symmetry

MDPI

Article

Bi-SANet—Bilateral Network with Scale Attention for Retinal Vessel Segmentation

Yun Jiang †, Huixia Yao *,†, Zeqi Ma and Jingyao Zhang

College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China; jiangyun@nwnu.edu.cn (Y.J.); 2019221847@nwnu.edu.cn (Z.M.); 2019221868@nwnu.edu.cn (J.Z.)
* Correspondence: 2019211759@nwnu.edu.cn
† These authors contributed equally to this work.

Abstract: The segmentation of retinal vessels is critical for the diagnosis of some fundus diseases. Retinal vessel segmentation requires abundant spatial information and receptive fields with different sizes while existing methods usually sacrifice spatial resolution to achieve real-time reasoning speed, resulting in inadequate vessel segmentation of low-contrast regions and weak anti-noise interference ability. The asymmetry of capillaries in fundus images also increases the difficulty of segmentation. In this paper, we proposed a two-branch network based on multi-scale attention to alleviate the above problem. First, a coarse network with multi-scale U-Net as the backbone is designed to capture more semantic information and to generate high-resolution features. A multi-scale attention module is used to obtain enough receptive fields. The other branch is a fine network, which uses the residual block of a small convolution kernel to make up for the deficiency of spatial information. Finally, we use the feature fusion module to aggregate the information of the coarse and fine networks. The experiments were performed on the DRIVE, CHASE, and STARE datasets. Respectively, the accuracy reached 96.93%, 97.58%, and 97.70%. The specificity reached 97.72%, 98.52%, and 98.94%. The F-measure reached 83.82%, 81.39%, and 84.36%. Experimental results show that compared with some state-of-art methods such as Sine-Net, SA-Net, our proposed method has better performance on three datasets.

Keywords: deep convolutional neural work; retinal vessel segmentation; scale attention; bilateral network

Citation: Jiang, Y.; Yao, H.; Ma, Z.; Zhang, J. Bi-SANet—Bilateral Network with Scale Attention for Retinal Vessel Segmentation. *Symmetry* **2021**, *13*, 1820. https://doi.org/10.3390/sym13101820

Academic Editor: Tomohiro Inagaki

Received: 21 August 2021
Accepted: 26 September 2021
Published: 29 September 2021

1. Introduction

At present, cataract, glaucoma, and diabetic retinopathy are the main diseases leading to blindness [1]. More than 418 million people worldwide suffer from these diseases, according to a report (https://www.who.int/publications/i/item/world-report-on-vision accessed on 13 September 2021) published by the World Health Organization. Patients suffering from eye diseases often do not notice the aggravation of asymptomatic conditions, so early screening and treatment of eye diseases is necessary [2]. The asymmetry of the fundus capillaries adds complexity to the physician's diagnosis of the disease. However, extensive screening or diagnosis is time-consuming and laborious. Therefore, the automatic segmentation method is particularly helpful for doctors to diagnose and predict eye diseases. In recent years, many scholars have studied retinal automatic segmentation algorithms, which are mainly divided into two categories: unsupervised methods and supervised methods [1].

Unsupervised blood vessel segmentation algorithms at home and abroad include the matched filter method [3], multi-threshold blood vessel detection [4], morphological-based blood vessel segmentation [5], the region growth method, the B-COSFIRE filter method [6], the multi-scale layer decomposition and local adaptive threshold blood vessel segmentation method [7], the finite element based binary level set method [8] and fuzzy clustering [9], etc. In [10], the multi-scale 2D Gabor wavelet transform and morphological reconstruction

were used to segment the fundus vessels. In [11], a combination of level set and shape preference approach was presented.

Compared with the unsupervised method, the supervised method uses manually labeled data to train the classifier to segment the fundus vascular image. A typical supervised method is the deep convolutional neural network (CNN) [12–14]. However, CNN cannot make the structured prediction, and the fully connected neural network (FCN) came into being, providing an end-to-end blood vessel segmentation scheme. Therefore, it is rapidly applied to blood vessel segmentation. In [15], an FCN with a side output layer called DeepVessel was proposed. The work in [16,17] proposed an FCN with a down-sample and up-sample layer to solve the class imbalance between blood vessels and background.

In addition, encoder and decoder structures are widely used in fundus image segmentation due to their excellent feature extraction capability, especially U-Net [18]. In [19], it proposed a multi-label architecture based U-Net. A side output layer was used to capture multi-scale features.

On the basis of encoders and decoders, a dual-branch network is another method to segment fundus images. One branch of the dual-branch network is used as a coarse segmentation network, and another fine segmentation network is used as an aid to the coarse segmentation network. The work in [20] proposed a multi-scale two-branch network (MS-NFN) to improve the performance of capillary segmentation. Both branches had similar U-Net structures. The work in [21] proposed a coarse-to-fine segmentation network (CTF-Net) for fundus vascular segmentation. Moreover, a multi-scale network is another important direction of fundus image segmentation. In [22], it proposed VesselNet based on a multi-scale method. In [23], a cross-connected convolution neural network (CcNet) for blood vessel segmentation was proposed, which also adopted a multi-scale method. In order to improve the segmentation ability of the network, the attention mechanism is gradually applied to retinal vessel segmentation. In [24], an attention guiding network (AG-Net) was proposed.

From the CNN and FCN to the U-shaped network architecture, the basic network for blood vessel segmentation has developed abundantly; the network with U-Net structure has especially become more and more popular. However, in the existing methods, there are still the following challenges: (1) The contrast and extraneous noise of fundus images (see Figure 1) make it difficult to segment blood vessels in fundus images, especially the segmentation of low-contrast regions and lesion regions. (2) The general structure of a two-branch network used two U-Net networks as the backbone, which led to the increase in the computational cost and the prolongation of the segmentation time of the model. This is not friendly to clinical diagnosis and medical treatment. (3) There is still space for improvement in the accuracy of blood vessel segmentation. Specifically, it is necessary to further improve the sensitivity while maintaining specificity and accuracy. (4) Abundant spatial information and receptive fields with different sizes cannot be satisfied at the same time.

To address the above problems, this paper proposes a retinal vessel segmentation model based on a bilateral network with scale attention (Bi-SANet). The main contributions of this paper include the following contents:

- We propose a bilateral network containing coarse and fine branches, and the two branches are responsible for the extraction of semantic information and spatial information, respectively. Meanwhile, multi-scale input is carried out on the network to improve the feature extraction ability of the model for images of different scales.
- In order to improve the network's ability to extract vessel semantic information in low-contrast regions, a multi-scale attention module is introduced at the end of down-sampling of the coarse network. This can improve the pertinence of information recovery in up-sampling.
- The U-shaped fine network is replaced by a module. This module mainly uses convolution layers with different dilation rates to make up for the lost spatial information of the coarse network. It not only improves the network segmentation ability but also

reduces its computational complexity. Finally, the feature fusion module is used to aggregate different levels of information of the coarse and fine networks.

