designed a multi-scale dense network that is more suitable for retinal vessel segmentation. In [18], Liu et al. presented a lightweight network that enables accurate segmentation of retinal vessels in an attempt to ad dress the problems of semantic information loss and restricted receptive fields. Wei et al. [19] proposed a robust orientation and context entanglement network, which can extract complex orientation and context information from blood vessels and maintain good continuity in thin blood vessels. To bridge the semantic gap between encoder and decoder features, Arsalan et al. [20] proposed a prompt deep lightweight network to improve segmentation performance. Li et al. [21] designed a multi-scale attention-guided fusion network, which can extract contextual information more efficiently for accurate segmentation of retinal vessels. Considering the issues of varying retinal vessel thickness and blurred boundaries, while the segmentation algorithms still have challenges in fine segmentation of capillaries and edges, Xu et al. [22] proposed a deep neural network with shared decoder and pyramid-like loss. Li et al.[23] presented GDF-Net, which facilitates accurate segmentation of retinal vessels by constructing segmentation networks and fusion networks to extract more effective global contextual information. To decrease the computational complexity of the model and improve the sensitivity and accuracy of the algorithm, Khursheed et al. [24] designed a lightweight deep learning model. Experiments show that the model can operate on low-end devices and achieve better results. Based on the U-Net network, Geetha et al. [25] designed two subnets for venous and arterial vascular detection, and the final predicted image is a combination of the two.

Although all the above methods are committed to improving retinal vessel segmentation and have achieved superior segmentation results, there is still much room for improvement. To this end, a lightweight high-resolution network, called LHR-Net, is proposed to achieve more accurate segmentation of retinal vessel. We summarize the main contributions of this paper as follows:

(1) A lightweight high-resolution network is proposed, which can maintain high-resolution representation throughout the entire process, while achieving features fusion between low and

high-resolution. The spatial prediction maps of blood vessels will be more precise.

(2) Before-activation residual block (BRB) is introduced to obtain better performance and accelerate the convergence of the network. To optimize the information flow and extract the multi-scale features of blood vessel, the multi-scale feature extraction module (MFEM) is added.

(3) A parallel channel attention mechanism (PCAM) is proposed, which accomplishes the recalibration of the channel information of different resolution feature maps before fusion. It improves the sensitivity of the network to important information and enhances the segmentation of microvessels.

(4) A series of experimental results show that the proposed method has higher performance indicators than most of the competing methods, which proves its superiority in retinal vessel segmentation task.

The subsequent sections of this paper are organized as follows: Section 2 introduces the proposed LHR-Net, MFEM and PCAM, respectively. Section 3 presents the experimental data and evaluation criteria. Section 4 presents the experimental results and analysis. The last section presents the conclusion.

2 Methodology

2.1 Proposed LHR-Net

The majority of existing deep learning-based retinal vessel segmentation methods use high-to-low and low-to-high network architectures. Fig. 1 displays the overall architecture of the proposed LHR-Net, which keeps high resolution consistently throughout the input-to-output process.

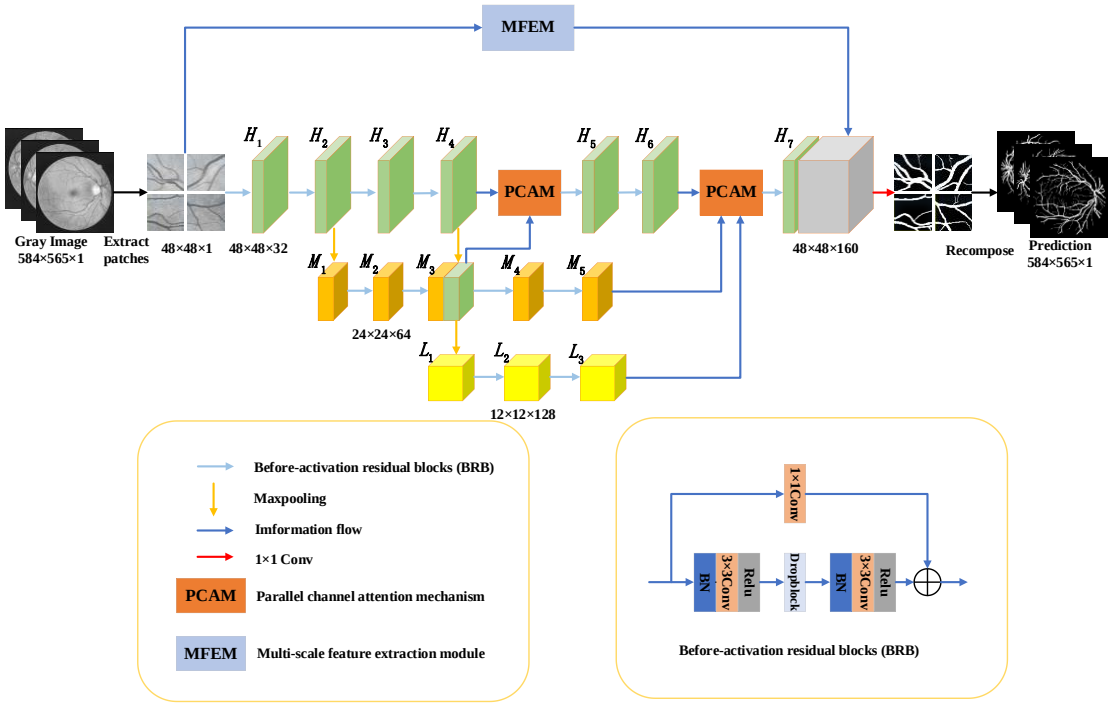


Fig.1 The LHR-Net architecture

LHR-Net is composed of a high-resolution main path, two low-resolution branches and a multi-scale feature extraction branch. Images are first fed into the high-resolution main path, to extract more effective vessel information and obtain better segmentation accuracy, the BRB is

