Noname manuscript No.
(will be inserted by the editor)

A Multi-scale Channel Attention Network for Retinal Vessel Segmentation

Xuejian Li[1] · Jiaqi Ding[1] · Jijun Tang[1,3,4] · Fei Guo[2,*]

Received: date / Accepted: date

Abstract Retinal diseases can be found timely by observing retinal fundus images. Blood vessels are extracted from retinal images is an important part. Most of previous methods based on deep learning cared more about accuracy and ignored complexity of the model for segmenting retinal vessels, which makes these methods difficult to apply to medical equipment. Besides, due to the great difference in the width of retinal vessels, some methods can't extract blood vessels of different widths at the same time well. So we propose a new lightweight network, called Res2Unet. It applies a multi-scale strategy to extract blood vessels of different widths and integrates this strategy into the channel to greatly reduce parameters and computation resource. Res2Unet also applies channel attention mechanism to promote the communication between channels and recalibrate channel features. Then, we propose two postprocessing methods to excavate hidden blood vessels and improve detection accuracy respectively. On DRIVE dataset, the Dice, IOU and AUC of our Res2Unet reach 0.8186, 0.6926 and 0.9772 respectively better than Unet with 0.8109, 0.6817 and 0.9751. Importantly, the calculation amount of Res2Unet is about one third of Unet. It means that Res2Unet has smaller hardware requirements.

Keywords Retinal vessel segmentation · multi-scale strategy · lightweight network · channel attention

1 Introduction

According to the statistics on International Diabetes Federation (IDF) [1], about 450,000 people are blind every year in the world, and a new case of blindness appears almost every minute. In many blinding diseases, retinal vascular disease is one of the most important causes. Diseases, such as diabetic retinopathy, arteriosclerosis and leukemia, will affect the length, width and angle of the fundus vessel, shown in Fig.1. At present, doctors use retinal fundus images to manually segment vessels for screening and diagnosis. However, manual segmentation depends on the technical experience of the medical experts, and the efficiency is low. With the development of artificial intelligence, computer-assisted segmentation of retinal vessels becomes necessary.

At present, many researches [2–7] have been carried out on the segmentation of retinal vessels. A number of effective methods have been proposed. Compared with other traditional methods, machine learning methods can unearth latent features that are not found. Currently, it achieves the highest accuracy and the best segmentation effect.

Machine learning methods can be divided into the learning methods of supervised and unsupervised. Supervised learning methods: In 2016, Qiaoliang Li et

Xuejian Li
E-mail: lixuejian@tju.edu.cn
Jiaqi Ding
E-mail: 13256683086@163.com
✉Jijun Tang
E-mail: tangjijun@tju.edu.cn
✉Fei Guo
E-mail: guofeieileen@163.com

[1] School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin 300350, China.
[2] School of Computer Science and Engineering, Central South University, Changsha 410083, China
[3] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China.
[4] Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208, USA

Fig. 1: **The retinal fundus images.** From left to right: the retinal fundus color image of healthy people, the retinal fundus color image of diabetic patients, the result of manual vascular segmentation for healthy retinal image.

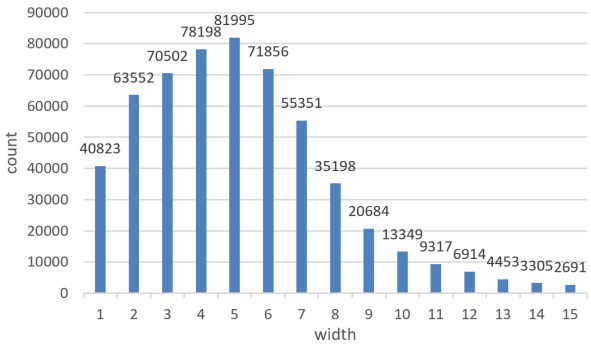

Fig. 2: **The number of retinal blood vessels with different widths.** We use pixel as the basic unit to calculate the blood vessel width. We set the blood vessel width greater than 15. Vessels with a width less than 4 are considered thin vessels.

al.[8] transformed the segmentation task into a cross-modal conversion problem. First, the denoising autoencoder was used to explore the relationship between the retinal images and the labels, and then the pre-trained weights were used as the initial weights of the neural network. Sheng et al. [9] used superpixels as the basic unit of segmentation and used the minimum spanning superpixel tree to explore the structure of vessels. Rodrigues et al. [10] extracted multi-modal features based on grey level and vessel connectivity. And then pixels were classified by Multiple classifier and region growing method. Ma et al. [11] used the coarse segmentation network to generate the primary map and used the fine network to optimize it. Wang et al. [12] employed hard attention. They used three decoder to distinguish hard and easy segmentation regions and segment the hard and easy regions respectively. Li et al. [13] used attention to focus on hard segmentation regions. More methods like this, such as: [14–19]. Unsupervised learning methods: Xie et al. [20] proposed a genetic algorithm combined with Fuzzy C-means Clustering (FCM) method in 2013. They used genetic algorithms to cluster analysis and determine the initial value of FCM. Then, they used the FCM to find the global optimum. Later, Sohini et al. [21] used the adaptive thresholding method to extract the initial vessel after enhancing vessel images. Then, they eliminated the existing vessels from the enhanced vessel images and then used the same method to identify new vessels iteratively. More similar methods: [22–24].

Above methods have got good results on the DRIVE and STARE datasets. Especially, the top-ranked methods have achieved satisfying results recently, such as: [25–29]. However, these methods paid more attention to improving accuracy and ignored the complexity of the proposed model to a certain extent. This makes these methods difficult to apply to medical equipment in future. So we apply the idea of classic lightweight networks [30,31], that is, using group convolution instead of standard convolution. This reduce model params and computation greatly. Besides, we count the number of retinal vessels of different widths, shown in Fig.2. We can see that the width of the blood vessels is quite different, and the thin vessels account for a large proportion. So, the features of vessels with different widths can't be obtained at the same time by using a single-scale convolutional neural network (CNN). If a single-scale CNN acquires a large receptive field, it is helpful to extract the features of thick vessels, but the information of thin vessels will be lost. If the receptive field is small, it is beneficial to extract the features of thin vessels, but it is bad for segmenting thick vessels. It is the reason why most of methods can't segment retinal vessels well. Therefore, we adopt a multi-scale strategy to extract retinal vessels. However, the classic multi-scale branch structure of Inception [32] requires manual design, and the calculation amount is large. So we apply Res2Net [33], a multi-scale strategy on the channels. Res2Net applied the idea of group convolution, and different groups can get a different receptive fields. So it can extract the features of retinal vessels with different widths at the same time. Besides, we apply the channel attention mechanism to learn the channel features of different scales in Res2Net for recalibrating the channels. Besides, due to Res2Net used group convolution, resulting in a lack of communication between groups. So the channel attention mechanism is also used to promote the information communication between groups. Compared with other methods, our method has less params and computation, and can perform better segment retinal vessels of different widths.

Our contribution can be summarized as follows: first, we combine the Unet [34], Res2Net [33] and channel attention mechanism [35] and propose a lightweight net-

work for segmenting retinal vessels, called Res2Unet. It can segment retinal vessels of different widths well with less params and computation. Second, we propose two postprocessing methods for different purposes. One is used to extract more retinal vessels, the other is applied to remove false positives and improve overall performance.

Our paper can be divided into five parts. In the first part, we introduce some classic methods for segmenting retinal vessels and our motivation. In the second part, we introduce the development of deep learning in retinal vessels and the researches related to our methods. In the third part, our network, preprocessing and postprocessing methods are introduced. In the fourth part, we introduce datasets, training parameters and evaluation metrics. Experimental results and discussion are shown in the fifth part.

