Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

ECU-Net: an efficient cascade U-Net for retinal vessel segmentation

Xin Yang (yangxin@nuaa.edu.cn)
Nanjing University of Aeronautics and Astronautics https://orcid.org/0000-0003-0445-6497
Zhiqiang Li
Nanjing University of Aeronautics and Astronautics
Yingqing Guo
Nanjing University of Aeronautics and Astronautics
Dake Zhou
Nanjing University of Aeronautics and Astronautics
Tao Li
Nanjing University of Aeronautics and Astronautics

Research Article

Keywords: Retinal vessel segmentation, deep learning, deep forest, convolutional neural network (CNN), attention mechanism

Posted Date: November 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-327331/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

ECU-Net: an efficient cascade U-Net for retinal vessel segmentation

Xin Yang* Zhiqiang Li Yingqing Guo Dake Zhou Tao Li

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Corresponding author：Xin Yang

yangxin@nuaa.edu.cn

(03/13/2021)

Tel: 86-13770510498 Fax: 86-25-58784479

Abstract: Early detection of retinal vessel lesions by fundus image examination is essential to prevent and screen the related diseases. However, due to retinal vessels' complex structure, it is time-consuming and laborious to extract retinal vessels manually by professionals. Therefore, retinal vessel automatic segmentation has great application value in modern medicine. In recent years, deep learning technology has achieved remarkable results in medical image segmentation, and the segmentation accuracy and speed are significantly improved compared with traditional methods. In this paper, we propose an efficient retinal vessel segmentation model ECU-Net. In this model, a feature extraction module combining edge attention module and efficient channel attention mechanism is designed to improve the vessel structure's recognition ability. To improve the fusion efficiency of the texture information in the shallow layers and the semantic information in the deep layers, ECU-Net constructs an up-sampling fusion module. In addition, we design a multi-scale channel attention module to connect U-Net's subnetwork to achieve efficient feature extraction. Compared with the state-of-the-art models, our ECU-Net has reached the highest level in some classical evaluation criteria and has a certain application value in the medical field.

Keywords: Retinal vessel segmentation, deep learning, deep forest, convolutional neural network (CNN), attention mechanism.

1. Introduction

The morphology and structure of retinal vessels in the fundus are related to many human diseases. It is the only vessel that can be observed directly and noninvasively. Diseases such as glaucoma, diabetic retinopathy (DR), hypertension, and arteriosclerosis can make the structure and morphology of retinal vessels abnormal [1], so it can be used as an important basis for rapid diagnosis of related diseases. Retinal vessel segmentation can be used in the diagnosis, follow-up treatment, and feasibility analysis of ophthalmic surgery of some ophthalmic diseases. In order to obtain the distribution information of vessel morphology quickly for further diagnosis and treatment, it is necessary to design a faster and higher-performance retinal vessel segmentation model. Some of the existing UNet-based models have achieved excellent segmentation performance on the open retinal vessel segmentation dataset, but some problems cannot be ignored if they are directly applied to clinical diagnosis. Therefore, we propose an efficient cascade retinal vessel segmentation model ECU-Net (efficient cascade U-Net) in this paper. The contributions of ECU-Net are as follows:

Firstly, we propose an edge attention module to enhance the model's activation in the striped vessel regions. In addition, combined with the efficient channel attention module and residual structure, we further propose a RA (residual attention) module for feature extraction. It can make the model have a stronger recognition ability to retinal vessels.

Secondly, we design a UF (up-sampling fusion) module for the information fusion between the decoding path and the skip connection path in U-Net's subnetwork, significantly improving the texture information's fusion efficiency in the shallow layers and the semantic information in the deep layers.

Thirdly, we use a one-dimensional convolution to design a MECA (multi-scale efficient channel attention) module to connect two U-Net subnetworks to achieve efficient feature extraction and reduce model over-fitting.

We test the ECU-Net on the public retinal vessel segmentation datasets and compare it with some state-of-the-art retinal vessel segmentation methods. Experimental results show the effectiveness of ECU-Net has a good effect on retinal vessel segmentation tasks.

2. Related work

Before the development of deep learning technology, retinal vessel segmentation is mainly based on traditional methods. It can be divided into five categories: vessel tracking, matched filtering, morphological processing, deformation model and machine learning methods [2]. The traditional retinal vessel segmentation method involves an extensive range, and has a strictly theoretical basis. Therefore, it has a very intuitive physical meaning. In addition, different kinds of methods can be used in combination to achieve better performance of vessel segmentation. For example, Can et al. [3] propose a real-time vessel segmentation method using a directional template for recursive tracking of blood vessels; Dash et al. [4] use morphological operator and kirsch template to extract retinal vein. However, these traditional methods have some disadvantages that cannot be ignored, such as high sensitivity to light changes, low segmentation accuracy of retinopathy region, and the need to manually extract retinal image features. It makes it difficult for traditional methods to be applied in clinical diagnosis. In recent years, deep learning technology has developed rapidly, and its related technology has also achieved remarkable results in medical image segmentation. Especially since FCN has been successfully applied to image segmentation tasks, the performance and speed of medical image segmentation have improved significantly compared with traditional methods. For example, U-Net [5] can consider global context information and local texture information at the same time and only needs a small amount of data to achieve good segmentation performance. Compared with the natural image semantic segmentation task, the medical image segmentation model generally needs to consider less training data and unbalanced training samples. Retinal vessel segmentation is a typical task with insufficient training data and a serious imbalance of positive and negative samples. In addition, different from other medical image segmentation, such as lung parenchyma segmentation, pancreas segmentation, liver segmentation, and so on, the structure of retinal vessels is more complex, the degree of vessel deformation is greater, and there is a lot of noise interference.

The overall performance of the retinal vessel segmentation method based on deep learning is much better than traditional methods. The deep learning model can automatically extract the most suitable features for retinal vessel segmentation according to retinal vessels' morphological structure. It has strong robustness to illumination changes and noise. In addition, the deep learning model has excellent scale invariance, translation invariance, and rotation invariance in a certain range. These advantages make deep learning based retinal

vessel segmentation more and more popular. Therefore, many high-performance algorithms have been proposed. For example, Alom et al. [6] proposed that the R2U-Net network embeds the recurrent neural network into the U-Net network and combines the residual connection to construct an image segmentation model. This model performs very well in retinal blood vessel segmentation, skin cancer segmentation, and lung organ segmentation. Li et al. [7] transform the retinal vessel segmentation into the problem of trans-modal data conversion from retinal image to vessel map, and design a new retinal vessel segmentation network. NGO et al. [8] input multiple scale features of retinal image into a CNN, and use the Dropout to enhance the generalization ability of the model. The CCNET [9] perform multiple cross-fusion of features between the network main path and the secondary path so that the model could effectively integrate multi-level feature information. Wu et al. [10] construct a retinal vessel segmentation network based on U-Net and further used the multi-scale deep supervision [11] to retain rich multi-scale deep features. Araujo et al. [12]cascaded a typical segmentation network with a variational automatic encoder. It can correct topology inconsistency by learning rich and compact structure space. Yu et al. [13] propose an unsupervised generative adversarial network (Sc-GAN) based on knowledge transfer and consistent shape for tubular artery segmentation, effectively reducing medical image annotation costs. Zhang et al. [14] propose a new retinal vessel segmentation network using U-Net structure and residual connection. It converted the original task into a multi-class segmentation task by adding new labels in the boundary area. The MAU-Net [15] integrates the deformable convolution and attention mechanism into the U-Net network to improve the segmentation ability of retinal vessels. Wang et al. [16] construct a dual encoder retinal vessel segmentation network (DEUNet). It uses the HR (high-resolution) spatial path to retain spatial information, uses multi-scale convolution to capture semantic information, and uses a feature fusion module to combine different levels of features. LadderNet [17] cascades two original U-Net models through multiple paths to build a deeper cascaded model. Zhao et al. [18] build a novel retinal vessel segmentation model using generative adversarial networks. Biyani et al. [19] construct a blood vessel segmentation algorithm based on an improved fuzzy minimum-maximum neural network supervised classifier. Kromm et al. [20] present a method for retinal vessel segmentation and vessel centerline extraction by combining capsule network and perception structure. Compared with the latest deep convolution neural network, it has fewer parameters and achieves good generalization performance without data enhancement.

