



## Research article

## Segmentation of retinal vessels based on MRANet

Sanli Yi, Yanrong Wei<sup>\*</sup>, Gang Zhang, Tianwei Wang, Furong She, Xuelian Yang*School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, Yunnan, China*

## ARTICLE INFO

**Keywords:**

U-net network  
MSR block  
MLF block  
Attention block  
Retinal vessel segmentation

## ABSTRACT

The segmentation of retinal vessel takes a crucial part in computer-aided diagnosis of diseases and eye disorders. However, the insufficient segmentation of the capillary vessels and weak anti-noise interference ability make such task more difficult. To solve this problem, we proposed a multi-scale residual attention network (MRANet) which is based on U-Net network. Firstly, to collect useful information about the blood vessels more effectively, we proposed a multi-level feature fusion block (MLF block). Then, different weights of each fused feature are learned by using attention blocks, which can retain more useful feature information while reducing the interference of redundant features. Thirdly, multi-scale residual connection block (MSR block) is constructed, which can better extract the image features. Finally, we use the DropBlock layer in the network to reduce the network parameters and alleviate network overfitting. Experiments show that based on DRIVE, the accuracy rate and the AUC performance value of our network are 0.9698 and 0.9899 respectively, and based on CHASE\_DB1 dataset, they are 0.9755 and 0.9893 respectively. Our network has a better segmentation effect compared with other methods, which can ensure the continuity and completeness of blood vessel segmentation.

## 1. Introduction

The change of the retinal blood vessel's structure can provide an important basis for disease diagnosis [1]. Retinal vessel segmentation highlights vascular morphological information, which helps doctors make an early diagnosis of lesions.

In the past decades, an abundance of retinal vessel segmentation methods have been proposed by researchers, mainly divided into unsupervised method and supervised method. The unsupervised method is to segment blood vessels without any prior labeling information, such as matched filtering, morphology, blood vessels tracking methods, etc. Many unsupervised methods have been studied: Upadhyay et al. [2] adopt a new algorithm which is based on rule to better segment the blood vessels of the retina. Palanivel et al. [3] according to the vessels' multi-fractal features proposed a segmentation algorithm for the retinal vasculature, which can minimize the noise of the image and obtain better results. Tian et al. [4] used traditional Frangi filtering and mathematical morphology methods to construct an improved segmentation algorithm for the extraction of the vascular. In [5], a new segmentation method which can better extract the retinal vessel based on fundus image is proposed. And in order to better segment the blood vessels, Khan et al [6] presented the width bifurcation method. The above algorithms are helpful to doctors to a certain extent, but due to the time consuming and susceptible to human error and subjective experience, ideal results cannot be achieved when faced with a large number of fundus images. Thus, finding a method to segment the vessels of the retinal automatically is quite necessary.

Other than unsupervised method, the supervised method which is based on a huge amount of data with physician annotations, can

<sup>\*</sup> Corresponding author.

*E-mail address:* [20202204025@stu.kust.edu.cn](mailto:20202204025@stu.kust.edu.cn) (Y. Wei).

automatically extract features from the images. Recently, deep learning technology has become a trend in the field of retinal segmentation. At first, Convolutional Neural Network (CNN), which can automatically extract image features, is proposed. Later, Fully Convolutional Network (FCN) [7] achieved great performance due to its end-to-end feature learning. U-Net [8] has been employed extensively in recent years, which uses skip connections to build an encoder-decoder structure to make information transfer more efficient. However, the existing U-Net network has difficulties such as the gradient dispersion and explosion with the deepening of the network, so He et al. [9] proposed the Residual Network (ResNet). Afterward, Hu et al. [10] introduced an attention mechanism which can increase the useful feature information's weight and reduce the redundant information's interference, and therefore further improve the network's expression. With the improvement of CNN, its application in retinal blood vessel segmentation is also deepening. Such as, Lin et al. [11] proposed a multi-path-scale high-resolution representation network (MPS-Net) for retinal segmentation, which can improve the performance of extracting the vessels of retinal. However, the proposed network is not ideal for the tiny vessels' segmentation. Tchinda et al. [12] adopted a vessel segmentation approach, which performed better in the retinal image, but fails to obtain satisfactory boundary structure information of blood vessels. Alom et al. [13] proposed the residual U-Net structure by combining residual connection and U-Net network, which can avoid the degradation of the network with deep layers and improve the capabilities of segmentation of small blood vessels. Zhao et al. [14] adapted an attention mechanism and a residual module called AttentionResU-Net, which better highlights the pixel information of the thick vessels and thin vessels, but can not segment the small vessels so well. Aiming to optimise the use of the contextual information of vessels in retinal images to segment fine vessels more accurately, Zhang et al [15] designed the U-net involving context. Deng et al. [16] introduced a new segmentation algorithm by using multi-scale attention mechanism to better segment the capillary and better to ensure vascular connectivity.

In this paper, to extract retinal blood vessels more efficiently, a multi-scale residual attention network (MRANet) is proposed. The new method is an extension of U-Net and constitutes by integrating multi-level feature fusion block (MLF block), attention block, and multi-scale residual connection block (MSR block). The contributions of our work are: (1) The different parts of the network and the architecture of the proposed MRANet are detailed in Section 2. (2) The experimental process and the evaluation of the performance are illustrated in Section 3. (4) The conclusion of this paper and the next steps are discussed in Section 4.