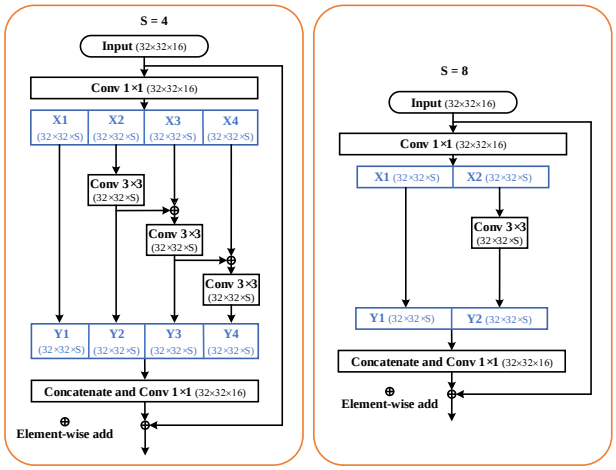

Fig. 3: **The structure of Res2Net.** Supposing the shape of input feature is $32 \times 32 \times 16$, which represent height, width and channel respectively. S is defined by us represents S channels as a group. The number of groups depends on the input channels and S.

2 Related Work

2.1 U-Net

In 1998, LeCun proposed the LeNet [36]. LeNet used convolution layers, pooling layers and fully connected layers. The basic architecture of CNN was decided from then on. In 2012, Hinton and Alex Krizhevsky proposed AlexNet [37]. They used Relu as activation function in CNN for the first time and proposed LRN layer to create a competitive mechanism for the activities of local neurons. In 2014, VGG network [38] was proposed and had an important effect on deep learning. VGG used 3×3 convolution kernel and 2×2 pooling kernel, which reduced computation and improved the ability to extract details and the generalization ability. In 2015, FCN [39] was proposed as the first end-to-end semantic segmentation network. In order to restore the feature maps to the size of input images. FCN used the deconvolution to up-sample the feature maps. So, the pixel-level classification could be realized. Based on FCN, Ronneberger proposed Unet. Unet can be divided into three parts: encoding, decoding, and skip connection. Unet extracted and screened features in encoding part and restored the encoded features to the same size as the input images in decoding part. The skip connection was used to compensate for the loss of information caused by encoding. In recent years, Unet has been used in the field of medical image segmentation widely. Most of networks were proposed based on Unet, such as [40–43].

2.2 Res2Net

With the GoogLeNet [44] was proposed, Inception structure attracted people's attention gradually. GoogLeNet was composed with a plurality of Inception connected in series. Inception was formed by the different sizes of the convolution kernel and pooling kernel in parallel. It could extract different scale features. The deeper the model, the more information the network extracted, but it caused the problems of gradient explosion and gradient disappearance at the same time. ResNet [45] solved this problem successfully. Even the network with hundreds of layers still had a good performance. Based on the Inception and ResNet idea, Res2Net was proposed. Res2Net applied the multi-scale strategy along the channel. The structure of Res2Net was shown in Fig.3. Firstly, the features were grouped by channels, and then the convolution operation was performed on each group of features. And the result of the convolution was combined with the next group of features. The results of every convolution were concatenated, and then 1×1 convolution is used to fusion. Finally, the fusion result was combined with the original input feature to get the final output. Whenever passed the 3×3 convolution, receptive field was expanded correspondingly. Thereby it realized multi-scale strategy on different groups. We define S in details in section 3.2.

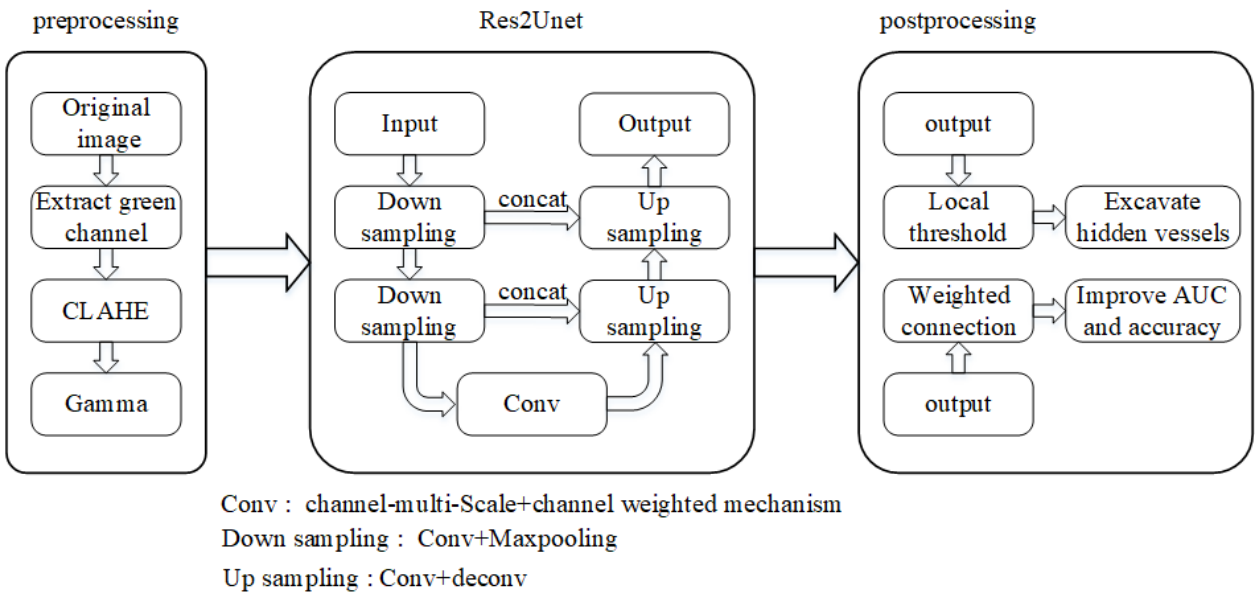

Fig. 4: **The overview of our proposed methods.** Our method is consisted with preprocessing, proposed network and postprocessing.

3 Methods

The overview of our proposed methods are shown in Fig.4. Firstly, the CLAHE and gamma algorithms are used to preprocess input data. Then, we design a novel lightweight network based on Unet. Our network combines channel-multi-scale strategy and channel attention mechanism to improve the ability of extracting retinal vessels. Final, We use two postprocessing methods for different purposes. This weighted connection method (WCM) can improve the overall effect of our model, but it also loses some blood vessels at the same time. We can adopt different postprocessing methods for different tasks. If the overall effect is pursued, the WCM method can be adopted. If some tasks need to extract more retinal blood vessels, the local threshold method (LTM) can be used. It is also possible to combine WCM and LTM to achieve better overall results.

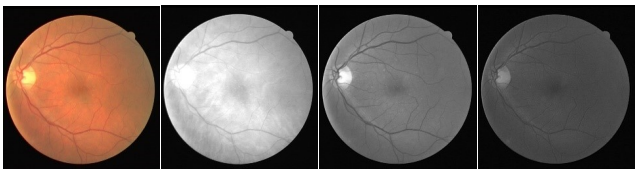

Fig. 5: **color image and R, G, B image.** From left to right: color image, Red channel image, Green channel image and Blue channel image.

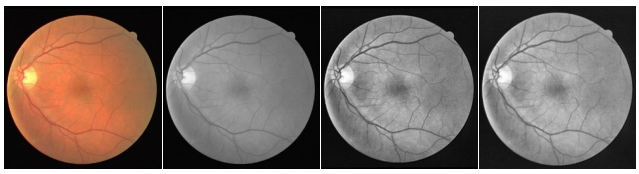

Fig. 6: **Original image and preprocessed images.** From left to right: Original image, Green channel image, the image processed by CLAHE and Gamma

3.1 Preprocessing

In order to accelerate the training, we preprocess the original images. As shown in Fig.5, compared with Red image and Blue image, the green channel image has a great contrast and the retinal vessels is clearer. Besides, we conduct experiments in section 5.3 to compare color image and green channel image. Experimental results show that the results obtained by using the two images as input are similar. Because most of preprocessing method is used for gray image, each channel of color image needs to be preprocessed separately. So, the time consumption of color image in preprocessing is three times that of green channel image, which increases burden. So, we choose green channel image for subsequent processing. No standardized image data is uneven. It is difficult for training. Therefore, we standardize the green channel images as follows.

$$X_{stand} = \frac{X_{input} - \mu}{\sigma} \tag{1}$$

$$X_{norm} = \frac{X_{stand} - X_{stand\_min}}{X_{stand\_max} - X_{stand\_min}} \quad (2)$$

where X_{input} represents the input images, μ represents the mean of all the images, σ is the standard deviation of all the images. X_{norm} is the normalized images. Afterwards, the CLAHE algorithm is used to process the standardized images for improving the quality and contrast. Finally, the gamma algorithm is used to enhance the contrast and dynamic range of images. The preprocessed images are shown in Fig.6. It can be seen that the grayscale distribution of the preprocessed image is more uniform, and the retinal vessels are clearer.