introduced into the whole stage of the network. Compared to other residual blocks, BRB provides superior performance compared to other residual blocks with the batch normalization layer placed before the activation unit [26]. To obtain advanced semantic information, 2 subnets from high-resolution to low-resolution are gradually added after down-sampling. Since high-resolution feature maps have detailed spatial information, low-resolution feature maps have rich semantic information, for this purpose, PCAM is utilized to fuse different resolution feature maps at multiple scales to guarantee that the high-resolution main path receives information from other low-resolution subnets, producing a rich high-resolution representation. By recalibrating the channel information of different resolution feature maps and merging the generated feature maps to transmit them to the subsequent layers of the high-resolution main path, PCAM can effectively highlight the boundary information of vessel ends and capillaries while suppressing noise. Retinal vessels have different morphologies and scales, whether the network has the ability to capture multi-scale features is crucial for segmentation results. Retinal vessels have different morphologies and scales, and whether the network can extract multi-scale features is crucial for segmentation results. Accordingly, the MFEM module is embedded into the multiscale feature extraction branch, which uses four atrous convolutions with different dilatation rates to expand the receptive field of the extracted features without increasing the computational complexity.

2.2 MFEM

Atrous convolution has been widely used in recent semantic segmentation tasks due to its excellent performance, which yields a larger receptive field without changing the computational and convolutional kernel sizes. In addition, atrous convolution can capture more valuable semantic features and overcome the issue of information loss in pooling operations. Atrous convolution is defined as follows:

$$y[i]=\sum_{k=1}^{K} x[i+r\cdot k]\cdot \omega[k]$$

Where r is the dilation rate, $w[k]$ denotes the kth parameter of the filter. K is the size of the filter, and $x[i]$ and $y[i]$ represent the input and output features, respectively. For a filter with a convolution kernel size of $k\times k$, the resulting dilation filter size is $K_r\times K_r$ when its dilation rate is r, where $K_r=k+(k-1)\times(r-1)$. Therefore, those convolutions with large dilation rates also have large receptive field.

To enhance the network's capability of feature extraction and representation and improve the accuracy of segmentation results. Meanwhile, to maximize the retention of information in the original image, a multi-scale extraction branch is constructed using MFEM, and its structure is shown in Fig. 2. Firstly, considering the resolution of the input image is 48×48, four parallel atrous convolutions with different dilation rates are placed in the MFEM, with dilation rates of 2, 4, 8 and 16 from top to bottom. secondly, to reduce the computational complexity, a 1×1 convolution is added before each atrous convolution, and the number of convolution kernels of both of them is set to 1/2 of the number of channels of the feature maps in the high-resolution main road. Furthermore, another 1×1 convolution is inserted after each atrous convolution to restore the number of channels of the feature map to 32. Finally, the feature maps of the four branches are concatenated to obtain the multiscale features of the blood vessels.

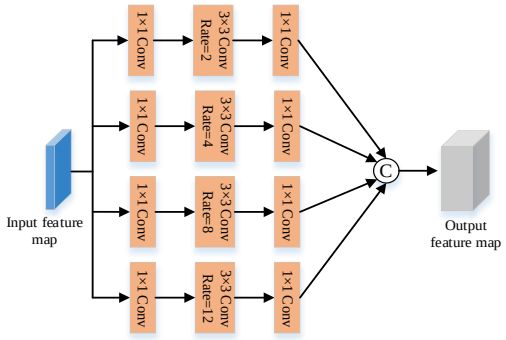


Fig. 2 Structure diagram of MFEM

2.3 PCAM

In LHR-Net, high-resolution feature maps have rich spatial information, which can accurately locate the position of blood vessels. Low-resolution feature maps have deep semantic information, which helps to recognize capillaries correctly. To fuse shallow spatial information with deep semantic information to facilitate better prediction of blood vessels. The proposed PCAM is embedded into the high-resolution main path, which is prompted to receive information from the low-resolution subnet to generate a rich high-resolution representation.

The structure of PCAM is shown in Fig. 3, which consists of three parallel channel attention. In this structure, the feature maps L_3 and M_5 from the low-resolution subnets are first up-sampled by a sampling factor of 4 and 2, respectively, to produce images with the same scale as the high-resolution feature maps. Next, three feature maps with the same scale are fed into 1×1 convolution to adjust the number of channels, and the size of the adjusted image is $48\times48\times32$. Then the global information of each channel is obtained by Global Average Pooling (GAP), which generates a tensor of $1\times1\times C$. After GAP, the Sigmoid activation function is used to acquire the attention weight coefficient α_n ($\alpha_n \in [0,1]$) of each channel, and it is multiplied with the channel of the corresponding feature map to get the channel-weighted attention map. To accelerate network convergence and facilitate training, skip connections are used. Finally, the three weighted feature maps are concatenated to generate the final channel attention map. The process is represented as follows:

$$\boldsymbol{O}_1 = \left(\sigma\left(GAP\left(W_1 \times up\left(\boldsymbol{L}_3,4\right)+b\right)\right)\right)\times up\left(\boldsymbol{L}_3,4\right)+up\left(\boldsymbol{L}_3,4\right)$$

$$\boldsymbol{O}_2 = \left(\sigma\left(GAP\left(W_2 \times up\left(\boldsymbol{M}_5,2\right)+b\right)\right)\right)\times up\left(\boldsymbol{M}_5,2\right)+up\left(\boldsymbol{M}_5,2\right)$$

$$\boldsymbol{O}_3 = \left(\sigma\left(GAP\left(W_3 \times \boldsymbol{H}_6+b\right)\right)\right)\times \boldsymbol{H}_6+\boldsymbol{H}_6$$

$$\boldsymbol{Y} = \psi\left(\boldsymbol{O}_1,\boldsymbol{O}_2,\boldsymbol{O}_3\right)$$

where O_1, O_2, and O_3 are the channel attention maps corresponding to L_3, M_5, and H_6, respectively. (W_1,b), (W_2,b), and (W_3,b) are the parameters of the corresponding

convolutional layers. $up(\cdots,k)$ is the upsampling with a sampling factor of k, and GAP is the Global Average Pooling. σ represents the Sigmoid activation function, ψ represents the concatenation operation, and Y denotes the final output of the module.

