PCRTAM-Net: A Novel Pre-activated Convolution Residual and Triple Attention Mechanism Network for Retinal Vessel Segmentation

Huadeng Wang, Zizheng Li, Xipeng Pan, Zhenbing Liu, Rushi Lan and Xiaonan Luo
Guangxi Key Laboratory of Image and Graphic Intelligent Processing,
Guilin University of Electronic Technology
Guilin, 541004,China
whd@guet.edu.cn

Idowu Paul Okuwobi
School of Artificial Intelligence, Guilin University of Electronic Technology
Guilin, 541004, China
paulokuwobi@guet.edu.cn

Bingbing Li
Department of PathologyGanzhou Municipal Hospital
Ganzhou, 341000, China
libingbing19932021@126.com

Abstract

Retinal images play an important role in early diagnosis of ophthalmic diseases. Automatic segmentation of retinal vessels in color fundus images is challenging due to the morphological differences between the retinal vessels and the low contrast background, while models struggle to capture representative and discriminative retinal vascular features. To fully utilize the structural information of the retinal blood vessels, we propose a novel deep learning network termed Pre-activated Convolution Residual and Triple Attention Mechanism Network (PCRTAM-Net). The proposed model uses the pre-activated dropout convolution of the residual method to improve the feature learning ability of the network. Residual atrous convolution spatial pyramid is integrated into both ends of the network encoder to extract multiscale information and improve blood vessel information flow. A triple attention mechanism is proposed to extract the structural information between vessel contexts and to learn long range feature dependencies. We evaluated PCRTAM-Net on four publicly available DRIVE, CHASEBD1, STARE, and HRF datasets. Our model achieves an ACC of 97.10%, 97.70%, 97.68% and 97.14%, and an F1 of 83.05%, 82.26%, 84.64% and 81.16%, respectively. The results show that the proposed PCRTAM-Net can extract more detailed retinal vessels and performed better than the state-of-the-art methods in retinal vessel segmentation.

Keywords: retinal image segmentation, triple attention mechanism, atrous convolution, residual network, multiscale feature

1. Introduction

There are a certain number of capillaries in the human retina, and their morphological changes are not only closely related to ophthalmic dis-eases but also reflect the symptoms of a variety of other cardio vessel diseases, such as diabetes, hypertension, arteriosclerosis, etc [10]. The retinal fundus image is an important tool for doctors to diagnose various ophthalmic diseases and other related diseases. However, due to the complex morphology of the retinal blood vessels and the lack of a high definition of fine blood vessels, the efficiency of manual diagnosis is relatively low, and subjectivity is likely to exist [12]. Therefore, automatic segmentation of retinal fundus blood vessel images has good research significance and clinical application value. However, retinal image segmentation is still a challenging task due to several constraints. Firstly, the morphology and size of the blood vessels in retinal images vary greatly. For example, blood vessel in fundus images usually vary between 1 and 20 pixels. Secondly, the retinal vessel tree has many closely connected tiny blood vessels, which are generally difficult to separate from other non-vessel structures. Thirdly, factors such as noise

during retinal image acquisition and exudates produced by lesions making the segmentation task a difficult process. Faced with these challenges, researchers have put in a lot of effort [7] to overcome these issues. The initial research approach [5] is to segment retinal vessels using hand-crafted features. Although these methods have achieved good results in the reported research works, but traditional image processing using hand-crafted features technique cannot represent the complex semantics of retinal blood vessels. Consequently, in a relatively large datasets and data with multiple complex situations, these techniques are liable to perform poorly. In recent years, deep convolution neural network (DNN) methods [13] have achieved remarkable results in medical image segmentation tasks. These networks have become very popular and useful in medical image segmentation problems. U-Net [16] and its various variants, have been improved from a fully convolution network for semantic segmentation, to a network that further improved retinal vessel segmentation performance. These variants are based on a symmetric encoder-decoder structure, where convolution layers and down sampling layers are continuously stacked to obtain retinal vessel features. Although these U-Net variants achieve good performance, but they are insufficient for the fundus image segmentation challenge. Two main disadvantages limit their application in modern medical applications. Ideally, due to factors such as noise, low resolution, and poor contrast, and the general U-shaped variant structure cannot stably segment all the blood vessel features. In addition, there is a lack of multiscale information from the retinal images of complex blood vessels. It is very challenging to develop a single vessel structure model suitable for robust extraction of multi-source vessel images under interference factors. To overcome the above limitations and further improve the performance of retinal vessel segmentation, we propose a retinal vessel segmentation network with a pre-residual attention mechanism to extract vessel structures from retinal images, termed Pre-activated Convolution Residual and Triple Attention Mechanism Network (PCR-TAM-Net). Our network is implemented based on the encoder-decoder structure and then consists of three core modules. Firstly, to better extract the boundary vessel features, we study the convolution layer of the basic network and propose a residual method based on pre-activated dropout convolution (Res-PDC) to capture more microvessel features to assist in the segmentation of vessel structure. Secondly, to extract multiscale information and improve blood vessel information flow, the residual atrous convolution spatial pyramid method (Res-ACSP) is used at both ends of the encoder. Finally, to fully learn the vessel features, as well as the structural information between vessel contexts, various attention mechanisms are investigated, and a triple attention mechanism (TAM) is proposed. The proposed PCRTAM-Net is validated on four publicly available retinal vessel datasets, and the results show that the proposed PCRTAM-Net performed better than the state-of-the-art methods. In summary, the main contributions of our paper are as follows.

(1) A retinal vessel segmentation network with a pre-residual attention mechanism is proposed, which extracts adequate vessel tree features from fundus images with complex vessel structures.

(2) A residual method based on pre-activated dropout convolution (Res-PDC) is proposed, which replaces the convolution block in the deep learning network and enhances the generalization ability by discarding random parts of the vessel structure so that the information on the vessel features can be fully extracted.

(3) To effectively utilize the multiscale information of complex blood vessels, the residual atrous convolution spatial pyramid method (Res-ACSP) is used at both ends of the encoder, so that the extracted information flow is improved. The channel and spatial attention modules are studied between the connected layers of the decoder, and a triple attention mechanism (TAM) is proposed to effectively utilize the multi-channel space for vessel feature normalization so that the background and vessel structure can be classified more effectively.

In the remainder of this paper, Section II presents the overall approach of the retinal vessel segmentation. Section III describes our network in detail. Section IV discusses the experimental results of our network on the publicly available datasets. Finally, conclusions are presented in Section V.