3.2 Res2Unet

Our network applies Unet as the backbone, shown in Fig.7. It can be divided into three parts: encoder, decoder, and skip connection. The encoder part uses maxpooling layers to downsample for expanding receptive field and extracting rich semantic features. The number of convolution kernels are doubled after each maxpooling layers. On one hand, because downsampling causes the receptive field of the convolution kernel to increase, more convolution kernels are required to extract features. On the other hand, as the network is deepened, more convolution kernels are needed to express high-level semantic information. The decoder part uses deconvolution to upsample for restoring the feature maps to the size of input images. The skip connection is used between encoder and decoder to compensate for the loss of information caused by maxpooling.

The original Unet applies single-scale convolution, so it can't well extract the features of retinal vessels with different widths at the same time. Therefore, we use Res2Net instead of single-scale convolution. As shown in Fig.3, Res2Net uses a group convolution strategy, and the latter group features combines the previous group features for convolution. Whenever passed the 3×3 convolution, receptive field was expanded correspondingly. Therefore, each group of Res2Net gets different scale receptive fields, which can extract multi-scale information. It is advantageous for extracting retinal blood vessels with different widths. The S is defined by us represents S channels as a group. Therefore, the number of groups can be counted as follows:

$$Group\_num = \frac{channel\_num_{input}}{S} \quad (3)$$

where $channel\_num_{input}$ represents the number of input channels, $Group\_num$ represents the number of groups. For example, as shown in Fig.3, $S = 4$ means to divide the 16 channels into 4 groups, $S = 8$ means to divide the 16 channels into 2 groups. Different S have different effects, we discuss it in section 5.1. Besides, because Res2Net uses group convolution, it has less params and computation. But, there is a lack of information exchange between different groups. So, the channel attention mechanism is used to learn the relationship between channels. On one hand, it can promote the communication between them, and on the other hand, it can recalibrate the channel features. In the original Res2Net paper, they tried to add SE block [35] to model the relationship between channels for image classification tasks. SE block learns the relationship between channels through fully connected layers and recalibrates the channel features. In our task, each group of Res2Net extracts retinal vessel features of different scales, so SE block can well promote the exchange of information at different scales. This is very suitable for our task. Therefore, we use SE block as the channel attention mechanism. We discuss the role of SE block for segmenting retinal vessels in the section 5.2.

Our Res2Unet combines Unet, Res2Net and SE block can greatly reduce params and computation while maintaining the effect. It means smaller hardware requirements and is beneficial for applications in medical equipments in future. In addition, our Res2Unet applies a multi-scale strategy, which is beneficial for extracting retinal vessels of different widths and diagnosing diseases.

3.3 Channel attention mechanism

SE Block has achieved satisfying results for other tasks recently, such as [46,47]. The structure of SE Block is shown in Fig.8. The format of input feature is $(C \times H \times W)$, where C, H and W are the depth, height and width of the feature, respectively. SE Block uses three steps to recalibrate the input features. First, using global average pooling to compress the features in order to obtain the global distribution of the response on the channel. After that, two fully connected layers are used as a bottleneck to model the correlation between channels. Using two fully connected layers has a great non-linearity and can better fit the correlation between channels. The first fully connected layer is used to compress dimension (we compress it 16 times) for reducing the params and computation. Finally, the result of fully connected layers get through sigmoid to obtain the weight of each channel feature. All weights are element-wise multiplied by the input features to weight every channel feature.

SE Block can recalibrate the input features and make the network to learn the importance of each channel features by itself. Besides, due to Res2Net uses group

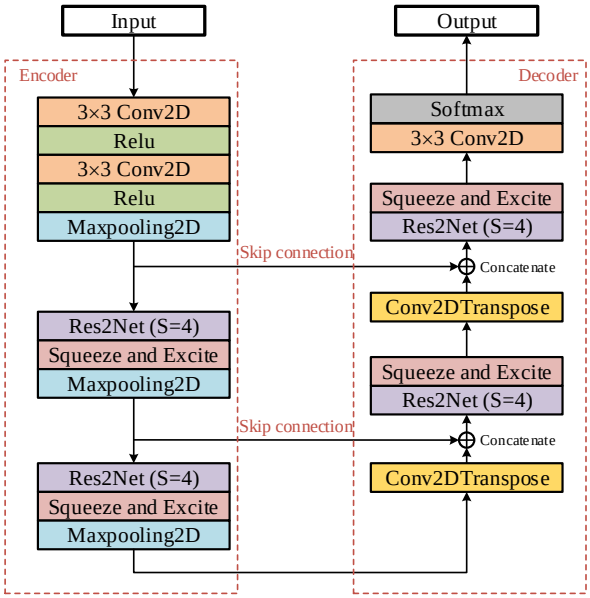

Fig. 7: **The structure of Res2Unet.** Our network is consisted with encoder, decoder and skip connection. Squeeze and Excite means channel attention mechanism. S represents s channels as a group. Our Res2Unet applies Res2Net and channel attention mechanism to extract features instead of standard convolution.

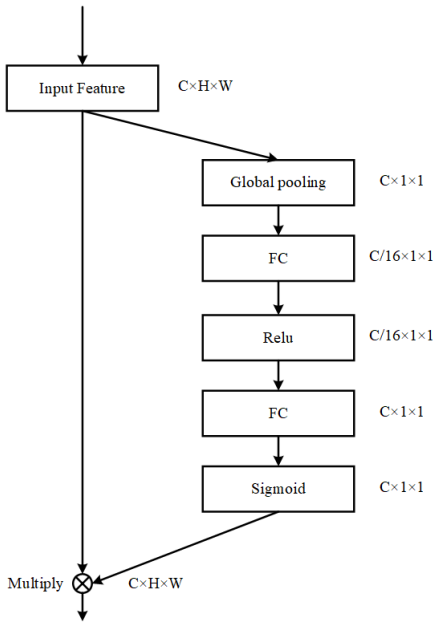

Fig. 8: **The structure of SE Block.** SE Block can be divided into two parts. One consists of global average pooling, fully connected layers, Relu and Sigmoid for generating weights. And the other applies shortcut. Finally, a element-wise multiplication is used to weight.

convolution, there is a lack of information communication between groups. So, the Se block is also used to promote the communication between them.

3.4 Postprocessing

We propose two postprocessing methods for different segmentation purposes. One method is used to search the hidden vessels that have not been found previously. And the other is applied to improve the accuracy and remove false positives. We discuss them separately.

3.4.1 Local threshold method

In order to excavate hidden retinal vessels, we propose the local threshold method (LTM). We use a fixed threshold (it is set as 0.5 and named P_0) to binarize the probability maps that are output by Res2Unet. But the retinal images have some low-contrast areas, which may get a low probability. When we use P_0 to binarize the probability maps, these areas will don't be segmented well. Therefore, many retinal vessels that have been segmented by the previously proposed methods may be broken. Especially, the thin vessels (the width of vessels less than 4 is classified as thin vessels in our methods) can't be segmented completely. So, we propose the LTM to search hidden retinal vessels.

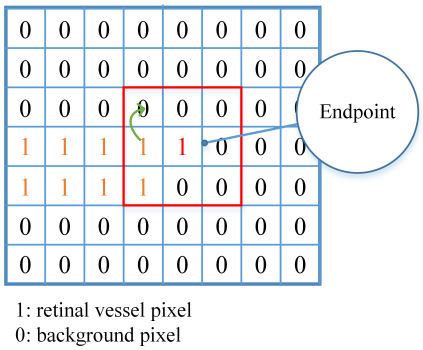

Fig. 9: **The method of finding endpoint.** 1 represents retinal vessel pixel; 0 represents background pixel; 1 marked in red is the candidate endpoint. In the eight fields of the candidate endpoint, the green arrow represents clockwise transitions from 1 to 0.