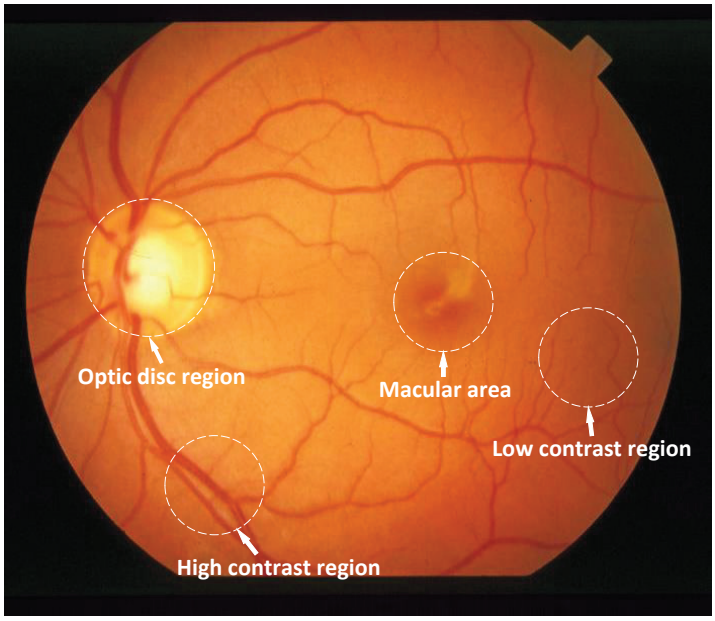

Figure 1. A retina fundus image on the STARE dataset.

The remainder of this article is organized as follows. Section 2 describes the proposed method in detail, including the network structure, multi-scale attention module, fine network structure, and feature fusion module. Section 3 introduces the datasets, experimental setting, and evaluation index. In Section 4, our experimental results are discussed and compared. Finally, the conclusion is drawn in Section 5.

2. Methods

2.1. The Network Structure of Bi-SANet

Compared with the encoder–decoder structure, the dual-branch network has different characteristics. The encoder–decoder structure uses the encoder network to extract features, the up-sample operation in the decoder restores the original resolution and connects the decoder with the feature map of the encoder.

In this paper, a two-branch network is used to segment fundus images: one is a coarse network (CoarseNet), the other is a fine network(FineNet). The network structure of this paper is shown in Figure 2. Among them, the coarse network is responsible for extracting the semantic information of fundus images, and the fine network makes up for the lost spatial information of the coarse network. As the backbone of the network, the improved U-Net is used in the coarse network to extract feature information. However, some spatial information cannot be captured in the process of down-sampling. In order to make up for the lack of spatial information, we a use fine network to further extract fine semantic information. Finally, the feature fusion module is used to aggregate the information of CoarseNet and FineNet.

In the CoarseNet, in order to extract the semantic information of feature maps with different scales, we introduced a multi-scale attention module at the end of the encoder. In the FineNet, because the spatial detail information needs to be saved, only one down-sampling operation is used as the encoder module. The spatial information module is used to extract spatial detail information.

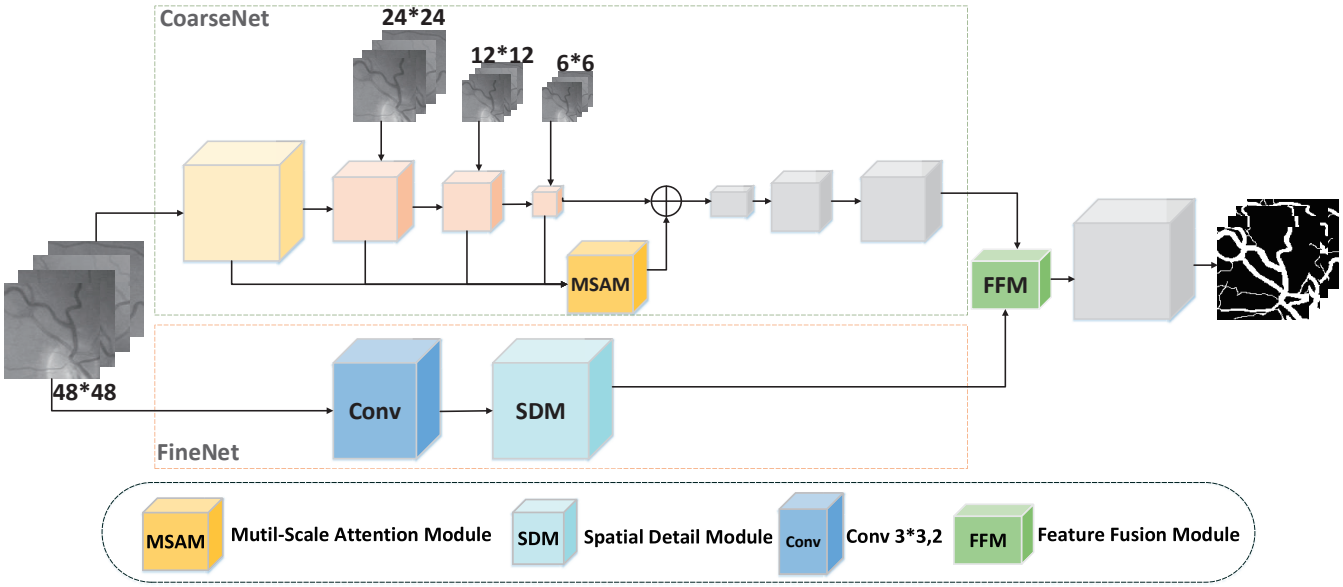

Figure 2. The Bi-SANet architecture.

2.2. Coarse Network (Coarse)

As shown in Figure 2, the basic network of the coarse network structure is a multi-scale U-Net. The multi-scale input is divided into four branches. The input image channels of the four branches are all 1. The 48 × 48, 24 × 24, 12 × 12, and 6 × 6 pixels are the input sizes for each branch.

The network structure of the coarse network is mainly composed of two parts: the first part is the encoder–feature extraction part, and the second part is the decoder–up-sampling part (gray part of Figure 2). The proposed multi-scale attention module is embedded in the encoder. The use of attentional mechanisms allows the network to focus on information related to blood vessels, to suppress noise in the background, and to alleviate information loss caused by down-sampling. It can enhance context semantic information. The details of the multi-scale attention module are as follows.

Multi-Scale Attention Module

Multi-scale U-Net obtains feature mappings of different scales in the process of down-sampling. In order to deal with objects of different scales better, a multi-scale attention module is used to combine these feature maps. This can alleviate the loss of semantic information during down-sampling.

The feature maps with different scales have different correlations with the input of the network, so we add an attention mechanism to calculate the correlation weight for each feature map with different scales. In this way, the network can pay more attention to the feature map with a higher correlation with the input image. This module is used at the end of the encoder.