2. Related work

Over the past few decades, many methods have been proposed for retinal vessel segmentation in fundus images. Previously, fundus images were segmented based on conventional image processing techniques, such as morphological operations [19] or threshold segmentation [5]. These methods need to be adjusted again in different situations to achieve better segmentation performance. These learning based methods [15] are not robust enough for this task because the hand-crafted features are misled by lesion regions and low-contrast microvessels. Compared with the above methods, the depthwise convolution methods [13] have better advantages in dealing with the specificity of retinal blood vessels. Sheng et al. [17] improved the detection of low-contrast and narrow blood vessels by using an efficient minimum generation superpixel tree to detect and capture global and local structures of retinal images, but

segmented blood vessels have the limitation of unsmooth boundaries. Yin et al. [29] developed a new method for accurately extracting blood vessels in nonfluorescein fundus images using direction aware detectors, which can filter out background noise near pathological and non-vascular structures, but the resulting accuracy of segmenting blood vessels is low. Dai et al. [1] developed a deep learning system called DeepDR that can detect the pretopoststages of diabetic retinopathy. Wang et al. [22] proposed a hard attention network (HAnet) that consists of one encoder and three decoders, while introducing an attention mecha-nism to enhance the features of the blood vessels in hard regions. Sun et al. [20] proposed a network integrating atrous convolution modules, which obtained a larger receptive field, and improved the thickness of the blood vessels and the perception of details to a certain extent. Jin et al. [8] proposed deformable convolution (DUNet) to replace ordinary convolution for the vessel segmentation. Mou et al. [14] embedded densely dilated convolution blocks into a U-shaped network for retinal blood vessel detection and used a probabilistic regularized walker algorithm to patch the breakage in detection. Wei et al. [23] proposed an automatically designed genetic U-Net, which can achieve better retinal vessel segmentation and solve the problems of overfitting and high computational complexity caused by many parameters. In recent years, attention mechanisms have been applied to the image domain and combined with convolution neural networks. Fu et al. [3] proposed a dual-attention mechanism convolution network (DANet) to solve the image segmentation task by capturing rich contextual dependencies based on a self-attention mechanism. Yang et al. [28] proposed an attention aware multiscale fusion network (AMF-Net), which perceives microvessels through dense convolution, and utilizes a channel attention module to fuse multiscale features with adaptive weights and utilized the location attention module to captures the distance spatial relationship of features to improve performance. Wu et al. [25] proposed a multiscale channel attention network based on the encoder-decoder structure, which extracted the multiscale structural information of blood vessels in the encoder part and fused the channel attention module in the decoder part to improve the vessel segmentation in the fundus images. However, due to the different morphology of retinal vessels, the low contrast between retinal vessels and the background, and the influence of lesions and equipment noise, the state-of-the-art methods for the segmentation of microvessels still need to be improved.

3. Our method

3.1. PCRTAM-Net

The proposed PCRTAM-Net is an encod-er-decoder network for retinal vessel segmentation, and the overall architecture is shown in Fig.1. The encoder of PCRTAM-Net network consists of Res-PDC module, Res-ACSP module and downsampling. The decoder consists of dual pre-activation dropout convolution (Dual-PDC), TAM module and upsampling. Finally, apply 1x1 convolution and sigmoid operation to binarize the vessel probability map.

3.2. Residual method based on pre-activated dropout convolution

The underlying convolution block based on encoder-decoder network plays an important role in segmenting complex structure of the retinal vessels. In traditional U-Net, the internal struc-ture of each convolution block consists of two base layers (3×3 conv + Relu). Furthermore, several U-Net variants consist of modified con-volution blocks such as in Genetic U-Net [23], DenseNet [6] and ResNet [4] convolution block. Inspired by the above work, we propose a residual method based on Res-PDC to replace the tradi-tional U-Net convolution method. The traditional 3×3 convolution layer is replaced by a $1 \times 1 + 3 \times 3 + 1 \times 1$ convolution layer, and pre-activation is added before the convolution, and dropout layer is added after the 3×3 convolution layer. Pre-activation is composed of batch normalization (BN) and rectified liner units (Relu). The purpose of the Pre-activation is to optimize the identity map, and BN pre-activation improves the regularization of the model. Dropout and BN in our Res-PDC convolution block are used together. Because dropout can effectively prevent overfitting problems in convolution networks. Also, the use of dropout slows down the network training speed, so BN is introduced to speed up the network. To avoid degradation problem that often affect the model prediction as the nujmber of layer increases, residual connections are introduced to form Res-PDC, as shown in Fig.1.

3.3. Triple attention mechanism

To extract structural information between vessel contexts, channel and spatial attention mechanisms are fully utilized to learn long-range feature dependencies. This paper proposes a TAM, including a Channel and Spatial Attention Module (CSAM) and Dual Convolution Block Attention Module (DCBAM). The output of the feature by the encoder is input to the decoder, and then to the CSAM module to generate channel-spatial attention aware representation features. The DCBAM module is used to multiply the attention map with the input feature map for adaptive feature optimization. Through the proposed TAM in this work. Important blood vessel information in the channel and space domains and large amount of information generated by continuous pooling and convolution operations of the feature map can be fully extracted.

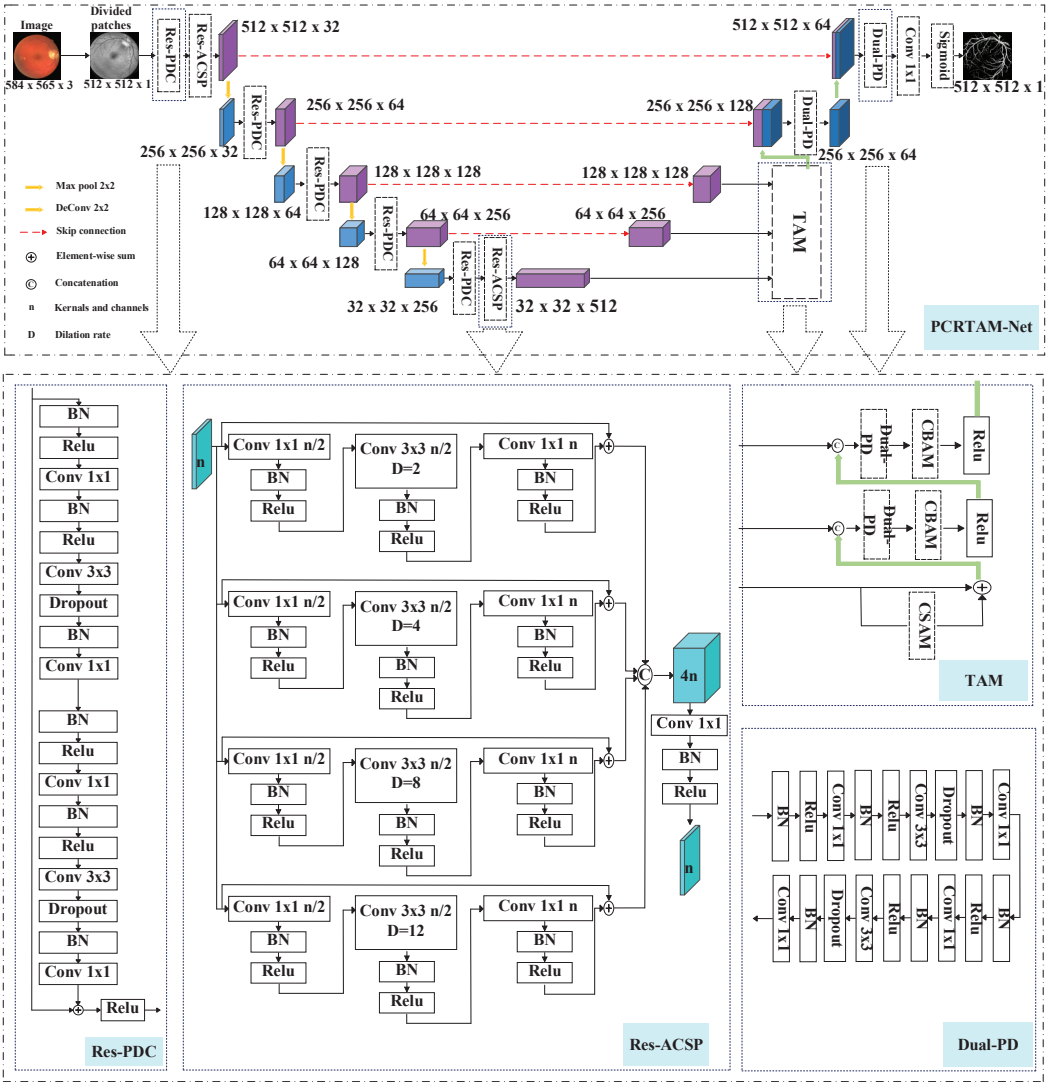

Figure 1. Overview architecture of proposed PCRTAM-Net.

