Wang HD, Li ZZ, Okuwobi IP *et al.* PCRTAM-Net: A novel pre-activated convolution residual and triple attention mechanism network for retinal vessel segmentation. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 38(3): 567-581 May 2023. DOI: 10.1007/s11390-023-3066-4

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

Hua-Deng Wang[1, 2] (汪华登), *Member, CCF*, Zi-Zheng Li[2] (李紫正), *Student Member, CCF*
Idowu Paul Okuwobi[3, *] (保　罗), Bing-Bing Li[4, *] (黎兵兵), Xi-Peng Pan[2] (潘细朋), *Member, CCF*
Zhen-Bing Liu[1, 2] (刘振丙), *Member, CCF*, Ru-Shi Lan[1, 2] (蓝如师), *Member, CCF*, and
Xiao-Nan Luo[1, 2] (罗笑南)

[1] *Guangxi Key Laboratory of Image and Graphic Intelligent Processing, Guilin 541004, China*

[2] *School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China*

[3] *School of Artificial Intelligence, Guilin University of Electronic Technology, Guilin 541004, China*

[4] *Department of Pathology, Ganzhou Municipal Hospital, Ganzhou 341000, China*

E-mail: whd@guet.edu.cn; 20032202024@mails.guet.edu.cn; paulokuwobi@guet.edu.cn; libingbing19932021@126.com
pxp201@guet.edu.cn; zbliu@guet.edu.cn; rslan@guet.edu.cn; luoxn@guet.edu.cn

Received December 30, 2022; accepted May 22, 2023.

Abstract Retinal images play an essential role in the 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. At the same time, automated 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 called Pre-Activated Convolution Residual and Triple Attention Mechanism Network (PCRTAM-Net). PCRTAM-Net uses the pre-activated dropout convolution residual method to improve the feature learning ability of the network. In addition, the 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 evaluate the proposed PCRTAM-Net on four publicly available datasets, DRIVE, CHASE_DB1, STARE, and HRF. Our model achieves state-of-the-art performance of 97.10%, 97.70%, 97.68%, and 97.14% for ACC and 83.05%, 82.26%, 84.64%, and 81.16% for $F1$, respectively.

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

1 Introduction

There are certain numbers of capillaries in the human retina, and their morphological changes are closely related to different ophthalmic diseases. These morphological changes are the associated symptoms of a variety of other cardio vessels diseases, such as diabetes, hypertension, and arteriosclerosis[1]. The retinal fundus image is an important tool for doctors to diagnose various ophthalmic diseases and other related diseases. However, due to the complex nature of the retinal blood vessels and the lack of quality imag-

Regular Paper

Special Section of CVM 2023

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

*Corresponding Author (Idowu Paul Okuwobi worked on the investigation and writing and provided computing power, and Bing-Bing Li was responsible for the data curation and methodology.)

ing of the blood vessels, the efficiency of manual diagnosis is relatively low, and subjective[2]. Therefore, automatic segmentation of retinal fundus blood vessels has good research significance and clinical value. However, retinal image segmentation is still a challenging task due to several constraints. Firstly, the nature and size of the blood vessels in retinal images vary greatly. For example, blood vessels in fundus images usually vary between 1 pixel and 20 pixels. Secondly, a retinal vessel tree has many closely connected tiny blood vessels, which are generally difficult to be separated from other non-vessel structures. Thirdly, factors such as noise during retinal image acquisition and exudates produced by lesions make the segmentation task a difficult process. Facing these challenges, researchers have put in a lot of effort[3] to overcome these issues. The initial research approach[4] is to segment retinal vessels using hand-crafted features. Traditional image processing using hand-crafted features techniques lacks the ability to represent the complex semantics of the retinal blood vessels. Consequently, in relatively large datasets with multiple complex scenarios, these techniques are liable to perform poorly. In recent years, deep convolution neural network (DNNs)[5] have achieved remarkable results in medical image segmentation tasks. These networks have become very popular and useful in medical image segmentation problems. U-Net[6] and its various variants have been improved from a full convolution network for semantic segmentation, to a network that further improves 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, they are insufficient for the fundus image segmentation challenge. Ideally, due to factors such as noise, low resolution, and poor contrast, 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 deep learning network (PCRTAM-Net) with pre-activated convolution residual and triple attention mechanism to segment blood vessels from retinal fundus images. Our network is based on an encoder-decoder structure and consists of three core modules for extracting boundary blood vessel features, multiscale information of blood vessels, and structural information between vascular contexts, respectively. The proposed PCRTAM-Net is validated on four publicly available retinal vessel datasets, and the results show that the proposed PCRTAM-Net performs 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, PCRTAM-Net, 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, 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 the 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 2 presents the overall approach to retinal vessel segmentation. Section 3 describes our network in detail. Section 4 discusses the experimental results of our network on publicly available datasets. Finally, conclusions are presented in Section 5.

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[7] and threshold segmentation[4]. These methods need modification for different situations to achieve better segmentation performance, and they are not robust enough for this task because the hand-crafted features are misled by lesion regions and low-contrast microvessels[8]. Compared with these conventional

methods, the deep learning-based methods[5] have better advantages in dealing with the specificity of retinal blood vessels. Sheng *et al.*[9] improved the detection of low-contrast and narrow blood vessels by using an efficient minimum generation super pixel 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.*[10] developed a new method for accurately extracting blood vessels in nonfluorescent 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.*[11] developed a deep learning system called DeepDR that can detect diabetic retinopathy at different stages. Wang *et al.*[12] proposed a hard attention network (HAnet) that consists of one encoder and three decoders, while introducing an attention mechanism to enhance the features of the blood vessels in hard regions. Sun *et al.*[13] proposed a network that integrates atrous convolution modules, enabling a larger receptive field and enhancing the thickness of blood vessels while improving the perception of details to a certain extent. Jin *et al.*[14] proposed a deformable convolution network (DUNet) to replace ordinary convolution for vessel segmentation. Mou *et al.*[15] 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.*[16] proposed an automatically designed genetic U-Net, which can achieve better retinal vessel segmentation and solve 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.*[17] 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.*[18] 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. It utilizes the location attention module, to capture the distance-spatial relationship of features to improve performance. Wu *et al.*[19] proposed a multiscale channel attention network based on the encoder-decoder structure, which extracts the multiscale structural information of the 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[20] for the segmentation of microvessels still need to be improved.

3 Our Method

3.1 PCRTAM-Net

The proposed PCRTAM-Net is an encoder-decoder network for retinal vessel segmentation, and the overall architecture is shown in Fig.1. The encoder of PCRTAM-Net consists of a residual method based on a pre-activated dropout convolution (Res-PDC) module, a residual atrous convolution spatial pyramid (Res-ACSP) module, and down-sampling. The decoder consists of dual pre-activation dropout convolution (Dual-PDC), a triple attention mechanism module (TAM), and upsampling. Finally, 1×1 convolution and sigmoid operation are applied to binarize the vessel probability map.