Our proposed multi-scale attention module is shown in Figure 3. First, we use bilinear interpolation to up-sample the feature maps F_s of different scales obtained by the decoder to the original image size. In order to reduce the computational cost, we use 1 × 1 convolution to compress these feature maps into four channels. Additionally, the compression results from different scales are superimposed together through channels to form a feature map F. The mixed feature maps are pooled globally and maximally to obtain the correlation weight coefficient β of each channel. Then, we use element multiplication to distribute the obtained correlation weight coefficients to four channels to obtain the attention mixed feature map F'. In order to alleviate the problems of gradient disappearance and gradient explosion, we then added skip connection. Finally, y_{MA} is obtained from the mixed feature map and the attention mixed feature map by element addition.

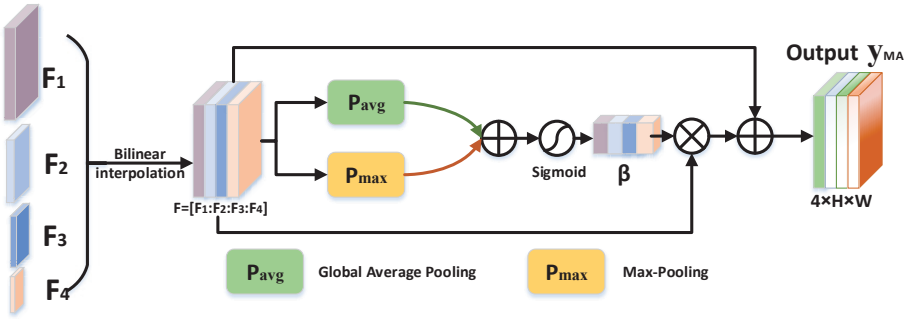

Figure 3. Structure of multi-scale attention module.

The calculation formula of the multi-scale attention module is defined as follows:

$$y_{MA} = F \oplus F \otimes \left(\delta\left(P_{\text{avg}}\left(F\right) \oplus P_{\max}(F)\right)\right) \quad (1)$$

Among them, δ represents sigmoid activation function, and P_{avg} and P_{max} represent the global average pooling and maximum pooling, respectively.

2.3. Fine Network (FineNet)

In order to capture more accurate boundary information, the FineNet only downsamples the feature map once with a convolution kernel of 3. The result of the down-sampling is input to the spatial information module for processing. There is no decoder module because of the need to save the spatial details in the FineNet. First, it reduced the size of the feature map by using one-step down-sampling convolution. Then, it applied a spatial information module to capture spatial details.

Details of the spatial information module are shown below.

2.3.1. Spatial Detail Module

The advantage of dilation convolution over traditional convolution operations is the ability to achieve larger receptive fields without increasing the number of parameters and maintaining the same feature resolution. The model can better understand the global context information. The thickness of blood vessels in retinal images is different. In order to segment blood vessels with different thicknesses more accurately, we use dilated convolution with different dilation rates in the spatial detail module to capture multi-scale feature information. This can improve the segmentation accuracy of blood vessel edges and tiny blood vessels. The structure of the spatial detail module is shown in Figure 4.

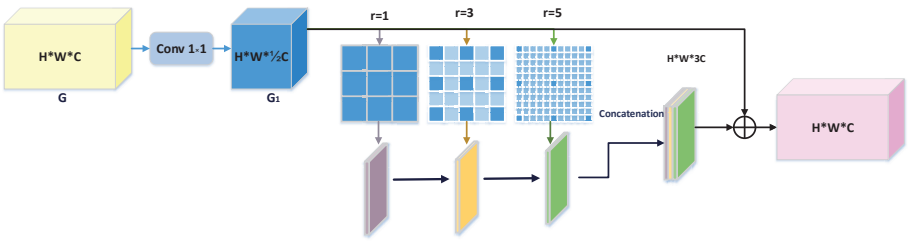

Figure 4. Structure of the spatial detail module.

In the spatial detail module, the number of channels of the input feature map is first halved using the 1×1 convolution method to obtain the feature map G_1. Then, three parallel convolution layers are used to capture the spatial feature information of the G_1. The number of channels is halved in order to reduce the number of parameters and computations to $1/2$, thus increasing the efficiency of the model to segment the fundus vessels. Three convolution layers with different dilation rates can capture multi-scale context information. The feature maps output by the three convolution layers is concatenated by channel. The skip connection is used to preserve the feature information

of the original scale. Finally, the feature is upsampled to a feature map size that can match the output of the CoarseNet.

2.3.2. Feature Fusion Module

As the FineNet is downsampled only once, the feature map obtained by the FineNet contains detailed information in fundus vessels, which is called low-level semantic information.

The feature map obtained by the coarse network contains rich context semantic information, but the detailed information in the retinal image is seriously lost after multiple downsamples, especially the boundary information of small blood vessels. The information obtained from the CoarseNet is called high-level semantic information. Low-level features are rich in details but lack semantic information. Therefore, we proposed a feature fusion module to integrate high-level features and low-level features, to enrich the spatial information in the coarse network, and to eliminate background noise from low-level features. The calculation process of high-level and low-level information aggregation is represented by Equation (2).

$$F_c = F^L \oplus F^H \tag{2}$$

Among them, F_L represents low-level features, F_H represents high-level features, and $\oplus$ is the element-level addition. After the above operations, in order to make the network pay more attention to the target area, we use global average pooling to process the fused features. In short, the final refined output is computed as Equation (3).

$$I_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i,j) \tag{3}$$

where $X_c(i,j)$ is the pixel value of the feature map F_c at the *(i, j)* position on the *C* channel.

3. Datasets and Evaluation

3.1. Datasets

The method in this paper is validated on three public datasets, DRIVE [25], CHASE [26] and STARE [27]. In the DRIVE data set, there are 40 retinal images, corresponding ground-truth images, and masks images. The size of each image is 565 × 584 pixels. The training and test sets are half of the images in the DRIVE dataset.

The CHASE data set consists of 28 retinal images, corresponding ground-truth images, and mask images, each of which is 1280 × 960 pixels. For the CHASE data set, we adopt the partition method proposed by Zhuang et al. [28] to train the first twenty images, and tested and evaluated the remaining eight images.

The STARE data set contains 20 retinal images, corresponding ground-truth images, and mask images. Each image is 700 × 605 pixels. We used the leave-one method to generate a training set and a test set. Each image is tested once. We averaged the experimental results of 20 images to obtain the evaluation results for the whole STARE dataset.

In addition, we expand the training data set by using the random patch method [29] on the image, which is very important to improve the accuracy of segmentation, prevent over-fitting and improve the robustness of the network.