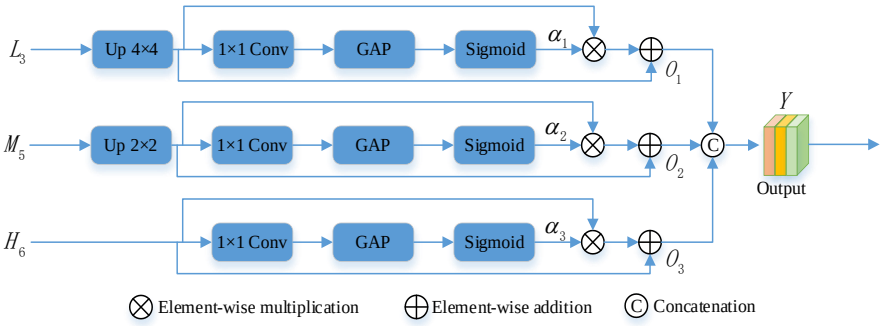


Fig. 3 The structure of PCAM

3 Dataset and evaluation metrics

3.1 Dataset

To verify the overall performance of the proposed LHR-Net, 2 commonly available public datasets of fundus images, DRIVE [27] and STARE [28], were used. The DRIVE dataset contains 40 color retinal images, each with its matching gold standard and mask, and the resolution of the images is 565×584. The training and testing sets have been divided, where the images numbered 21-40 are the training set and the images numbered 1-20 are the test set. The STARE dataset consists of 20 color retinal images, half of which contain pathological indications and the rest are from healthy subjects, each with a resolution of 700×605. Here, the first 10 images from this data are used as the training set and the remaining 10 images are used as the test set. In this paper, the images manually annotated by the first expert in the above 2 datasets are used as the reference standard for the experiments.

3.2 Data preprocessing and patch cropping

Most of the retinal images in the dataset suffer from noise interference, uneven illumination and low contrast. It is necessary to preprocess retinal images appropriately, which can improve image quality and thus enhance segmentation accuracy. In data preprocessing, the initial retinal image is first converted to a grayscale image with high foreground and background contrast. Then, data normalization and contrast limited adaptive histogram equalization (CLAHE) are used to improve the grayscale details and further enhance the contrast. Finally, gamma transformation is utilized to suppress the illumination inhomogeneity.

The limited number of retinal images in the dataset makes it difficult to meet the training demand of deep convolutional neural networks. Thus, the strategy of patch cropping is applied to model training and testing, which can effectively reduce the risk of overfitting and improve the model performance. Specifically, in model training and testing, a patch with a resolution of 48×48 is randomly cropped from an image in a step of 5 using a sliding window. The strategy produces a large number of image patches for the purpose of data augmentation. For the DRIVE and STARE training sets, 10,000 and 8,000 patches were cropped for training, respectively.

3.3 Evaluation metrics

To quantitatively evaluate the comprehensive performance of LHR-Net, four commonly used evaluation metrics were introduced, including Accuracy (Acc), Sensitivity (Se), Specificity (Sp), and F1-score ($F1$), as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Se = \frac{TP}{TP + FN}$$

$$Sp = \frac{TN}{TN + FP}$$

$$F1 = \frac{2TP}{2TP + FP + FN}$$

Where TP (True Positive) represents vessel pixels correctly classified by the classifier. TN (True Negative) represents non-vessel pixels correctly classified by the classifier. FP (False Positive) denotes vessel pixels incorrectly classified by the classifier. FN (False Negative) denotes non-vessel pixels incorrectly classified by the classifier. $Accuracy$ represents the proportion of correctly classified pixels and reflects the overall segmentation accuracy of the model. Se represents the proportion of correctly classified vessel pixels, the closer it is to 1, the fewer unsegmented vessel pixels there are, and the stronger the model's ability to effectively extract blood vessels. Sp represents the percentage of correctly classified non-vessel pixels. $F1$ is a very important metric in binary classification models, which is used to evaluate the similarity between segmentation results and the ground truth. Additionally, the area under curve (AUC) of receiver operating characteristic (ROC) is adopted to evaluate the performance of the model.

4 Experiment and analysis

4.1 Implementation details

The computer configuration used in this experiment is Intel(R) Core(TM) i5-12400F CPU, 16 GB RAM, NVIDIA GeForce RTX 3060ti GPU. The operating system is Linux, the programming language for building the network is Python 3.7, and the main library packages are Pytorch 1.2, OpenCV 4.5.1 and Numpy 1.19.2, etc. All experiments are conducted in this experimental environment. The Adam function is used as the optimizer of the network, and the initial learning rate is set to 0.001. The batch size of training is 18, and the epoch is 50. The cross-entropy loss function is adopted to calculate the difference between the predicted image and the ground truth, which is defined as follows.

$$Loss_{CE}(g, p) = -\sum g_i \log p_i + (1 - g_i)\log(1 - p_i)$$

where g_i and p_i are the true and predicted labels of pixel i, respectively.

4.2 Comparison between different models

The segmentation performance of LHR-Net is tested on DRIVE and STARE datasets, and compared with the experimental results of U-Net, U-Net++ [29], MFI-Net [30] and Ladder-Net [31]. Considering the validity of the comparison, the same experimental environment and methodology are used for the training and testing of these five models, and the parameters is tuned to the optimum.

4.2.1 Comparison of segmentation results

Fig. 4 presents the segmentation results of LHR-Net with the other four models. In Fig. 4, the first to third rows show the retinal images of the DRIVE dataset, while the fourth and fifth rows show the retinal images of the STARE dataset. Among them, (a) and (b) are the original image and the ground truth, respectively, and (c)~(g) are the segmentation images of LHR-Net, U-Net, U-Net++, MFI-Net and Ladder-Net, respectively.