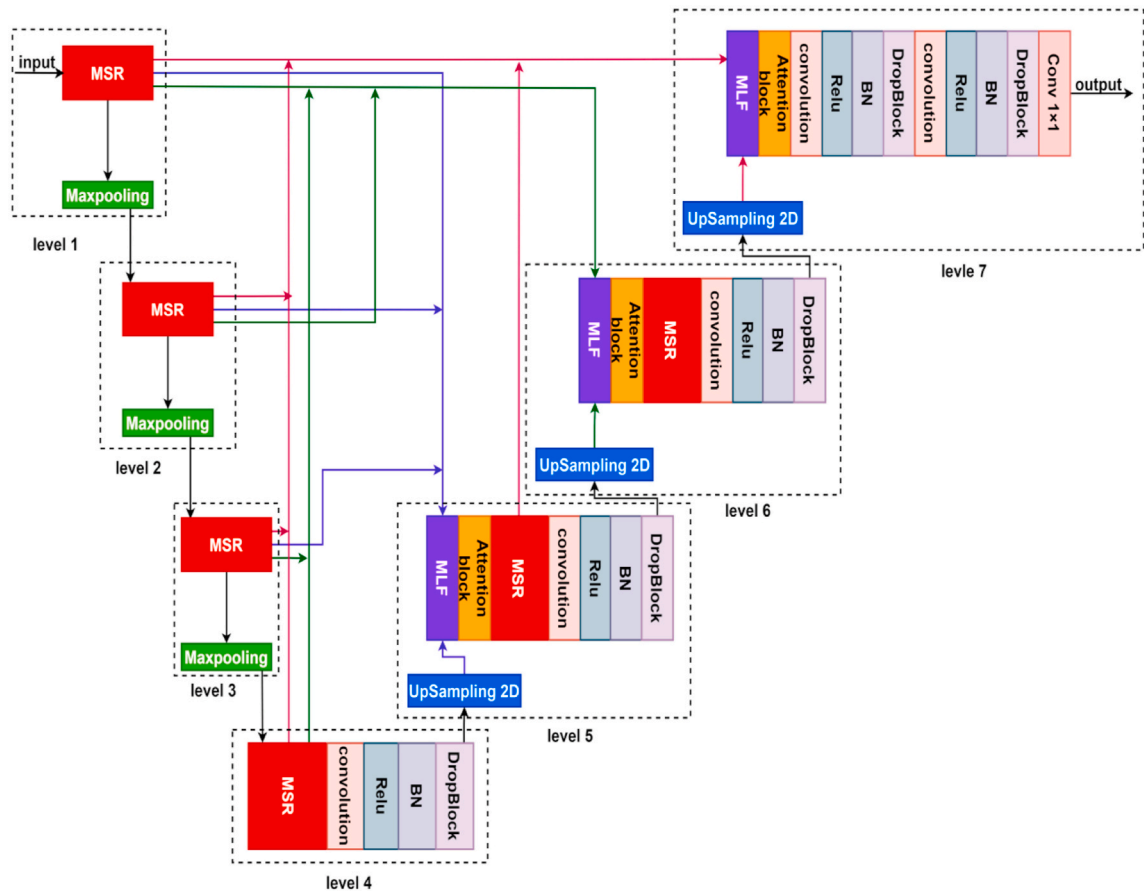


Figure 1. MRANet network structure.

## 2. Methodology

MRANet is based on U-Net network. And some function blocks are adopted to make the network realize more significant representation. Firstly, to solve the limits on the number of information flow paths and increase the utilization rate of information, MLF block is applied. Then MSR block is used to help deeper networks to obtain more complex information. The architecture of this network is shown in Figure 1.

From Figure 1, the MRANet network is composed of two parts: the encoding part and the decoding part. The encoding part includes four layers, each layer consists of one MSR block and one  $2 \times 2$  maxpooling function. The image patches are sent to the network as input, and each subsequent layer's output is then forwarded for the next layer's input. At the fourth encoding layer,  $3 \times 3$  convolution, Relu, BN, and DropBlock are additionally added, which produce high semantic information. Meanwhile, all the information of the encoding layers is sent to the decoding layers. These encoding layers ensure that the network is able to better extract image features.

The decoding part consists of three layers. In each layer: First, the newly proposed block, MLF block is used to avoid information loss while making full use of all information. Second, attention blocks are used to enhance the important features and the location relationship of vascular pixels and reduce the interference from useless features. Third, the proposed MSR block is adopted to strengthen the network's ability of multi-scale feature extraction. In the U-net, the encoding part's output is connected with the corresponding feature maps of the decoding part by using a copying and cropping procedure. Different from the original connection way, a new way to aggregate shallow fine information and deep rough information is proposed. Through the max-pooling operation and transposed convolution operation in MRANet, the multi-scale feature information of the MSR block and the information of the corresponding up-sampling layer are integrated, so that more global information can be extracted for improving the network's ability of feature utilization. Then, the attention block is adopted to extract useful vascular information and then input to the MSR block to extract more feature images. The MSR block is followed by the  $3 \times 3$  convolution, Relu, BN and DropBlock. The output map for each layer is twice as large as the original input, at the same time the channels are half as many as the original. In the third decoding layer, a  $2 \times 2$  up-sampling is utilized to restore features and a  $1 \times 1$  convolution is using for the mapping of each component feature vector to obtain the required amount of categories and to get a better blood vessel segmentation map. Table 1 lists the parameters used in MRANet.

MRANet is constructed by using the above functional blocks:MLF block, attention block and MSR block. And the details of these function blocks are presented as follows.

### 2.1. MLF block

To better aggregate the up-sampling feature information and the information of the MSR block in each layer, MLF block is used in the decoding path, which allows for maximum reuse of the functionality and thus reduce the loss of detail. The structure of MLF block is shown in Figure 2.

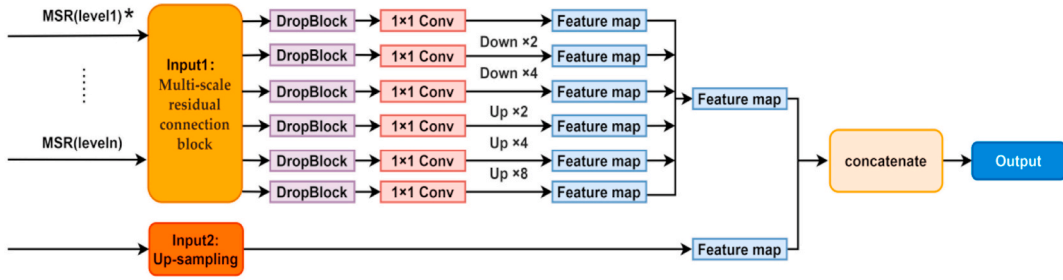
As shown in Figure 2, there are two kinds of input: input1 and input2. The processing of input 1 is as follows: firstly, the information of all the MSR blocks in the levels that before current level goes through the DropBlock layers, which are randomly discarded regions of adjacent elements in the feature map by blocks. DropBlock can effectively keep the convolutional network from over-fitting. Then, for reducing the channel's dimension, the  $1 \times 1$  convolution is used after DropBlock layers. However, because of the different resolution of previous MSR block features, the network cannot directly transfer the information from the shallow layer to the deep nodes. To make all the previous MSR blocks and the corresponding up-sampling feature maps of input 2 at the same resolution levels, the asynchronous max-pooling operation and transpose convolution operation for the above different input features are adopted. Finally, their fusion feature maps are output.