The local threshold method mainly focuses on the broken areas of retinal thin vessels, because most of broken areas are from thin vessels. Firstly, we use the P_0 to binarize the probability maps for obtaining the binarized segmentation maps. Then we seek all endpoints of

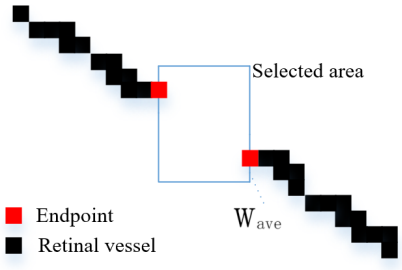

Fig. 10: **Local threshold method.** All points marked in red mean endpoints. The blue area is selected to search hidden blood vessels. In this area, we use a threshold that is less than 0.5 (it is set as 0.4) to excavate hidden vessels.

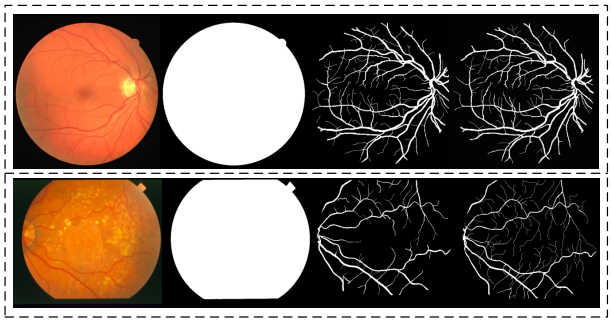

Fig. 11: **Sample images from DRIVE and STARE datasets.** From top to bottom: the DRIVE dataset and STARE dataset. From left to right: the color image, FOV mask, first expert label and second expert label.

thin vessels that are named V_0. The method of searching the endpoint is shown in Fig.9. We calculate the number of clockwise transitions from 1 to 0 in eight fields of the candidate point. If the number of transitions is less than 2, the candidate point will be classified into endpoint. After searching all the endpoints, we calculate the Euclidean distance between two endpoints. When the distance is less than the distance threshold (we set the distance threshold to 20), we suppose that the hidden vessels may exist between the two endpoints. Then we calculate the width of vessels at two endpoints, which is named W_0 and W_1. After that we calculate the average width W_ave. We define a rectangular area by the Wave. In this area, we use a threshold that is less than 0.5 (we recommend that the threshold range is 0.4 ± 0.05, which is discovered through our many observations and tests on different probability maps. we choose 0.4 in this paper) to excavate hidden vessels. Pixels with a probability greater than this threshold are regarded as hidden vessels. Local threshold method is shown in Fig.10.

3.4.2 Weighted correction method

In order to improve the accuracy and segmentation performance of our network, we propose the Weighted Correction Method (WCM). There is a serious imbalance between the background and vessels in retinal images, the ratio is about 9 : 1. Besides, we count the number of retinal blood vessels with different widths, as shown in Fig.2. It shows that the widths of retinal vessel also have a great difference. However, Unet adopted single-scale convolution strategy. So Unet can only obtain features of one scale in a convolution process. Due to the high proportion of background pixels, Unet tends to learn more background features and ignore blood vessel features. Res2Unet applies a multi-scale convolution strategy. So, Res2Unet has different scales of receptive field, which can extract the features of blood vessels with different widths. But Res2Unet will produce some false positives inevitably. Similarly, Unet will also inevitably produce many false negatives. Therefore, we combine their probability maps, and take the average of the two probability maps as the final probability maps. The calculation process is as follows:

$$P_{final} = \frac{P_{Unet} + P_{Res2Unet}}{2.0} \tag{4}$$

where P_{final} represents the probability maps calculated by the WCM method, P_{Unet} represents the probability maps output by Unet, and $P_{Res2Unet}$ represents the probability maps output by Res2Unet. The WCM method makes full use of the advantages of two networks and fuses the probability maps output by the two networks to achieve learning from each other.

4 Experiment

4.1 Datasets

DRIVE dataset was often used for segmenting retinal images, which contained 40 color fundus images with the size of 565×584. DRIVE dataset has 20 images for training and 20 images for testing. As shown in Fig.11, each image had a binarized FOV mask for showing the effective area. There are no blood vessels outside the mask, so we think is an invalid area. So, our training, testing and calculation metrics are all inside the FOV mask. In addition, the training set had the label marked by the first expert, and the test set contained the labels marked by the first expert and the second expert. We choose the label marked by the first expert as the gold standard.

STARE dataset contained 20 retinal fundus images, which included 10 pathological images. The size of 20

images was 700×605. STARE dataset didn't specify the set of training and testing. Similar to previous method, we select 15 images for training and 5 images for testing. This dataset doesn't provide FOV masks, so we apply the masks provided by [48]. We also select the label marked by the first expert as the gold standard.

4.2 Training Parameters

We set the clahe_equalized and tileGridSize of the CLAHE algorithm to 2.0 and 8.0, respectively. We set the gamma to 1.0. These parameters come from [49]. CNN requires a number of training data to train the model, but the retinal fundus images are less, so we expand the training data firstly. We randomly select 4500 patches with the size of 64×64 on each fundus image. Each patch is segmented independently. In test, we first fill the test images to 640×640, and then divide each image into 100 patches with the size of 64×64. So, there is no overlap in test patches. Finally, we concatenate the segmentation maps of the test patches into a complete image for evaluation. Besides, we select the training images of 90% as a training set, the 10% as the validation set. Batch size is set to 32. Epoch is set to 100, about $250,000$ iterations. The validation set is evaluated once per epoch. If the loss on the validation set still is not dropped after 20 epochs, the learning rate will be decreased to the original 0.1 times. We use categorical crossentropy as loss function. We apply Adam as optimizer network and set initial learning rate to $3e^{-4}$. Dropout is set to 0.2. Finally, our training was completed on the Google Colab platform and applied TESLA P100.

4.3 Evaluation Metrics

In order to evaluate the performance of our model objectively, we introduce four evaluation metrics. TP: The actual label is vessel pixel, and the prediction is vessel pixel; FN: The actual label is vessel pixel, but the prediction is background; TN: The actual label is background, and the prediction is background; FP: The actual label is background, but the prediction is vessel pixel.

Based on above four metrics, we introduce specificity (Sp), sensitivity (Se), Dice and intersection over union(IOU) to evaluate our methods. Evaluation metrics are defined as follows:

$$Se = \frac{TP}{TP + FN} \tag{5}$$

$$Sp = \frac{TN}{TN + FP} \tag{6}$$

$$Pr = \frac{TP}{TP + FP} \tag{7}$$

$$Dice = \frac{2.0 \times Pr \times Se}{Pr + Se} \tag{8}$$

IOU is defined as follows:

$$IOU = \frac{P \cap L}{P \cup L} \tag{9}$$

where P represents binarized segmentation maps, and L represents labels. In addition, we plot the ROC curve according to the correlation between Se and 1-Sp, and the area of under the ROC curve (AUC) reflects the quality of our model. AUC is closer to 1, which indicates the best segmentation performance of network. We pool all the predicted images together, and then calculate these metrics if there are no special instructions.

5 Results And Discussion

Firstly, we discuss the number of groups in Res2Net and its impact on model performance. The more groups there are, the more scale features Res2Net can extract. It is beneficial for segmenting retinal blood vessels of different widths. But it reduces the communication and fusion between groups. Secondly, we discuss the role of the channel attention mechanism. Thirdly, we discuss the importance of preprocessing. Fourthly, we compare our Res2Unet with Unet and classic improved Unet to illustrate the advantages of our model. Fifthly, we discuss the effectiveness of our proposed postprocessing method. Finally, we compare Res2Unet with previous classic methods. Because our train, test and calculated metrics are all inside the FOV mask, we only compare the method that use the same strategy. In the following experiments, we use the same preprocessing, experimental parameters, and no postprocessing if there is no special statement.