3.2 Residual Method Based on Pre-Activated Dropout Convolution

The underlying convolution block based on the encoder-decoder network plays an important role in segmenting the complex structure of the retinal vessels. In the traditional U-Net, the internal structure of each convolution block consists of two base layers (3×3 convolution + rectified linear units (Relu)[21]). Furthermore, several U-Net variants consist of modified convolution blocks such as in Genetic U-Net[16], DenseNet[22], and ResNet[23]. Inspired by the above work, we propose a residual method based on Res-PDC to replace the traditional 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, pre-activation is added before the convolution, and a dropout layer is added after the 3×3 convolution layer. The pre-activation is composed of batch normalization (BN)[24] and rectified linear units. The purpose of the pre-activation is to optimize the identity map, and the BN pre-activation improves the regularization of our model. Dropout and BN in our Res-PDC convolution block are used together. Dropout can effectively prevent overfitting in convolution net-

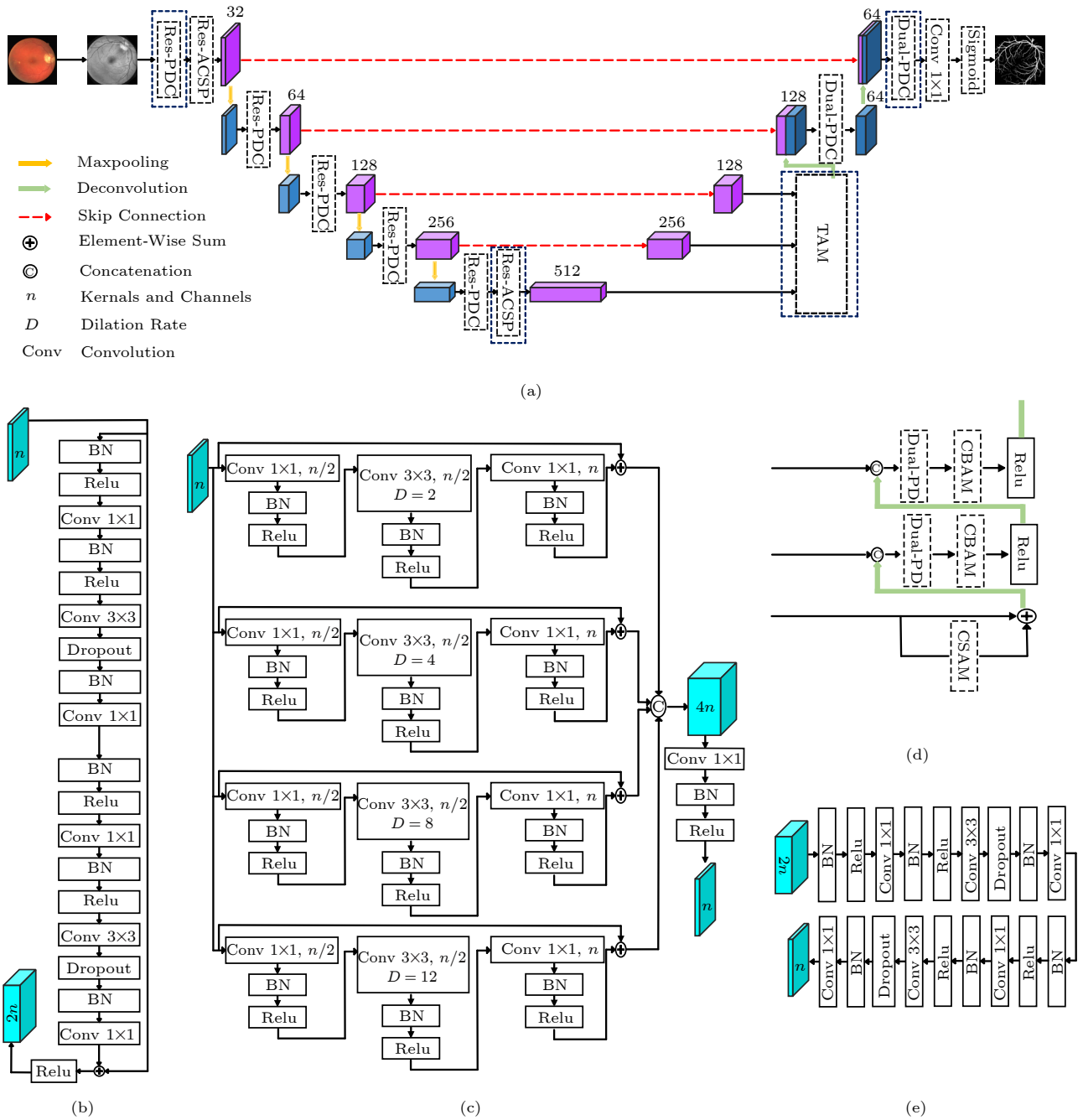

Fig.1. Architecture of our model. (a) PCRTAM-Net. (b) Res-PDC. (c) Res-ACSP. (d) TAM. (e) Dual-PDC.

works. Also, the use of dropout slows down the network training speed; therefore BN is introduced to speed up the network. To avoid the degradation problem that often affects the model prediction as the number of layers 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 a spatial attention module (CSAM)[25] and a dual convolution block attention module (DCBAM). The output of the feature by the encoder is an input to the decoder, and then to the CSAM module to generate channel-spatial attention aware representation features. The DCBAM 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 a large amount of information generated by continuous pooling and convolution operations of the feature map can be fully extracted.

3.3.1 *Channel and Spatial Attention Module*

The channel and spatial attention module is composed of two parallel attention modules. The spatial attention block (SAB) selectively aggregates the features of each space by weighting the features of all spatial locations, which enables the model to capture long-term feature dependencies that are related to each other regardless of the distance. Meanwhile, the channel attention block (CAB) enhances the contrast of the features in different channels by using the full space domain for representation and normalization, which can lead to higher discriminative power.

1) *Spatial Attention Block.* The difference between the spatial attention block module and the position attention module in DANet[17] is that the latter operates directly on the original features, while the former operates on the new transposed features and then sums. There are one 3×1 convolution + BN + Relu and one 1×3 convolution + BN + Relu, which are used to extract the edge features of the vessel structure in the horizontal and vertical directions respectively. The low-level information is fused by skip connections, and the lost spatial information is compensated. Overall, the input features $\boldsymbol{F} \in \mathbb{R}^{C\times H\times W}$ go through 3×1 and 1×3 convolution layers, to generate two new feature maps $\boldsymbol{E}_y \in \mathbb{R}^{C\times H\times W}$ and $\boldsymbol{M}_x \in \mathbb{R}^{C\times H\times W}$, where C represents the input feature dimension, H and W are the height and width of the input image, respectively, and $\boldsymbol{E}_y$ and $\boldsymbol{M}_x$ 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. Therefore, the features captured at $\boldsymbol{E}$ and $\boldsymbol{M}$ can be applied to transposed matrix multiplication to obtain the spatial correlation of features, as shown in (1). Through SAB, the global context feature map is captured, and the context features can be aggregated according to the spatial attention map outputted by SAB, which improves the accuracy of the blood vessel segmentation process.

$$S_{(x,y)} = \frac{\exp(\boldsymbol{E}_y^{\mathrm{T}} \times \boldsymbol{M}_x)}{\sum_{x'=1}^{N} \exp(\boldsymbol{E}_y^{\mathrm{T}} \times \boldsymbol{M}_{x'})}. \tag{1}$$

2) *Channel Attention Block.* The channel attention block is obtained by applying a softmax layer on the channel similarity map between the input features $\boldsymbol{F}$ (named as $\boldsymbol{F}_x \in \mathbb{R}^{C\times H\times W}$) and their transposed features (named as $\boldsymbol{F}^{\mathrm{T}} \in \mathbb{R}^{C\times H\times W}$), and the process is as shown in (2). T represents the transpose operator in (2). Performing such operations on each pixel can enhance the contrast between class-related features and help improve the expressiveness of features.

$$C_{(x,y)} = \frac{\exp(\boldsymbol{F}_x \times \boldsymbol{F}_y^{\mathrm{T}})}{\sum_{x'=1}^{N} \exp(\boldsymbol{F}_{x'} \times \boldsymbol{F}_y^{\mathrm{T}})}. \tag{2}$$

3.3.2 *Dual Convolution Block Attention Module*

The dual convolution block attention module (DCBAM) is a simple yet effective attention module for feed forward convolution neural networks. Given an intermediate feature map, DCBAM 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 DCBAM 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 processes of both the channel and the spatial attention is shown in (3) and (4) respectively.

$$M_{\mathrm{c}}(\boldsymbol{F}) = \sigma(MLP(AvgPool(\boldsymbol{F})) + MLP(MaxPool(\boldsymbol{F}))), \tag{3}$$

$$M_{\mathrm{s}}(\boldsymbol{F}) = \sigma(f^{7\times 7}((AvgPool(\boldsymbol{F}); MaxPool(\boldsymbol{F})))), \tag{4}$$

where σ is the sigmoid operation, MLP is the multilayer perceptron, $AvgPool$ is the global average pooling, $MaxPool$ is the maximum pooling, and $f^{7\times 7}$ represents a convolution kernel of size 7×7.