### 2.2. Attention block

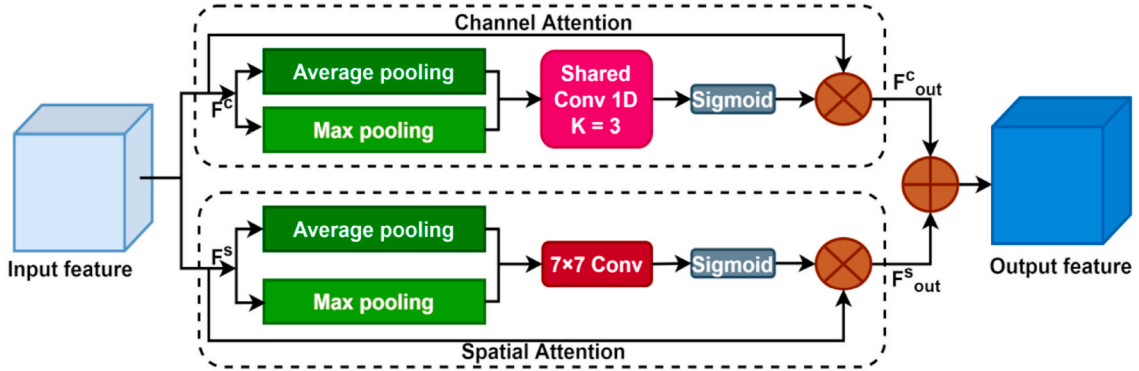
The attention block includes channel attention part and spatial attention part. In this paper, an approach with parallel structure that connect the features in space and channel is adopted. Through extracting the information of the space and the channel at the same time, we can obtain both vascular pixels and non-vascular pixels, and the relative positions of different features are also obtained. The block's structure is presented in Figure 3.

**Table 1**  
The layers and layer parameters of the proposed network.

Block	Output Shape	Trainable Parameters
Block 1	[32,32,16]	8646
Block 2	[16,16,32]	44,053
Block 3	[8,8,64]	165,829
Block 4	[16,16,256]	872,202
Block 5	[32,32,32]	667,434
Block 6	[64,64,16]	152,549
Block 7	[64,64,1]	32,902



**Figure 2.** The structure of MLF. \* For MLF blocks at different levels, the number of inputs is different. For example, for the MLF block at level 7 the inputs are from MSR(level 1) to MSR(level 5), while for the MLF block at level 6 the inputs are from MSR(level 1) to MSR(level 4).



**Figure 3.** Attention block.

### 2.2.1. Channel attention

The main function of channel attention part is to preserve the structural information between feature channels. Common channel attention such as SENet [10] and GSOP-Net [17] are widely used in deep learning. However, they dedicated to generating channel attention maps by learning the weight of each channel, which inevitably increase the complexity of the network. Recently proposed ECANet [18] achieves superior performance, mainly because it avoids the dimensionality reduction operation and uses cross-channel information interaction. Therefore, ECANet is applied to channel attention in this paper.

In the Channel attention structure in Figure 3, firstly, the input  $F^c \in R^{H \times W \times C}$  is used for the asynchronous max-pooling and the average-pooling and gets the channel's descriptions of  $F_{mp}^c$  and  $F_{ap}^c$  respectively. Since the features extracted between different channels of the image have local periodicity, a 1D convolution of size K is used to allow the information to flow between adjacent channels instead of the traditional FC layer. Next, all the obtained features are added to get more effective integration information. Lastly, a weight map of the channels can be generated by using Sigmoid activation function, that produces an output map  $F_{out}^c$  by multiplying it with the original input feature map. Briefly, the formula is as Eqs. (1), (2) and (3):

$$F_{mp}^c = \text{Max}(F^c(i, j)), 0 < c < C, 0 < i < H, 0 < j < W \quad (1)$$

$$F_{ap}^c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (F^c(i, j)), 0 < c < C \quad (2)$$

$$F_{out}^c = F^c \cdot F_{scale} \left( \sigma \left( \text{Conv1D} \left( F_{mp}^c \right) + \text{Conv1D} \left( F_{ap}^c \right) \right) \right) \quad (3)$$

In these equations,  $\sigma$  represents the activation function of Sigmoid,  $F_{scale}$  represents the weight map of each channel after one-dimensional convolution. To be mentioned, the kernel of the 1D convolution in this paper is set as 3 ( $k = 3$ , that is, there are three neighbors participating in the attention prediction of this channel).

### 2.2.2. Spatial attention

The spatial attention can notice the location of key information and enhance the ability of useful feature's extraction.

In the Spatial attention part of Figure 3, the feature map  $F^s \in R^{H \times W \times C}$  first generates  $F_{mp}^s$  and  $F_{ap}^s$  along the channel, among which  $F_{mp}^s$  is input to the max-pooling,  $F_{ap}^s$  is input to the average-pooling. Then a convolutional layer of  $7 \times 7$  is followed by the Sigmoid activation function layer, output of which is the spatial map. At last, it is multiplied with the original feature map, then a new feature

map  $F_{out}^s$  is obtained. The formula is as Eqs. (4), (5) and (6):

$$F_{mp}^s = \text{Max}(F^s(i, j)), 0 < c < C, 0 < i < H, 0 < j < W \tag{4}$$

$$F_{ap}^s = \frac{1}{C} \sum_{i=1}^C (F^s(i, j)), 0 < i < H, 0 < j < W \tag{5}$$

$$F_{out}^s = F^s \cdot F_{scale} \left( \sigma \left( f^{7 \times 7} \left( \left[ F_{mp}^s; F_{ap}^s \right] \right) \right) \right) \tag{6}$$

Where  $\sigma$  represents the Sigmoid activation function, the convolution kernel of  $7 \times 7$  is used to extract more important spatial features and obtain more location information of the target image.