3. The proposed model

3.1 Overall structure of ECU-Net

Our model is a two-stage cascaded U-Net structure, as shown in Figure 1. In this paper, we design three novel modules and apply them to cascaded U-Net architecture. It ensures the reasoning speed of the model and further improves the model performance in the retinal

vessel segmentation tasks. After the preprocessed retinal image data is input into the network, the primary encoding and decoding are carried out through the primary U-Net's subnetwork. Firstly, a conventional convolution is used for preliminary feature extraction to obtain a 16 channel feature map, and then the feature is extracted by the RA module. A convolution with a step size of 2 is used to down-sampling the feature map, and the number of channels is doubled simultaneously.

After three times of down-sampling, the primary encoding is finished, and the primary decoding is started. The UF module is used to fuse the output feature map of the encoder with the primary coding feature map of double scale, and then the RA module is used to extract features. It recurs until the resolution of the feature map is the same as that of the input image. MECA module is used to add the multi-scale feature map obtained from primary decoding and coding feature map so as to enhance the information transfer efficiency between U-Net's subnetworks. The secondary U-Net's subnetwork compresses the channel by a 1×1 convolution. Finally, we use the Sigmoid activation function to get the prediction probability at the pixel level.

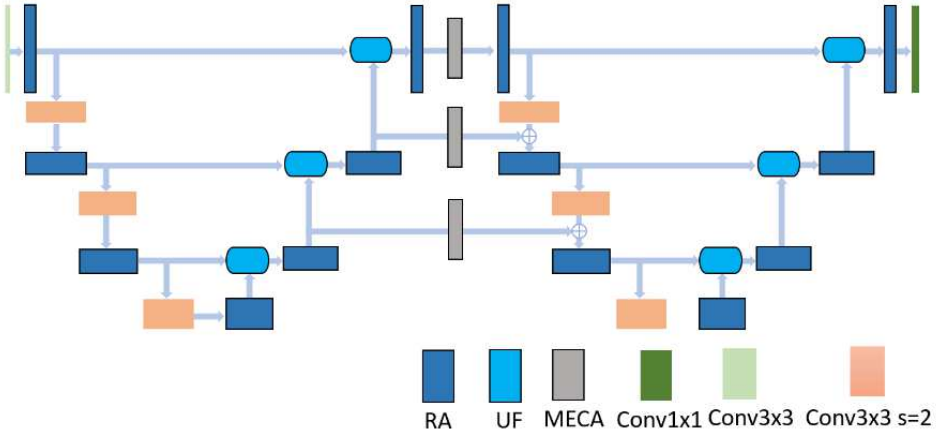


Figure 1. Overall structure of ECU-Net

3.2 Residual attention feature extraction module

Conventional 1×1 convolution cannot fit the object with geometric deformation well because its convolution kernel is square. Retinal vessels have very complex branches, so conventional 1×1 convolution is not the best choice for retinal vessel segmentation. Therefore, in this paper, we design a BA (border attention) module that can extract the spatial features of strip structure, as shown in Figure 2. For the input feature map, the 1×5 and 5×1 two-dimensional convolutions are used to extract the stripe pattern, and the 3×3 convolution is used to extract the conventional pattern. The feature maps are spliced in the channel dimension, and then the 1×1 convolution and Sigmoid activation function are used to get the spatial attention weight. Finally, the weight is multiplied by the input feature map to get the output result. BA module can adapt to the distribution of vessel structure and improve the model's recognition ability for complex vessel structure.

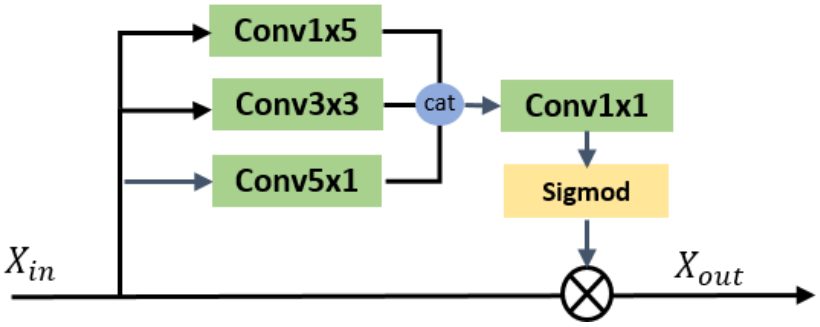


Figure 2. Structure of BA module

CA (Channel attention) mechanism can obtain the importance of each channel of the feature map, and use it to weight the different channels of the feature map. However, the use of fully connected neural networks is faced with too large parameters and over-fitting. The most effective improvement to the original channel attention is the ECA (efficient channel attention) module proposed in ECANet [21]. On this basis, we design the IECA (improved efficient attention) module, which adds ReLU activation function and one-dimensional convolution to improve the nonlinear mapping ability and information fusion between channels. Figure 3 shows the structure of IECA. As shown in the figure, for the input feature map, the global average pooling operation is performed to get each channel's original weight . Then, two one-dimensional convolutions with the size of 3 and the ReLU activation function are used for nonlinear mapping. Finally, the Sigmoid activation function is used to obtain the channel attention weight, which is multiplied by the input feature map to get the output.

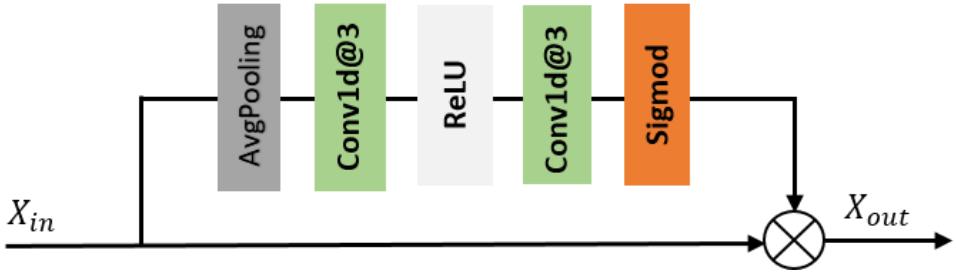


Figure 3. Structure of IECA module

We use the Ba module and IECA module to design the residual attention (RA) module for feature extraction, as shown in Figure 4. For the ECA module, we introduce the RA module and BA module and use batch normalization, Dropblock and residual connection to improve the generalization ability and training convergence speed of the model.

Let the input feature map of RA be X_{in} . After a 3×3 convolution operation, the normalization and the activation, we can get the feature map F_1 as follows:

$$F_1 = \sigma_R\left[f_{bn}\left(f_{conv(3\times3)}(X_{in})\right)\right] \tag{1}$$

where $f_{bn}(\square)$ and $f_{conv(3\times3)}(\square)$ represent the normalization and the convolution operations with the convolution kernel size of 3×3, respectively, and σ_R represents ReLU activation.

The ECA module is used to weight the feature map, and then the Dropblock is used to get the feature map F_2 as

$$F_2 = D_r\left(f_{ECA}(F_1)\right), \tag{2}$$

where D_r and $f_{ECA}(\square)$ represent the Dropblock and ECA module, respectively.

Then, a 3×3 convolution and BN layer are used to learn the feature, and then the feature is sent to the BA module. Finally, we add the output of the BA module to the original input to get the output X_{out} as

$$X_{out} = \sigma_R\left\{X_{in} \oplus f_{BA}\left[f_{bn}\left(f_{conv(3\times3)}(F_2)\right)\right]\right\}, \tag{3}$$

where $f_{BA}(\square)$ is the BA module, and $\oplus$ represents the element summation operation

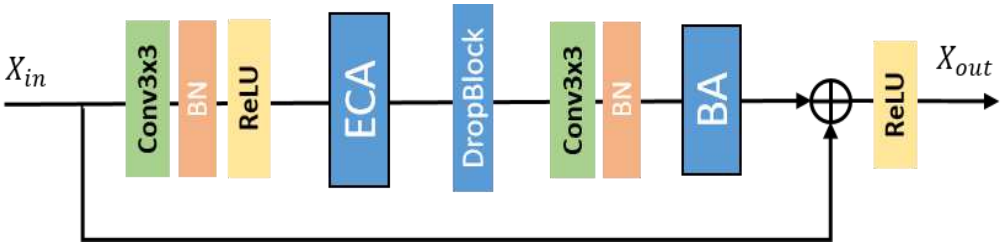


Figure 4. Structure of RA module

3.3 Up-sampling fusion module

The information fusion between the decoding path and skip connection of the U-Net subnetwork is very important. It involves how to unify the HR texture information in shallow layers with the LR (low-resolution) semantic information in deep layers. The commonly used strategy is to up-sampling the LR feature map by the transposed convolution or the interpolation, and then directly adding it to the HR feature map to get the fusion result. However, the effect of such a simple fusion operation is not ideal. Therefore, we design an up-sampling fusion (UF) module to fuse different resolution feature maps, which is used for information fusion between decoding path and skip connection of U-Net subnetwork.