3.4 Multiscale Vessel Feature Extraction

Atrous convolution helps to extract the multiscale features of the image. The process of atrous convolution operations is shown in (5).

$$\boldsymbol{y}(i) = \sum\nolimits_k \boldsymbol{x}(i + r \times k) \times w(k), \tag{5}$$

where $\boldsymbol{x}$ is the input feature, w is the filter, $\boldsymbol{y}$ is the output feature, and r denotes the dilation rate.

Some recent studies have shown that residual atrous spatial pyramid pooling[26] combines residual connections[23] with atrous convolution, to improve in-

formation flow and extract multiscale features. However, a residual atrous spatial pyramid pooling method uses the structures 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. To expedite network training and optimize vascular information extraction, BN + Relu is incorporated after every convolution operation. Res-ACSP is composed of four branches, each consisting of a residual module aimed at enhancing the information flow and a convolution module responsible for extracting vessel features using atrous convolutions. The convolution module consists of three convolution operations. The primary objective of the first convolution operation in each branch is to reduce the feature dimension and decrease computational complexity. The second convolution operation utilizes dilated convolutions with different dilation rates to capture multi-scale features of thick and thin blood vessels. The third convolution operation is employed to restore the feature dimension. Finally, the outputs of the four branches are concatenated to obtain enriched vessel features. The main processes of multiscale feature extraction are shown in (6)–(10). Relu[21] is the rectified linear unit, which enables a network to efficiently acquire sparse representations. BN[24] refers to Batch Normalization, which is a linear transformation layer inserted between the convolutional layer and the activation function layer. *Conv* represents a convolutional operation.

$$\boldsymbol{C}_1 = Relu(BN(Conv\left(\frac{n}{2}\right)(\boldsymbol{x}))), \tag{6}$$

$$\boldsymbol{C}_2 = Relu(BN(Conv\left(\frac{n}{2}, D_i\right)(\boldsymbol{C}_1))), \tag{7}$$

$$\boldsymbol{C}_3 = Relu(BN(Conv(n)(\boldsymbol{C}_2))), \tag{8}$$

$$\boldsymbol{y}_j = \boldsymbol{C}_3 + \boldsymbol{x}, \tag{9}$$

$$\boldsymbol{y} = Relu(BN(Conv(n)(cat(\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \boldsymbol{y}_4)))), \tag{10}$$

where $\boldsymbol{x}$ represents the input feature, *cat* represents the feature concatenation, D_i represents the dilation rate, and $i \in \{2, 4, 8, 12\}$. $\boldsymbol{C}_1$, $\boldsymbol{C}_2$, and $\boldsymbol{C}_3$ correspond to the convolution process in each branch, $\boldsymbol{y}_j$ represents the output feature of each branch, and $j \in \{1, 2, 3, 4\}$, and $\boldsymbol{y}$ represents the total output feature of Res-ACSP.

3.5 Loss Function

In this work, we use the binary cross-entropy (BCE) loss as the objective function for network training to directly evaluate the distance between expert annotations and the predictions. The BCE loss function is mathematically expressed in (11).

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

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 Experiment and Analysis

4.1 Datasets

Experiments are conducted on four publicly available datasets, namely DRIVE[7], STARE[4], CHASE_DB1[27], and HRF[28], as detailed in Table 1. The DRIVE, CHASE_DB1, and STARE datasets contain annotations from two experts, with the first expert's annotations typically used. The HRF dataset, on the other hand, only includes annotations from a single expert. The DRIVE and CHASE_DB1 datasets consist of 40 and 28 images, respectively, with 20 images used for training and the rest for testing. The STARE dataset includes 20 images, and the leave-one-out method is employed to split the dataset. The HRF dataset is divided into three categories, healthy, diabetic retinopathy, and glaucoma, each containing 15 images. The first 10 images of each category are used for training, while the rest were used for testing.

Table 1. Descriptions of Datasets

Dataset	Quantity	Training-Testing Split	Resolution
DRIVE[7]	40	20/20	565 × 584
CHASE_DB1[27]	28	20/8	999 × 960
STARE[4]	20	Leave one out	700 × 605
HRF[28]	45	30/15	3 504 × 2 336

4.2 Data Preprocessing

Due to data insufficiency, it is necessary to increase the number of samples to prevent over-fitting. Furthermore, the datasets consist of different image sizes, and we set the patch sizes of DRIVE 512 × 512, CHASE_DB1 960 × 960, and STARE 592 × 592.

Each image is rotated at an interval of 10 degrees and then mirrored. Also, each image is moved randomly between 20 pixels and 50 pixels towards each of the four corners. Finally, four corners of each im-

age are clipped. We also enhance 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, the network generalization ability is enhanced, and the over-fitting problem is prevented. The enhanced image is shown in Fig.2.

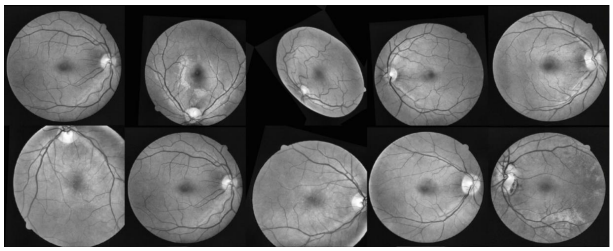

Fig.2. Grayscale image after data augmentation.

4.3 Evaluation Metrics

We use the accuracy (ACC), sensitivity (SE), specificity (SP), $F1$-score ($F1$) and area under curve (AUC) to evaluate the proposed PCRTAM-Net.

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

$$SE = \frac{TP}{TP + FN},$$

$$SP = \frac{TN}{TN + FP},$$

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

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. For the calculation of metrics, the area of interest is the whole image including the fundus area and the black background.

4.4 Experimental Setup

The implementation of our proposed PCRTAM-Net is based on the PyTorch platform and trained on NVIDIA RTX3090 GPU. We use the Adam algorithm with an initial learning rate of 0.000 1 as the optimization method, and the learning rate decays to 0 after 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 effectiveness of each component. The visual results and statistical comparison of different components are shown in Fig.3 and Table 2. The order is the same with that in Table 2. As described in Section 3, TAM consists of three components: one CSAM[25] and two CBAMs[29]. We further experiment and verify the effectiveness of the attention mechanism in the feature extraction process. In this ablation experiment, 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 "backbone + Res-PDC" to the DRIVE dataset. As shown in Fig.3, three samples of blood vessel segmentation results indicate 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 + Res-PDC" improves ACC, SE, $F1$, and AUC from 96.81%, 78.73%, 81.17%, and 98.31% to 97.00%, 77.67%, 81.85%, and 98.65%, respectively. This indicates the importance of the Res-PDC convolution block as a key factor in improving the accuracy of the vessel segmentation network.

4.5.2 *Ablation Study of TAM Module*

We investigate the effectiveness of TAM. 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% respectively, indicating that the triple attention mechanism is effective. Compared with "backbone+TAM", "backbone +CSAM" has lower ACC, SE, and $F1$, indicating that the channel and spatial attention in the decoder can effectively extract features. Our experimental results demonstrate the importance of CBAM in the proposed TAM module.