### 2.3. MSR block

The traditional residual block is composed of 2 stacked  $3 \times 3$  convolutions and skip-connections [9], which can reduce the risk of network degradation and gradient disappearance. But due to the structure of its convolutional core is too simple, its feature extraction capability is limited. Therefore, to improve the network’s abilities of extracting and transferring image features, an MSR block with multi-scale residual structure is designed. Figure 4 presents the detailed structure of MSR.

In Figure 4, the MSR block consists of three branches. Among them, the first branch is composed of 2 depth-wise over-parameterized convolutional layers (DO-Conv) [19], which allows convolutional layers to be enhanced by using additional depth-separable convolutions with different 2D kernels for each input channel. Meanwhile, such layer could enhance the accuracy of the network without increasing the network’s computational complexity. The second branch consists of two  $3 \times 3$  convolutions with a dilation rate of 3, which can expand the perceptual range to extract more image features without increasing the number of effective units. The third branch consists of a  $3 \times 3$  convolution and asymmetric convolution block which is composed of  $1 \times 3$  convolution and  $3 \times 1$  convolution. The asymmetric convolution block can suppress the overfitting of the network while improving the nonlinear scalability of the network. Therefore, it can extract more spatial information with multivariate features, and its extraction process is more stable. The output of the three branches in parallel are added, and then pass through the DropBlock layer and the attention block layer. At last, the output of the attention block is combined with the original input through skip-connection. The DropBlock layer can effectively prevent overfitting, the attention block layer can extract more useful information through recalibrating the features of three branches,

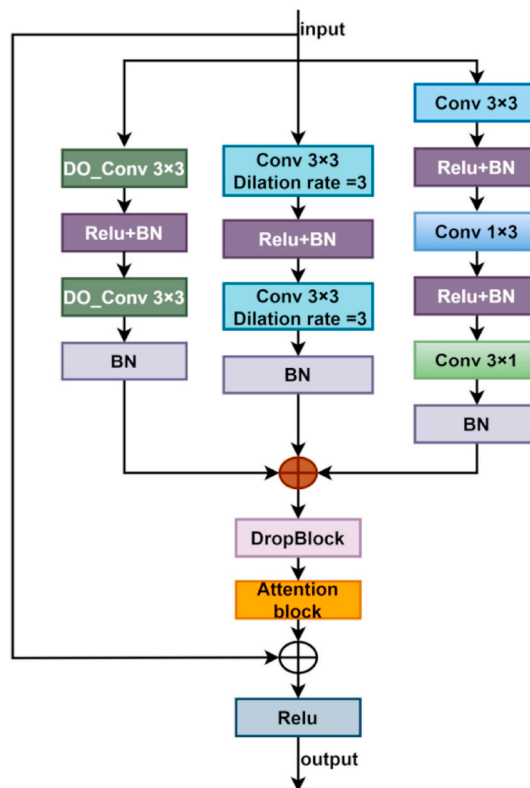


Figure 4. MSR block.

and the skip-connection of original input can avoid the information's loss during the process of the network's forward propagation.

### 3. Experiment

#### 3.1. Dataset

For the DRIVE [20] dataset and CHASE\_DB1 [21] dataset having sufficient amount of image data, more clearer annotation and better image quality, they are employed in this paper. The DRIVE dataset has 40 colour fundus images, measuring  $586 \times 565$ , of which 20 are for training and 20 are for testing. The CHASE\_DB1 dataset has 28 colour fundus images, the size of which is  $999 \times 960$ . We use its first 14 images to be the training set and its last 14 images to be the testing set. Meanwhile, the binary vascular maps can be used as ground truths, which are manually segmented by experts.

To reduce the influences of background noise, uneven illumination, and other factors on the image, the extraction of green channel, equalization of the histogram, standardization process, and gamma to transform pre-processing are used. In Figure 5, we can see that the vascular contour information is more highlighted in the pre-processed images. Then, to augment the data, a sliding window with step size of 5 is adopted, with which the original image can be randomly cut into 200,000 image patches. Of these patches, 80% are used for training and the rest are for testing. The size of the patch, i.e. the input data for the network, is  $64 \times 64$ . Figure 5(a) the example of the original image, Figure 5(b) illustrate the example of the image being pre-processed.

#### 3.2. Implementation detail

For both datasets, we choose the cross-entropy loss function and use Gradient Descent (SGD) for parameter optimization, for which the initial learning rate is 0.01, the decay rate is  $10^{-4}$  and the momentum is 0.9. The batch\_size of network training is 8 and the number of iterations is 50. The simulation platform is PyCharm, using the public Keras with Tensorflow as the backend. All experiments are run on Intel(R)Core(TM) i7-11700K CPU@3.60 GHz, 64.0 GB RAM, NVIDIA GeForce GTX 3060.

#### 3.3. Evaluation metrics

To analyze the segmentation effect of the network more objective, we employ the metrics as: accuracy, Kappa coefficient, sensitivity, specificity, F1-score, Matthews correlation coefficient (MCC), and area AUC under ROC, which are calculated as Eqs. (7), (8), (9), (10), (11), (12) and (13):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$p_e = \frac{(TP + FN)(TP + FP) + (TN + FN)(TN + FP)}{(TP + TN + FP + FN)^2} \quad (8)$$

$$Kappa = \frac{Accuracy - p_e}{1 - p_e} \quad (9)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (10)$$

$$Specificity = \frac{TN}{TN + FP} \quad (11)$$

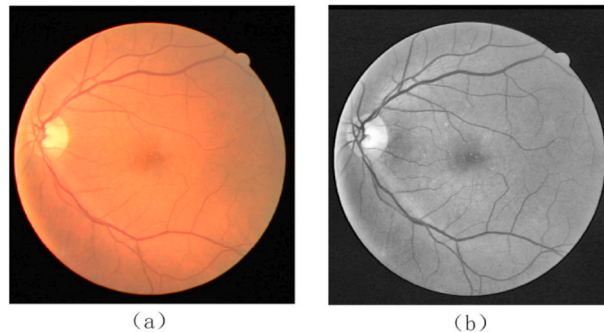


Figure 5. (a) DRIVE original image; (b) Pre-processed image.

$$F1 - score = \frac{2TP}{2TP + FN + FP} \quad (12)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (13)$$

Where TP (true positive) presents the true positive area, which means that the predicted algorithm segmentation result and the corresponding expert manual annotation of the same total number for the target blood vessel pixels, otherwise it is FP (false positive); For background pixels, TN (true negative) presents the total number of predicted algorithm segmentation results identical to the corresponding expert manual annotation, otherwise it is FN (false negative).