3.3.1 Channel and spatial attention module

The Channel and Spatial Attention Module (CSAM) is composed of two parallel attention modules. The Spatial Attention Block (SAB) se-lectively aggregates the features of each space by weighting the features of all spatial locations, which enables the model to capture long-term features dependencies that are related to each other regardless of the distance. Meanwhile, the Channel Attention Block (CAB) enhances the contrast of features in different channels by using the full space domain for representation and normalization, which can lead to higher discrimina-tive power.

(1) Spatial attention block The difference between the SAB module and the position attention module in DANet [3] is that the latter operates directly on the original features, while the former operates on new features, and sums. There is one 3 × 1 convolution + BN + Relu and one 1 × 3 convolution + BN + Relu, which are used to extract the edge features of thevessel structure in the horizontal and vertical directions. The low-level information is fused by skip con-nections, and the lost spatial information is compensated. Overall, the input features $F \in R^{C\times H\times W}$ go through 3 × 1 and 1 × 3 convolution layers, to generate two new feature maps $E_y \in R^{C\times H\times W}$ and $M_x \in R^{C\times H\times W}$ where C represents the input feature dimension, H and W are the height and width of the input image, respectively, where E_y and M_x represent the features of the extracted vessel structures in the vertical and horizontal directions, respectively. The extracted new feature map is then reshaped,

where N is the number of features. So the features captured at E and K can be applied to transposed matrix multiplication to obtain the spatial correlation of features, as shown in Eq.1. Through SAB, the global context feature map is captured, and the context features can be aggregated according to the spatial attention map out-put by SAB, which improved the accuracy of the blood vessel segmentation process.

$$S_{(x,y)} = \frac{exp(E_y^T \cdot M_x)}{\sum_{x'=1}^{N} exp(E_y^T \cdot M_{x'})} \tag{1}$$

(2) Channel attention block The channel attention map is obtained by applying a softmax layer on the channel similarity map between the input features and their transposed features, as shown in Eq.2. Performing such operations on each pixel can enhance the contrast between class related features and help improve the expressiveness of features.

$$C_{(x,y)} = \frac{exp(F_x \cdot F_y^T)}{\sum_{x'=1}^{C} exp(F_{x'} \cdot F_y^T)} \tag{2}$$

3.3.2 Dual convolution block attention module

Dual Convolution Block Attention Module (DCBAM) is a simple yet effective attention module for feed forward convolution neural net-works. Given an intermediate feature map, the DCBAM module sequentially infers the attention map along two independent channel dimensions and space dimensions and then multiplies the inferred attention map with the input feature map for adaptive feature optimization. Because the DCBAM module is a lightweight general module, the computational cost of this module can be ignored and can be embedded behind the convolution layers of the decoder. The calculation process of both the channel and spatial attentions are shown in Eq.3-4.

$$\begin{aligned} M_{(c)}(F) &= \sigma(MLP(AvgPool(F)) \\ &\quad + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F^c_{avg})) + W_1(W_0(F^c_{max}))) \end{aligned} \tag{3}$$

$$\begin{aligned} M_{(s)}(F) &= \sigma(f^{7\times7}([AvgPool(F); MaxPool(F)])) \\ &= \sigma(f^{7\times7}([F^s_{avg}); F^s_{max})])) \end{aligned} \tag{4}$$

σ is the sigmoid operation, MLP is the shared fully connected layer, AvgPool is the global average pooling, and MaxPool is the maximum pooling, W_0 and W_1 are the respective parameters of the two convolutional layers in the two-layer convolutional network MLP, $f^{7\times7}$ represents a convolution kernel of size 7×7.

3.4. Multiscale vessel feature extraction meth-od based on residual atrous convolution spatial pyramid method

Atrous convolution helps to extract the multiscale features of the image. The process of atrous convolution operations is shown in Eq.5, where the input feature x and filter w, which generates output y, and r represents the dilation rate. Some recent studies have shown that residual atrous spatial pyramid pooling [18] combines residual connections[21] with atrous convolution, to improve information flow and extract multiscale features. However, residual atrous spatial pyramid pooling method uses the structure of 1 × 1 convolution, 3 × 3 convolution + BN + Relu and 1 × 1 convolution. The increase in the number of network layers slows the training process of the network. The main processes of multiscale feature extraction are shown in the Eq.6-10. Where x is the input feature, (Eq.6), (Eq.7) and (Eq.8) are the corresponding convolution process, y is the output, D_i is the dilation rate, and $i \in 2, 4, 8, 12$. The main processes are as follows.

(1) Input Res-PDC into RES-ACSP and di-vide into four branches, each of which includes a residual module for improving the information flow and a convolution operation for extracting the corresponding vessel features based on atrous convolution.

(2) In the convolution operation of each branch on the left side of the Res-ACSP, first add BN and Relu after 1 × 1 convolution of each con-volution nucleus n/2 to reduce the computational complexity in order to divide the blood vessels of different standards. The four different atrous convolution containing 2, 4, 8, and 12 are used to achieve multiscale characteristics of thick and thin blood vessels.

To reduce the loss of detailed information, the output of each branch of the residual module on the left side of the Res-ACSP is added to the output of the convolution operation.

$$y[i] = \sum_k x[i + r \cdot k] \cdot w[k] \tag{5}$$

$$C_1 = Relu(BN(Conv(1, \frac{n}{2})(x))) \tag{6}$$

$$C_2 = Relu(BN(Conv(3, \frac{n}{2}, D_i)[C_1])) \tag{7}$$

$$C_3 = Relu(BN(Conv(1, n)[C_2])) \tag{8}$$

$$y_i = C_3 + x \tag{9}$$

$$y = Rleu(BN(Conv(1, n)[cat[y_2, y_4, y_8, y_12]])) \tag{10}$$

3.5. Loss Function

In this work, we used the Binary Cross-Entropy (BCE) loss as the objective function for network training to di-

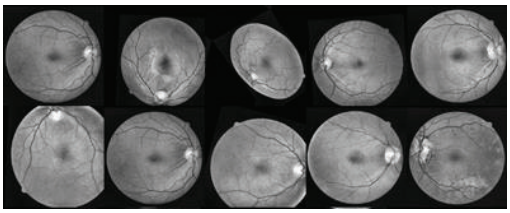

Figure 2. Grayscale image after data augmentation.