4.5.3 *Ablation Study of Res-ACSP Module*

To extract multiscale vessel information and im-

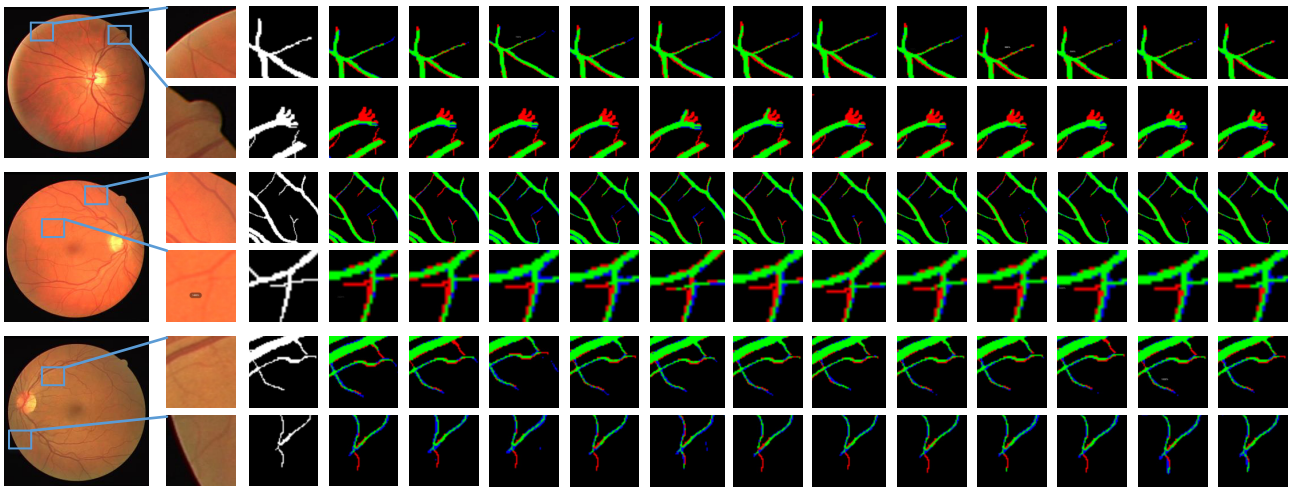

Fig.3. Visualized segmentation results for ablation experiments on DRIVE[7]. (a) Original image. (b) Magnified image. (c) Ground truth. (d) Backbone. (e) Backbone + Res-PDC. (f) Backbone + Res-ACSP. (g) Backbone + CSAM. (h) Backbone + Single CBAM. (i) Backbone + Double CBAM. (j) Backbone + CSAM + Single CBAM. (k) Backbone + TAM. (l) Backbone + Res-PDC + Res-ACSP. (m) Backbone + Res-PDC + TAM. (n) Backbone + Res-ACSP + TAM. (o) PCRTAM-Net.

Table 2. Performance Comparison of Ablation Studies on DRIVE[7]

Method	ACC	SE	SP	$F1$	AUC
Backbone	0.968 1	0.787 3	0.985 8	0.811 7	0.983 1
Backbone + Res-PDC	0.970 0	0.776 7	0.988 8	0.818 5	0.986 5
Backbone + Res-ACSP	0.969 2	0.806 5	0.985 0	0.820 2	0.984 2
Backbone + CSAM[25]	0.969 3	0.786 0	0.987 0	0.816 6	0.984 7
Backbone + Single CBAM[29]	0.968 2	0.779 5	0.986 5	0.810 3	0.983 0
Backbone + Double CBAM[29]	0.968 7	0.770 6	0.987 9	0.810 7	0.981 7
Backbone + CSAM[25] + Single CBAM[29]	0.969 4	0.779 9	0.987 8	0.810 7	0.983 0
Backbone + TAM	0.969 9	0.783 2	0.988 0	0.819 0	0.985 5
Backbone + Res-PDC + Res-ACSP	0.970 6	0.803 0	0.986 9	0.826 1	0.986 9
Backbone + Res-PDC + TAM	0.970 6	0.811 8	0.986 0	0.827 7	0.988 1
Backbone + Res-ACSP + TAM	0.970 3	0.784 5	0.988 3	0.821 5	0.987 2
Backbone + Res-PDC+ Res-ACSP + TAM	0.971 0	0.815 8	0.986 1	0.830 5	0.988 0

prove information flow, the Res-ACSP module is also added to our network. As shown in Fig.3, it can be observed that our model obtains finer segmentation results due to the Res-ACSP module. As shown in Table 2, when Res-ACSP is added to backbone, the $F1$ value of the model increases by 0.85%. From the visualization results and statistical analysis of the ablation experiment, it can be seen that the three modules proposed by us, Res-PDC, TAM and Res-ACSP, can better segment the microvessels. In addition, we combine the ROC curve to further evaluate the ability of each component to improve the network, as shown in Fig.4.

4.6 Performance Comparison with Other Methods

We compare our method with several recently published state-of-the-art methods. The overview of the state-of-the-art methods and their corresponding performance metrics on DRIVE, CHASE_DB1, and STARE datasets is shown in Table 3.

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, DUNet[14], and PCR-TAM-Net on the DRIVE and CHASE_DB1 datasets. We compare the four models on the ACC, SE, $F1$, and AUC metrics, and the results are shown in Table 4 and Table 5. More importantly, we evaluate the model using the ROC curve, as shown in Fig.5. The closer the ROC curve is to the upper left boundary, the more accurate the model is.

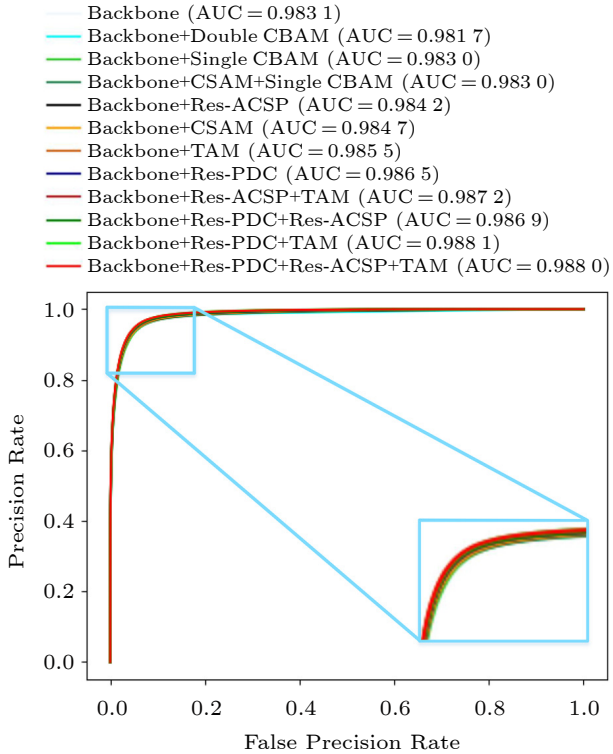

Fig.4. Comparison of ROC curve for ablation experiments on DRIVE.

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 Table 5 also show that PCRTAM-Net obtains 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.6 and Fig.7, where PCRTAM-Net produces better vessel segmentation results. The microvessels and occluded vessels lost in the U-Net, CE-Net, and DUNet are detected.

Fig.8 shows a locally enlarged view of the same location of the visual segmentation results for the four models, from which multiple micorvessels can be seen connected to each other in an intricate manner. Due to the complex nature of the blood vessels, U-Net, CE-Net, and DUNet can only extract rough blood vessel information due to the limitations of the networks. For the tiny blood vessels, U-Net, CE-Net, and DUNet 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, PCRTAM-Net segments better the vessel tree structure which is difficult for the other segmentation algorithms to accurately segment the complex structure.