### 3.4. Experimental results and discussion

Finding the pixel points in the retinal image that belong to the vessel region is the main target of vessel segmentation. To justify the higher efficiency of our MRANet in the segmentation of retinal vessels, we conducted experiments on datasets introduced in section 3.1.

#### 1) Comparison of the results from different existing methods

Table 2 illustrates that MRANet's results are compared with those of the comparison network by using the DRIVE dataset, and Table 3 shows these comparisons by using the CHASE\_DB1 dataset.

From Tables 2 and 3, we can glean useful information. The accuracy, sensitivity, specificity and AUC of the proposed MRANet are 0.9698, 0.8488, 0.9907 and 0.9899, respectively, on the DRIVE dataset. And these parameters are 0.9755, 0.8533, 0.9856 and 0.9893 on CHASE\_DB1 dataset, respectively. Due to the interference of background, the MRANet network is only slightly less specific, but the differences are not significant. Two tables show that MRANet performs better than the existing algorithms. Overall, the proposed network can realize the task of vessel segmentation better, thus demonstrating the effectiveness of its ability to segment retinal vessels.

To further highlight the reliability of MRANet, the ROC curve trend chart is given in Figure 6. The curve represent the relationship between sensitivity and specificity, of which the horizontal coordinate indicates the false positive rate (FPR), the true positive rate (TPR) indicated by the longitudinal coordinate. The higher the AUC value, the more effective the network is in segmenting the vessels. It can be seen that the general performance of MRANet is better. Figure 6(a) is the ROC chart of DRIVE dataset, and Figure 6(b) is that of the CHASE\_DB1 dataset.

#### 2) Comparison of ablation experiments

To testify the efficiency of the new blocks used in MRANet, the ablation experiments based on two datasets are conducted in this section. In this experiment, U-Net, MRNet and MRANet are compared, among which U-Net is the basic network, MRNet is the combination of U-Net, attention block and MLF block, and the proposed network, MRANet, is the combination of U-Net, MLF block, attention block and MSR block.

In Table 4, we can see that based on the DRIVE dataset, the accuracy, sensitivity, specificity, f1-score, MCC, Kappa and AUC values of U-Net are 0.9619, 0.7789, 0.9775, 0.8193, 0.7813, 0.7849 and 0.9798 respectively, all of these are lower than that of MRNet. The reason of the better performance of MRNet is that it adopts MLF block and attention block, which can improve its segmentation ability of the blood vessel edge. As for MRANet, the accuracy, Kappa, sensitivity, f1-score, specificity, MCC and AUC values are 0.9698, 0.8102, 0.8488, 0.8231, 0.9907, 0.8059 and 0.9899, respectively, all these indicators are higher than MRNet due to the addition of MSR block, which has a strong ability to distinguish blood vessel details and branching structures. Based on above statements, we can see that MRANet performs better than others on DRIVE dataset, and same conclusion can be draw based on CHASE\_DB1 dataset.

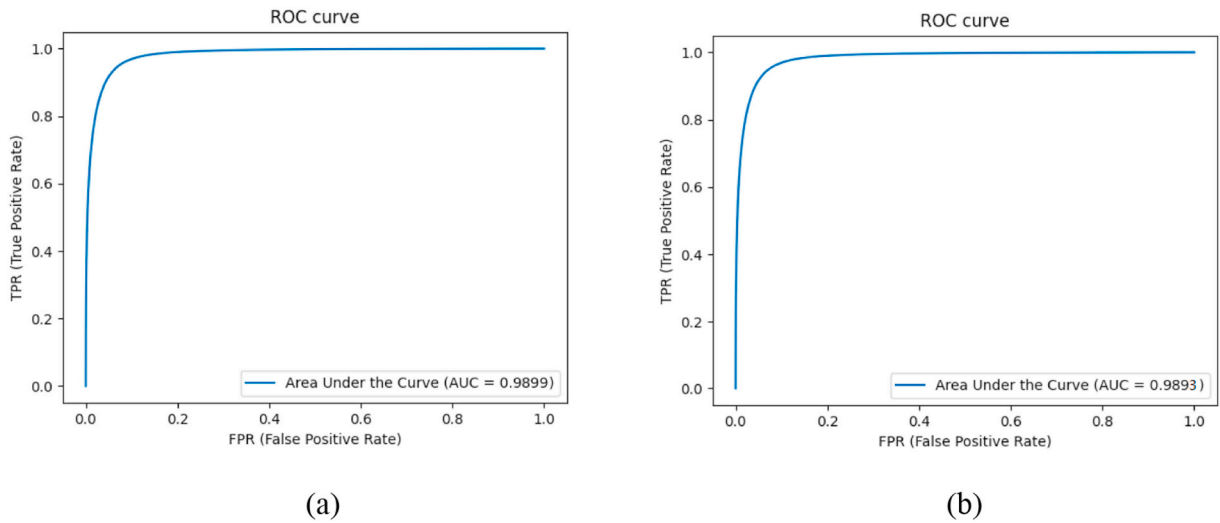
#### 3) p-Value analysis

**Table 2**  
Results of various algorithms on DRIVE.

Methods	Year	Acc	Sen	Spe	AUC
Shi [22]	2020	0.9676	0.8065	0.9826	-
Tchinda [12]	2021	0.9480	0.7352	0.9775	0.9678
Deng [16]	2022	0.9683	0.8363	0.9811	-
Khan [6]	2022	0.9610	0.8125	0.9763	-
Zhang [15]	2022	0.9565	0.7853	0.9818	0.9834
Wang [24]	2022	0.9611	0.8386	0.9867	0.9829
Dong [25]	2022	0.9586	0.7954	-	0.9830
Xu [26]	2022	0.9630	<b>0.8745</b>	0.9823	0.9670
<b>MRANet</b>	<b>2022</b>	<b>0.9698</b>	<b>0.8488</b>	<b>0.9907</b>	<b>0.9899</b>