3.2. Experimental Environment and Parameter Settings

Our network structure implementation and training method followed that of Jiang et al. [30]. It was implemented on a server with Quadro RTX 6000 on Ubuntu64, based on the open source package Pytorch. We trained the model method with the random patch method in [29]. The patch size was set to 48 × 48 pixels. During the training period, the epochs were set to 100, the batch size was set to 256, and the number of patches extracted from each image was 10,480. The initial learning rate was set to 0.001. The learning rate was updated using the step decay method. The decay coefficient and the weight decay

coefficient were set to 0.01 and 0.0005, respectively. The optimizer we use for the model was Adam, where the momentum was 10^{-8}.

The loss function used a cross-entropy loss function. It is defined as follows:

$$\text{Loss}_{ce}(y, \hat{y}) = -\sum y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \tag{4}$$

where y_i means the real label and $\hat{y}_i$ represents the predicted label.

3.3. Performance Evaluation Indicator

In this paper, Sensitivity, Specificity, Accuracy, F-measure, and other evaluation indicators were calculated by a confusion matrix. Additionally, the performance of retinal image segmentation was analyzed. Each evaluation index is defined by the following:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \tag{5}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{6}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{7}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{F-measure} = \frac{2 \times \text{ Precision } \times \text{ Sensitivity}}{\text{Precision} + \text{ Sensitivity}} \tag{9}$$

where TP is the number of correctly divided blood vessel pixels, TN is the number of correctly divided background pixels, FP is the background pixel incorrectly divided into blood vessel pixels, and FN is the blood vessel pixel incorrectly marked as the background pixel.

4. Experiment Results and Analysis

4.1. Discussion of Model Performance at Different Dilation Rates

Convolution combinations with different dilation rates have different receptive fields. In order to verify the influence of different receptive fields, we compared the experimental results of the model under different dilation rates on the DRIVE data set. From the experimental results in Table 1, it can be seen that, when the dilation rate is (1, 3, 5) in the spatial detail module, the model has the strongest segmentation ability. The bolded data indicate that the value is the optimal value for the indicator.

Table 1. Model performance at different dilation rates.

(d1, d2, d3)	F-Measure	Sensitivity	Specificity	Accuracy
(1, 1, 1)	0.8274	0.8042	0.9866	**0.9706**
(1, 2, 3)	0.8266	0.8107	0.9855	0.9702
(2, 2, 2)	0.8282	0.8092	0.9861	**0.9706**
(1, 2, 4)	0.8293	0.8318	0.9832	0.9700
(3, 3, 3)	0.8253	0.7926	**0.9876**	**0.9706**
(1, 3, 5)	**0.8382**	**0.8890**	0.9772	0.9693
(5, 5, 5)	0.8270	0.8039	0.9865	0.9705

4.2. Structure Ablation

In order to verify the effectiveness of the multi-scale attention module (MA), we compared the experimental results before and after adding this module to the model. In Tables 2 and 3, MUet represents our original baseline network, which is a multi-scale input UNet, MA represents a multi-scale attention module, FineNet represents the fine network, and FFA represents a feature fusion module. MU-Net and MA make up CoarseNet.

From the experimental results in Tables 2 and 3, it can be seen that the accuracy of the model on DRIVE and CHASE data sets has been improved to some extent after adding the multi-scale attention module (MA). As MA can effectively capture multi-scale information, the model can better segment retinal vessels with different thicknesses.

Table 2. Performance evaluation of four models on the DRIVE dataset.

Methods	Sensitivity	Specificity	Accuracy	F-Measure	Params Size (MB)
MU-Net	0.8579	**0.9799**	0.9691	0.8320	**44.49**
MU-Net+MA	0.8664	0.9787	0.9687	0.8340	44.76
MU-Net+MA+FineNet	0.8846	0.9782	0.9693	0.8376	45.46
MU-Net+MA+FineNet+FFA	**0.8890**	0.9772	**0.9699**	**0.8382**	46.17

Table 3. Performance evaluation of four models on the CHASE dataset.

Methods	Sensitivity	Specificity	Accuracy	F-Measure	Params Size (MB)
MU-Net	0.8006	**0.9871**	0.9753	0.8039	**44.49**
MU-Net+MA	0.8082	0.9868	0.9755	0.8058	44.76
MU-Net+MA+FineNet	0.8082	**0.9871**	0.9758	0.8083	45.46
MU-Net+MA+FineNet+FFA	**0.8371**	0.9852	**0.9759**	**0.8139**	46.17

In Tables 2 and 3, we compared the influence of the FineNet on the accuracy. From the experimental results, we can see that the addition of FineNet improved the accuracy, sensitivity, F-measure, and AUC curve area of the model to varying degrees. Additionally, the addition of FineNet does not bring a particularly large number of parameters to the network while improving segmentation performance. The FineNet used few down-sampling modules and does not contain any up-sampling and few pooling operations. It retained the edge details of the image and made up for the loss of information due to more convolution and pooling in the CoarseNet.

In order to verify the effectiveness of our added module, in this section, we compared the ROC (Receiver Operating Characteristic) and PR (Precision Recall) curves of different models on the DRIVE and CHASE data sets. In Figures 5 and 6, MUet represents our original baseline network, which is a multi-scale input U-Net; MA represents a multi-scale attention module; FineNet represents a fine network; and FFA represents a feature fusion module. From Figure 5, we find that the ROC and PR values of the model increase for each module added. The model with the highest ROC and PR values is the model that contained all three modules. Compared with the baseline network, its ROC value is increased by 0.3% and the PR value is increased by 1.69%. The closer the ROC curve is to the upper left corner of the coordinate, the more accurate the model is.

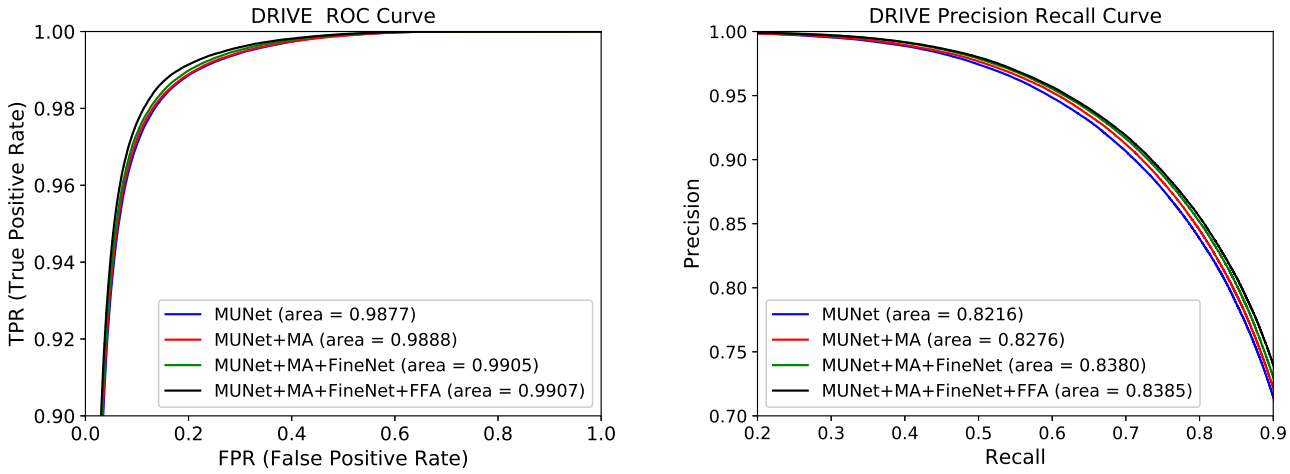

Figure 5. ROC curve and PR curve for four models on the DRIVE dataset.