5.1 Hyper-Parameters Evaluation

Res2Net uses the strategy of channel-multi-scale, which not only reduces the params and computation of our model, but also facilitates the extraction of retinal vessels of different widths. The smaller S is, the more groups there are, and the more receptive fields of different scales can be obtained by Res2Net, which is more beneficial for extracting retinal vessels of different widths. But it reduces the communication of features between groups. We discuss the influence of S.

In order to ensure that the number of groups is an integer. So the values of S are set to $[2, 4, 8, 16]$. On the DRIVE dataset, the impact of different S is showed in

Table 1: Results of Res2Unet with different S on the DRIVE dataset.

Methods	Dice	IOU	Se	Sp	AUC	Params
Res2Unet ($S = 2$)	0.8143	0.6865	0.7981	**0.9764**	0.9748	**0.136M**
Res2Unet ($S = 4$)	**0.8168**	**0.6900**	**0.8162**	0.9734	**0.9762**	0.136M
Res2Unet ($S = 8$)	0.8161	0.6890	0.8098	0.9745	0.9755	0.142M
Res2Unet ($S = 16$)	0.8147	0.6869	0.8128	0.9734	0.9756	0.155M

Table 2: Comparison of Res2Unet with SE Block and without SE Block on the DRIVE dataset.

Methods	Dice	IOU	Se	Sp	AUC
without SE Block	0.8135	0.6852	0.7879	**0.9782**	0.9742
with SE Block	**0.8168**	**0.6900**	**0.8162**	0.9734	**0.9762**

Table 3: Res2Unet for different input.

Input Images	Dice	IOU	Se	Sp	AUC	Params	GFlops
Color images	**0.8169**	**0.6901**	0.7986	**0.9771**	**0.9763**	0.137 M	0.306
Green channel images	0.8168	0.6900	**0.8162**	0.9734	0.9762	**0.136 M**	**0.302**

Table 4: The importance of preprocessing. Res2Unet as the segmentation network.

Methods	Dice	IOU	Se	Sp	AUC
Without preprocessing	0.8146	0.6868	0.8083	**0.9743**	0.9738
With preprocessing	**0.8168**	**0.6900**	**0.8162**	0.9734	**0.9762**

TABLE 1. The Se of $S = 4$ reaches 0.8168, which is best than other S. This shows that $S = 4$ can segment more blood vessels, which is beneficial for the diagnosis of retinal diseases. The Sp of $S = 4$ is 0.9734 lower than other S, which shows that it also produces some false positives. But the Dice, IOU, and AUC of $S = 4$ reach 0.8168, 0.6900 and 0.9762 respectively, which are the best. In addition, the params of $S = 4$ is less, which is more beneficial for applications in medical equipment in future. We use $S = 4$ in the following all experiments.

5.2 Evaluation of channel attention mechanism

We evaluate the channel attention mechanism on the DRIVE dataset. Results are shown in TABLE 2. It is observed that the Dice, IOU, AUC of Res2Unet with SE Block reach 0.8168, 0.6900, 0.9762 respectively better than without SE block. This means that although Res2Net can extract more retinal blood vessels, the lack of information communication between channels causes the retinal vessel features of different scales to not be fused in time. SE block effectively alleviates this problem. So, we apply the SE Block after Res2Net is very significant. In this paper, Res2Unet includes SE block by default.

5.3 Preprocessing Evaluation

First, We discuss the case of color images and green channel images as our input. We apply CLAHE and gamma algorithm for the two input images to enhance the images. We use our Res2Unet as the segmentation network. The experimental results are shown in 3, the Dice, IOU and AUC of color images reach 0.8169, 0.6901 and 0.9763 respectively, which are simliar as green channel images. This is because the quality of red channel images and blue channle images are poor. It can't provide more information for segmenting retinal vessels. The Se of green channel images reach 0.8162 higher than color images, indicating that green channel images are more sensitive to retinal vessels structure. It is beneficial for diagnosis. Besides, the methods of preprocessing are used for gray images, so each channel of color images needs to be preprocessed separately. The time consumption of color images in preprocessing is three times that of green channel images, which increases the burden. So we finally choose green channel images as our input.

Then, we discuss the importance of preprocessing. We use Res2Unet as the segmentation network. As shown in 4, the Se of Res2Unet with preprocessing reach 0.8162 is better than without preprocessing. This shows that

Table 5: Compare Res2Unet with Unet and improved Unet.

Methods	Dice	IOU	Se	Sp	AUC	Params	GFlops
Unet	0.8109	0.6817	0.7689	**0.9814**	0.9751	0.466M	0.793
Unet++ [50]	0.8163	0.6890	0.7916	0.9784	0.9757	2.205M	2.180
Attention Unet [51]	0.8138	0.6857	0.7849	0.9790	0.9756	0.487M	0.810
CENet [52]	0.8135	0.6854	0.7855	0.9787	0.9749	0.737M	0.958
R2Unet	0.8124	0.6838	0.7946	0.9764	0.9745	1.022M	1.590
ResUnet	0.8111	0.6819	0.7704	0.9812	0.9752	0.487M	0.836
Res2Unet	**0.8168**	**0.6900**	**0.8162**	0.9734	**0.9762**	**0.136M**	**0.302**

Table 6: Stability analysis for Res2Unet and Unet.

	Average			Standard deviation		
Methods	Dice	IOU	AUC	Dice	IOU	AUC
Unet	0.8102	0.6812	0.9761	0.0226	0.0323	0.0064
Res2Unet	**0.8161**	**0.6893**	**0.9771**	**0.0151**	**0.0217**	**0.0058**

Table 7: Res2Unet with postprocessing on the DRIVE dataset.

Methods	Dice	IOU	Se	Sp	AUC
Res2Unet	0.8168	0.6900	0.8162	0.9734	0.9762
Res2Unet+LTM	0.8168	0.6900	**0.8175**	0.9731	-
Res2Unet+WCM	0.8183	0.6921	0.7931	**0.9788**	0.9772
Res2Unet+LTM+WCM	**0.8186**	**0.6926**	0.7945	0.9786	-

Table 8: Res2Unet with postprocessing on the STARE dataset.

Methods	Dice	IOU	Se	Sp	AUC
Res2Unet	0.8038	0.6708	**0.8234**	0.9760	0.9802
Res2Unet+WCM	0.8120	0.6825	0.7947	**0.9827**	**0.9818**
Res2Unet+LTM+WCM	**0.8121**	**0.6827**	0.7958	0.9826	-

preprocessing can enhance the retinal vessels and make it easier to be segmented. Besides, the Dice, IOU and AUC of Res2Unet with preprocessing reach 0.8168, 0.6900 and 0.9762 respectively are higher than without preprocessin. Therefore, it is important that we perform preprocessing to improve image quality.

5.4 Comparison with Unet and improved Unet

In order to better evaluate our Res2Unet, we train the Unet and classic improved Unet under the same settings. As shown in TABLE 5, the Se, IOU of Res2Unet reach 0.8162, 0.6900 best than others. It means that Res2Unet has a stronger ability to extract retinal vessels of different widths. In other words, the channel-multi-scale strategy is meaningful. For the diagnosis of retinal diseases, the distribution of retinal vessels plays an important role. So, our Res2Unet is more beneficial for diagnosing. Besides, the Dice, AUC of Res2Unet reach 0.8168, 0.9762 respectively. It shows that our Res2Unet has a better overall performance. Importantly, the params, GFlops of Res2Unet are 0.136 million, 0.302 respectively. They are lowest than other classic networks. It means that our network has low requirements for hardware, which is beneficial for applications to medical equipment in future.

Besides, we verify the stability of Unet and Res2Unet. Before, We pool all the predicted images together, and then calculate these metrics. Now, we independently calculate the Dice, IOU and AUC of each image, and calculate their average and standard deviation, shown in TABLE 6. The average Dice, IOU and AUC of Res2Unet reach 0.8161, 0.6893 and 0.9771 higher than Unet. Besides, The standard deviation is lower than Unet for Dice, IOU and AUC. It means that Res2Unet is more stable.

We visualize the experimental results that are predicted by the Unet and Res2Unet on the DRIVE dataset, shown in Fig.12. Due to the difference in the width of

Table 9: Comparison with existing methods on the DRIVE and STARE datasets.