Dataset	Total	Train	Test	Resolution
DRIVE	40	20	20	565×584
CHASEDB1	28	20	8	999×960
DRIVE	20	19	1	700×605
DRIVE	45	30	15	3504×2336

Table 1. Overview of the experimental datasets.

rectly evaluate the distance between expert annotations and the pre-dictions. The BCE loss function is mathematical-ly express in Eq.11 as follows.

$$Loss_{(BCE)} = -\frac{1}{N}\sum_{i=1}^{N} g_i \cdot log(p_i) + (1-g_i) \cdot log(1-p_i) \tag{11}$$

where g is the ground truth, p is the model prediction, n is the total number of samples, and i is the i^{th} sample.

4. Electronic Supplementary Material

4.1. Datasets

We conducted experiments on DRIVE, CHASEDB1, STARE and HRF publicly available fundus image datasets. Table 1 summarizes the total number of images, training and test splits and image size (width × height) for the four publicly available datasets.

4.2. Data preprocessing

Due to the data insufficiency, it is necessary to increase the number of samples to prevent overfitting. Furthermore, the datasets consist of different image size, we set the patch size of DRIVE as 512×512, CHASEDB1 as 960×960 and STARE as 592×592. Each image is rotated at an interval of 10 degrees and then mirrored. Also, each image is moved randomly between 20 and 50 pixels to-wards each of the four corners. Finally, four cor-ners of each image are clipped. We also enhanced the contrast, brightness, chroma, and saturation of the image to reduce the interference of external noise factors. Through the above data enhancement method, the data capacity is increased, and the network generalization ability is enhanced, and the overfitting problem is prevented. The enhanced image is shown in Fig.2.

4.3. Evaluation metrics

We used the Accuracy (ACC), Sensitivity (SE), Specificity (SP), F1-score and Area under Curve (AUC) to evaluate the proposed PCRTAM-Net.

$$ACC = \frac{TP+TN}{TP+FP+TN+FN} \tag{12}$$

$$SE = \frac{TP}{TP+FN} \tag{13}$$

$$SP = \frac{TN}{TN+FP} \tag{14}$$

$$F1 = \frac{2 \cdot TP}{2 \cdot TP+FP+FN} \tag{15}$$

Where TP, FN, TN and FP represent true positive, false negative, true negative, and false positive, respectively. The AUC measures the segmentation performance based on recall and precision.

4.4. Experimental setup

The implementation of our proposed PCRTAM-Net is based on the PyTorch platform and trained on NVIDIA RTX 3090 GPU. We use the Adam algorithm with an initial learning rate of 1e-3 as the optimization method, where the learning rate is set to decay to zero for 600 epochs. In the experiments, the batch size is set to 4.

4.5. Ablation studies

To demonstrate the effectiveness of our proposed PCRTAM-Net, ablation experiments are performed to verify the effect of each component. The visual results and statistical comparison of different components are shown in Fig.3 and Table 2. In Fig.3, (a) original image, (b) ROI of original image (blue rectangle), (c) Ground truth image, and (d-o) represents the vessels visualization. The order is the same as that in Table 2. As described in Section 3, the TAM module consists of three components: one CSAM and two CBAMs. We further experiment and verify the effectiveness of attention in feature extraction. In this ablation experiments, we adopt a U-shaped network consisting of encoders and feature decoders of five residual blocks as the backbone of our proposed network.

4.5.1 Ablation study of Res-PDC module

We replace the traditional convolution block with the Res-PDC module (referred to as 'backbone + Res-PDC') and apply it to the DRIVE dataset. As shown in Fig.3, three typic examples of blood vessel segmentation results in fundus images are shown, indicating that our Res-PDC module can effectively segment various blood vessels and improve the performance of the backbone network. As shown in Table 2, compared with 'backbone', 'backbone +

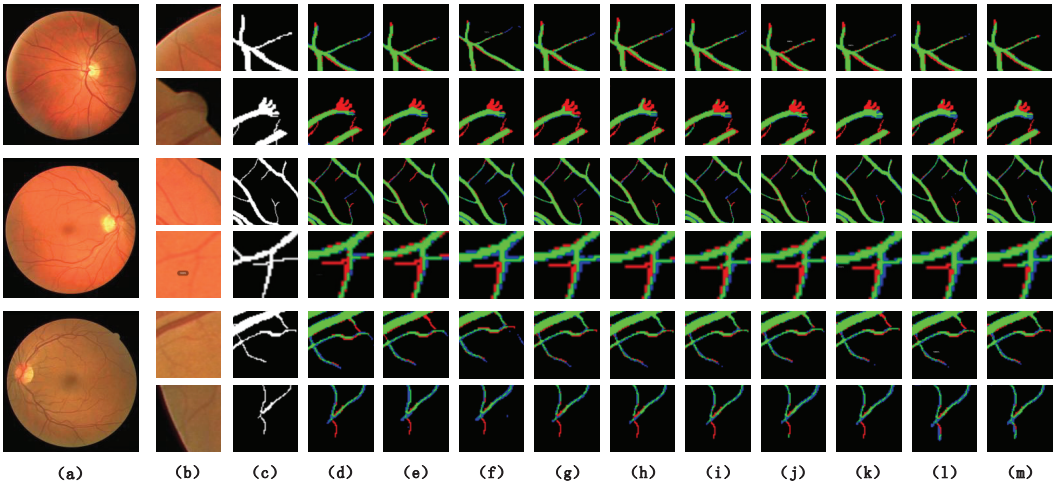

Figure 3. Visualized segmentation results for ablation experiments on DRIVE and CHASEDB1.

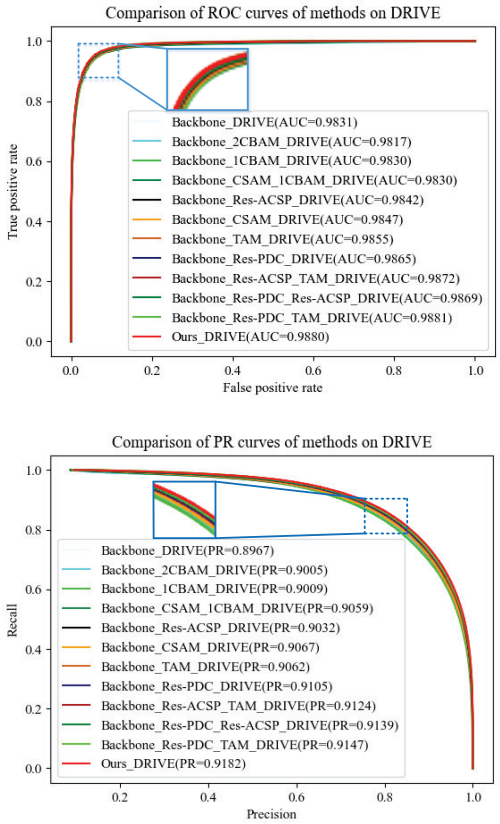

Figure 4. Comparison of ROC and PR curves for ablation experiments on DRIVE.

Res-PDC' improves ACC, SE, F1 and AUC from 96.81%, 78.73%, 81.17% and 98.31% by 97.00%, 77.67%, 81.85% and 98.65%. This indicates the importance of the Res-PDC convolution block as a key factor in improving the accuracy of vessel segmentation network.