The structure of UF module is shown in Figure 5. In the figure, X_l and X_h represent the low resolution semantic feature map in deep layers and the detailed HR texture feature map in shallow layers, respectively. Firstly, the resolution of X_l is doubled by transpose convolution, and then it is spliced with X_{h1} processed by a 3×3 convolution in channel dimension. The specific formula is given as

$$\begin{aligned} X_{h1} &= f_{conv(3\times3)}(X_h) \\ X_{l1} &= f_{cat}\left(X_{h1}, f_{deconv}(X_l)\right) \end{aligned}, \tag{4}$$

where $f_{deconv}(\square)$ is transposed convolution and $f_{cat}(\square)$ is splicing operation.

We can obtain X_{l2} by a 3×3 convolution and ReLU activation. By fusing X_{l2} and X_{h1}, we get X_{h2} as

$$\begin{aligned} X_{l2} &= \sigma_R\left(f_{conv(3\times3)}(X_{l1})\right) \\ X_{h2} &= f_{cat}(X_{h1}, X_{l2}) \end{aligned} \tag{5}$$

By inputting X_{h2} into a 3×3 convolution and Sigmoid activation, we can obtain the weights of each channel of the HR texture feature map X_h in the shallow layer. Then we multiply the weights by X_h to get X_{h3} as

$$X_{h3} = X_h \otimes \sigma_S\left(f_{conv(3\times3)}(X_{h2})\right), \tag{6}$$

where σ_R is the Sigmoid activation function.

Finally, the model splices X_{h3} and X_{l2} of the two paths and introduces them into a 1 × 1 convolution to compress the channel number to obtain the module output X_{out} as

$$X_{out} = f_{conv(1\times1)}\left(f_{cat}(X_{l2}, X_{h3})\right) \tag{7}$$

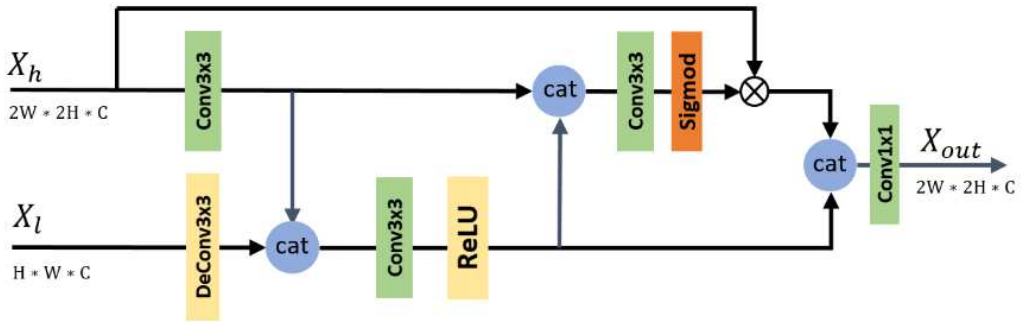


Figure 5. Structure of UF module

UF module uses the HR texture information in the shallow layers to improve the semantic feature map sampling effect in the deep layers, and uses the low resolution semantic information in the deep layers to weight the channel of the texture feature map in shallow layers. Finally, we use splicing and convolution to fuse the feature maps, so as to complete effective information fusion.

3.4 Multi-scale efficient channel attention module

In order to achieve effective information fusion between U-Net subnets, we propose a MECA (multi-scale efficient channel attention) module to connect U-Net subnets. As shown in Figure 6. In the MECA module, the global average pooling and maximum pooling are used to generate the channel's original attention weight, and then three adaptive one-dimensional convolutions are used to obtain the multi-scale attention weight of the channel. Finally, MECA adds the attention weights of three channels with different scales and uses Sigmoid activation function to activate them nonlinearly to get the final channel attention weights. In addition, MECA uses residual connection to further enhance the information exchange between input and output. Due to the use of one-dimensional convolution instead of the original channel attention fully connected neural network, the module has less parameters and better generalization ability.

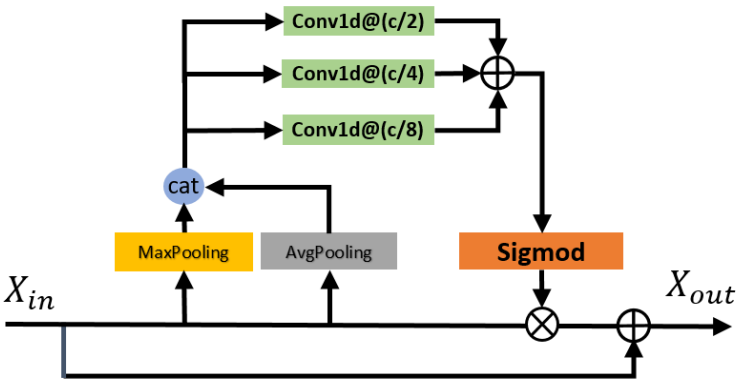


Figure 6. Structure of MECA module

MECA first reduces the dimension of the input feature map by using the maximum pooling $P_{Max}(\square)$ and average pooling $P_{Avg}(\square)$. By the splicing operation f_{cat}, we can obtain the F_1 as

$$F_1 = f_{cat}\left(P_{Max}(X_{in}), P_{Avg}(X_{in})\right). \tag{8}$$

MECA uses three different sizes of one-dimensional convolutions (the convolution kernel size is $c/2$, $c/4$, and $c/8$, where c is the number of channels in the input feature map) to get the three features of different scales and sum them. By the Sigmoid activation function, we obtain the channel attention weight F_2 as

$$F_2 = \sigma_R\left[\Sigma\left(f_{conv1d(c/2)}(F_1), f_{conv1d(c/4)}(F_1), f_{conv1d(c/8)}(F_1)\right)\right], \tag{9}$$

where $f_{conv1d(\square)}(\square)$ is the one-dimensional convolution operation, $f_{cat}(\square)$ is the splicing operation of the feature map in the channel dimension, and $\Sigma(\square)$ is the summation operation of multiple items.

Finally, we multiply the channel attention weight by the input, and add the result to the input element by element. Then, we have

$$X_{out} = (X_{in} \otimes F_2) \oplus X_{in}, \tag{10}$$

where $\otimes$ and $\oplus$ are multiplication and addition operations, respectively.

4. Experimental results and analysis

4.1 Public data sets and evaluation indicators

At present, there are few public data sets of retinal vessel segmentation, and each data set contains a small number of samples. Most methods are tested on three public datasets: DRIVE [22], CHASE_DB1 [23] and STARE [24]. In this paper, we also use these three retina datasets to verify the effectiveness of our model.

DRIVE is the most commonly used data set in retinal vessel segmentation. It contains 40 color fundus retinal images, which are collected from 453 subjects aged 31 to 86 in the Dutch diabetic retinopathy screening program. In DRIVE, Canon Cr5 non astigmatic 3-CCD camera is used to capture the image at 45 ° FOV (field of view). The resolution of each image is 565

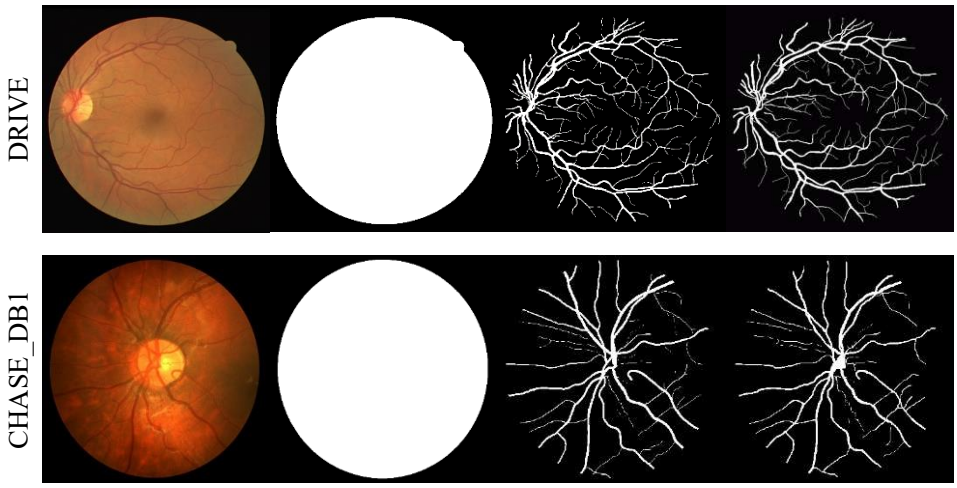
× 584, and the corresponding FOV is a circle with a diameter of 540 pixels. DRIVE has been divided into a training set with 20 images and a test set with 20 images, and each image contains a binary mask image of FOV.