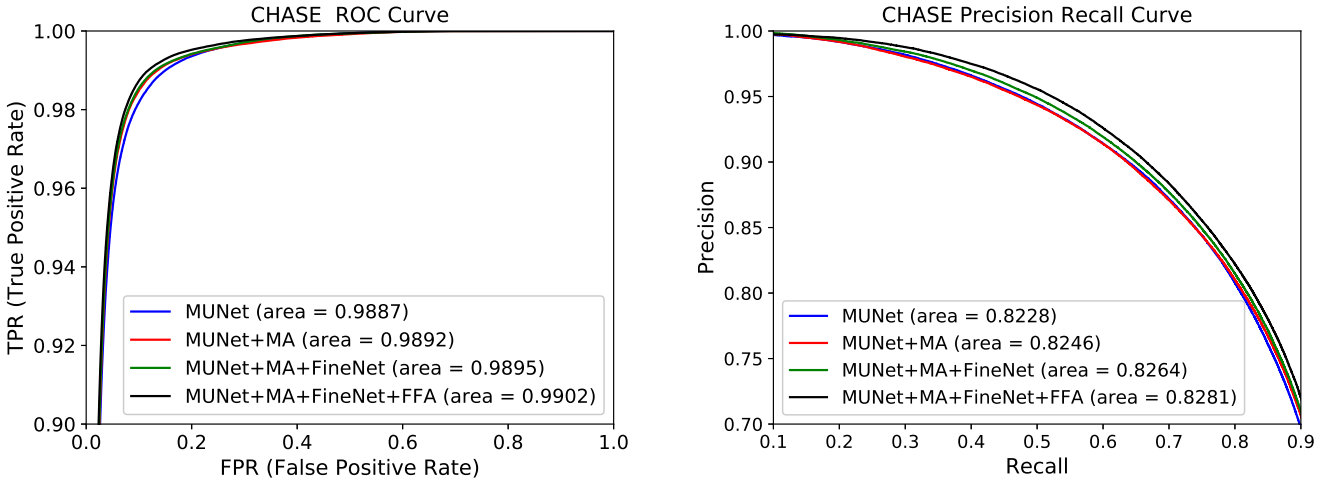

Figure 6. ROC curve and PR curve for four models on the CHASE dataset.

4.3. Attention Module Ablation

The multi-scale attention module (MA) was designed to enhance the network segmentation capability. In order to verify the effectiveness of the MA module and to verify that it had better performance than other attention modules, we designed a set of ablation experiments. The MU-Net+FineNet+FFA in the structural ablation section was used as a baseline and only the attention module embedded in the CoarseNet was changed to verify the advanced nature of the MA module proposed in this paper. We have selected three classic, lightweight attention modules to compare with the MA module. Among them, the first is the efficient channel attention (ECA) module of ECA-Net [31], which is often used in object detection and instance segmentation tasks. It was empirically shown that avoiding dimensionality reduction and appropriate cross-channel interaction are important to learn effective channel attention. The second attention module is the squeeze-and-excitation (SE) block in SENet [32], which improves the performance of model classification by modeling the inter-dependencies between feature channels. The last attention module is the expectation-maximization attention (EMA) block in EMANet [33] for semantic segmentation. It iterates through the expectation maximization (EM) algorithm to produce a compact set of bases on which to run the attention mechanism, thus greatly reducing the complexity.

Tables 4 and 5 show the experimental results of the MA module with the ECA module, SE block, and EMA module added into the baseline network, respectively. Although three attention modules, ECA, SE, and EMA, improved the performance of the model to a certain extent, from the evaluation indicators of F-Measure, the overall segmentation results of the MA were higher than those of the three attention modules embedded in the baseline network. Compared with the other three methods, the MA module has the best performance. This is because MA can automatically obtain the importance of each feature map by learning and can then promote the useful features and suppress the features that are not useful for the current task according to this importance. Meanwhile, MA brings minimal parameters to the network, which is friendly for clinical diagnosis.

Table 4. Experimental results of different attention modules on the DRIVE dataset.

Methods	Sensitivity	Specificity	Accuracy	F-Measure	Params Size (MB)
ECA [31]+Baseline	0.8153	0.9851	0.9702	0.8276	46.50
SE [32]+Baseline	0.7979	**0.9867**	0.9702	0.8245	46.43
EMA [33]+Baseline	0.8107	0.9856	**0.9703**	0.8273	46.32
MA+Baseline(Ours)	**0.8890**	0.9772	0.9693	**0.8382**	**46.17**

Table 5. Experimental results of different attention modules on the CHASE dataset.

Methods	Sensitivity	Specificity	Accuracy	F-Measure	Params Size (MB)
ECA [31]+Baseline	0.8175	0.9854	0.9748	0.8037	46.50
SE [32]+Baseline	0.8269	0.9847	0.9748	0.8056	46.43
EMA [33]+Baseline	0.7923	**0.9874**	0.9751	0.8008	46.32
MA+Baseline(Ours)	**0.8371**	0.9852	**0.9759**	**0.8139**	**46.17**

4.4. Model Parameter Quantity and Computation Time Analysis

The total parameters of this model are 46.17 MB. Our model took 6 s, 9.61 s, and 10.62 s to segment a complete retinal vessel image on DIRVE, CHASE, and STARE datasets. The U-Net model in [34] took 4 s to segment a complete image of the DRIVE data set, and the F-measure reached 0.8142. The F-measure of our model on the DRIVE data set is 0.8382. Compared with U-Net, our model takes 2 s more to segment images from DRIVE data sets, but our model is higher than U-Net in all indexes, especially the sensitivity is 13.53% higher than U-Net.

4.5. Visual Comparison with Different Methods

We compared our method with the methods proposed by U-Net [18] and WA-Net [35]. Figures 7 and 8 present a visualization of DRIVE and STARE data sets, respectively. In Figure 8, column (c) represents the result of segmentation using the slice method in [23].