4.5.2 Ablation study of TAM module

We investigated the effectiveness of the TAM module. As shown in Table 2, compared with 'backbone', 'backbone + TAM' improves ACC, SE, F1 and AUC to 96.99%, 78.32%, 81.90% and 98.55%, which indicates the necessity for multiple attention in feature extraction. Compared with 'backbone + TAM', it is observed that 'backbone + CSAM' performance reduces in ACC, SE, F1 to 96.93%, 78.60%, 81.66% and 98.47%, indicating that channel and spatial attention in the decoder are paired to extract features. Our experimental results demonstrated the importance of CBAM in the proposed TAM module.

4.5.3 Ablation study of Res-ACSP module

To extract multi-scale vessel information and improve information flow, the Res-ACSP module is also added to our network. As shown in Fig.3, it can be observed that our model obtained finer segmentation results due to the Res-ACSP module. It is higher than the backbone in terms of F1 by 0.85%, as shown in Table 2. From the visualization and statistics of the ablation experiments, we observed that our model Res-PDC, TAM and Res-ACSP, which proves that our model can handle this problem.

In addition, we combine the ROC curve and PR curve to further evaluate the ability of each component to improve the network, as shown in Fig.4. According to the ROC curve and PR curve, AUC and PR calculated from the DRIVE dataset, we observed that our PCRTAM-Net obtains the highest AUC and PR value.

Methods	ACC	SE	SP	F1	AUC
Backone	0.9681	0.7873	0.9858	0.8117	0.9831
Backbone + Res-PDC	0.9700	0.7767	**0.9888**	0.8185	0.9865
Backbone + Res-ACSP	0.9692	0.8065	0.9850	0.8202	0.9842
Backbone + CSAM	0.9693	0.7860	0.9870	0.8166	0.9847
Backbone + Single CBAM	0.9682	0.7795	0.9865	0.8103	0.9830
Backbone + Double CBAM	0.9687	0.7706	0.9879	0.8107	0.9817
Backbone + CSAM + Single CBAM	0.9694	0.7799	0.9878	0.8107	0.9830
Backbone + TAM	0.9699	0.7832	0.9880	0.8190	0.9855
Backbone + Res-PDC + Res-ACSP	0.9706	0.8030	0.9869	0.8261	0.9869
Backbone + Res-PDC + TAM	0.9706	0.8118	0.9860	0.8277	**0.9881**
Backbone + Res-ACSP + TAM	0.9703	0.7845	0.9883	0.8215	0.9872
Backbone + Res-PDC+ Res-ACSP + TAM	**0.9710**	**0.8158**	0.9861	**0.8305**	0.9880

Table 2. Performance comparison of ablation studies on DRIVE.

4.6. Performance comparison with state-of- the-art methods

We compare our method with several recently published state-of-the-art methods. Table 3 highlight the state-of-the-art methods and their performance on the DRIVE, CHASEDB1, and STARE datasets.

4.7. Performance comparison of four U-Net variants

Under the same experimental parameters and training methods, we run the publicly provided network codes of U-Net, CE-Net, DU-Net and PCRTAM-Net on DRIVE and CHASEDB1 datasets. We compare the four models on the ACC, SE, F1 and AUC metrics, and the results are shown in Table 4 and 5. As observed from the table, the performance of our model is optimal. The global accuracies of U-Net, CE-Net, DU-Net and PCRTAM-Net are 0.9665, 0.9679, 0.9681, 0.9710 on DRIVE and 0.9715, 0.9738, 0.9745, 0.9770 on CHASEDB1. More importantly, we evaluated the model using the ROC curve, as shown in Fig.6. The closer the ROC curve is to the upper left boundary in the ROC coordinates, the more accurate the model is. Considering the high imbalance problem, we also used the PR curve to evaluate the model, as shown in Fig.6. The closer the PR curve is to the upper right boundary in the PR coordinates, the better the performance of the model. These results show that among the four models, the PCRTAM-Net curve is the most complete, while the U-Net curve is the lowest. In addition, the results in Table 4 and 5 also show that PCRTAM-Net obtained the largest area under the ROC curve (AUC). To further observe the segmentation results of these four models, the probability maps of the blood vessel segmentation in fundus images are shown in Fig.8 and 9. From the figures, PCRTAM-Net produces better vessel segmentation results. The micro-vessels and occluded vessels that were lost in the U-Net, CE-Net and DU-Net were detected. The details of the segmentation results for the four models are shown in Fig.5, which shows a local magnified view of the vessel junction, where multiple vessels are intricately connected, and the micro-vessels of DRIVE and CHASEDB1. Due to the complexity of the vessel tree structure, it is difficult for the segmentation algorithm to accurately segment the complex structure. At the connections between blood vessels, U-Net, CE-Net and DU-Net extract rough blood vessel information due to the limitations of the network. For the tiny blood vessels, U-Net, CE-Net and DU-Net show limitations in processing details. However, PCRTAM-Net achieves satisfactory segmentation results at these tiny vessels. The experimental results show that among the four models, the PCRTAM-Net model has a more ideal performance in dealing with complex and micro-vessel structures.

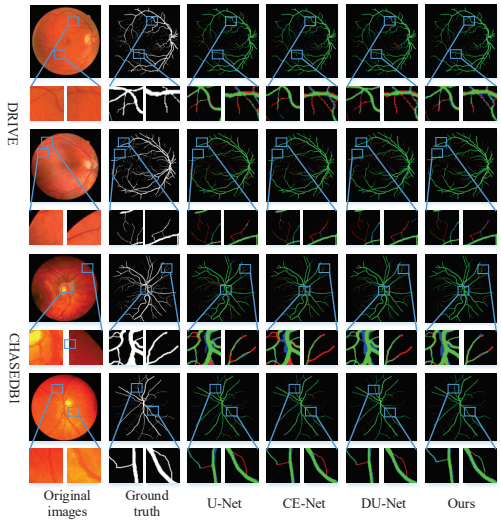

Figure 5. Comparison of visualization results on DRIVE and CHASEDB1.