CHASE_DB1 contains 28 color fundus retinal images from 14 children under the age of 10. The resolution of each image is 999×960, and the approximate FOV size is 950×900. CHASE_DB1 also contains the segmentation image of blood vessels annotated by two observers, but it does not provide the binary mask of FOV. So we need to extract the FOV mask manually for model training and testing.

STARE was created by Michael in 1975 and then published by Hoover. STARE includes 20 fundus images, 10 of which have signs of mild diabetic retinopathy. It is captured by the TOPCON trv-50 fundus camera at 35 ° field of vision. The resolution of each image is 700×605, and the approximate FOV size is 650×500. STARE does not provide the binary mask of FOV, which needs to be extracted manually for model training and testing.

Figure 7 shows the sample images of the three datasets. It can be seen that the data in different datasets are obviously different, which is mainly caused by the different types of data acquisition equipment and parameter settings. These factors also increase the difficulty of designing a general retinal vessel segmentation algorithm. In addition, unlike DRIVE, which has divided all images into the training set and test set, STARE and CHASE_DB1 do not distinguish between the training set and test set. Therefore, in order to evaluate the model performance more accurately, this paper chooses the cross validation method to evaluate the model performance on the STARE and CHASE_DB1 datasets respectively.

In order to evaluate the performance of the model accurately and compare with other methods fairly, this paper takes the manual marking result of the first expert (observer) as the standard result (the first gold standard) and only conducts the training and evaluation in the FOV.



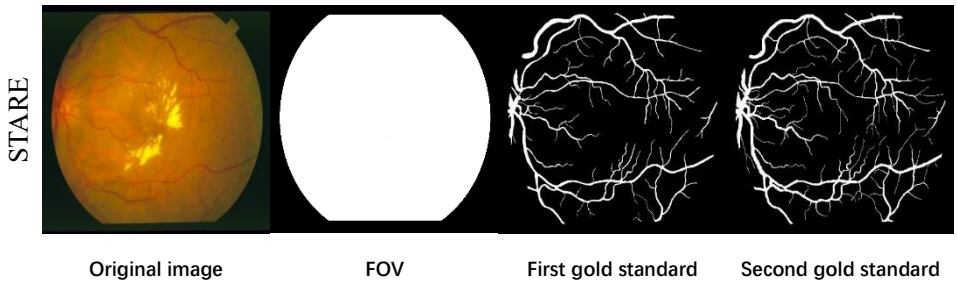


Figure 7. Example of public data set

4.2 Evaluation criteria

In this paper, we use five criteria to evaluate the performance of the model: accuracy (ACC), sensitivity (SE), specificity (SP), F1-Score and AUC (area under the ROC curve). SE measures the ability of the model to distinguish between foreground (blood vessels). SP measures the ability of the model to distinguish background (non-vessel). ACC, F1-Score, and AUC measure the overall segmentation performance of the model. The larger the value of these five criteria is, the better the performance is. The calculation formula is given as

$$AC = \frac{TP + TN}{TP + TN + FP + FN}, \tag{11}$$

$$SE = \frac{TP}{TP + FN}, \tag{12}$$

$$SP = \frac{TN}{TN + FP}, \tag{13}$$

$$Precision = \frac{TP}{TP + FP}, \tag{14}$$

$$Recall = \frac{TP}{TP + FN}, \tag{15}$$

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \tag{16}$$

where TP (true positive) represents the number of foreground (blood vessel) pixels with correct classification; TN (true negative) represents the number of background (non-vessel) pixels with correct classification; FP (false positive) represents the number of background (non-vessel) pixels mistakenly classified as foreground (vessel) pixels; FN (false negative) denotes the number of foreground (vessel) pixels mistakenly classified as background (non-vessel) pixels.

In addition, AUC is obtained by calculating the area under the ROC curve. In order to ensure the rationality of the evaluation, all the evaluation criteria are calculated in the FOV.

The above five evaluation criteria can be used to comprehensively evaluate the performance of retinal vessel segmentation algorithm. Generally, AUC is the area under the ROC curve by default. In order to measure the performance of the model more comprehensively, we also draw the PR (precision recall) curve and calculate the area under the PR curve. However, AUC of PR curve is not used as the main evaluation criteria in many

literatures, so it is not used as the main performance evaluation criterion in this paper.

4.3 Experimental setup

Image preprocessing: In the color retinal image, the contrast between the blood vessel and the background is low, and the uneven illumination of the image is obvious. Especially at the end of blood vessels, it is difficult for human eyes to distinguish the blood vessels. These factors will affect the accuracy of model segmentation, so preprocessing operation is needed to enhance the vessel region in the original data. The preprocessed image helps the convolution neural network to learn more appropriate features, so the retinal vessel segmentation performance of the model is effectively improved. Our preprocessing strategies are as follows:

① Image graying

The information of the three channels is fused by graying the original image. Compared with the original image, the obtained gray image has a clearer vessel area. The formula for converting RGB image to grayscale image is as follows:

$$X_{Gary} = \sqrt[2.2]{\frac{R^{2.2} + (1.5G)^{2.2} + (0.6B)^{2.2}}{1 + 1.5^{2.2} + 0.6^{2.2}}} \tag{17}$$

② Image normalization

Firstly, the gray image is standardized, then the maximum and minimum normalization operation is carried out, and finally the data distribution range is adjusted to 0 ~ 255. The calculation formula is as follows:

$$\overline{X} = \frac{X_{Gary} - \mu}{\sigma} \tag{18}$$

$$X_{norm} = \frac{\overline{X} - min(\overline{X})}{max(\overline{X}) - min(\overline{X})} \cdot 255 \tag{19}$$

where, X_{Gary} is the input gray image, μ and σ are the mean and variance of the input gray image, $\overline{X}$ is the standardization of input gray image; $min(\overline{X})$ and $max(\overline{X})$ are the maximum and minimum values of the image $\overline{X}$, and X_{norm} is the output result of normalization processing. Through the image normalization operation, the adjusted image's whole gray distribution is more uniform, and the image features can be further enhanced.

③ CLAHE enhancement

The illumination intensity of the gray image near the optic cup is higher than that at the FOV edge. So the contrast of the normalized image is adjusted by the CLAHE [26] algorithm to further improve the contrast between the blood vessel and the background.

④ Gamma correction

By further using gamma correction ($\gamma = 1.2$ in this paper) to enhance the global contrast

of the image enhanced by the CLAHE algorithm, we can get a clearer image of the vessel region. The calculation formula of gamma correction is as follows:

$$X_{\gamma} = (X_{clahe} / 255)^{1/\gamma} \cdot 255 \tag{20}$$

Training data extraction: Due to the small amount of data in the public data set, it is necessary to further sample and enlarge the preprocessed image data. For the DRIVE data set, there are 20 fundus images in the training set. We randomly select 10000 image blocks with the size of 64×64 from each image in the training set. Then 200000 training samples can be extracted from DRIVE. We take 10% of them as the verification set. CHASE_DB1 has 28 fundus images, but it is not divided into the training set and the test set, so we use 4 fold cross-validation to evaluate the model. We take seven images each time as test data and the other 21 images as training data. The training set uses a training sample extraction method similar to DRIVE. We can get 210000 training samples by extracting 10000 64×64 image blocks from each image. Similarly, we take 10% of them as the verification set. There are 20 images in the STARE. We also use 4-fold cross-validation and extracted training data in the same way.