	DRIVE				STARE			
Methods	Dice	Se	Sp	AUC	Dice	Se	Sp	AUC
2nd Observer	0.7891	0.7746	0.9725	-	0.7530	0.9196	0.9443	-
Azzopardi [53]	-	0.7655	0.9704	0.9614	-	0.7716	0.9701	0.9563
Marin [48]	0.7690	0.7067	0.9801	0.9588	0.7531	0.6944	0.9819	0.9769
Zhao [54]	-	0.7420	0.9820	0.8620	-	0.7800	0.9780	0.8740
Zhou [55]	0.7942	0.8078	0.9674	-	0.8017	0.8065	0.9761	-
Dasgupta [49]	0.8074	0.7691	0.9801	0.9744	-	-	-	-
Fraz [56]	0.7465	0.7406	0.9807	0.9747	0.7747	0.7548	0.9763	0.9768
Yan [57]	0.8097	0.7653	0.9818	0.9752	-	0.7581	0.9846	0.9801
Zengqiang [58]	-	0.7631	**0.9820**	0.9750	-	0.7737	**0.9857**	**0.9833**
Unet	0.8109	0.7689	0.9814	0.9751	0.7958	0.7629	0.9836	0.9773
Res2Unet	0.8168	0.8162	0.9734	0.9762	0.8038	0.8234	0.9760	0.9802
Res2Unet+LTM	0.8168	**0.8175**	0.9731	-	0.8032	**0.8239**	0.9758	-
Res2Unet+WCM	0.8183	0.7931	0.9788	**0.9772**	0.8120	0.7947	0.9827	0.9818
Res2Unet+LTM+WCM	**0.8186**	0.7945	0.9786	-	**0.8121**	0.7958	0.9826	-

retinal blood vessels, it is difficult to extract the features of blood vessels with different widths at the same time. And most of the previous methods focused on segmenting the retinal background in order to obtain an better accuracy. Res2Unet can find the retinal vessels that are unfound by Unet. Especially Res2Unet can search more thin vessels. Besides, Res2Unet is able to connect the broken vessels and keeps the completeness of vessels. It shows that Res2Unet has the better ability to extract retinal vessel features.

5.5 Performance of postprocessing

Our postprocessing methods are tested on the DRIVE and STARE datasets, which are shown in TABLE 7 and TABLE 8. The LTM can improve Se obviously while keeping the original accuracy. As shown in Fig.13, The LTM can excavate more retinal blood vessels that haven't been discovered before. We mark the blood vessels in red, which is found by postprocessing. In some areas, we can even connect broken vessels. It plays an import role in the diagnosis of retinal diseases. In addition, we can improve the Dice, IOU and AUC of our network by the WCM method. This is more obvious on the STARE dataset. Dice and IOU increased about 1%. The AUC reach 0.9818 higher than without WCM. Combining LTM and WCM can achieve the best overall performance.

5.6 Compared with existing methods

We compare the existing methods for retinal vessel segmentation in recent years, shown in TABLE 9. The DRIVE dataset provides two manual segmentation maps of retinal vessels. For our Res2Unet, Dice, Se, SP, AUC are 0.8168, 0.8162, 0.9734, 0.9762 respectively when we use the first observer as the gold standard on the DRIVE dataset. These metrics are higher than the second observer significantly. It indicates that our methods can replace the manual segmentation of retinal vessels completely. On the STARE dataset, although the second observer's Se reaches 0.9196 higher than Res2Unet, our Dice reaches 0.8038 significantly better than the second observer's 0.7530. We have better overall performance.

Compared with previous methods, Res2Unet has the highest Se on the two datasets. Our network is greatly sensitive to retinal vessels. This means that it is meaningful for us to apply the channel-multi-scale strategy to extract retinal vessels of different widths. In addition, the Dice, AUC of Res2Unet reach 0.8168, 0.9762 respectively best than other methods on the DRIVE dataset. On the STARE dataset, our Res2Unet is second only to [58] for AUC. Finally, we propose two postprocessing methods. If we focus on extracting more retinal vessels, LTM algorithm can be used. If we focus on improving Dice and AUC, the WCM algorithm can be used. Combining LTM and WCM, our Dices can reach 0.8186 and 0.8121 respectively on the two datasets, which are best than the others.

6 Conclusion

Because there is a great difference in the width of retinal vessels. So most of the previous methods can't segment vessels of different widths well. For solving the problem, we use Res2Net, which applies a multi-scale convolution strategy and integrates the strategy on the channel. Small-scale receptive fields can well retain thin vessels, and large-scale can extract thick vessels well.

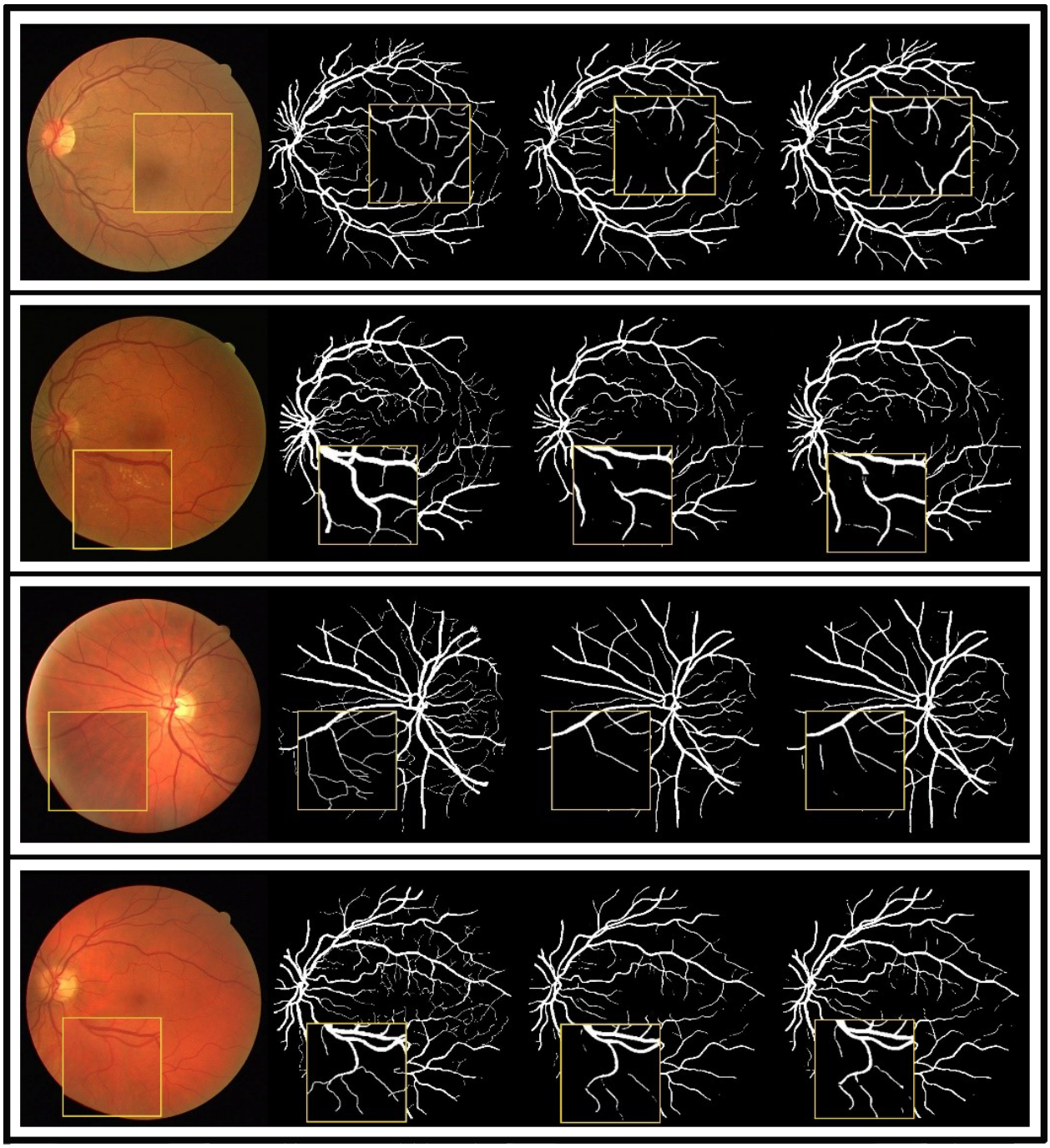

Fig. 12: **Predictions of Res2Unet and Unet on the DRIVE dataset.** From left to right: Original image, Ground-truth, Unet, Res2Unet. We show a local enlargement with a yellow box.