4.8 Performance of Our Method on High-Resolution Datasets

To validate the performance of our method on High-Resolution Fundus (HRF) images, we crop the HRF[28] dataset into patches of size 960 × 960. Table 6 summarizes the comparison results of the proposed method and the methods on HRF. For the division of training and testing images, we take the same view as Soomro *et al.*[37]. As shown in Table 6, the overall per-

Table 3. Performance Comparison on Three Datasets

Method	DRIVE[7]				CHASE_DB1[27]				STARE[4]			
	ACC	SE	*F*1	AUC	ACC	SE	*F*1	AUC	ACC	SE	*F*1	AUC
FR-CRF[8]	-	0.789 7	0.785 7	-	-	0.727 7	0.733 2	-	-	0.768 0	0.764 4	-
HAnet[12]	0.958 1	0.799 1	0.829 3	0.982 3	0.967 0	0.823 9	0.819 1	0.987 1	0.967 3	0.818 6	0.837 9	0.988 7
CIEU-Net[13]	0.967 1	0.793 3	0.822 7	0.977 8	0.975 1	0.798 8	0.807 3	0.968 8	0.971 4	0.827 3	-	0.823 0
DUNet[14]	0.956 6	0.796 3	-	0.980 2	-	-	-	-	0.964 1	0.759 5	-	0.983 2
DDNet[15]	0.960 7	0.813 2	-	-	-	-	-	-	0.969 8	0.839 8	-	-
MD-Net[26]	0.967 6	0.806 5	-	-	0.973 1	0.750 4	-	-	0.973 2	0.829 0	-	-
SCS-Net[20]	0.969 7	0.828 9	-	0.983 7	0.974 4	0.836 5	-	0.986 7	0.973 6	0.820 7	-	0.987 7
SID2Net[30]	0.952 0	-	0.816 3	0.975 4	-	-	-	-	0.962 0	-	0.823 3	0.982 4
NFN+[31]	0.958 2	0.799 6	-	0.983 0	-	-	-	-	0.967 2	0.796 3	-	0.987 5
Khan *et al.*[32]	0.961 0	0.812 5	-	-	0.957 8	0.801 2	-	-	0.958 6	0.807 8	-	-
Li *et al.*[33]	0.952 7	0.756 9	-	0.973 8	0.958 1	0.750 7	-	0.971 6	0.962 8	0.772 6	-	0.987 9
Wu *et al.*[34]	0.956 7	0.784 4	-	0.980 7	0.963 7	0.753 8	-	0.982 5	-	-	-	-
MPC-EM[35]	0.957 4	0.808 3	-	0.982 2	0.965 4	0.813 8	-	0.985 0	0.969 5	0.816 2	-	0.989 8
CRAUNet[36]	0.958 6	0.795 4	0.830 2	0.983 0	0.965 9	0.825 9	0.815 6	0.986 4	-	-	-	-
PCRTAM-Net	0.971 0	0.815 8	0.830 5	0.988 0	0.977 0	0.847 3	0.822 6	0.991 4	0.976 8	0.857 1	0.846 4	0.990 5

Table 4. Comparison of Performance on DRIVE[7]

Method	ACC	SE	$F1$	AUC
U-Net	0.966 5	0.769 8	0.799 9	0.980 6
CE-Net	0.967 9	0.784 0	0.809 8	0.982 5
DUNet	0.968 1	0.784 1	0.810 7	0.981 6
PCRTAM-Net	0.971 0	0.815 8	0.830 5	0.988 0

Table 5. Comparison of Performance on CHASE_DB1[27]

Method	ACC	SE	$F1$	AUC
U-Net	0.971 5	0.735 7	0.764 4	0.981 1
CE-Net	0.973 8	0.722 6	0.776 2	0.987 2
DUNet	0.975 2	0.826 4	0.807 3	0.989 8
PCRTAM-Net	0.977 0	0.847 3	0.822 6	0.991 4

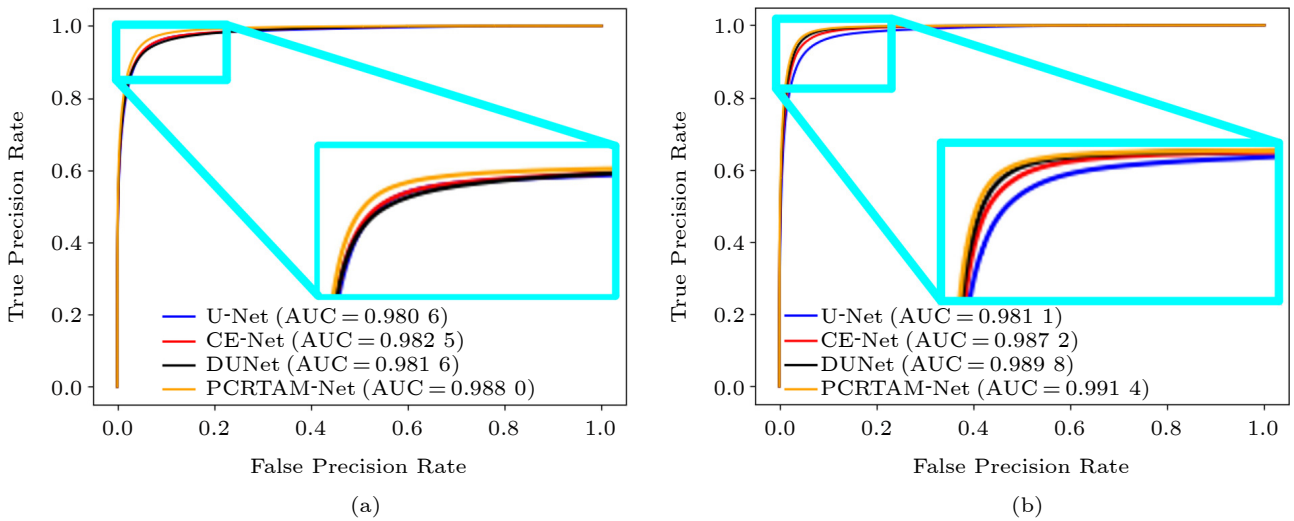

Fig.5. ROC curves of different models. (a) DRIVE dataset. (b) CHASE_DB1 dataset.

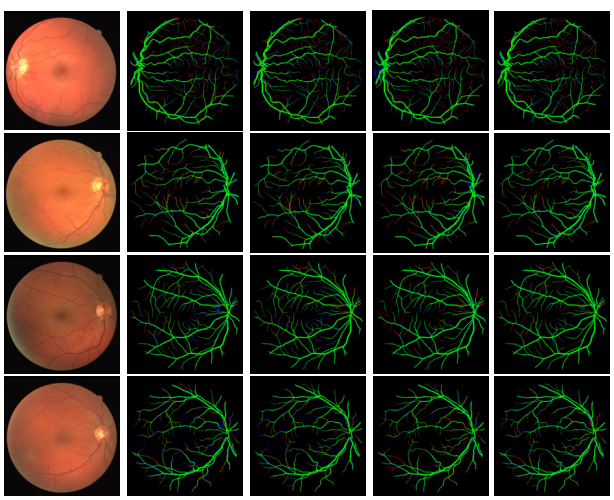

Fig.6. Comparison of probability maps on DRIVE. (a) Original image. (b) U-Net. (c) CE-Net. (d) DUNet. (e) PCRTAM-Net.