Model training parameter setting: During the training, the training samples are randomly cut to the blocks with the size of 48×48, and the data augmentation strategies of the random horizontal, vertical flip and rotation are adopted. The Adam optimizer is selected, and the default parameters (β_1=0.9 , β_2=0.999 , ε=1×10^{-8}) are used. The initial learning rate is set to 0.0001, and the cosine function learning rate decay strategy is used to adjust the learning rate dynamically. The random inactivation rate of Dropblock is set to 0.25. The batch size is set to 64. The epoch is set to 50, and the training is terminated when the performance of the verification set is not improved within ten epoch. The experimental program is based on Python 3.6 and python deep learning framework and is trained and tested by calling NVIDIA gtx2080ti GPU on Ubuntu 18.04.

4.4 Experimental results

We test the performance of ECU-Net on DRIVE, CHASE_DB1, and STARE datasets. In all experiments, the result marked by the first expert (or observer) in the data set is regarded as the standard segmentation. During the test, we conduct a block test on the image using the sliding window of a certain step length and then splice the block prediction to get the final experimental results. The sliding window size is set to 96 × 96, and the sliding step is set to 8. We only evaluate within the FOV. Table 1 shows the test results of ECU-Net on each data set. It can be seen from the test results in Table 1 that the F1-Score, SP, and ACC of ECU-Net on the three data sets are higher than the manual segmentation results of the second observer. The F1-Score is more than 83%. In addition, the AUC values of ECU-Net on the three data sets are higher than 98%. Therefore, the experimental data show that the overall

segmentation ability of ECU-Net is excellent.

Table 1

Comprehensive experimental results of ECU net

Dataset	*Method*	*F1-Score*	*SE*	*SP*	*ACC*	*AUC*
DRIVE	2nd Human observer	-	0.7760	0.9725	0.9473	-
	Proposed method	0.8299	0.8169	0.9779	0.9574	0.9819
CHASE_DB1	2nd Human observer	-	0.7705	0.9778	0.9561	-
	Proposed method	0.8323	0.8193	0.9834	0.9670	0.9879
STARE	2nd Human observer	-	0.8956	0.9381	0.9346	-
	Proposed method	0.8337	0.8127	0.9840	0.9661	0.9877

In order to intuitively evaluate the effectiveness of the ECU-Net in retinal vessel segmentation task, we provide its ROC curve and PR curve on three data sets, as shown in Figures 8, 10, and 12, and the visual segmentation effects on three data sets, as shown in Figures 9, 11, and 13. As can be seen from these figures, ECU-Net has excellent segmentation performance. Whether it is the main blood vessel or the smaller blood vessel, ECU-Net can accurately identify them.

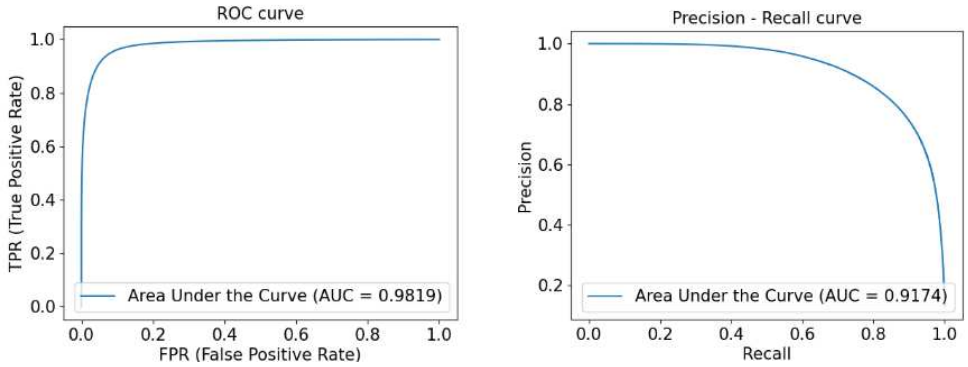
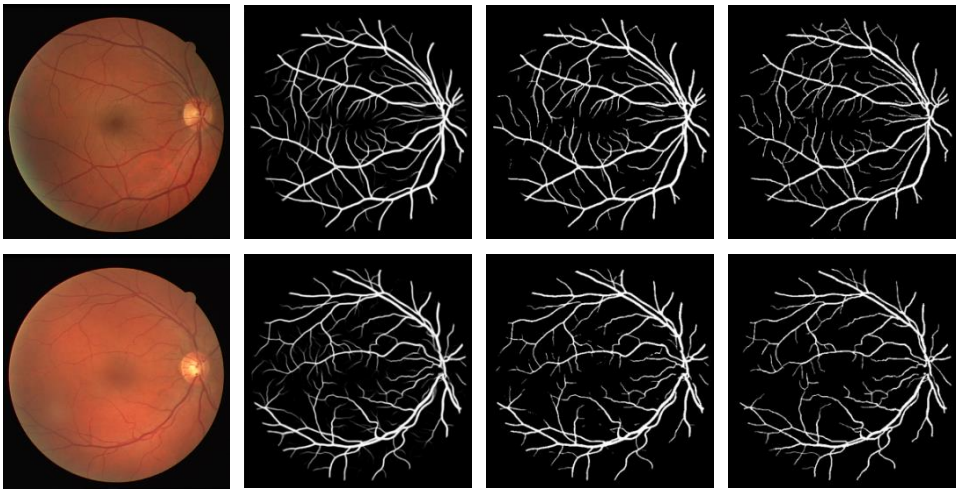


Figure 8. ROC curve and PR curve of ECU-Net on DRIVE



Original Image | Segmentation Probability Map | Segmentation Binary Map | Ground Truth