**Table 3**  
Results of various algorithms on CHASE\_DB1.

Methods	Year	Acc	Sen	Spe	AUC
Zhang [27]	2016	0.9452	0.7626	0.9661	0.9606
Roychowdhury [28]	2014	0.9530	0.7201	0.9824	0.9532
Fraz [29]	2014	0.9524	0.7259	0.9770	0.9760
Shi [22]	2020	0.9731	0.7504	<b>0.9889</b>	-
Cheng [23]	2020	0.9488	0.7672	0.9834	0.9793
Tchinda [12]	2021	0.9452	0.7279	0.9658	0.9681
Deng [16]	2022	0.9714	0.8541	0.9794	-
Khan [6]	2022	0.9578	0.8012	0.9730	-
Zhang [15]	2022	0.9667	0.8132	0.9840	0.9893
Wang [24]	2022	0.9662	0.7958	0.9659	0.9873
Dong [25]	2022	0.9659	0.8259	-	0.9864
Xu [26]	2022	0.9694	<b>0.8916</b>	0.9794	0.9677
<b>MRANet</b>	<b>2022</b>	<b>0.9755</b>	<b>0.8533</b>	<b>0.9856</b>	<b>0.9893</b>



**Figure 6.** ROC curve chart: (a) DRIVE dataset; (b) CHASE\_DB1 dataset.

**Table 4**  
Comparison of ablation experiments with different datasets.

Datasets	Methods	Acc	Sen	Spe	AUC	F1-score	MCC	Kappa
DRIVE	U-Net	0.9619	0.7789	0.9775	0.9798	0.8193	0.7813	0.7849
	MRNet	0.9648	0.8392	0.9879	0.9839	0.8124	0.7968	0.7962
	<b>MRANet</b>	<b>0.9698</b>	<b>0.8488</b>	<b>0.9907</b>	<b>0.9899</b>	<b>0.8231</b>	<b>0.8059</b>	<b>0.8102</b>
CHASE_DB1	U-Net	0.9633	0.8457	0.9789	0.9839	0.7985	0.7814	0.7851
	MRNet	0.9692	0.8466	0.9896	0.9854	0.8114	0.7972	0.7926
	<b>MRANet</b>	<b>0.9755</b>	<b>0.8534</b>	<b>0.9856</b>	<b>0.9893</b>	<b>0.8281</b>	<b>0.8046</b>	<b>0.8035</b>

The dependent t-test for paired samples is used to check if the difference between our proposed results and other approaches is significant. As shown in Table 5.

Based on the p-values in Table 5, we can see the difference between our MRANet and U-Net, i.e. ours and U-Net have different levels of minimum significance when reject the hypothesis of no difference on metrics. In the DRIVE dataset, the hypothesis of no difference in accuracy, MCC and AUC are all 0.000, presenting the lowest level of significance rejected, while the hypothesis of no difference in specificity is 0.290, presenting the highest level of significance rejected. In CHASE\_DB1 dataset, the hypothesis of no difference in accuracy, MCC and AUC are all 0.000, presenting the lowest level of significance rejected, and the hypothesis of no difference of specificity is 0.239, presenting the highest level of significance rejected. As can be seen in Table 5, when the significance level of  $\alpha = 0.05$ , six of the seven metrics for DRIVE and CHASE\_DB1 are significant. When setting a more significant level of significance of  $\alpha = 0.01$ , DRIVE has four indicators that are highly significant and CHASE\_DB1 has five indicators that are significant. When the significance level of  $\alpha = 0.05$ , the specificity indicators in the DRIVE and CHASE\_DB1 is not significant, it is because when the retinal images are trained, the interference of the image background makes the specificity index not change significantly during the test, so when the



**Table 5**  
The Statics p-values on Metrics of MRANet and U-Net.

data	Metrics	p-value	$\alpha = 0.05$	$\alpha = 0.01$
DRIVE	Acc	0.000	√	√
	Sen	0.042	√	×
	Spe	0.290	×	×
	AUC	0.000	√	√
	f1-score	0.002	√	×
	MCC	0.000	√	√
	Kappa	0.001	√	√
CHASE_DB1	Acc	0.001	√	√
	Sen	0.003	√	×
	Spe	0.239	×	×
	AUC	0.000	√	√
	f1-score	0.000	√	√
	MCC	0.002	√	√
	Kappa	0.002	√	√

significance analysis is performed, the result of insignificant specificity is obtained.

#### 4) Comparison of segmentation experiment

By using the datasets introduced in section 3.1, the results of the ablation experiments were compared to further demonstrate the advantages of our network. The details of comparison can be seen in Figures 7 and 8.

For the DRIVE dataset of Figure 7, in the first line, compared the details of the red frames, we can see that the corresponding small blood vessels of Figure 7(d) and (e) are not segmented completely, while capillaries can be extracted well without loss of detailed feature information in Figure 7(c), which is almost the same as Figure 7(b). Similarly, in the second line, in the red frames of Figure 7(d) and (e), it can be observed that the developed method is incapable of detecting sufficiently blood vessels correctly. And in the third line, it is easy to cause small vessel incompleteness at the vessel ends of Figure 7(d) and (e), while the continuity of the blood vessels of Figure 7(c) can be guaranteed compared with Figure 7(b).

For the CHASE\_DB1 dataset, in the red frames of Figure 8, due to the influence of background and illumination, the contour information of Figure 8(d) is not clear. In the first line of Figure 8(d) and (e), the preservation of small vessels is not complete compared with Figure 8(b). In the second line, there is a phenomenon of rupture at vessel bifurcations of Figure 8(d) and (e). And in the third line, the blood vessels are not smooth enough, however, thick and thin blood vessels can be accurately segmented under uneven illumination in the line Figure 8(c).