It can be seen from the first and second rows of healthy retinal images in Fig. 4 that LHR-Net has better robustness in microvessel segmentation compared to the other four models. its segmented microvessels are closer in shape and number to those in the ground truth, while the other four models all show more microvessel breaks. Observing the pathological retina images in the third row, more pathological information mis-segmentation occurred in the segmented images of U-Net, U-Net++ and Ladder-Net, which led to the increase of false negatives, and there were also more blood vessel breaks. In contrast, LHR-Net and MFI-Net can better suppress the missegmentation of pathological information. The optic disc region in the fourth row of healthy retinal images is affected by uneven illumination. Except for the proposed method, which can segment the blood vessels around the optic disc accurately, the remaining four methods show vessel breakage. The contrast of the pathological retinal image in the fifth row is low, and the main blood vessels in the U-Net, U-Net++ and MFI-Net segmentation images all show signs of fracture. Influenced by artefacts and pathological regions, more noise appeared in the Ladder-Net segmentation results. Although our method produced a small amount of mis-segmentation, it outperformed the other methods in both the segmentation of the main vessels and the microvessels.

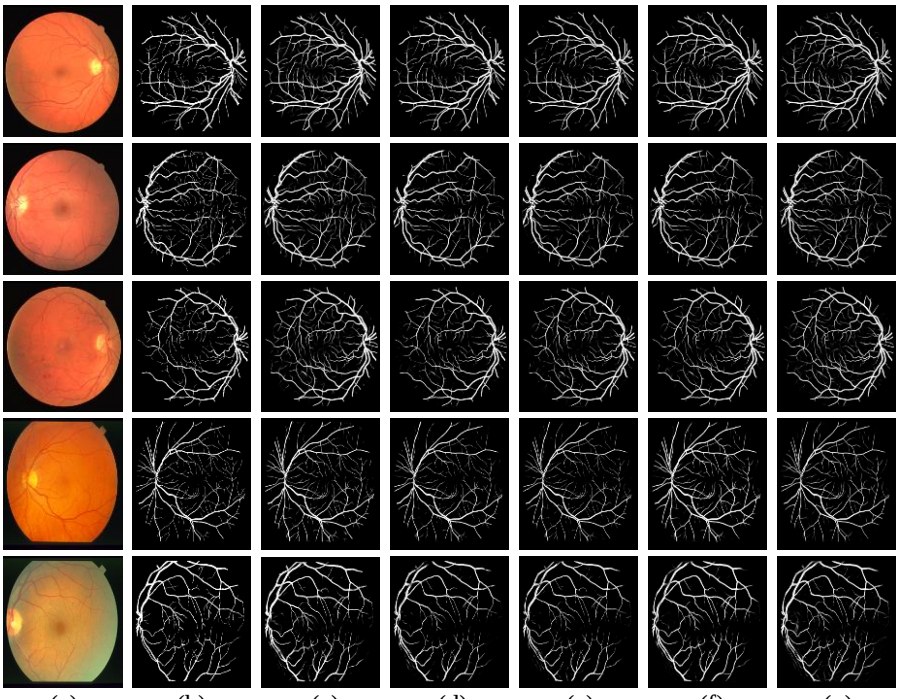


Fig. 4 Segmentation results of 5 models on DRIVE and STARE datasets. (a) Original image. (b) Ground truth. (c) LHR-Net. (d) U-Net. (e) U-Net++. (f) MFI-Net. (g) Ladder-Net.

To more clearly illustrate the superior segmentation results of LHR-Net. The segmentation

details of the five models on the DRIVE and STARE datasets are show in Fig. 5. Where (a)~(c) are the original image, original details and ground truth details, respectively, (d)~(h) are LHR-Net, U-Net, U-Net++, MFI-Net and Ladder-Net details, respectively.

The vessels around the optic disc in the first retinal image in Fig. 5 have complex topology and different scales, which makes accurate segmentation difficult. The other four models exhibited varying degrees of fracture when segmenting the intersecting blood vessels around the optic disc, and demonstrated limited recognition of microvessels. In contrast, our method can accurately detect the aforementioned vessels. This is mainly attributed to MFEM, which can effectively extract the multi-scale features of blood vessels to obtain more detailed information. The analysis of the second retinal image shows that the proposed LHR-Net has a stronger capability of identifying microvessels in low contrast compared to other models, indicating that PCAM improves the robustness of the model to the segmentation of microvessels by recalibrating and fusing the information of the feature map channels in different paths. The segmentation details of the third image were analyzed. Affected by the low-contrast region, the other four models are fractured or absent in segmenting both the main blood vessels and the microvessels at the intersection, while the segmentation results of LHR-Net are clearly better, showing superior vessel details.

Through the analysis of the visualisation results, the proposed LHR-Net achieves more accurate segmentation of blood vessels with different shapes and scales, and has good robustness in the segmentation of microvessels. Meanwhile, it overcome the effects of factors such as low-contrast and noise in fundus images. Its overall performance is superior to the comparison model.