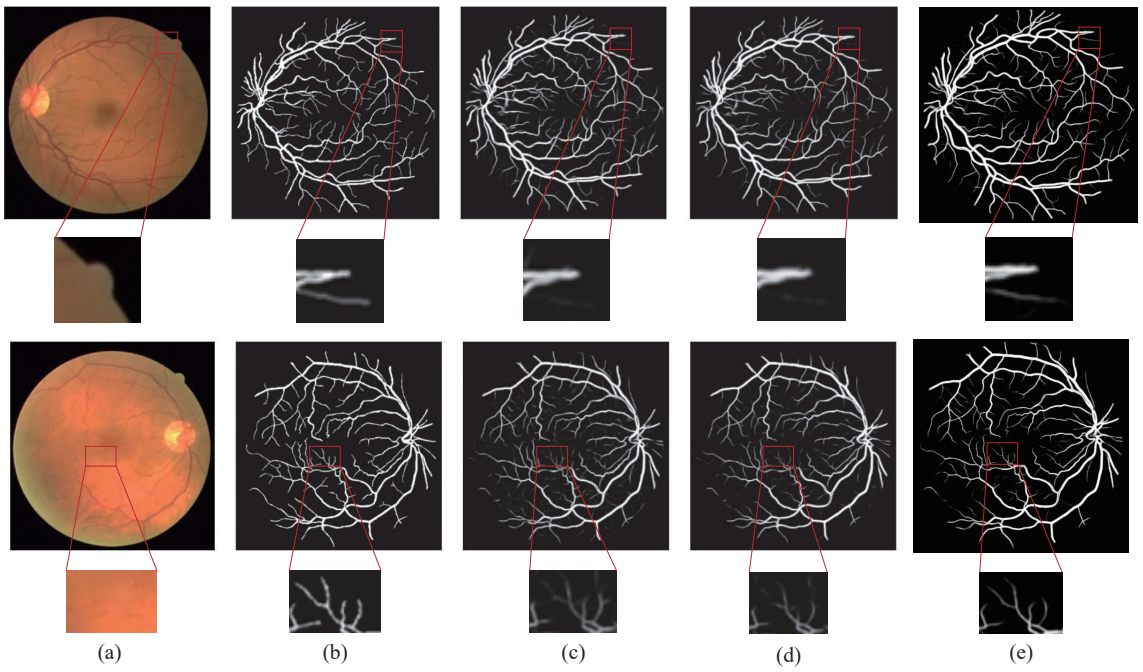

Figure 7. Visual comparison of segmentation results on the DRIVE dataset. Column (**a**): image; column (**b**): ground truth; column (**c**): U-Net [18]; column (**d**): WA-Net [35]; and column (**e**): our method.

The retinal vessels segmented by U-Net [18] contained more noise, and the background was mistakenly segmented into blood vessels. The small vessels at low contrast region are not clearly segmented and appear to have broken blood vessels. Although the retinal vessels segmented by WA-Net [35]contained less noise, the problem of unclear small vessels remained. Cc-Net [23] is susceptible to the lesion area, resulting in a lot of noise in the segmentation results. The segmented blood vessels are discontinuous.

The method proposed in this paper uses CoarseNet to extract rich context semantic information. The FineNet makes up for the spatial information, enables the network to distinguish the foreground and background regions well, and reduces wrong segmentations.

The segmentation results of our method contain less noise, especially the segmentation of small vessels in low-contrast regions. As shown in the red boxed regions of Figure 7 and 8. In these figures, it can be seen that the noise in the retinal vessel image segmented by our method is less and that the segmentation of small blood vessels is more comprehensive, clearer, and has better robustness and accuracy.

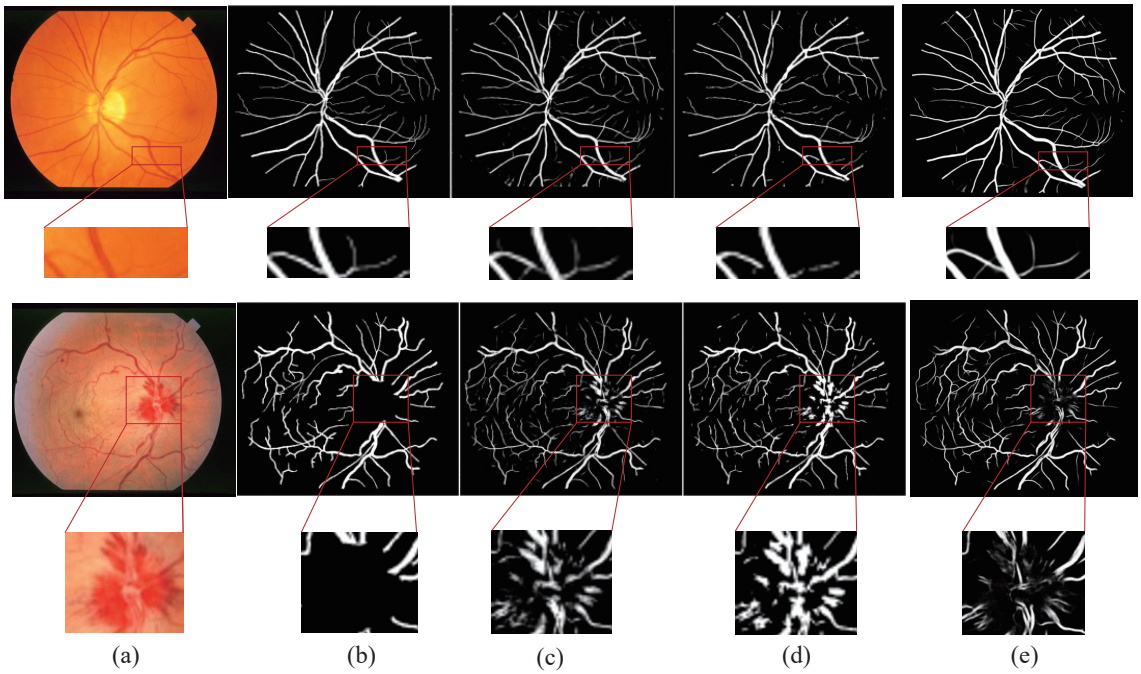

Figure 8. Visual comparison of segmentation results on CHASE dataset. Column (**a**): image; column (**b**): ground truth; column (**c**): using the slice method [23]; column (**d**): Cc-Net [23]; and column (**e**): our method

4.6. Comparison of Segmentation Results with Different Methods

In order to further verify the effectiveness of the proposed algorithm for retinal vessel segmentation, the proposed method is compared with some unsupervised and supervised methods in accuracy, sensitivity, specificity, and F-measure on the three data sets of DRIVE, CHASE, and STARE. From the experimental results in Tables 6–8, compared with unsupervised methods, supervised methods generally have better performance on retinal vessel segmentation. For the DRIVE data set, the F-measure of retinal vessel segmentation in this method reaches 83.82%, which is 0.93% higher than that in [36], and the sensitivity is 2.51% higher than that in [37].

Table 6. Comparison of proposed methods with other methods in the DRIVE database.