From the segmentation results it can be concluded that MRANet can more clearly distinguish the vessels from the background and can reduce the rate of misjudgment and missed judgments of blood vessels. Meanwhile, under the condition of low contrast and noise interference, MRANet not only can ensure vascular connectivity and integrity, but also have a good level at vessel bifurcation and small vessel connections. Therefore, it can be seen that the proposed network has a remarkable segmentation ability on the complex vascular morphology.

## 4. Conclusion

We propose a multi-scale residual attention network (MRANet), that is an enhanced version of the U-Net and consists of MLF blocks, attention blocks and MSR blocks. The advantages of the structure are as follows: (1) With the MLF block, the image details and spatial location information of the shallow features can be fully used in the decoding part. (2) The attention block strengthens the results of the network's feature extraction of the blood vessel area. (3) The MSR block is used in the whole network to reduce the gradient disappearance and learn more information. To verify the network's effectiveness, experiments are carried out by using the datasets of DRIVE and CHASE\_DB1. The results of the experiments shown that our network outperforms previous networks.

Although our network achieved satisfied results in the segmentation of retinal vessel, it still has some limitations. Due to the multi-branch structure and the addition of operations such as concatenate and skip-connection, the network has a high memory requirement, which leads to a slow operation during the training process of the network. In the next stage of our work, we plan to make efforts to reduce the memory requirements of the network.

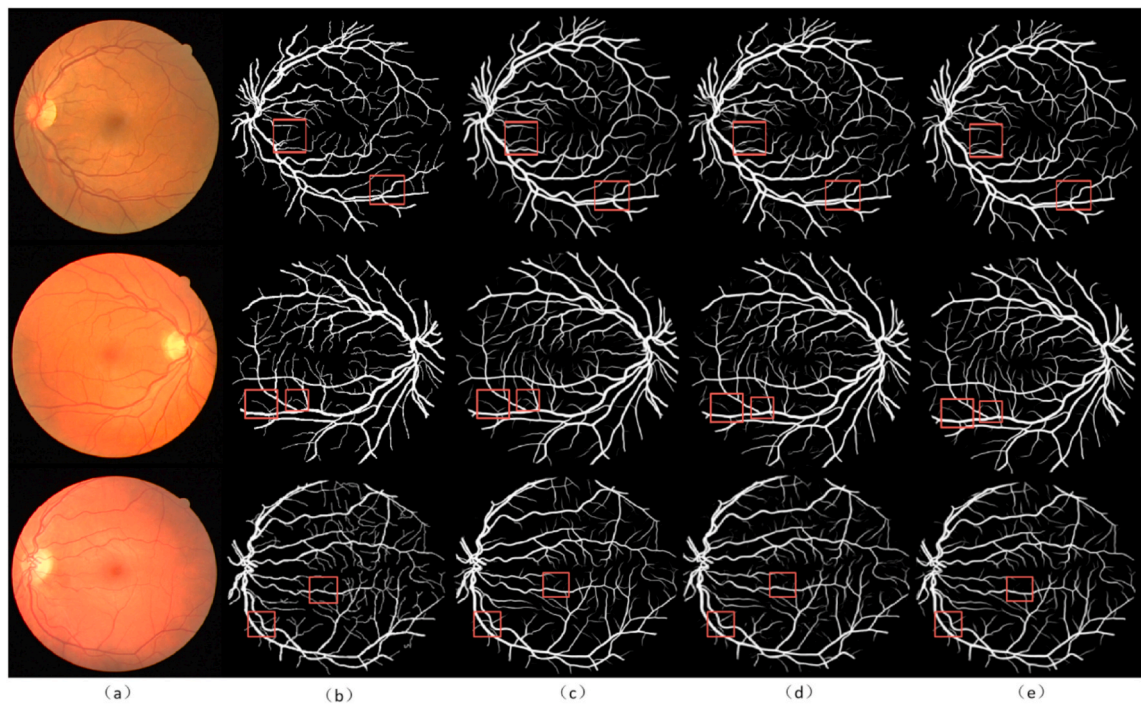
## Declarations

### Author contribution statement

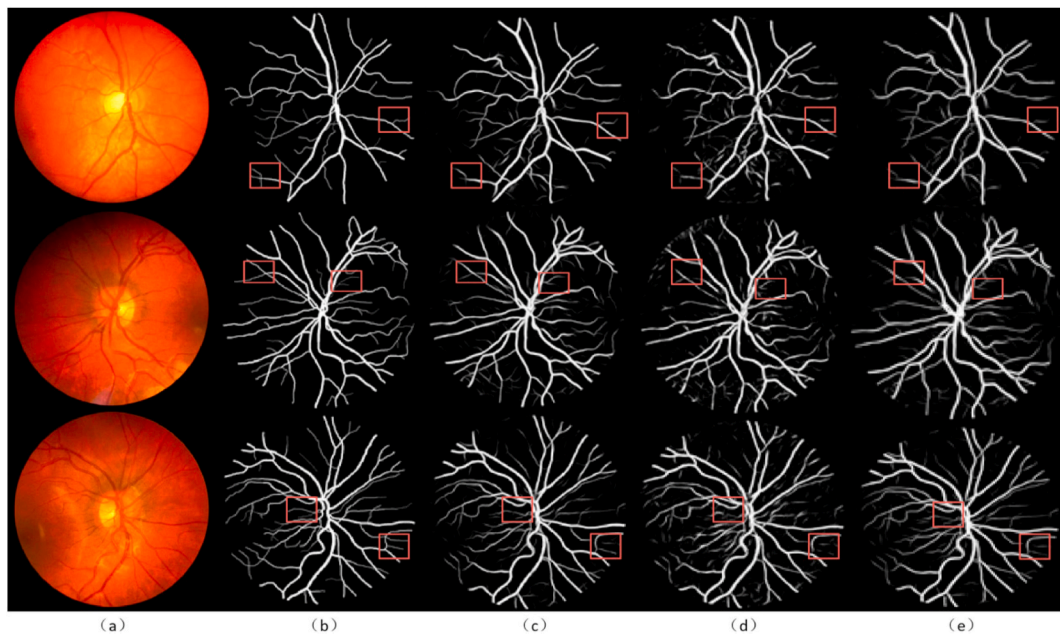
Sanli Yi: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Yanrong Wei: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Gang Zhang; Tianwei Wang: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.