Since Res2Net uses group convolution, the features of different groups lack communication. Therefore, we add SE block after Res2Net. On the one hand, it promotes the communication between channel features. On the other hand, it recalibrates the channel features. Finally, we combine Res2Net, SE block and Unet to propose Res2Unet, a lightweight network. Our Res2Unet can extract more retinal blood vessels. Importantly, Res2Unet can greatly reduce params and computation while keeping performance. This is beneficial for applications to medical equipment in future. In addition, we propose two postprocessing methods. LTM is used to extract more vessels, and WCM is used to remove false positives and improve the overall performance of our model.

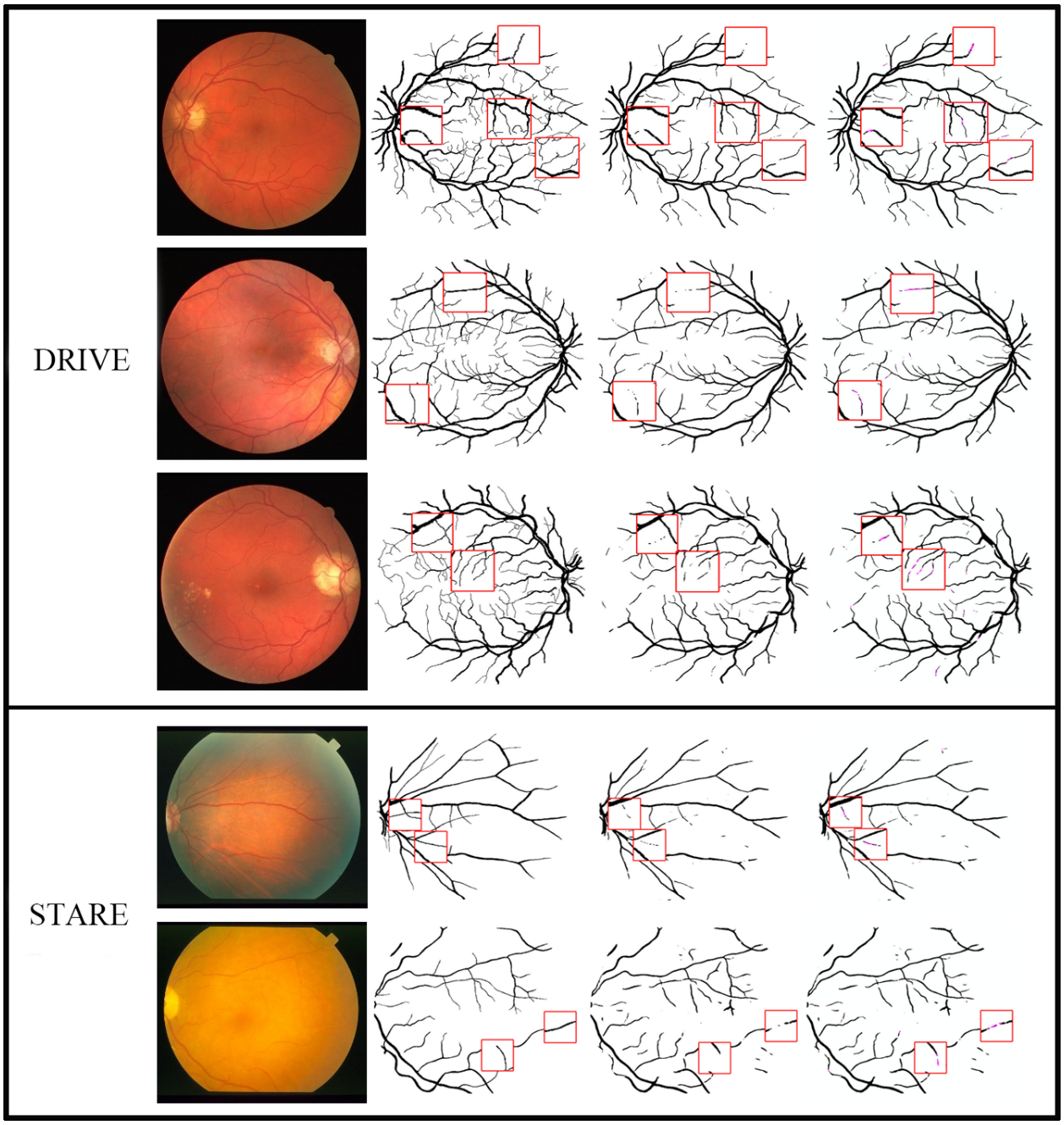

Fig. 13: **The WCM and LTM for postprocessing.** From left to right: original image, ground-truth, Res2Unet with WCM, Res2Unet with WCM and LTM. We show a local enlargement with a red box. The red vessels are the hidden ones we search by LTM.

But we don't pay more attention to comparing different preprocessing methods. What we use is a common preprocessing method for retinal fundus images. More methods, such as: [59,60] and so on. In our future work, we will discuss more preprocessing methods to achieve better results.

Acknowledgements This work is supported by a grant from the National Natural Science Foundation of China (NSFC 62172296) and National Key R&D Program of China (2020YFA0908400).

References

1. International Diabetes Federation. Idf diabetes atlas. Technical report, 6thed, Brussels, Belgium: International Diabetes Federation, 2014.