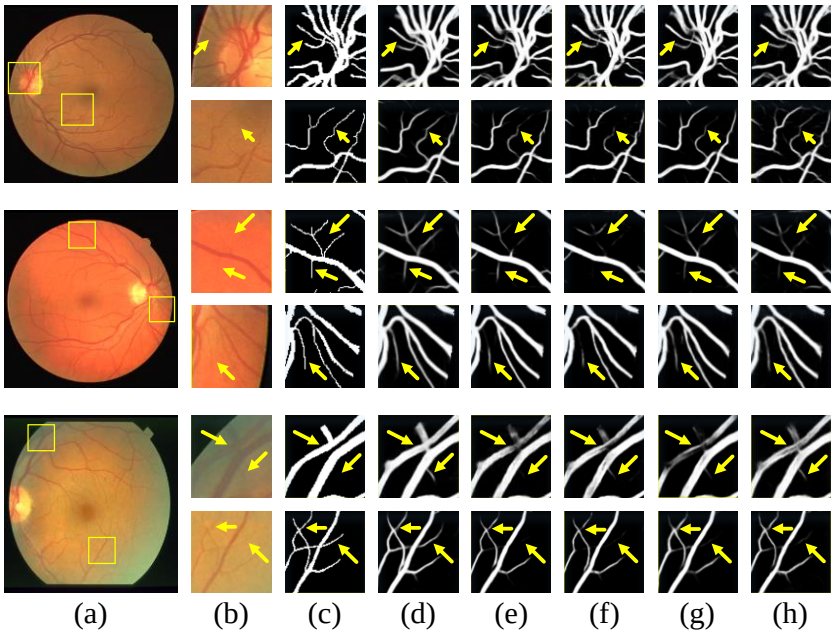


Fig. 5 Segmentation details of the 5 models on the DRIVE and STARE datasets. (a) Original image. (b) Original image details. (c) Ground truth details. (d) LHR-Net details. (e) U-Net details. (f) U-Net++ details. (g) MFI-Net details. (h) Ladder-Net details.

4.2.2 Quantitative comparison

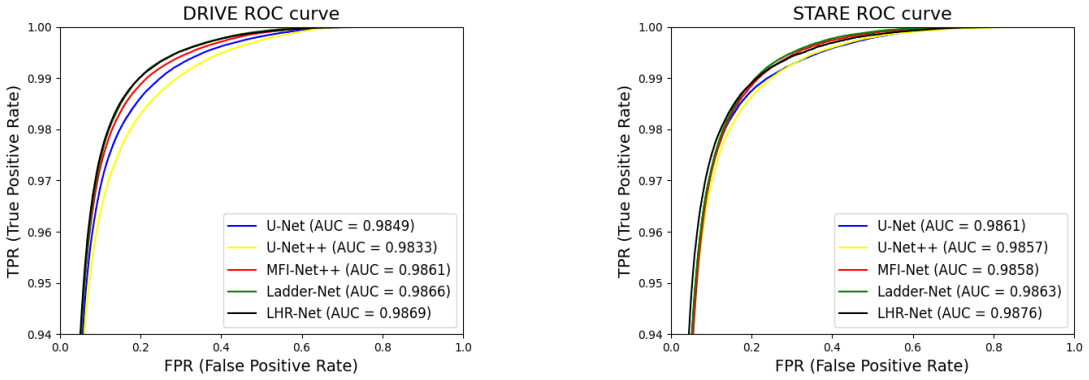
To quantitatively analyse the experimental results, Table 1 presents the performance

comparison of LHR-Net with other four models on DRIVE and STARE datasets, where the numbers marked in bold are the optimal values of this metric. In the DRIVE dataset, the proposed LHR-Net outperforms all the compared models in most of the metrics. In particular, *Se* reaches 84.70%, which is a 3.28% improvement over the second-highest Ladder-Net. Its *Acc* and AUC achieve 96.96% and 98.69%, respectively. In the STARE dataset, LHR-Net is higher than all comparison models in all metrics except *Sp*. Compared with U-Net++, its *Se* and *F*1 improved by 12.19% and 2.77%, respectively. Also its *Acc* and AUC improved from 97.30% and 98.57% to 97.40% and 98.76%, respectively. In terms of network parameters, the network parameters of LHR-Net are only 1.0M, which are much less than those of U-Net++ and MFI-Net. It shows that the proposed LHR-Net is more lightweight, with lower algorithmic complexity, which is easy to be applied in low-configuration hardware systems to complete retinal vessel segmentation.

Table 1 Performacnce comparison of 5 models on DRIVE and STARE datasets

Models	Parameters/M	DRIVE					STARE				
		Acc/%	Se/%	Spe/%	F1	AUC/%	Acc/%	Se/%	Spe/%	F1	AUC/%
U-Net	7.8	96.90	78.65	98.66	81.69	98.49	97.34	75.15	99.19	80.88	98.61
U-Net++	36.1	96.86	77.54	**98.72**	81.25	98.33	97.30	72.41	**99.36**	80.34	98.57
MFI-Net	27.6	96.94	79.50	98.62	82.05	98.61	97.15	79.73	98.60	80.17	98.58
Ladder-Net	1.5	96.92	81.42	98.41	82.26	98.66	97.38	79.41	98.82	81.68	98.63
LHR-Net	1.0	**96.96**	**84.70**	98.05	**82.62**	**98.69**	**97.40**	**84.60**	98.44	**82.51**	**98.76**

As shown in Fig. 6, the ROC curves of five models on DRIVE and STARE reflect the information between false positive samples (background pixels incorrectly segmented as vessel pixels) and true positive samples (correctly segmented vessel pixels). As can be seen from (a) and (b) in Fig. 6, the area AUC value under the ROC curve of LHR-Net is closer to 1, demonstrating its excellent segmentation performance for retinal vessels.

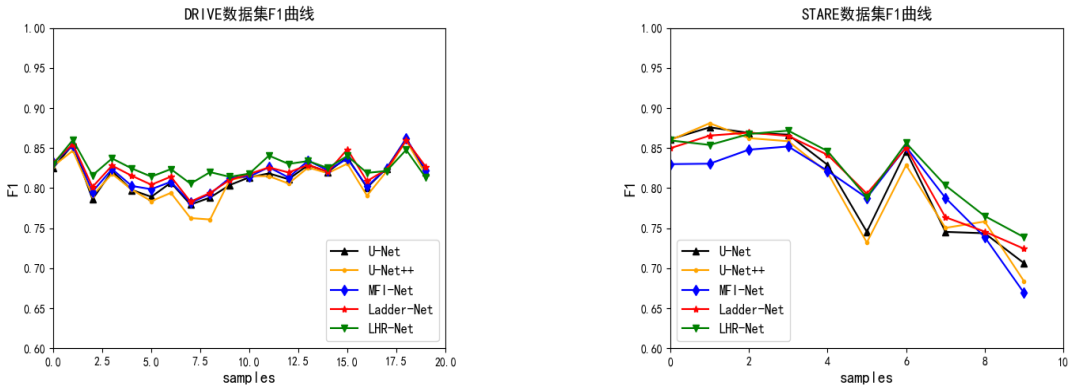


(a) (b)

Fig. 6 ROC curves of retinal vessel segmentation for five models. (a) ROC curve of DRIVE. (b) ROC curve of STARE.