Figure 9. ROC curve and PR curve of ECU-Net on DRIVE

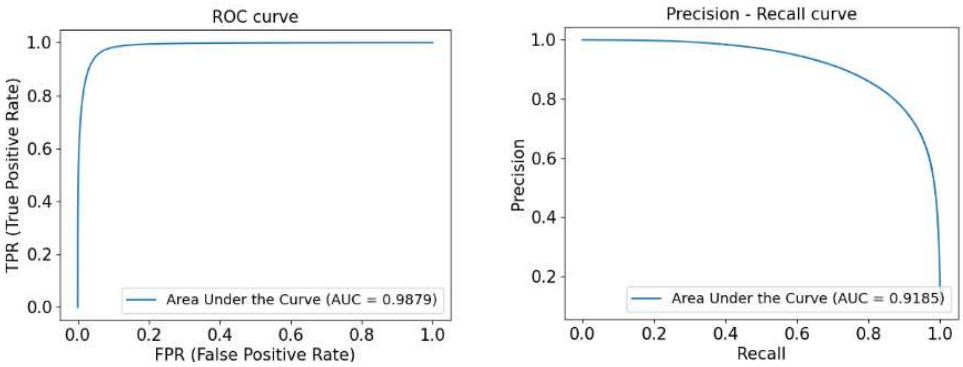


Figure 10. ROC curve and PR curve of ECU-Net on CHASE_DB1

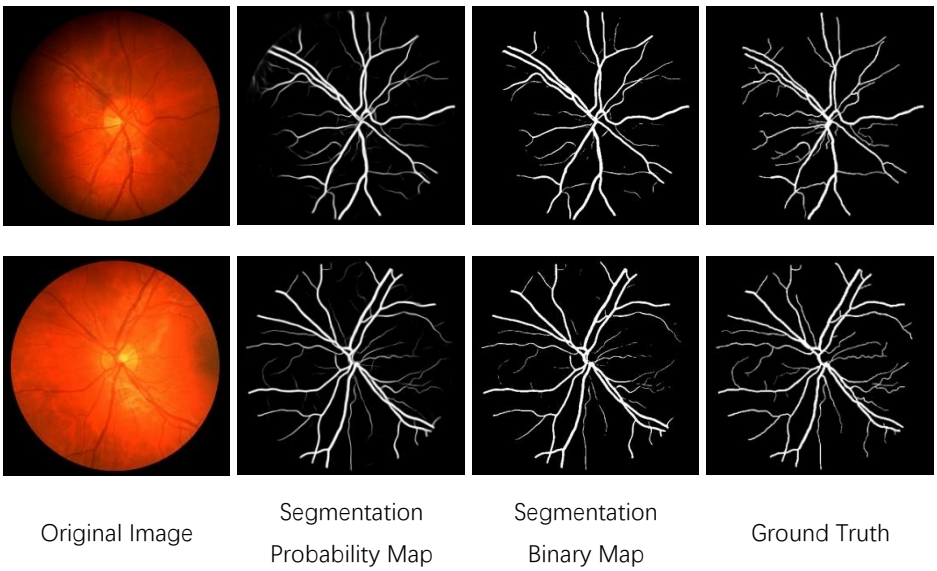


Figure 11. Segmentation example of ECU-Net on CHASE_DB1

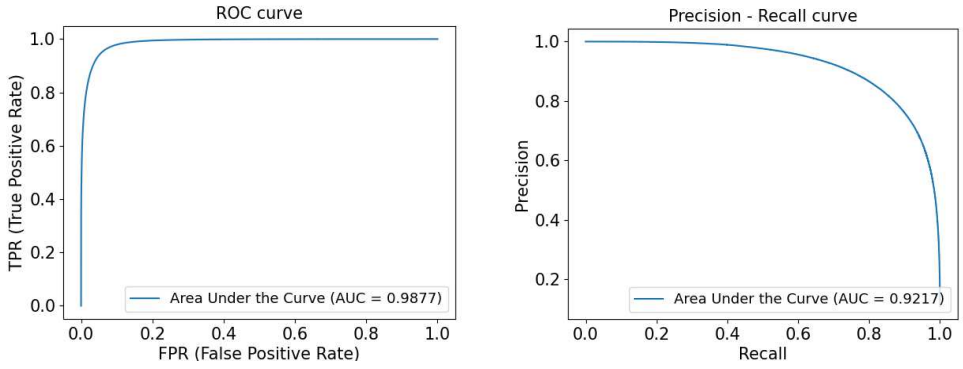


Figure 12. ROC curve and PR curve of ECU-Net on STARE

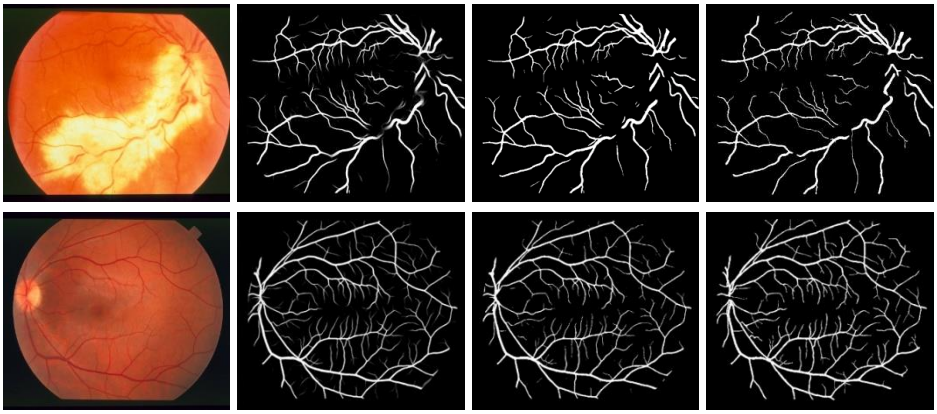
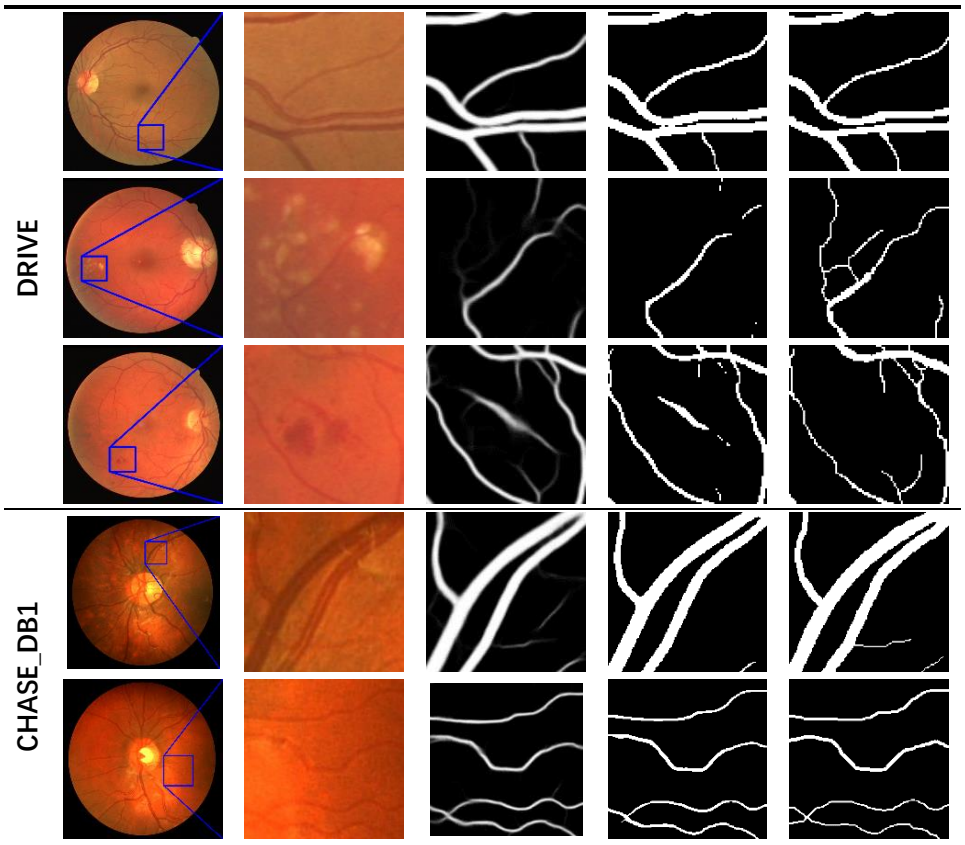


Figure 13. Segmentation example of ECU-Net on STARE

In order to intuitively analyze the segmentation performance of ECU-Net in detail, we locally enlarge the segmentation results on three datasets, as shown in Figure 14. Through careful observation, we can see that ECU-Net has an excellent performance in vessel segmentation without retinopathy. Smaller vessels can also be accurately segmented, and vessel discontinuities rarely occur. In the region with retinopathy, the frequency of identifying lesion spots as blood vessels and vessel discontinuities is also significantly reduced.