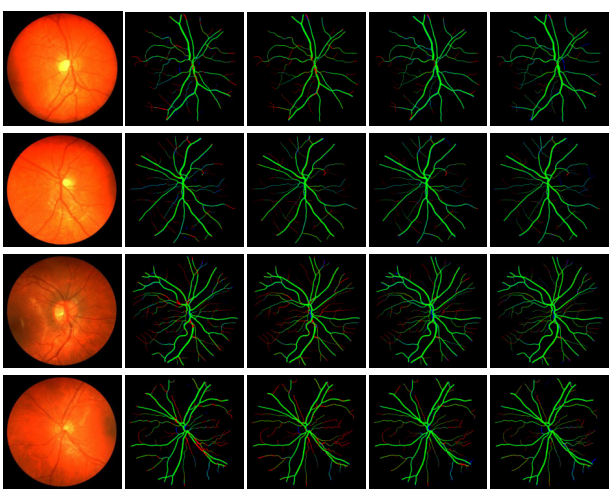

Fig.7. Comparison of probability maps on CHASE_DB1. (a) Original image. (b) U-Net. (c) CE-Net. (d) DUNet. (e) PCRTAM-Net.

formance of our method is higher than that of existing methods. Through the experimental results, we observe that our method achieves the best performance on the HRF dataset, and thus PCRTAM-Net is suitable for the segmentation of blood vessels in high-resolution fundus images. Fig.9 shows the segmentation results on the HRF dataset.

4.9 Generalization Ability Verification Based on Cross-Training Evaluation

To further investigate the generalization ability of the proposed model, we perform a cross-training process on the STARE and DRIVE datasets. We use the cross-training process, which refers to the evaluation

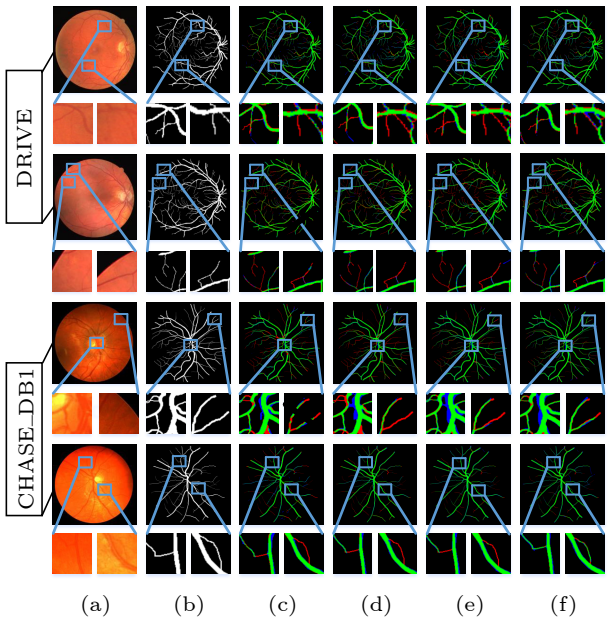

Fig.8. Comparison of visualization results on DRIVE and CHASE_DB1. (a) Original image. (b) Ground truth. (c) U-Net. (d) CE-Net. (e) DUNet. (f) PCRTAM-Net

Table 6. Performance Comparison with Other Methods on HRF[28]

Method	ACC	SE	$F1$	AUC
FR-CRF[8]	-	0.787 4	0.715 8	-
HAnet[12]	0.965 4	0.780 3	0.807 4	0.983 7
DUNet[14]	0.965 1	0.746 4	-	0.983 1
SCS-Net[20]	0.968 7	0.811 4	-	0.984 2
MPC-EM[35]	0.963 1	0.778 2	-	0.984 3
PCRTAM-Net	0.971 4	0.798 1	0.811 6	0.987 1

method of testing pre-trained models on unseen datasets. Without fine-tuning, we apply the PCRTAM-Net trained on one dataset to the other datasets and evaluated it. To facilitate the training and testing operation, the original images are cropped to a dimension of 512 × 512 pixels. Table 7 and Table 8 present the performance of several existing methods and our PCRTAM-Net. Our PCRTAM-Net achieves better ACC, SE, $F1$, and AUC metrics compared with other models on the testing DRIVE dataset. Since our model is trained on the STARE dataset, there are more thick blood vessels in the STARE dataset, while there were more microvessels in the DRIVE dataset, which lead to the phenomenon of blood vessel rupture when our model is tested on the DRIVE dataset. Our PCRTAM-Net achieves better ACC, $F1$, and AUC metrics and also reaches suboptimal performance in the testing on the dataset. Since the DRIVE dataset mainly contains microvessels and the STARE is relatively complex, the test performance is slightly lower. Our proposed PCRTAM-Net has a good generalization ability for blood vessel segmentation in fundus images. Fig.10 shows the segmentation results on the DRIVE and STARE datasets.

4.10 Computation Complexity

In order to make a fair comparison and exclude the influence of different platforms, we compare the number of model parameters, as shown in Table 9.

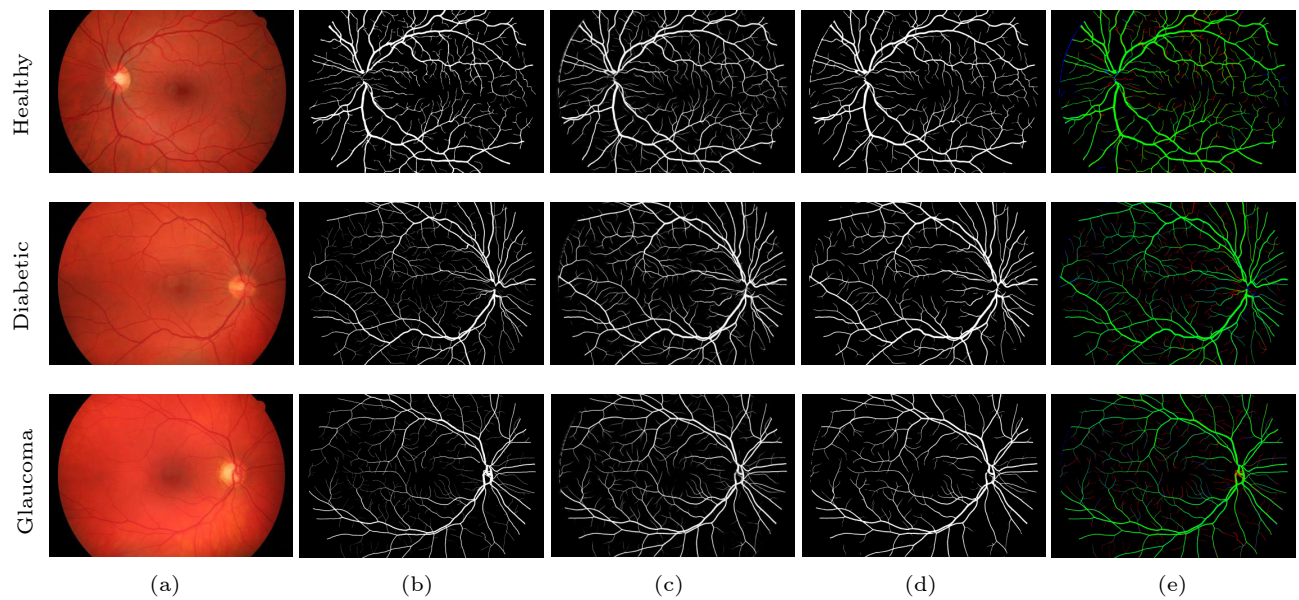

Fig.9. Probability maps of three types of retinal images on HRF. (a) Image. (b) Ground truth. (c) Probability map. (d) 0/1 map. (e) Difference map.