Type	Methods	Year	Sensitivity	Specificity	Accuracy	F-Measure
Unsupervised methods	Zhang [3]	2010	-	-	0.9382	-
	Wang [10]	2015	-	-	0.9457	-
	Miao [38]	2015	0.7481	0.9748	0.9597	-
Supervised methods	Marín [39]	2010	0.7607	0.9801	0.9452	-
	Aslani [40]	2016	0.7545	0.9801	0.9513	-
	Feng [17]	2017	0.7811	**0.9839**	0.9560	-
	U-Net [34]	2018	0.7537	0.9820	0.9531	0.8142
	IterNet [41]	2019	0.7735	0.9838	0.9573	0.8205
	Tian [37]	2019	0.8639	0.9690	0.9580	-
	Sine-Net [42]	2020	0.8260	0.9824	0.9685	-
	CTF-Net [21]	2020	0.7849	0.9813	0.9567	0.8241
	SA-Net [36]	2021	0.8252	0.9764	0.9569	0.8289
	CA-Net [43]	2021	0.8082	0.9858	**0.9703**	0.8261
	Ours	2021	**0.8890**	0.9772	0.9693	**0.8382**

Table 7. Comparison of proposed methods with other methods in the CHASE database.

Type	Methods	Year	Sensitivity	Specificity	Accuracy	F-Measure
Unsupervised methods	Azzopardi [6]	2015	0.7655	0.9704	0.9442	-
Supervised methods	Mo [44]	2017	0.7661	0.9816	0.9599	-
	Yan [45]	2018	0.7641	0.9806	0.9607	-
	U-Net [34]	2018	0.8288	0.9701	0.9578	0.7783
	DUNet [46]	2019	0.8155	0.9752	0.9610	0.7883
	Tian [37]	2020	**0.8778**	0.9680	0.9601	-
	IterNet [41]	2020	0.7970	0.9823	0.9655	0.8073
	Sine-Net [42]	2021	0.7856	0.9845	0.9676	-
	CA-Net [43]	2021	0.8138	0.9867	0.9758	0.8093
	Ours	2021	0.8371	**0.9852**	**0.9759**	**0.8139**

Table 8. Comparison of proposed methods with other methods in the STARE database.

Type	Methods	Year	Sensitivity	Specificity	Accuracy	F-Measure
Unsupervised methods	Azzopardi [6]	2015	0.7716	0.9701	0.9497	-
	Miao [38]	2015	0.7298	0.9831	0.9532	-
	Wang [10]	2015	-	-	0.9451	-
Supervised methods	Mo [44]	2017	0.8147	0.9844	0.9674	-
	U-Net [34]	2018	0.8270	0.9842	0.9690	0.8373
	IterNet [41]	2019	0.7715	0.9886	0.9701	0.8146
	Sine-Net [42]	2020	0.6776	0.9946	0.9711	-
	HANet [47]	2020	0.8186	0.9844	0.9673	0.8379
	Ours	2021	**0.8290**	**0.9894**	**0.9770**	**0.8436**

Table 9. Results of the leave-one-out in the STARE database.

Image	Accuracy	Sensitivity	Specificity	F-Measure
0	0.9725	0.8365	0.9843	0.8290
1	0.9772	0.7824	0.9911	0.8207
2	0.9826	0.8180	0.9931	0.8493
3	0.9686	0.6457	0.9945	0.7532
4	0.9649	0.7661	0.9847	0.7979
5	0.9776	0.8816	0.9849	0.8462
6	0.9794	0.9150	0.9850	0.8770
7	0.9806	0.8516	0.9910	0.8678
8	0.9833	0.8756	0.9925	0.8920
9	0.9747	0.8831	0.9827	0.8486
10	0.9797	0.8644	0.9886	0.8586
11	0.9829	0.9305	0.9873	0.8937
12	0.9777	0.8560	0.9895	0.8721
13	0.9801	0.8837	0.9897	0.8896
14	0.9767	0.8495	0.9887	0.8629
15	0.9637	0.7248	0.9908	0.8029
16	0.9761	0.8705	0.9865	0.8669
17	0.9858	0.8053	0.9955	0.8519
18	0.9846	0.7650	0.9945	0.8111
19	0.9709	0.7745	0.9849	0.7803
Average	**0.9770**	**0.8290**	**0.9894**	**0.8436**

From Tables 6–8, it can be seen that the specificity, accuracy, and F-measure of our method on different data sets are the highest in the table. Although the sensitivity of [37] on the CHASE data set is higher than that of our method, the segmentation effect on small blood vessels is not very good, and sometimes, a fracture occurs. Moreover, our method

has the highest F-measure, the specificity remains relatively stable, and the noise contained in the segmented image is relatively small. On the STARE data set, our method is 0.2% and 0.63% higher than that in [34] in sensitivity and F-measure, respectively. Therefore, from the evaluation results of blood vessel segmentation in Tables 6–8, it can be seen that the method in this paper is superior to other supervised blood vessel segmentation methods in segmenting retinal vessels and backgrounds and extracting different features. Table 9 shows test results per image on STARE data sets using the leave-one method.

5. Conclusions

The segmentation of retinal vessels is a key step in the diagnosis of ophthalmic diseases. In this paper, we proposed a two-branch model with a scale attention mechanism, which can automatically segment blood vessels in fundus images. The coarse network of this model takes a multi-scale U-Net as the backbone to capture more semantic information and to generate high-resolution features. At the same time, a multi-scale attention module is used to obtain enough reception fields. The other branch is a fine network, which used the residual block of a small convolution kernel to make up for the deficiency of spatial information from the coarse network. Finally, the feature fusion module is used to aggregate the information of coarse and fine branches. We validated this method on the DRIVE, STARE, and CHASE data sets, and the experimental results showed that our method has better performance in retinal vessel segmentation than some latest algorithms, such as WA-Net [35] and Sine-Net [42].

Several experimental results showed that our model has good results on these three data sets, which indicates that it has practical application potential in the screening and diagnosis system of ophthalmic diseases. The visualization results show that our method has good performance on small vessels in low-contrast areas. The imbalance in the number of foreground and background pixels in fundus images is also a problem that hinders vessel segmentation. In the future, we will alleviate the above problem by designing an auxiliary loss function.

Author Contributions: Data curation, J.Z.; methodology, H.Y.; resources, Y.J.; software, Z.M.; supervision, H.Y.; writing—review and editing, H.Y. and Y.J.; funding acquisition, Y.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (61962054), in part by the National Natural Science Foundation of China (61163036), in part by the Northwest Normal University Major Research Project Incubation Program (nwnu-LKZD2021_06), and in part by the Northwest Normal University's Third Phase of Knowledge and Innovation Engineering Research Backbone Project (nwnu-kjcxgc-03-67).

Institutional Review Board Statement: Ethical review and approval are not applicable for this paper.

Informed Consent Statement: An informed consent statement is not applicable.