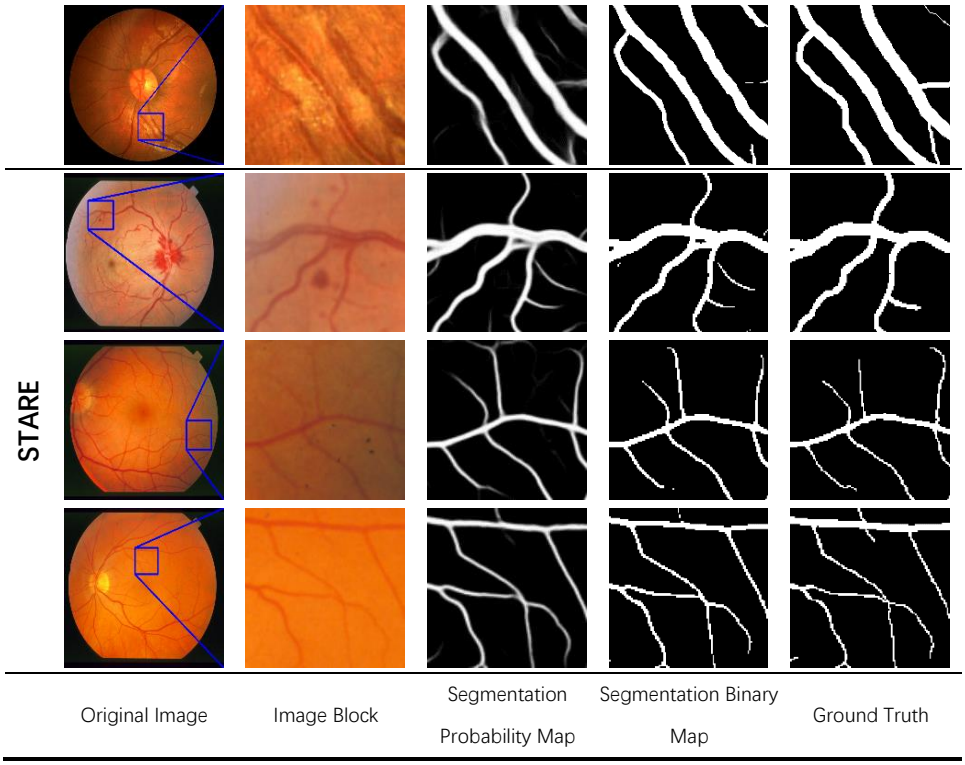


Figure 14. Local enlarged comparison of segmentation results of ECU-Net

4.5 Ablation experiment

This paper designs the RA module, UF module, and MECA module based on the cascaded U-Net structure. To test the contribution of these three modules to the overall model performance, we conduct ablation experiments on the drive dataset. The experimental results are shown in Table 2. The baseline is the benchmark model without any modules. Its F1-Score, Se, SP, ACC, and AUC are 0.8168, 0.7778, 0.9815, 0.9556 and 0.9772, respectively, and its parameter number is 0.8771M. Then we add the EA module, UF module, and MECA module to the benchmark model for testing. The results show that F1-Score has been significantly improved by about 1%, SE by about 2% ~ 3%, and ACC by a small margin. AUC increased to more than 98%. It shows that the three modules proposed in this paper can improve the model performance. After we add all three modules to the benchmark model, the performance of the final ECU-Net is the best. F1-Score, Se, SP, ACC, and AUC were 0.8299, 0.8169, 0.9779, 0.9574, and 0.9819, respectively, and the number of parameters increases to 1.6993M. Compared with the benchmark model, the final model's F1-Score increased by 1.3%, SE increased by 3.91%, ACC increased by 0.18%, AUC increased by 0.47%, and SP decreased by 0.36%. On the whole, the performance improvement of the model with three modules is very obvious.

Table 2

Ablation Experiment of ECU-Net

Method	*RA*	*UF*	*MECA*	*F1-Score*	*SE*	*SP*	*ACC*	*AUC*	*Param(M)*
Baseline				0.8168	0.7778	**0.9815**	0.9556	0.9772	0.8771
Baseline+RA	√			0.8282	0.8144	0.9778	0.9570	0.9811	1.3826
Baseline+UF		√		0.8256	0.7996	0.9799	0.9570	0.9807	1.1936
Baseline+MECA			√	0.8225	0.8031	0.9782	0.9559	0.9796	0.8773
ECU-Net	√	√	√	**0.8299**	**0.8169**	0.9779	**0.9574**	**0.9819**	1.6993

In order to more intuitively illustrate the influence of the three modules on the model performance, we provide the ROC curve and PR curve of the ablation experiment, as shown in Figures 15 and 16. It can be clearly seen from the figures that adding any one of these three modules to the benchmark model can improve the model performance to a certain extent. Adding all three modules to the benchmark model can improve the performance to a great extent. Therefore, we can conclude that our ECU-Net with three modules is very effective in retinal vessel segmentation.

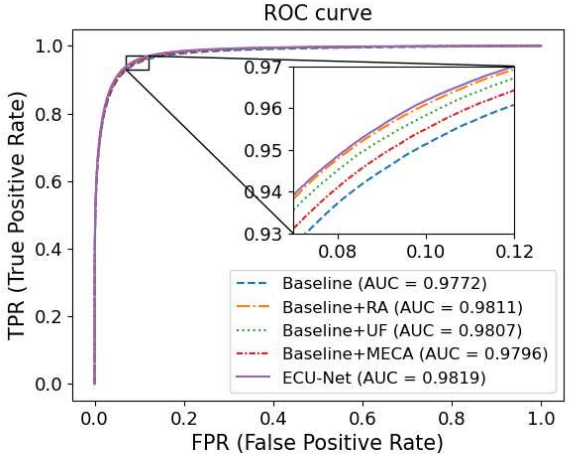


Figure 15. ROC curves of different models on DRIVE in ECU-Net's ablation experiment

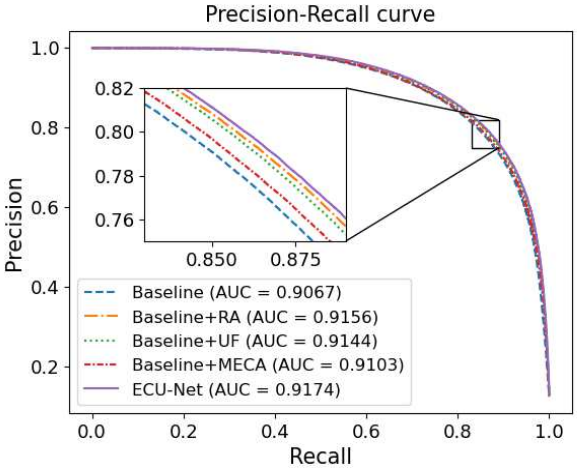


Figure 16. PR curves of different models on DRIVE in ECU-Net's ablation experiment

4.6 Comparison experiment

This section compares our ECU-Net with some state-of-the-art retinal vessel segmentation

algorithms in the recent five years on three datasets. These methods are based on deep learning and have great advantages in performance and speed compared with traditional methods. The experimental results are shown in Table 3.

From the Table, The F1-Score, Se, SP, ACC and AUC of ECU-Net are 0.8299, 0.8169, 0.9779, 0.9574 and 0.9819 on DRIVE, 0.8323, 0.8193, 0.9834, 0.9670 and 0.9879 on CHASE_DB1 and 0.8337, 0.8127, 0.9840, 0.9661 and 0.9877 on STARE, respectively. We can find that on DRIVE, the SP of ECU-Net is slightly lower than that of 0.9820 in literature[28], while the other performance criteria of ECU-Net are better than those of other models. On the CHASE_DB1 dataset, the ECU-Net's F1-Score, ACC, and AUC are the highest, the ECU-Net's SE is lower than 0.8288 of U-Net in literature [5], and the ECU-Net's SP is lower than 0.9847 of MS-NFN [31]. On STARE, ECU-Net's F1-Score and SE are the highest, while other criteria of ECU-Net are slightly lower than those of other algorithms. Because of the reciprocal and mutually restrictive relationship between SE and SP, it is not reasonable to compare SE or SP separately.

On the whole, compared with the state-of-the-art, ECU-Net's F1-Score achieves the highest performance on all three datasets, ACC and AUC achieve the highest performance on DRIVE and CHASE_DB1. It fully proves that our ECU-NET algorithm is effective and advanced in the retinal vessel segmentation.

Table 3

Performance comparison between ECU-Net and some classical and state-of-the-art algorithms

Dataset	*Method*	*Year*	*F1-Score*	*SE*	*SP*	*ACC*	*AUC*
DRIVE	Melinščak et al.[25]	2015	-	0.7276	0.9785	0.9466	0.9749
	Li et al.[7]	2015	-	0.7569	0.9816	0.9527	0.9738
	Liskowski et al.[27]	2016	-	0.7520	0.9806	0.9515	0.9710
	U-Net [5]	2018	0.8142	0.7537	**0.9820**	0.9531	0.9755
	R2U-Net [6]	2018	0.8171	0.7792	0.9813	0.9556	0.9784
	Residual U-Net [6]	2018	0.8149	0.7726	**0.9820**	0.9553	0.9779
	Recurrent U-Net [6]	2018	0.8155	0.7751	0.9816	0.9556	0.9782
	MS-NFN [31]	2018	-	0.7844	0.9819	0.9567	0.9807
	LadderNet [17]	2018	0.8202	0.7856	0.9810	0.9561	0.9793
	Yan et al.[28]	2018	-	0.7631	**0.9820**	0.9538	0.9750
	DEU-Net [16]	2019	0.8270	0.7940	0.9816	0.9567	0.9772
	BCDU-Net (d=3) [29]	2019	0.8224	0.8007	0.9786	0.9560	0.9789
	IterNet [30]	2019	0.8205	0.7735	0.9838	0.9573	0.9816
	Kromm et al.[20]	2020		0.7651	0.9818	0.9547	0.9750
	MAU-Net[15]	2020	0.8192	0.7890	0.9799	0.9557	0.9774
	Proposed(ECU-Net)	2020	**0.8299**	**0.8169**	0.9779	**0.9574**	**0.9819**
CHASE_DB1	Li et al.[7]	2015	-	0.7507	0.9793	0.9581	0.9716
	U-Net [5]	2018	0.7783	**0.8288**	0.9701	0.9578	0.9772