The *F*1 performance curves of the five models in the DRIVE and STARE datasets are shown in Fig. 7, so that we can further analyze the performance of the models in segmenting a single image. As shown in Fig. 7 (a), from the *F*1 curve of the DRIVE dataset, it can be seen that the *F*1 curve of LHR-Net has less fluctuation and is relatively stable compared with the other four models, which proves that it can extract feature information stably and has strong robustness in segmenting healthy and pathological retinal images. Observing the *F*1 curve of the

STARE dataset in Fig.7 (b), in the first five images, the $F1$ values of all models basically showed a stable trend. In the last five retinal images, four of them have severe pathological information and low contrast, so it is difficult to extract blood vessel features accurately. $F1$ curve of LHR-Net has smaller variation amplitude compared with other four models, which verifies its better segmentation performance for pathological retinal images.



(a) (b)

Fig. 7 $F1$ performance curve of retinal vessel segmentation for five models. (a) $F1$ curve of DRIVE dataset. (b) $F1$ curve of STARE dataset.

4.3 Ablation experiment analysis

To investigate the effect of different modules in LHR-Net on the segmentation performance, ablation experiments are conducted. Here, the high-resolution network without any module is taken as the baseline network, and the LHR-Net is split into four networks in total. Subnet 1 is the baseline network. Subnet 2 represents subnet 1 joining BRB. subnet 3 represents subnet 2 joining MFEM. subnet 4 represents subnet 3 joining PCAM, i.e., LHR-Net. The ablation experiments are conducted on the DRIVE dataset, and the experimental results are shown in Table 2.

As can be seen from the experimental data in Table 2, the sensitivity and accuracy of the Subnet 1 are low, indicating that the ability of ordinary convolution to extract feature information is weak. As the depth of the network increases, ordinary convolution causes some important information to be ignored in the transmission process, so the sensitivity and accuracy are lower. Subnet 2 has some overall improvement, which shows that joining BRB can effectively improve the model's ability to extract feature information. Subnet 3 has a greater improvement in accuracy, sensitivity and AUC compared to subnet 1. It suggests that MFEM can prompt the network to learn more detailed information by capturing the multi-scale features of blood vessels, which improves the model segmentation accuracy. When the three modules were simultaneously incorporated into subnet 1 to form LHR-Net, it obtained the best segmentation results. From the data of subnet 3 and subnet 4, the sensitivity is improved by 3.01% after incorporating PCAM. It is demonstrated that PCAM effectively improves the ability of the model to correctly classify vessel pixels by recalibrating the feature channel information and fusing feature maps with different resolutions, and also better suppresses the effect of noise in fundus images.

Table 2 Results of LHR-Net ablation experiments on the DRIVE dataset

Networks	Acc/%	Se/%	Spe/%	F1/%	AUC/%
Subnet 1	96.75	78.65.	**98.74**	81.69	98.46
Subnet 2	96.82	80.29	98.64	82.53	98.62

Subnet 3	96.91	81.69	98.46	82.58	98.67
Subnet 4	**96.96**	**84.70**	98.05	**82.62**	**98.69**

4.4 Comparison with other state-of-the-art algorithms

To further illustrate the advantages of LHR-Net and compare it with the advanced models in recent three years, Table 3-4 show the quantitative comparison of different retinal vessel segmentation models on DRIVE and STARE, and the optimal values of each evaluation metrics are marked in bold.

Table 3 Comparison with state-of-the-art algorithms on the DRIVE dataset

Method	Year	Acc/%	Se/%	Spe/%	F1/%	AUC/%
Liang et al [10]	2022	95.67	83.48	97.42	82.98	98.11
Bridge-Net [32]	2022	95.65	78.53	98.18	82.03	98.34
CRAUNet [33]	2022	95.86	79.54	--	83.02	98.30
CSAUNet [34]	2022	96.76	83.40	98.10	--	97.58
CFCNet [35]	2022	96.82	80.63	98.35	--	--
Wave-Net [18]	2023	95.61	81.64	97.64	82.54	--
OCE-Net [19]	2023	95.81	80.18	98.26	83.02	98.21
PLVS-Net [20]	2023	96.78	82.50	98.37	--	98.15
MAGF-Net [21]	2023	95.78	82.62	97.83	**83.07**	98.19
GDF-Net [23]	2023	96.22	82.91	98.52	83.02	98.59
FSE-Net [36]	2024	96.27	81.50	98.34	--	98.39
RCAR-UNet [37]	2024	95.37	74.87	98.36	--	--
ResMU-Net [38]	2024	96.85	78.43	**98.64**	81.49	98.42
CaFFNet [39]	2024	96.13	83.33	98.21	--	98.61
LF-UNet [40]	2024	95.75	80.19	98.02	82.77	98.16
Wang et al [41]	2024	95.72	79.15	98.14	82.50	98.12
LHR-Net (Ours)	2024	**96.96**	**84.70**	98.05	82.62	**98.69**