Dataset	Methods	Year	ACC	SE	SP	F1	AUC
DRIVE	Li et al. [11]	2016	0.9527	0.7569	0.9816	-	0.9738
	FR-CRF [15]	2017	-	0.7897	0.9684	0.7857	-
	Wu et al. [26]	2018	0.9567	0.7844	0.9819	-	0.9807
	MPC-EM [21]	2019	0.9574	0.8083	0.9796	-	0.9822
	DUNet [8]	2019	0.9566	0.7963	0.9800	-	0.9802
	SID2Net [30]	2020	0.9520	-	-	0.8163	0.9754
	NFN+ [27]	2020	0.9582	0.7996	0.9813	-	0.9830
	DDNet [14]	2020	0.9607	0.8132	0.9783	-	-
	HAnet [22]	2020	0.9581	0.7991	0.9813	0.8293	0.9823
	CIEU-Net [20]	2021	0.9671	0.7933	-	0.8227	0.9778
	MD-Net [18]	2021	0.9676	0.8065	0.9826	-	-
	SCS-Net [24]	2021	0.9697	**0.8289**	0.9838	-	0.9837
	Khan et al. [9]	2022	0.9610	0.8125	0.9763	-	-
	CRAUNet. [2]	2022	0.9587	0.7954	-	0.8302	0.9830
	PCRTAM-Net	2022	**0.9710**	0.8158	**0.9860**	**0.8305**	**0.9880**
CHASEDB1	Li et al. [11]	2016	0.9581	0.7507	0.9793	-	0.9716
	FR-CRF [15]	2017	-	0.7277	0.9712	0.7332	-
	Wu et al. [26]	2018	0.9637	0.7538	0.9847	-	0.9825
	MPC-EM [21]	2019	0.9654	0.8138	0.9807	-	0.9850
	HAnet [22]	2020	0.9670	0.8239	0.9813	0.8191	0.9871
	CIEU-Net [20]	2021	0.9751	0.7988	-	0.8073	0.9688
	MD-Net [18]	2021	0.9731	0.7504	**0.9889**	-	-
	SCS-Net [24]	2021	0.9744	0.8365	0.9839	-	0.9867
	Khan et al. [9]	2022	0.9578	0.8012	0.9730	-	-
	CRAUNet. [2]	2022	0.9659	0.8259	-	0.8156	0.9864
	PCRTAM-Net	2022	**0.9770**	**0.8473**	0.9858	**0.8226**	**0.9914**
STARE	Li et al. [11]	2016	0.9628	0.7726	0.9844	-	0.9879
	FR-CRF [15]	2017	-	0.7680	0.9738	0.7644	-
	MPC-EM [21]	2019	0.9695	0.8162	0.9869	-	0.9898
	DUNet [8]	2019	0.9641	0.7595	**0.9878**	-	0.9832
	SID2Net [30]	2020	0.9620	-	-	0.8233	0.9824
	NFN+ [27]	2020	0.9672	0.7963	0.9863	-	0.9875
	DDNet [14]	2020	0.9698	0.8398	0.9761	-	-
	HAnet [22]	2020	0.9673	0.8186	0.9844	0.8379	0.9887
	CIEU-Net [20]	2021	0.9714	0.8273	-	-	0.8230
	MD-Net [18]	2021	0.9732	0.8295	0.9866	-	-
	SCS-Net [24]	2021	0.9736	0.8207	0.9839	-	0.9877
	Khan et al. [9]	2022	0.9586	0.8078	0.9721	-	-
	PCRTAM-Net	2022	**0.9768**	**0.8571**	0.9864	**0.8464**	**0.9905**

Table 3. Comparing the performance of the three types of mathematics.

Methods	ACC	SE	F1	AUC
U-Net	0.9665	0.7698	0.7999	0.9806
CE-Net	0.9679	0.7840	0.8098	0.9825
DU-Net	0.9681	0.7841	0.8107	0.9816
PCR-TAM-Net	**0.9710**	**0.8158**	**0.8305**	**0.9880**

Table 4. Comparison of performance visualization results on DRIVE.

Methods	ACC	SE	F1	AUC
U-Net	0.9715	0.7357	0.7644	0.9811
CE-Net	0.9738	0.7226	0.7762	0.9872
DU-Net	0.9752	0.8264	0.8073	0.9898
PCR-TAM-Net	**0.9770**	**0.8473**	**0.8226**	**0.9914**

Table 5. Comparison of performance visualization results on CHASEDB1.

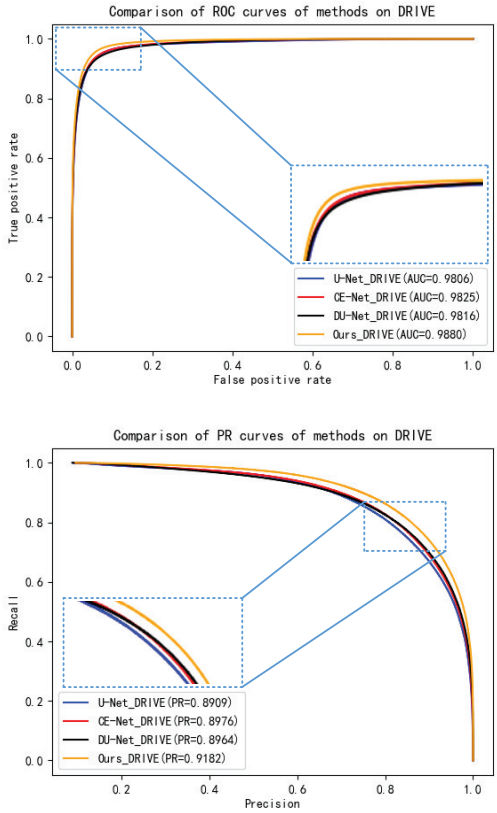

Figure 6. Comparison of ROC and PR curves of different methods on DRIVE.

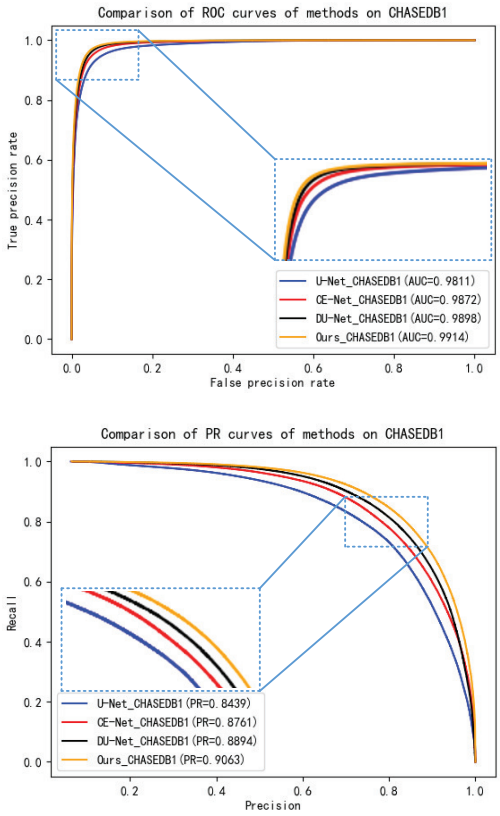

Figure 7. Comparison of ROC and PR curves of different methods on CHASEDB1.

Methods	ACC	SE	F1	AUC
FR-CRF [15]	-	0.7874	0.7158	-
MPC-EM [21]	0.9631	0.7782	-	0.9843
DUNet [8]	0.9651	0.7464	-	0.9831
HAnet [22]	0.9654	0.7803	0.8074	0.9837
SCS-Net [24]	0.9687	**0.8114**	-	0.9842
PCR-TAM-Net	**0.9714**	0.7981	**0.8116**	**0.9871**

Table 6. Performance comparison with other methods on HRF.

4.8. The performance of our method on high-resolution datasets

To validate the performance of our method on High-Resolution Fundus (HRF) images. We cropped the HRF dataset into patches of size 960 × 960. Table 6 summarizes the comparison of the proposed method with existing methods. For the division of training and test images, we take the same view as Soomro et al. [11]. The first ten images of each of the three images are used for training, and the rest are used for testing. As shown in Table 6, the overall performance of our method is higher than that of existing methods. Through the experimental results, we observed that our method achieves the best performance on the HRF

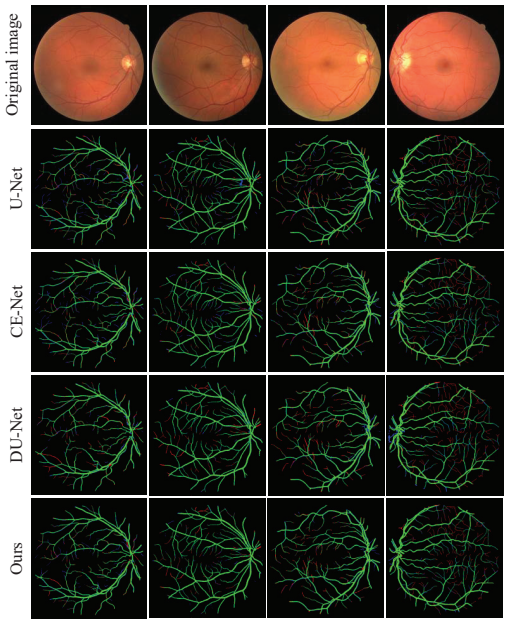

Figure 8. Comparison of probability maps on DRIVE.