	R2U-Net [6]	2018	0.7928	0.7756	0.9820	0.9634	0.9815
	Residual U-Net [6]	2018	0.7800	0.7726	0.9820	0.9553	0.9779
	Recurrent U-Net [6]	2018	0.7810	0.7459	0.9836	0.9622	0.9803
	MS-NFN [31]	2018	-	0.7538	**0.9847**	0.9637	0.9825
	LadderNet [17]	2018	0.8031	0.7978	0.9818	0.9656	0.9839
	Yan et al.[28]	2018	-	0.7641	0.9806	0.9607	0.9776
	DEU-Net [16]	2019	0.8037	0.8074	0.9821	0.9661	0.9812
	IterNet [30]	2019	0.8073	0.7970	0.9823	0.9655	0.9851
	MAU-Net[15]	2020	0.8037	0.7798	0.9822	0.9620	0.9791
	Proposed(ECU-Net)	2020	**0.8323**	0.8193	0.9834	**0.9670**	**0.9879**
STARE	Li et al.[7]	2015	-	0.7726	0.9844	0.9628	0.9879
	Yan et al.[28]	2018	-	0.7735	0.9857	0.9638	0.9833
	IterNet [30]	2019	0.8146	0.7715	**0.9886**	**0.9701**	**0.9881**
	MAU-Net[15]	2020	0.7826	0.7536	0.9808	0.9581	0.9721
	Proposed(ECU-Net)	2020	**0.8337**	**0.8127**	0.9840	0.9661	0.9877

5. Conclusion

Automatic segmentation of retinal vessels has great application value in modern medicine. Based on the existing retinal vessel segmentation methods and deep learning technology, this paper proposes an efficient retinal vessel segmentation model ECU-net. The main contributions are as follows:

1. We propose an edge attention module to enhance the model's activation degree to the strip distribution of blood vessels. Combined with the efficient channel attention module and residual structure, we design a residual attention module for feature extraction, making the model have a stronger recognition ability to retinal blood vessels.

2. We propose an up-sampling fusion module for information fusion, which significantly improves the fusion efficiency of low-level texture information and high-level semantic information.

3. We construct a multi-scale channel attention module to improve feature transfer efficiency to achieve efficient feature extraction and reduce the over-fitting risk.

Compared with other state-of-the-art retinal vessel segmentation models based on deep learning, our ECU-Net has better comprehensive performance on public retinal datasets.

There are still some problems to be solved. The segmentation accuracy of ECU-Net for the main blood vessels is very good, but there is still the problem of blood vessel fracture in the segmentation image, which is more obvious in the case of lesion interference. In the future work, we will try to combine the traditional vessel segmentation methods, such as vessel tracking and template matching, to reduce the problem of false recognition of the lesion area

and the fracture of the small vessel segmentation so as to further improve the segmentation accuracy of the model.

Declarations:

Funding:

This research was supported by the National Natural Science Foundation of China (61573182, 62073164), and by the Fundamental Research Funds for the Central Universities (NS2020025).

Conflicts of interest/Competing interests:

The authors claim no conflicts of interest.

Availability of data and material:

All data generated or analysed during this study are included in this submitted article.

Code availability:

Some or all code generated or used during the study are available from the corresponding author by request. (Xin Yang: yangxin@nuaa.edu.cn)

Authors' contributions:

Xin Yang constructed the ECU-Net model for retinal vascular segmentation and wrote the manuscript. Zhiqiang Li and Yingqing Guo carried out all the experiments and collected the experimental data. Dake Zhou and Tao Li gave useful opinions on the experimental method.

Ethics approval:

Not applicable.

Consent to participate:

Not applicable.

Consent for publication:

Not applicable.

Reference:

[1] Michael D. Abràmoff, MD, PhD; James C. Folk, MD; Dennis P. Han, MD; et al. Automated analysis of retinal images for detection of referable diabetic retinopathy: importance [J]. JAMA Ophthalmology, 2013, 131(3):351-357.

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

[3] Can A, Shen H, Turner J N, et al. Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms[J]. IEEE Transactions on information Technology in Biomedicine, 1999, 3(2): 125-138.

[4] Dash J, Bhoi N. Retinal Blood Vessel Extraction Using Morphological Operators and Kirsch' s Template[M]//Soft Computing and Signal Processing. Springer, Singapore, 2019: 603-611.

[5] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[6] Alom M Z, Hasan M, Yakopcic C, et al. Recurrent residual convolutional neural network based on U-Net (r2U-Net) for medical image segmentation[J]. arXiv preprint arXiv:1802.06955, 2018.

[7] Li Q, Feng B, Xie L P, et al. A cross-modality learning approach for vessel segmentation in retinal images[J]. IEEE transactions on medical imaging, 2015, 35(1): 109-118.

[8] Ngo L, Han J H. Multi-level deep neural network for efficient segmentation of blood vessels in fundus images[J]. Electronics Letters, 2017, 53(16): 1096-1098.

[9] Feng S, Zhuo Z, Pan D, et al. CcNet: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features[J]. Neurocomputing, 2020, 392: 268-276.

[10] Wu Y, Xia Y, Song Y, et al. Multiscale network followed network model for retinal vessel segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018: 119-126.

[11] Wu Y, Xia Y, Song Y, et al. Vessel-Net: retinal vessel segmentation under multi-path supervision[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 264-272.

[12] Araújo R J, Cardoso J S, Oliveira H P. A deep learning design for improving topology coherence in blood vessel segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 93-101.

[13] Yu F, Zhao J, Gong Y, et al. Annotation-free cardiac vessel segmentation via knowledge transfer from retinal images[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 714-722.

[14] Zhang Y, Chung A C S. Deep supervision with additional labels for retinal vessel segmentation task[C]//International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2018: 83-91.

[15] Li H, Wang Y K, Wan C, et al. MAU-Net: A Retinal Vessels Segmentation Method[C]//2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020: 1958-1961.

[16] Wang B, Qiu S, He H. Dual Encoding U-Net for Retinal Vessel Segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 84-92.

[17] Zhuang J. Laddernet: Multi-path networks based on U-Net for medical image segmentation[J]. arXiv preprint arXiv:1810.07810, 2018.

[18] Zhao H, Qiu X, Lu W, et al. High-quality retinal vessel segmentation using generative adversarial network with a large receptive field[J]. International Journal of Imaging Systems and Technology, 2020, 30(3): 828-842.

[19] Biyani R S, Patre B M, Kulkarni U V. Retinal vessel segmentation using enhanced fuzzy min-max neural network[J]. Multimedia Tools and Applications, 2019, 78(24): 35053-35073.

[20] Kromm C, Rohr K. Inception Capsule Network for Retinal Blood Vessel Segmentation and Centerline Extraction[C]//2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020: 1223-1226.

[21] Wang Q, Wu B, Zhu P, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11534-11542.

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

[23] Owen C G, Rudnicka A R, Mullen R, et al. Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program[J]. Investigative ophthalmology & visual science, 2009, 50(5): 2004-2010.

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

[25] Melinščak M, Prentašić P, Lončarić S. Retinal vessel segmentation using deep neural networks[C]//10th International Conference on Computer Vision Theory and Applications (VISAPP 2015). 2015.

[26] Zuiderveld, Karel. "Contrast limited adaptive histogram equalization." Graphics gems (1994): 474-485.

[27] Liskowski P, Krawiec K. Segmenting retinal blood vessels with deep neural networks[J]. IEEE transactions on medical imaging, 2016, 35(11): 2369-2380.

[28] Yan Z, Yang X, Cheng K T T. A three-stage deep learning model for accurate retinal vessel segmentation[J]. IEEE journal of biomedical and health informatics, 2018.

[29] Azad R, Asadi-Aghbolaghi M, Fathy M, et al. Bi-directional ConvLSTM U-net with Densley connected convolutions[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. 2019: 0-0.

[30] Li L, Verma M, Nakashima Y, et al. Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks[C]//The IEEE Winter Conference on Applications of Computer Vision. 2020: 3656-3665.

[31] Wu Y, Xia Y, Song Y, et al. Multiscale network followed network model for retinal vessel segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018: 119-126.