Table 4 Comparison with state-of-the-art on the STARE dataset

Method	Year	Acc/%	Se/%	Spe/%	F1/%	AUC/%
Liang et al [10]	2022	96.67	83.16	98.25	**83.95**	98.69
Bridge-Net [32]	2022	96.68	80.02	98.64	82.89	**99.01**
CSAUNet [34]	2022	97.28	83.04	98.26	--	98.76
CFCNet [35]	2022	97.37	82.35	98.80	--	--
Wave-Net [18]	2023	96.41	79.02	98.36	81.40	--
OCE-Net [19]	2023	96.72	80.12	98.65	83.41	98.76
PLVS-Net [20]	2023	97.27	81.90	98.74	--	97.06
MAGF-Net [21]	2023	96.49	80.93	98.44	83.64	98.98
GDF-Net [23]	2023	96.53	76.16	**99.57**	80.22	98.89
FSE-Net [36]	2024	97.10	68.10	98.18	--	98.76
RCAR-UNet [37]	2024	95.94	69.79	99.05	--	--
Feng et al [39]	2024	96.93	81.64	98.87	--	98.82
LF-UNet [40]	2024	95.72	71.97	98.43	77.53	97.69
Wang et al [41]	2024	96.54	73.68	99.01	80.62	98.62
LHR-Net (Ours)	2024	**97.40**	**84.60**	98.44	82.51	98.76

Based on the data presented in Table 3 and Table 4, it can be concluded that LHR-Net achieved the best segmentation performance on both the DRIVE and STARE datasets. On the DRIVE datasets, the accuracy, sensitivity, and AUC of LHR-Net are higher than other comparison

methods. In particular, the sensitivity is 1.22% higher than the second best method. The specificity of ResMU-Net [38] is 0.59% higher than ours, but our accuracy, sensitivity, F1 and AUC are 0.11%, 6.27%, 1.13% and 0.27% higher, respectively. The F1 of MAGF-Net [21] is 0.45% higher than our method, but its accuracy, sensitivity, specificity, and AUC are 1.18%, 2.08%, 0.22%, and 0.5% lower than ours, respectively.

As shown in Table 4, the accuracy, sensitivity, specificity, F1 and AUC of the proposed LHR-Net on STARE dataset are 97.40%, 84.60%, 98.44%, 82.51%, and 98.76%, respectively, which is the best among all the comparative methods in terms of accuracy and sensitivity, and the sensitivity is higher than the second-best method by 1.44%. The specificity and AUC of GDF-Net [23] are 1.13% and 0.13% higher than ours, respectively, but its accuracy, sensitivity, and F1 are 0.87%, 8.44%, and 2.29% lower than ours, respectively. Liang et al. [10] has a higher F1 than ours, but all other evaluation metrics are lower than ours. The specificity, F1 and AUC of Bridge-Net [32] are close to our method, but the accuracy and sensitivity of LHR-Net are 0.72% and 4.58% higher, respectively.

By quantitatively comparing with the above state-of-the-art methods on DRIVE and STARE datasets, the proposed LHR-Net has superior segmentation performance, especially with the best accuracy, sensitivity. It is fully demonstrated that LHR-Net can segment microvessels more accurately and achieve the correct classification of more vesse pixel points with better robustness.

5 Conclusion

We propose a lightweight high-resolution network (LHR-Net) for accurate segmentation of retinal blood vessels. Firstly, the proposed LHR-Net not only learns multilevel hierarchical feature information, but also maintains a high-resolution representation throughout the process, which contributes to the acquisition of rich spatial information. Ordinary convolution is replaced by BRB to enhance the network's ability to extract feature information. Secondly, PCAM is merged into the high-resolution primary path, which generates rich semantic and spatial information and facilitates the prediction of different types of blood vessels. In addition, the MFEM improves the accuracy of network segmentation by capturing the multi-scale feature information of blood vessels. Finally, experimental results show that LHR-Net achieves better performance than other state-of-the-art methods on two publicly available datasets. It can segment finer and more robust vessel structures, while its lightweight network structure is convenient to be applied in the clinical diagnosis of ophthalmic diseases.

In future work, we expect to apply the proposed network architecture to other medical image segmentation tasks and further optimise the performance of the model.

References

[1] Zana F, Klein J C. A multimodal registration algorithm of eye fundus images using vessels detection and Hough transform[J]. IEEE transactions on Medical Imaging, 1999, 18(5): 419-428.

[2] Staal J, Abràmoff M D, Niemeijer M, et al. Ridge-based vessel segmentation in color images of the retina[J]. IEEE transactions on medical imaging, 2004, 23(4): 501-509.

[3] Fraz M M, Remagnino P, Hoppe A, et al. Blood vessel segmentation methodologies in retinal images – a survey[J]. Computer methods and programs in biomedicine, 2012, 108(1): 407-433.

[4] Zhang B, Zhang L, Zhang L, et al. Retinal vessel extraction by matched filter with first-order derivative of Gaussian[J]. Computers in biology and medicine, 2010, 40(4): 438-445.

[5] Tian F, Li Y, Wang J, et al. Blood vessel segmentation of fundus retinal images based on improved frangi and mathematical morphology[J]. Computational and Mathematical Methods in Medicine, 2021, 2021: 1-11.

[6] Ramos-Soto O, Rodríguez-Esparza E, Balderas-Mata S E, et al. An efficient retinal blood vessel segmentation in eye fundus images by using optimized top-hat and homomorphic filtering[J]. Computer Methods and Programs in Biomedicine, 2021, 201: 105949.

[7] Faridoon F, Khan K, Umair M Y, et al. Unsupervised Vessel Segmentation Technique Using LTP and PCA[C]//2023 International Conference on Communication Technologies (ComTech). IEEE, 2023: 7-12.

[8] Zhang J, Li H, Nie Q, et al. A retinal vessel boundary tracking method based on Bayesian theory and multi-scale line detection[J]. Computerized Medical Imaging and Graphics, 2014, 38(6): 517-525.

[9] Zhao Y, Rada L, Chen K, et al. Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images[J]. IEEE transactions on medical imaging, 2015, 34(9): 1797-1807.