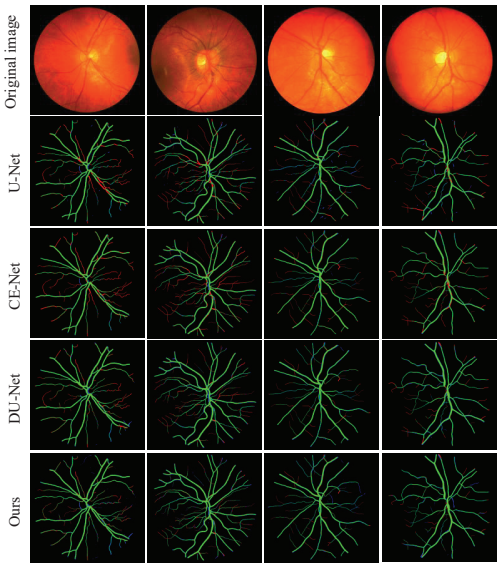

Figure 9. Comparison of probability maps on CHASEDB1.

dataset, and thus the PCRTAM-Net model is suitable for the segmentation of blood vessels in high-resolution fundus images. Fig.10 shows the segmentation results on HRF dataset.

4.9. Generalization ability verification based on cross-training evaluation

To further investigate the generalization ability of the proposed model, we performed a cross-training process on the STARE and DRIVE datasets. The cross-training refers to the evaluation method of testing pre-trained models on unseen datasets. Without finetuning, we applied the PCRTAM-Net model trained on one dataset to other datasets and evaluated it. For the convenience of training and testing, we crop to 512×512 patches. Table 7 and Table 8 present the performance of several existing methods and our PCRTAM-Net model. The experimental results better ACC, SE, F1 and AUC When tested on the DRIVE dataset. Because the STARE dataset mainly contains thick blood vessels and small blood vessels, when testing the DRIVE dataset, some blood vessels ruptured, making the overall slightly lower than the single training and testing strategy. Tested on the STARE dataset, our PCRTAM-Net model achieves a better ACC, F1 and AUC, and Sen is lower than the best performance. Since the DRIVE dataset mainly contains micro-vessels and the STARE is relatively complex, the test performance is slightly lower. Our proposed PCRTAM-Net model has a good generalization ability for blood vessel segmentation in fundus images. Fig.11 shows the segmentation results on DRIVE and STARE dataset.

Methods	ACC	SE	F1	AUC
Li et al. [11]	0.9486	0.7273	-	0.9677
MPC-EM [21]	0.9501	0.7652	-	0.9740
DUNet [8]	0.9481	0.6505	-	0.9781
SID2Net [30]	0.9499	-	0.8010	-
NFN+ [27]	0.9530	0.7187	-	0.9761
HAnet [22]	0.9530	0.7140	-	0.9758
PCR-TAM-Net	**0.9677**	**0.7981**	**0.8161**	**0.9839**

Table 7. Comparison of performance results for the cross- training on DRIVE.

Methods	ACC	SE	F1	AUC
Li et al. [11]	0.9545	0.7024	-	0.9671
MPC-EM [21]	0.9522	0.7447	-	0.9754
DUNet [8]	0.9445	**0.8419**	-	0.9690
SID2Net [30]	0.9569	-	0.7866	-
NFN+ [27]	0.9629	0.7704	-	0.9783
HAnet [22]	0.9543	0.8187	-	0.9648
PCR-TAM-Net	**0.9687**	0.8097	**0.7886**	**0.9804**

Table 8. Comparison of performance results for the cross-training on STARE.

Methods	U-Net	CE-Net	DU-Net	PCRTAM-Net
Params	3.35M	29.00M	0.06M	16.42M

Table 9. This article compares the complexity of the model with some other existing models.

4.10. Computation complexity

In this section, we will discuss and analyze the complexity of the model in detail. To make a fair comparison, and to rule out the impact of different platforms, we give the number of model parameters. We also select U-Net, CE-Net and DU-Net for comparison. As can be seen from Table 9, PCRTAM-Net has 16.42M parameters, which is not the highest parameter amount com-pared to other models. PCRTAM-Net has the highest performance scores in ACC, SE, F1 and AUC compared to other existing methods.

4.11. Limitations

On the one hand, when the contrast between blood vessels and background is extremely low, our model is difficult to fully segment blood vessels. On the other hand, for areas that are too noisy, our method may produce some false positives. In future work, we will further study and improve the model to accurately identify smooth fine blood vessels in the case of low contrast between blood vessels and background and too much noise.

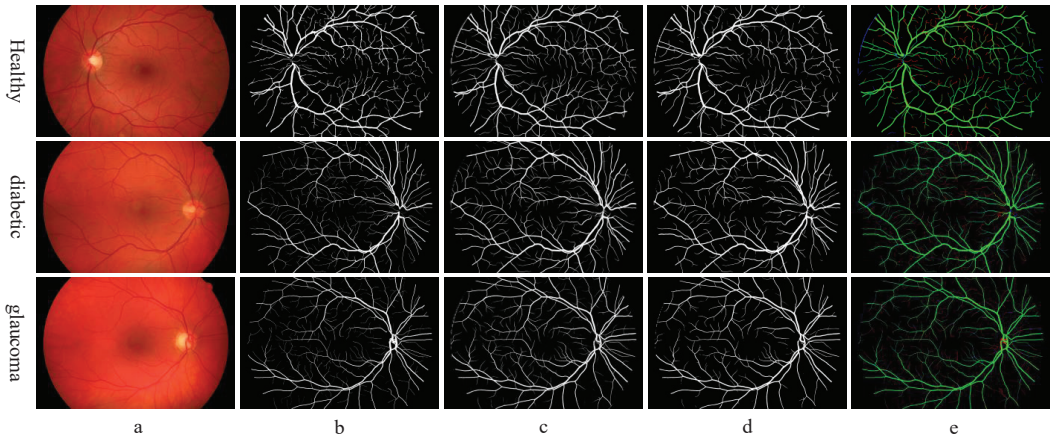

Figure 10. CProbability maps of three types of retinal images on HRF: (a) Image; (b) Ground Truth; (c) Probabil-ity map; (d) 0/1 map; (e) Difference map..