2. A. Salazar-Gonzalez, D. Kaba, Y. Li, and X. Liu. Segmentation of the blood vessels and optic disk in retinal images. *IEEE Journal of Biomedical and Health Informatics*, 18(6):1874–1886, 2014.
3. R. Annunziata, A. Garzelli, L. Ballerini, A. Mecocci, and E. Trucco. Leveraging multiscale hessian-based enhancement with a novel exudate inpainting technique for retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, 20(4):1129–1138, 2016.
4. José Ignacio Orlando, Elena Prokofyeva, and Matthew B. Blaschko. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE Transactions on Biomedical Engineering*, 64(1):16–27, 2016.
5. Y. Zhang and Acs Chung. Deep supervision with additional labels for retinal vessel segmentation task. *International Conference on Medical Image Computing & Computer-assisted Intervention*, pages 83–91, 2018.
6. Moreira Oliveira Américo Filipe, Mano Pereira Sérgio Rafael, and Batista Silva Carlos Alberto. Retinal vessel segmentation based on fully convolutional neural networks. *Expert Systems with Applications*, 112, 2018.
7. Yicheng Wu, Yong Xia, Yang Song, Yanning Zhang, and Weidong Cai. *Multiscale Network Followed Network Model for Retinal Vessel Segmentation*. Springer International Publishing, Cham, 2018.
8. Qiaoliang Li, Bowei Feng, Lin Pei Xie, Ping Liang, Huisheng Zhang, and Tianfu Wang. A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Transactions on Medical Imaging*, 35(1):109–118, 2015.
9. B. Sheng, P. Li, S. Mo, H. Li, X. Hou, Q. Wu, J. Qin, R. Fang, and D. D. Feng. Retinal vessel segmentation using minimum spanning superpixel tree detector. *IEEE Transactions on Cybernetics*, 49(7):2707–2719, 2019.
10. E. O. Rodrigues, A. Conci, and P. Liatsis. Element: Multi-modal retinal vessel segmentation based on a coupled region growing and machine learning approach. *IEEE Journal of Biomedical and Health Informatics*, 24(12):3507–3519, 2020.
11. Y. Ma, H. Hao, J. Xie, H. Fu, J. Zhang, J. Yang, Z. Wang, J. Liu, Y. Zheng, and Y. Zhao. Rose: A retinal oct-angiography vessel segmentation dataset and new model. *IEEE Transactions on Medical Imaging*, 40(3):928–939, 2021.
12. D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao. Hard attention net for automatic retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, 24(12):3384–3396, 2020.
13. X. Li, Y. Jiang, M. Li, and S. Yin. Lightweight attention convolutional neural network for retinal vessel image segmentation. *IEEE Transactions on Industrial Informatics*, 17(3):1958–1967, 2021.
14. Y. Lv, H. Ma, J. Li, and S. Liu. Attention guided u-net with atrous convolution for accurate retinal vessels segmentation. *IEEE Access*, 8:32826–32839, 2020.
15. Q. Fu, S. Li, and X. Wang. Mscnn-am: A multi-scale convolutional neural network with attention mechanisms for retinal vessel segmentation. *IEEE Access*, 8:163926–163936, 2020.
16. Shihao Zhang, Huazhu Fu, Yuguang Yan, Yubing Zhang, Qingyao Wu, Ming Yang, Mingkui Tan, and Yanwu Xu. Attention guided network for retinal image segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 797–805, Cham, 2019. Springer International Publishing.
17. Lei. Mou, Yitian. Zhao, and Li. Chen. Cs-net: Channel and spatial attention network for curvilinear structure segmentation. In *MICCAI 2019: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 721–730, Cham, 2019. Springer International Publishing.
18. Wenao Ma, Shuang Yu, Kai Ma, Jiexiang Wang, Xinghao Ding, and Yefeng Zheng. Multi-task neural networks with spatial activation for retinal vessel segmentation and artery/vein classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 769–778, Cham, 2019. Springer International Publishing.
19. D. Li, M. H. Bawany, A. E. Kuriyan, R. S. Ramchandran, and G. Sharma. A novel deep learning pipeline for retinal vessel detection in fluorescein angiography. *IEEE Transactions on Image Processing*, PP(99):1–1, 2020.
20. Songhua Xie and Hui Nie. Retinal vascular image segmentation using genetic algorithm plus fcm clustering. In *2013 Third International Conference on Intelligent System Design and Engineering Applications*, pages 1225–1228, 2013.
21. Sohini Roychowdhury, Dara D Koozekanani, and Keshab K Parhi. Iterative vessel segmentation of fundus images. *IEEE Transactions on Biomedical Engineering*, 62(7):1738–1749, 2015.
22. B. Liu, L. Gu, and F. Lu. Unsupervised ensemble strategy for retinal vessel segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 111–119, Cham, 2019. Springer International Publishing.
23. S. A. A. Shah, A. Shahzad, M. A. Khan, C. Lu, and T. B. Tang. Unsupervised method for retinal vessel segmentation based on gabor wavelet and multiscale line detector. *IEEE Access*, 7:167221–167228, 2019.
24. Yicheng Wu, Yong Xia, Yang Song, Donghao Zhang, and Weidong Cai. Vessel-net: Retinal vessel segmentation under multi-path supervision. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 264–272, Cham, 2019. Springer International Publishing.
25. Sharif Amit Kamran, Khondker Fariha Hossain, Alireza Tavakkoli, Stewart Lee Zuckerbrod, Kenton M. Sanders, and Salah A. Baker. RV-GAN: Segmenting Retinal Vascular Structure in Fundus Photographs using a Novel Multi-scale Generative Adversarial Network. *arXiv e-prints*, page arXiv:2101.00535, January 2021.
26. Y. Zhou, H. Yu, and H. Shi. Study group learning: Improving retinal vessel segmentation trained with noisy labels. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020"*, Cham, 2021. Springer International Publishing.
27. Changlu Guo, Márton Szemenyei, Yugen Yi, Wenle Wang, Buer Chen, and Changqi Fan. SA-UNet: Spatial Attention U-Net for Retinal Vessel Segmentation. *arXiv e-prints*, page arXiv:2004.03696, April 2020.
28. Liangzhi Li, Manisha Verma, Yuta Nakashima, Hajime Nagahara, and Ryo Kawasaki. Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 3645–3654, 2020.
29. Juntang Zhuang. LadderNet: Multi-path networks based on U-Net for medical image segmentation. *arXiv e-prints*, page arXiv:1810.07810, October 2018.
30. Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *2018*

IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4510–4520, 2018.
31. N. Ma, X. Zhang, H. T. Zheng, and J. Sun. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Computer Vision – ECCV 2018*, pages 122–138, Cham, 2018. Springer International Publishing.
32. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.
33. Shanghua Gao, Ming Ming Cheng, Kai Zhao, Xin Yu Zhang, and Philip H. S. Torr. Res2net: A new multi-scale backbone architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1, 2019.
34. Olaf Ronneberger, Philipp Fischer, editor="Navab Nassir Brox, Thomas", Joachim Hornegger, William M. Wells, and Alejandro F. Frangi. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.
35. Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99), 2017.
36. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
37. Alex Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25(2), 2012.
38. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *Computer ence*, 2014.
39. Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, 2015.
40. Zhijie Zhang, Huazhu Fu, Hang Dai, Jianbing Shen, and Ling Shao. Et-net: A generic edge-attention guidance network for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019"*, pages 442–450, Cham, 2019. Springer International Publishing.
41. Xiao Xiao, Shen Lian, Zhiming Luo, and Shaozi Li. Weighted res-unet for high-quality retina vessel segmentation. In *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, pages 327–331, 2018.
42. Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari. Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. *arXiv e-prints*, page arXiv:1802.06955, February 2018.
43. Lei Mou, Li Chen, Jun Cheng, Zaiwang Gu, Yitian Zhao, and Jiang Liu. Dense dilated network with probability regularized walk for vessel detection. *IEEE Transactions on Medical Imaging*, 39(5):1392–1403, 2020.
44. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, and Andrew Rabinovich. Going deeper with convolutions. *IEEE Computer Society*, 2014.
45. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
46. D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. D. Lange, P. Halvorsen, and H. D. Johansen. Resunet++: An advanced architecture for medical image segmentation. In *2019 IEEE International Symposium on Multimedia (ISM)*, pages 225–2255, 2019.
47. Z. Tang, X. Liu, Y. Li, P. Yap, and D. Shen. Multi-atlas brain parcellation using squeeze-and-excitation fully convolutional networks. *IEEE Transactions on Image Processing*, 29:6864–6872, 2020.
48. D. Marín, A. Aquino, M. E. Gegundez-Arias, and J. M. Bravo. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Transactions on Medical Imaging*, 30(1):146–158, 2011.
49. Avijit Dasgupta and Sonam Singh. A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. *IEEE*, pages 248–251, 2017.
50. Z. Zhou, Mmr Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6):1856–1867, 2020.
51. Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv e-prints*, page arXiv:1804.03999, April 2018.
52. Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE Transactions on Medical Imaging*, 38(10):2281–2292, 2019.
53. George Azzopardi, Nicola Strisciuglio, Mario Vento, and Nicolai Petkov. Trainable cosfire filters for vessel delineation with application to retinal images. *Medical Image Analysis*, 19(1):46–57, 2015.
54. Yitian Zhao, Lavdie Rada, Ke Chen, Simon P. Harding, and Yalin Zheng. Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images. *IEEE Transactions on Medical Imaging*, 34, 2015.
55. Lei Zhou, Qi Yu, Xun Xu, Yun Gu, and Jie Yang. Improving dense conditional random field for retinal vessel segmentation by discriminative feature learning and thin-vessel enhancement. *Computer Methods and Programs in Biomedicine*, 148:13–25, 2017.
56. M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 59(9):2538–2548, 2012.
57. Zengqiang Yan, Xin Yang, and Kwang Ting Cheng. Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 65(9):1912–1923, 2018.
58. Zengqiang Yan, Xin Yang, and Kwang-Ting Cheng. A three-stage deep learning model for accurate retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, 23(4):1427–1436, 2019.
59. K. S. Sreejini and V. K. Govindan. Improved multiscale matched filter for retina vessel segmentation using pso algorithm. *Egyptian Informatics Journal*, 16(3), 2015.
60. K. Kumar, D. Samal, and Suraj. Automated retinal vessel segmentation based on morphological preprocessing and 2d-gabor wavelets. In *Advanced Computing and In-*

telligent Engineering, pages 411–423, Singapore, 2020. Springer Singapore.