[10] Liang L, Feng J, Zhou L, et al. U-shaped Retinal Vessel Segmentation Based on Adaptive Aggregation of Feature Information[J]. Interdisciplinary Sciences: Computational Life Sciences, 2022, 14(2): 623-637.

[11] Wang J, Su Y, Yao J, et al. Apple rapid recognition and processing method based on an improved version of YOLOv5[J]. Ecological Informatics, 2023, 77: 102196.

[12] Liu S, Zhang Q, Huang L. Graphic image classification method based on an attention mechanism and fusion of multilevel and multiscale deep features[J]. Computer Communications, 2023, 209: 230-238.

[13] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[14] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.

[15] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.

[16] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with

deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848.
[17] Shi Z, Wang T, Huang Z, et al. MD-Net: A multi-scale dense network for retinal vessel segmentation[J]. Biomedical Signal Processing and Control, 2021, 70: 102977.
[18] Liu Y, Shen J, Yang L, et al. Wave-Net: A lightweight deep network for retinal vessel segmentation from fundus images[J]. Computers in Biology and Medicine, 2023, 152: 106341.
[19] Wei X, Yang K, Bzdok D, et al. Orientation and Context Entangled Network for Retinal Vessel Segmentation[J]. Expert Systems with Applications, 2023, 217: 119443.
[20] Arsalan M, Khan T M, Naqvi S S, et al. Prompt deep light-weight vessel segmentation network (PLVS-Net)[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2022, 20(2): 1363-1371.
[21] Li J, Gao G, Liu Y, et al. MAGF-Net: A multiscale attention-guided fusion network for retinal vessel segmentation[J]. Measurement, 2023, 206: 112316.
[22] Xu G X, Ren C X. SPNet: A novel deep neural network for retinal vessel segmentation based on shared decoder and pyramid-like loss[J]. Neurocomputing, 2023, 523: 199-212.
[23] Li J, Gao G, Yang L, et al. GDF-Net: A multi-task symmetrical network for retinal vessel segmentation[J]. Biomedical Signal Processing and Control, 2023, 81: 104426.
[24] Aurangzeb K, Alharthi R S, Haider S I, et al. An efficient and light weight deep learning model for accurate retinal vessels segmentation[J]. IEEE Access, 2022, 11: 23107-23118.
[25] Biswal B, Gandhi T K. SAF-NET: Split Attention Fusion Network for retinal vessel segmentation from fundus images[C]//2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2023: 1-7.
[26] Li D, Dharmawan D A, Ng B P, et al. Residual u-net for retinal vessel segmentation[C]//2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019: 1425-1429.
[27] Staal J, Abràmoff M D, Niemeijer M, et al. Ridge-based vessel segmentation in color images of the retina[J]. IEEE transactions on medical imaging, 2004, 23(4): 501-509.
[28] Hoover A D, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response[J]. IEEE Transactions on Medical imaging, 2000, 19(3): 203-210.
[29] Zhou Z, Siddiquee M M R, Tajbakhsh N, et al. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE transactions on medical imaging, 2019, 39(6): 1856-1867.
[30] Jiang Y, Wu C, Wang G, et al. MFI-Net: A multi-resolution fusion input network for retinal vessel segmentation[J]. Plos one, 2021, 16(7): e0253056.
[31] Zhuang J. LadderNet: Multi-path networks based on U-Net for medical image segmentation[J]. arXiv preprint arXiv:1810.07810, 2018.
[32] Zhang Y, He M, Chen Z, et al. Bridge-Net: Context-involved U-net with patch-based loss weight mapping for retinal blood vessel segmentation[J]. Expert Systems with Applications, 2022, 195: 116526.
[33] Dong F, Wu D, Guo C, et al. CRAUNet: A cascaded residual attention U-Net for retinal vessel segmentation[J]. Computers in Biology and Medicine, 2022, 147: 105651.
[34] Huang Z, Sun M, Liu Y, et al. CSAUNet: A cascade self-attention u-shaped network for precise fundus vessel segmentation[J]. Biomedical Signal Processing and Control, 2022, 75: 103613.

[35] Cai D, Fu Y, Wang T, et al. CFCNet: A coarse-to-fine cascade network for retinal vessel segmentation[C]//2022 China Automation Congress (CAC). IEEE, 2022: 2728-2733.
[36] Ni J, Mu W, Pan A, et al. FSE-Net: Rethinking the up-sampling operation in encoder-decoder structure for retinal vessel segmentation[J]. Biomedical Signal Processing and Control, 2024, 90: 105861.
[37] Ding W, Sun Y, Huang J, et al. RCAR-UNet: Retinal vessel segmentation network algorithm via novel rough attention mechanism[J]. Information Sciences, 2024, 657: 120007.
[38] Panchal S, Kokare M. ResMU-Net: Residual Multi-kernel U-Net for blood vessel segmentation in retinal fundus images[J]. Biomedical Signal Processing and Control, 2024, 90: 105859.
[39] Feng T, Wang J, Shen J, et al. Retinal Vessel Segmentation via Cross-attention Feature Fusion[C]//2024 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2024: 1-6.
[40] Zhu X, Zhang W, Li H. LF-UNet: An Attention-Based U-Net for Retinal Vessel Segmentation[C]//2024 IEEE 19th Conference on Industrial Electronics and Applications (ICIEA). IEEE, 2024: 1-6.
[41] Wang Z, Jia L V, Liang H. Partial class activation mapping guided graph convolution network cascaded U-Net for retinal vessel segmentation[J]. Computers in Biology and Medicine, 2024: 108736.

Ethics Statement

The research followed all applicable international and national ethical norms, was approved by the review board of Jiangxi College of Applied Technology, obtained written informed consent provided by all participants, and had the right to withdraw from the research at any time. Research data were collected and processed in accordance with privacy protection principles, and all personally identifiable information was deleted or encrypted.