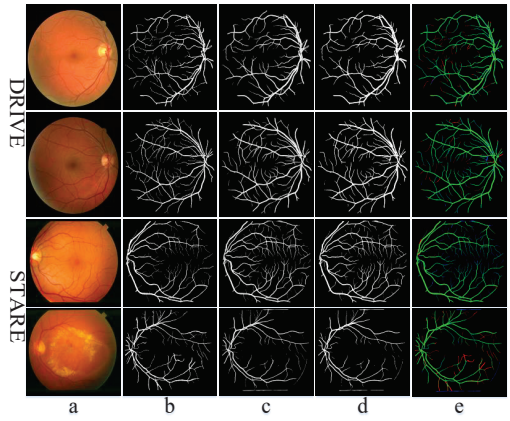

Figure 11. Visualization results on DRIVE and STARE: (a) Image; (b) Ground Truth; (c) Probability map; (d) 0/1 map; (e) Difference map.

5. Conclusions

This paper proposes a novel retinal vessel segmentation network termed PCRTAM-Net, which has addressed the low performance of retinal vessel segmentation. The PCRTAM-Net consists of three main parts, they are Res-PDC, Res-ACSP and TAM. In the encoder-decoder, we propose the Res-PDC method to extract more feature information by replacing the traditional convolution method. At both ends of the encoder, the Res-ACSP method is used to extract multi-scale information and improve information flow. In the decoder, the features are adaptively optimized by effectively utilizing channel and spatial attention through TAM. Comparative evaluations are performed on four public datasets namely DRIVE, CHASEDB1, STARE and HRF, which demonstrates that the proposed method outperforms the state-of-the-art methods. The proposed model has better generalization ability and is promising to be extended to other medical image segmentation tasks. Due to the complexity of the retinal vessel morphology and the influence of various interference factors, the segmentation of micro-vessels needs further study.

Acknowledgement

This work was partially supported by the Open Funds from Guilin University of Electronic Technology, Guangxi Key Laboratory of Image and Graphic Intelligent Processing (Grant No. GIIP2209), National Natural Science Foundation of China (Grant Nos. 62172120, 62002082), Guangxi Natural Science Foundation (Grant Nos.2019GXNSFAA245014,2020GXNSFBA238014).

References

[1] L. Dai, L. Wu, H. Li, C. Cai, Q. Wu, H. Kong, R. Liu, X. Wang, X. Hou, Y. Liu, et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nature communications*, 12(1):3242, 2021. 3

[2] F. Dong, D. Wu, C. Guo, S. Zhang, B. Yang, and X. Gong. Craunet: A cascaded residual attention u-net for retinal vessel segmentation. *Computers in Biology and Medicine*, 147:105651, 2022. 9

[3] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3146–3154, 2019. 3, 4

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[5] A. Hoover, V. Kouznetsova, and M. Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical imaging*, 19(3):203–210, 2000. 2

[6] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 3

[7] K. Huang and M. Yan. A region based algorithm for vessel detection in retinal images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006: 9th International Conference, Copenhagen, Denmark, October 1-6, 2006. Proceedings, Part I 9*, pages 645–653. Springer, 2006. 2

[8] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su. Dunet: A deformable network for retinal vessel segmentation. *Knowledge-Based Systems*, 178:149–162, 2019. 3, 9, 10, 11

[9] T. M. Khan, M. A. Khan, N. U. Rehman, K. Naveed, I. U. Afridi, S. S. Naqvi, and I. Raazak. Width-wise vessel bifurcation for improved retinal vessel segmentation. *Biomedical Signal Processing and Control*, 71:103169, 2022. 9

[10] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki. Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3656–3665, 2020. 1

[11] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang. A cross-modality learning approach for vessel segmentation in retinal images. *IEEE transactions on medical imaging*, 35(1):109–118, 2015. 9, 10, 11

[12] Z. Li, X. Zhang, H. Müller, and S. Zhang. Large-scale retrieval for medical image analytics: A comprehensive review. *Medical image analysis*, 43:66–84, 2018. 1

[13] P. Liskowski and K. Krawiec. Segmenting retinal blood vessels with deep neural networks. *IEEE transactions on medical imaging*, 35(11):2369–2380, 2016. 2

[14] L. Mou, L. Chen, J. Cheng, Z. Gu, Y. Zhao, and J. Liu. Dense dilated network with probability regularized walk for vessel detection. *IEEE transactions on medical imaging*, 39(5):1392–1403, 2019. 3, 9

[15] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE transactions on Biomedical Engineering*, 64(1):16–27, 2016. 2, 9, 10

[16] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 2

[17] B. Sheng, P. Li, S. Mo, H. Li, X. Hou, Q. Wu, J. Qin, R. Fang, and D. D. Feng. Retinal vessel segmentation using minimum spanning superpixel tree detector. *IEEE transactions on cybernetics*, 49(7):2707–2719, 2018. 2

[18] Z. Shi, T. Wang, Z. Huang, F. Xie, Z. Liu, B. Wang, and J. Xu. Md-net: a multi-scale dense network for retinal vessel segmentation. *Biomedical Signal Processing and Control*, 70:102977, 2021. 5, 9

[19] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509, 2004. 2

[20] M. Sun, K. Li, X. Qi, H. Dang, and G. Zhang. Contextual information enhanced convolutional neural networks for retinal vessel segmentation in color fundus images. *Journal of Visual Communication and Image Representation*, 77:103134, 2021. 3, 9

[21] P. Tang, Q. Liang, X. Yan, D. Zhang, G. Coppola, and W. Sun. Multi-proportion channel ensemble model for retinal vessel segmentation. *Computers in biology and medicine*, 111:103352, 2019. 9, 10, 11

[22] D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao. Hard attention net for automatic retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, 24(12):3384–3396, 2020. 3, 9, 10, 11

[23] J. Wei, G. Zhu, Z. Fan, J. Liu, Y. Rong, J. Mo, W. Li, and X. Chen. Genetic u-net: automatically designed deep networks for retinal vessel segmentation using a genetic algorithm. *IEEE Transactions on Medical Imaging*, 41(2):292–307, 2021. 3

[24] H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, and J. Qin. Scs-net: A scale and context sensitive network for retinal vessel segmentation. *Medical Image Analysis*, 70:102025, 2021. 9, 10

[25] T. Wu, L. Li, and J. Li. Mscan: Multi-scale channel attention for fundus retinal vessel segmentation. In *2020 IEEE 2nd International Conference on Power Data Science (ICPDS)*, pages 18–27. IEEE, 2020. 3

[26] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai. Multiscale network followed network model for retinal vessel segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*, pages 119–126. Springer, 2018. 9

[27] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai. Nfn+: A novel network followed network for retinal vessel segmentation. *Neural Networks*, 126:153–162, 2020. 9, 11

[28] Q. Yang, B. Ma, H. Cui, and J. Ma. Amf-net: Attention-aware multi-scale fusion network for retinal vessel segmentation. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3277–3280. IEEE, 2021. 3

[29] B. Yin, H. Li, B. Sheng, X. Hou, Y. Chen, W. Wu, P. Li, R. Shen, Y. Bao, and W. Jia. Vessel extraction from non-fluorescein fundus images using orientation-aware detector. *Medical image analysis*, 26(1):232–242, 2015. 3

[30] Z. Zhuo, J. Huang, K. Lu, D. Pan, and S. Feng. A size-invariant convolutional network with dense connectivity applied to retinal vessel segmentation measured by a unique index. *Computer methods and programs in biomedicine*, 196:105508, 2020. 9, 11