**Figure 7.** The results segmented by various algorithms on DRIVE: (a) Original image; (b) ground truth; (c) MRANet; (d) U-Net; (e) MRNet.



**Figure 8.** The results segmented by various algorithms on CHASE\_DB1: (a) Original image; (b) ground truth; (c) MRANet; (d) U-Net; (e) MRNet.

Furong She: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Xuelian Yang: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

**Funding statement**

This work was supported by National Natural Science Foundation of China [22174057 & 82160347].

**Data availability statement**

Data associated with this study has been deposited at <https://drive.grand-challenge.org/> and <https://blogs.kingston.ac.uk/retinal/chasedb1/>

**Declaration of interest's statement**

The authors declare no conflict of interest.

**Additional information**

No additional information is available for this paper.

**References**

- [1] T.Y. Wong, R. McIntosh, Hypertensive retinopathy signs as risk indicators of cardiovascular morbidity and mortality, *Br. Med. Bull.* 73 (1) (2005) 57–70.
- [2] K. Upadhyay, M. Agrawal, P. Vashist, Unsupervised multiscale retinal blood vessel segmentation using fundus images, *IET Image Process.* 14 (11) (2020) 2616–2625.
- [3] D.A. Palanivel, S. Natarajan, S. Gopalakrishnan, Retinal vessel segmentation using multifractal characterization, *Appl. Soft Comput.* 94 (2020), 106439.
- [4] F. Tian, Y. Li, J. Wang, et al., Blood vessel segmentation of fundus retinal images based on improved frangi and mathematical morphology, *Comput. Math. Methods Med.* (2021) 2021.
- [5] O. Ramos-Soto, E. Rodríguez-Esparza, S.E. Balderas-Mata, et al., An efficient retinal blood vessel segmentation in eye fundus images by using optimized top-hat and homomorphic filtering, *Comput. Methods Progr. Biomed.* 201 (2021), 105949.
- [6] T.M. Khan, M.A.U. Khan, N.U. Rehman, et al., Width-wise vessel bifurcation for improved retinal vessel segmentation, *Biomed. Signal Process Control* 71 (2022), 103169.
- [7] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.
- [8] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 2015, pp. 234–241.
- [9] K. He, X. Zhang, S. Ren, et al., Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [10] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.
- [11] Z. Lin, J. Huang, Y. Chen, et al., A high resolution representation network with multi-path scale for retinal vessel segmentation, *Comput. Methods Progr. Biomed.* 208 (2021), 106206.
- [12] B.S. Tchinda, D. Tchiotop, M. Noubom, et al., Retinal blood vessels segmentation using classical edge detection filters and the neural network, *Inform. Med. Unlocked* 23 (2021), 100521.
- [13] M.Z. Alom, M. Hasan, C. Yakopcic, et al., Recurrent Residual Convolutional Neural Network Based on u-net (r2u-net) for Medical Image Segmentation, 2018 arXiv preprint arXiv:1802.06955.
- [14] S. Zhao, T. Liu, B. Liu, et al., Attention residual convolution neural network based on U-net (AttentionResU-Net) for retina vessel segmentation, in: *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, 2020, 032138, 440(3).
- [15] Y. Zhang, M. He, Z. Chen, et al., Bridge-Net: context-involved U-net with patch-based loss weight mapping for retinal blood vessel segmentation, *Expert Syst. Appl.* 195 (2022), 116526.
- [16] X. Deng, J. Ye, A retinal blood vessel segmentation based on improved D-MNet and pulse-coupled neural network, *Biomed. Signal Process Control* 73 (2022), 103467.
- [17] Z. Gao, J. Xie, Q. Wang, et al., Global second-order pooling convolutional networks, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3024–3033.
- [18] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-net: Efficient Channel Attention for Deep Convolutional Neural Networks, in: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11531–11539.
- [19] Cao J, Li Y, Sun M, et al. DO-Conv: Depthwise Over-parameterized Convolutional Layer. arXiv 2020. arXiv preprint arXiv:2006.12030.
- [20] <https://drive.grand-challenge.org/>.
- [21] <https://blogs.kingston.ac.uk/retinal/chasedb1/>.
- [22] Z. Shi, T. Wang, Z. Huang, et al., MD-Net: A multi-scale dense network for retinal vessel segmentation, *Biomed. Signal Process Control* 70 (2021), 102977.
- [23] Y. Cheng, M. Ma, L. Zhang, et al., Retinal blood vessel segmentation based on Densely Connected U-Net[J], *Math. Biosci. Eng.* 17 (4) (2020) 3088–3108.
- [24] H. Wang, G. Xu, X. Pan, et al., Attention-inception-based U-Net for retinal vessel segmentation with advanced residual, *Comput. Electr. Eng.* 98 (2022), 107670.
- [25] F. Dong, D. Wu, C. Guo, et al., CRAUNet: a cascaded residual attention U-Net for retinal vessel segmentation, *Comput. Biol. Med.* (2022), 105651.
- [26] Y. Xu, Y. Fan, Dual-channel asymmetric convolutional neural network for an efficient retinal blood vessel segmentation in eye fundus images, *Biocybern. Biomed. Eng.* 42 (2) (2022) 695–706.
- [27] J. Zhang, B. Dashtbozorg, E. Bekkers, J. Pluim, R. Duits, T.H.R. Bart, Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores, *IEEE Trans. Med. Imag.* 35 (12) (2016) 1.
- [28] S. Roychowdhury, D. Koozekanani, K. Parhi, Blood vessel segmentation of fundus images by major vessel extraction and sub-image classification, *IEEE J. Biomed. Health Inf.* (2014) 1.
- [29] M.M. Fraz, A.R. Rudnicka, C.G. Owen, S.A. Barman, Delineation of blood vessels in pediatric retinal images using decision trees-based ensemble classification, *Int. J. Comput. Assist. Radiol. Surg.* 9 (5) (2014) 795–811.