Table 7. Comparison of Performance Results for the Cross-Training on DRIVE[7]

Method	ACC	SE	$F1$	AUC
HAnet[12]	0.953 0	0.714 0	-	0.975 8
DUNet[14]	0.948 1	0.650 5	-	0.978 1
SID2Net[30]	0.949 9	-	0.801 0	-
NFN+[31]	0.953 8	0.718 7	-	0.976 1
Li *et al.*[33]	0.948 6	0.727 3	-	0.967 7
MPC-EM[35]	0.950 1	0.765 2	-	0.974 0
PCRTAM-Net	0.967 7	0.823 8	0.816 1	0.983 9

Table 8. Comparison of Performance Results for the Cross-Training on STARE[4]

Method	ACC	SE	$F1$	AUC
HAnet[12]	0.954 3	0.818 7	-	0.964 8
DUNet[14]	0.944 5	0.841 9	-	0.969 0
SID2Net[30]	0.956 9	-	0.786 6	-
NFN+[31]	0.962 9	0.770 4	-	0.978 3
Li *et al.*[33]	0.954 5	0.702 4	-	0.967 1
MPC-EM[35]	0.952 2	0.744 7	-	0.975 4
PCRTAM-Net	0.968 7	0.809 7	0.788 6	0.980 4

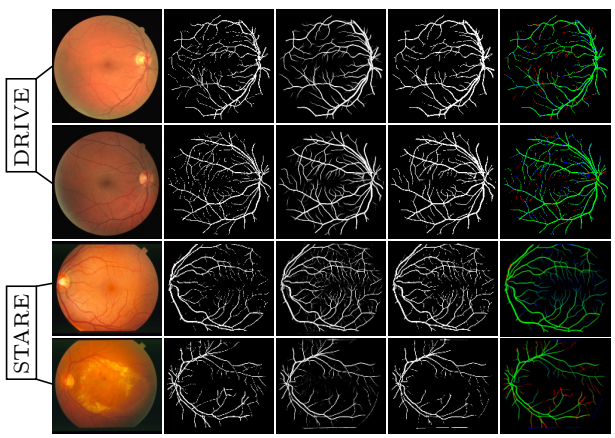

Fig.10. Visualization results on DRIVE and STARE. (a) Original image. (b) Ground truth. (c) Probability map. (d) 0/1 map. (e) Difference map.

Table 9. Comparison of Computational Complexity with Other Models

Method	Number of Model Parameters ($\times 10^6$)
U-Net	3.35
CE-Net	29.00
DUNet	0.06
PCRTAM-Net	16.42

We select U-Net, CE-Net, and DUNet for comparison. As it can be seen from Table 9, PCRTAM-Net has 16.42×10^6 parameters, and is 12.58×10^6 fewer than CE-Net. PCRTAM-Net has the highest performance scores in ACC, SE, $F1$, and AUC compared with other existing methods.

4.11 Limitations

When the contrast between blood vessels and background is extremely low, our model finds it difficult to segment blood vessels fully. In addition, for areas that are too noisy, our method may produce some false positives. In future work, we will focus on further studying and enhancing the model to achieve accurate identification of microvessels, particularly in challenging scenarios with low contrast between blood vessels and the background, as well as severe noise interference.

5 Conclusions

This paper proposed a novel retinal vessel segmentation network termed PCRTAM-Net, which addresses the low performance of retinal vessel segmentation. PCRTAM-Net consists of three main parts, Res-PDC, Res-ACSP, and TAM. The Res-PDC method extracts more feature information by replacing the traditional convolution method. At the both ends of the encoder, the Res-ACSP method is used to extract multiscale information and improve information flow. In the decoder, the features are adaptively optimized by effectively utilizing the channel and spatial attention through TAM. Our proposed method achieved an improvement in the ACC index of 0.13%, 0.19%, 0.32%, and 0.27% on the DRIVE, CHASE_DB1, STARE, and HRF datasets respectively, compared with the optimal method[20]. The proposed model requires further research in terms of its performance in microvessel segmentation and model lightweight.

References

[1] Li L Z, Verma M, Nakashima Y, Nagahara H, Kawasaki R. IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks. In *Proc. the 2020 IEEE Winter Conference on Applications of Computer Vision*, Mar. 2020, pp.3645–3654. DOI: 10.1109/WACV45572.2020.9093621.

[2] Li Z Y, Zhang X F, Muller H, Zhang S T. Large-scale retrieval for medical image analytics: A comprehensive review. *Medical Image Analysis*, 2018, 43: 66–84. DOI: 10.1016/j.media.2017.09.007.

[3] Huang K, Yan M. A region based algorithm for vessel detection in retinal images. In *Proc. the 9th International*

Conference on Medical Image Computing and Computer-Assisted Intervention, Oct. 2006, pp.645–653. DOI: 10.1007/11866565_79.

[4] Hoover A D, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Medical Imaging*, 2002, 19(3): 203–210. DOI: 10.1109/42.845178.

[5] Liskowski P, Krawiec K. Segmenting retinal blood vessels with deep neural networks. *IEEE Trans. Medical Imaging*, 2016, 35(11): 2369–2380. DOI: 10.1109/TMI.2016.2546227.

[6] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In *Proc. the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, Oct. 2015, pp.234–241. DOI: 10.1007/978-3-319-24574-4_28.

[7] Staal J, Abramoff M D, Niemeijer M, Viergever M A, Van Ginneken B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Medical Imaging*, 2004, 23(4): 501–509. DOI: 10.1109/TMI.2004.825627.

[8] Orlando J I, Prokofyeva E, Blaschko M B. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE Trans. Biomedical Engineering*, 2017, 64(1): 16–27. DOI: 10.1109/TBME.2016.2535311.

[9] Sheng B, Li P, Mo S J, Li H T, Hou X H, Wu Q, Qin J, Fang R G, Feng D D. Retinal vessel segmentation using minimum spanning superpixel tree detector. *IEEE Trans. Cybernetics*, 2018, 49(7): 2707–2719. DOI: 10.1109/TCYB.2018.2833963.

[10] Yin B J, Li H T, Sheng B, Hou X H, Chen Y, Wu W, Li P, Shen R M, Bao Y Q, Jia W P. Vessel extraction from non-fluorescein fundus images using orientation-aware detector. *Medical Image Analysis*, 2015, 26(1): 232–242. DOI: 10.1016/j.media.2015.09.002.

[11] Dai L, Wu L, Li H T, Cai C, Wu Q, Kong H Y, Liu R H, Wang X N, Hou X H, Liu Y X, Long X X, Wen Y, Lu L N, Shen Y X, Chen Y, Shen D G, Yang X K, Zou H D, Sheng B, Jia W P. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nature Communications*, 2021, 12: Article No. 3242. DOI: 10.1038/s41467-021-23458-5.

[12] Wang D Y, Haytham A, Pottenburgh J, Saeedi O, Tao Y. Hard attention net for automatic retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, 2020, 24(12): 3384–3396. DOI: 10.1109/JBHI.2020.3002985.

[13] Sun M Y, Li K Q, Qi X Q, Dang H, Zhang G H. Contextual information enhanced convolutional neural networks for retinal vessel segmentation in color fundus images. *Journal of Visual Communication and Image Representation*, 2021, 77: 103134. DOI: 10.1016/j.jvcir.2021.103134.

[14] Jin Q G, Meng Z P, Pham T D, Chen Q, Wei L Y, Su R. DUNet: A deformable network for retinal vessel segmentation. *Knowledge-Based Systems*, 2019, 178: 149–162. DOI: 10.1016/j.knosys.2019.04.025.

[15] Mou L, Chen L, Cheng J, Gu Z W, Zhao Y T, Liu J. Dense dilated network with probability regularized walk for vessel detection. *IEEE Trans. Medical Imaging*, 2020, 39(5): 1392–1403. DOI: 10.1109/TMI.2019.2950051.

[16] Wei J H, Zhu G J, Fan Z, Liu J C, Rong Y B, Mo J J, Li W J, Chen X J. Genetic U-Net: Automatically designed deep networks for retinal vessel segmentation using a genetic algorithm. *IEEE Trans. Medical Imaging*, 2022, 41(2): 292–307. DOI: 10.1109/TMI.2021.3111679.

[17] Fu J, Liu J, Tian H J, Li Y, Bao Y J, Fang Z W, Lu H Q. Dual attention network for scene segmentation. In *Proc. the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2019, pp.3141–3149. DOI: 10.1109/CVPR.2019.00326.

[18] Yang Q, Ma B Q, Cui H, Ma J Q. AMF-NET: Attention-aware multi-scale fusion network for retinal vessel segmentation. In *Proc. the 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, Nov. 2021, pp.3277–3280. DOI: 10.1109/EMBC46164.2021.9630756.

[19] Wu T F, Li L L, Li J B. MSCAN: Multi-scale channel attention for fundus retinal vessel segmentation. In *Proc. the 2nd IEEE International Conference on Power Data Science*, Dec. 2020, pp.18–27. DOI: 10.1109/ICPDS51559.2020.9332494.

[20] Wu H S, Wang W, Zhong J F, Lei B Y, Wen Z K, Qin J. SCS-Net: A scale and context sensitive network for retinal vessel segmentation. *Medical Image Analysis*, 2021, 70: 102025. DOI: 10.1016/j.media.2021.102025.

[21] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In *Proc. the 14th International Conference on Artificial Intelligence and Statistics*, Apr. 2011, pp.315–323.

[22] Huang G, Liu Z, Van Der Maaten L, Weinberger K Q. Densely connected convolutional networks. In *Proc. the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Jul. 2017, pp.2261–2269. DOI: 10.1109/CVPR.2017.243.

[23] He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In *Proc. the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2016, pp.770–778. DOI: 10.1109/CVPR.2016.90.

[24] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proc. the 32nd International Conference on Machine Learning*, Jul. 2015, pp.448–456.

[25] Mou L, Zhao Y T, Fu H Z, Liu Y H, Cheng J, Zheng Y L, Su P, Yang J L, Chen L, Frangi A F, Akiba M, Liu J. CS^2-Net: Deep learning segmentation of curvilinear structures in medical imaging. *Medical Image Analysis*, 2021, 67: 101874. DOI: 10.1016/j.media.2020.101874.

[26] Shi Z J, Wang T Y, Huang Z, Xie F, Liu Z H, Wang B L, Xu J. MD-Net: A multi-scale dense network for retinal vessel segmentation. *Biomedical Signal Processing and Control*, 2021, 70: 102977. DOI: 10.1016/j.bspc.2021.102977.

[27] Owen C G, Rudnicka A R, Mullen R, Barman S A, Mon-

ekosso D, Whincup P H, Ng J, Paterson C. Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (CAIAR) program. *Investigative Ophthalmology & Visual Science*, 2009, 50(5): 2004–2010. DOI: 10.1167/iovs.08-3018.

[28] Köhler T, Budai A, Kraus M F, Odstrčilik J, Michelson G, Hornegger J. Automatic no-reference quality assessment for retinal fundus images using vessel segmentation. In *Proc. the 26th IEEE International Symposium on Computer-Based Medical Systems*, Jun. 2013, pp.95–100. DOI: 10.1109/CBMS.2013.6627771.

[29] Woo S, Park J, Lee J Y, Kweon I S. CBAM: Convolutional block attention module. In *Proc. the 15th European Conference on Computer Vision*, Sept. 2018, pp.3–19. DOI: 10.1007/978-3-030-01234-2_1.

[30] Zhuo Z S, Huang J P, Lu K, Pan D R, Feng S T. A size-invariant convolutional network with dense connectivity applied to retinal vessel segmentation measured by a unique index. *Computer Methods and Programs in Biomedicine*, 2020, 196: 105508. DOI: 10.1016/j.cmpb.2020.105508.

[31] Wu Y C, Xia Y, Song Y, Zhang Y N, Cai W D. NFN+: A novel network followed network for retinal vessel segmentation. *Neural Networks*, 2020, 126: 153–162. DOI: 10.1016/j.neunet.2020.02.018.

[32] Khan T M, Khan M A U, Rehman N U, Naveed K, Afridi I U, Naqvi S S, Raazak I. Width-wise vessel bifurcation for improved retinal vessel segmentation. *Biomedical Signal Processing and Control*, 2022, 71: 103169. DOI: 10.1016/j.bspc.2021.103169.

[33] Li Q L, Feng B W, Xie L P, Liang P, Zhang H S, Wang T F. A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Trans. Medical Imaging*, 2016, 35(1): 109–118. DOI: 10.1109/TMI.2015.2457891.

[34] Wu Y C, Xia Y, Song Y, Zhang Y N, Cai W D. Multiscale network followed network model for retinal vessel segmentation. In *Proc. the 21st International Conference on Medical Image Computing and Computer-Assisted Intervention*, Sept. 2018, pp.119–126.

[35] Tang P, Liang Q K, Yan X T, Zhang D, Coppola G, Sun W. Multi-proportion channel ensemble model for retinal vessel segmentation. *Computers in Biology and Medicine*, 2019, 111: 103352. DOI: 10.1016/j.compbiomed.2019.103352.

[36] Dong F F, Wu D Y, Guo C Y, Zhang S T, Yang B L, Gong X Y. CRAUNet: A cascaded residual attention U-Net for retinal vessel segmentation. *Computers in Biology and Medicine*, 2022, 147: 105651. DOI: 10.1016/j.compbiomed.2022.105651.

[37] Soomro T A, Afifi A J, Gao J B, Hellwich O, Zheng L H, Paul M. Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation. *Expert Systems with Applications*, 2019, 134: 36–52. DOI: 10.1016/j.eswa.2019.05.029.

Hua-Deng Wang received his B.S. and M.S. degrees in computer science and technology from Huazhong University of Science and Technology, Wuhan, in 2002 and 2007 respectively, and his Ph.D. degree in information and communication engineering from Guilin University of Electronic Technology, Guilin, in 2022. He is currently a professor at the School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin. His current research interests include medical image processing.

Zi-Zheng Li received his B.S. degree in electronic information engineering from Dezhou University, Dezhou, in 2020. He is currently a Master student with the School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin. His current research interests include medical image segmentation.

Idowu Paul Okuwobi received his Ph.D. degree in computer science and technology from Nanjing University of Science and Technology, Nanjing, in 2019. He is currently an associate professor with the School of Artificial Intelligence, Guilin University of Electronic Technology, Guilin. He oversees the Vision and Image Processing Lab, Guilin University of Electronic Technology, where his current research objective is to develop new intelligent algorithms for medical diagnosis and treatment.

Bing-Bing Li received his B.S. degree in clinical medicine from Gannan Medical University, Ganzhou, in 2018, and his M.S. degree in clinical medicine from Southern Medical University, Guangzhou, in 2021. Since then, he has worked in the Department of Pathology, Ganzhou Municipal Hospital, Ganzhou, and served as the deputy director of the Department of Pathology. His current research interests include tumor pathology, molecular pathology, and computational pathology.

Xi-Peng Pan received his Ph.D. degree in control science and engineering from the Beijing University of Posts and Telecommunications, Beijing, in 2019. He is a Master supervisor in the School of Computer Science and Information Security of Guilin University of Electronic Technology, Guilin. His research interests include machine learning.

Zhen-Bing Liu received his Ph.D. degree in pattern recognition and intelligent system from Huazhong University of Science and Technology, Wuhan, in 2010. He is currently a professor at the School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin. His main research interests include image processing.

Ru-Shi Lan received his Ph.D. degree in soft engineering from the University of Macau, Macau, in 2016. He is currently an associate professor with the School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin. His research interests include image processing. He received the Guangxi Science Fund for Distinguished Young Scholars granted by the Science and Technology Department of Guangxi Zhuang Autonomous Region.

Xiao-Nan Luo received his Ph.D. degree in computational mathematics from the Dalian University of Technology, Dalian, in 1991. From 1995 to 2016, Dr. Luo was a professor at the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou. He is currently a professor at the School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin. His research interests include computer graphics and CAD. He received the National Science Fund for Distinguished Young Scholars granted by the National Natural Science Foundation of China.