Data Availability Statement: We used three public datasets to evaluate the proposed segmentation network, namely DRIVE [25], CHASE [26], and STARE [27].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Abràmoff, M.D.; Folk, J.C.; Han, D.P.; Walker, J.D.; Williams, D.F.; Russell, S.R.; Massin, P.; Cochener, B.; Gain, P.; Tang, L.; et al. Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmol.* **2013**, *131*, 351–357.
2. Robinson, B.E. Prevalence of Asymptomatic Eye Disease Prévalence des maladies oculaires asymptomatiques. *Rev. Can. D'Optométrie* **2003**, *65*, 175.
3. Zhang, B.; Zhang, L.; Zhang, L.; Karray, F. Retinal vessel extraction by matched filter with first-order derivative of Gaussian. *Comput. Biol. Med.* **2010**, *40*, 438–445.
4. Jiang, X.; Mojon, D. Adaptive Local Thresholding by Verification-Based Multithreshold Probing with Application to Vessel Detection in Retinal Images. *IEEE Comput. Soc.* **2003**, *25*, 131–137.
5. Zana, F.; Klein, J.C. Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE Trans. Image Process.* **2001**, *10*, 1010–1019

6. Azzopardi, G.; Strisciuglio, N.; Vento, M.; Petkov, N. Trainable COSFIRE filters for vessel delineation with application to retinal images. *Med. Image Anal.* **2015**, *19*, 46–57.
7. Wang, Y.; Ji, G.; Lin, P.; Trucco, E. Retinal vessel segmentation using multiwavelet kernels and multiscale hierarchical decomposition. *Pattern Recognit.* **2013**, *46*, 2117–2133.
8. Guo, Z.; Lin, P.; Ji, G.; Wang, Y. Retinal vessel segmentation using a finite element based binary level set method. *Inverse Probl. Imaging* **2017**, *8*, 459–473
9. Tolias, Y.A.; Panas, S.M. A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering. *IEEE Trans. Med. Imaging* **1998**, *17*, 263–273.
10. Wang, X.-H.; Zhao, Y.-Q.; Liao, M.; Zou, B.-J. Automatic segmentation for retinal vessel based on multiscale 2D Gabor wavelet. *Acta Autom. Sin.* **2015**, *41*, 970–980.
11. Liang, L.M.; Huang, C.L.; Shi, F.; Wu, J.; Jiang, H.J.; Chen, X.J. Retinal Vessel Segmentation Using Level Set Combined with Shape Priori. *Chin. J. Comput.* **2018**, *41*, 1678–1692.
12. Khalaf, A.F.; Yassine, I.A.; Fahmy, A.S. Convolutional neural networks for deep feature learning in retinal vessel segmentation. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 385–388.
13. Liskowski, P.; Krawiec, K. Segmenting retinal blood vessels with deep neural networks. *IEEE Trans. Med. Imaging* **2016**, *35*, 2369–2380.
14. Yu, L.; Qin, Z.; Zhuang, T.; Ding, Y.; Qin, Z.; Choo, K.R. A framework for hierarchical division of retinal vascular networks. *Neurocomputing* **2020**, *392*, 221–232.
15. Fu, H.; Xu, Y.; Lin, S.; Wong, D.W.K.; Liu, J. Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In *Lecture Notes in Computer Science, Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17–21 October 2016*; Springer: Cham, Switzerland, 2016; pp. 132–139.
16. Dasgupta, A.; Singh, S. A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentatio. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, Australia, 18–21 April 2017; pp. 248–251.
17. Feng, Z.; Yang, J.; Yao, L. Patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 1742–1746.
18. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science, Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241.
19. Zhang, Y.; Chung, A.C.S. Deep supervision with additional labels for retinal vessel segmentation task. In *Lecture Notes in Computer Science, Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018*; Springer: Cham, Switzerland, 2018; pp. 83–91.
20. Wu, Y.; Xia, Y.; Song, Y.; Zhang, Y; Cai, W. Multiscale network followed network model for retinal vessel segmentation. In *Lecture Notes in Computer Science, Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018*; Springer: Cham, Switzerland, 2018; pp. 119–126.
21. Wang, K.; Zhang, X.; Huang, S.; Wang, Q.; Chen, F. CTF-Net: Retinal Vessel Segmentation via Deep Coarse-To-Fine Supervision Network. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 1237–1241.
22. Wu, Y.; Xia, Y.; Song, Y.; Zhang, D.; Liu, D.; Zhang, C.; Cai, W. Vessel-Net: Retinal vessel segmentation under multi-path supervision. In *Lecture Notes in Computer Science, Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019*; Springer: Cham, Switzerland, 2019; pp. 264–272.
23. Feng, S.; Zhuo, Z.; Pan, D.; Tian, Q. CcNet: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features. *Neurocomputing* **2020**, *392*, 268–276.
24. Zhang, S.; Fu, H.; Yan, Y.; Zhang, Y.; Wu, Q.; Yang, M.; Tan, M; Xu, Y. Attention guided network for retinal image segmentation. In *Lecture Notes in Computer Science, Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019*; Springer: Cham, Switzerland, 2019; pp. 797–805.
25. Staal, J.; Abràmoff, M.D.; Niemeijer, M.; Viergever, M.A.; Ginneken B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **2004**, *23*, 501–509.
26. Owen, C.G.; Rudnicka, A.R.; Mullen, R.; Barman, S.A.; Monekosso, D.; Whincup, P.H.; Ng, J.; Paterson, C. Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (CAIAR) program. *Investig. Ophthalmol. Vis. Sci.* **2009**, *50*, 2004–2010.
27. Hoover, A.D.; Kouznetsova, V.; Goldbaum, M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imaging* **2000**, *19*, 203–210.
28. Zhuang, J. LadderNet: Multi-path networks based on U-Net for medical image segmentation. *arXiv*. **2018**, arXiv:1810.07810.
29. Jiang, Y.; Zhang, H.; Tan, N.; Chen, L. Automatic retinal blood vessel segmentation based on fully convolutional neural networks. *Symmetry* **2019**, *11*, 1112.

30. Jiang, Y.; Yao, H.; Wu, C.; Liu, W. A Multi-Scale Residual Attention Network for Retinal Vessel Segmentation. *Symmetry* **2021**, *13*, 24.
31. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020.
32. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
33. Li, X.; Zhong, Z.; Wu, J.; Yang, Y.; Lin, Z.; Liu, H. Expectation-maximization attention networks for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October– 2 November 2019; pp. 9167–9176.
34. Alom, M.Z.; Hasan, M.; Yakopcic, C.; Taha, T.M.; Asari, V.K. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv*. **2018**, arXiv:1802.06955.
35. Ma, Y.; Li, X.; Duan, X.; Peng, Y.; Zhang, Y. Retinal Vessel Segmentation by Deep Residual Learning with Wide Activation. *Comput. Intell. Neurosci.* **2020**, *2020*, 8822407.
36. Hu, J.; Wang, H.; Wang, J.; Wang, Y.; He, F.; Zhang, J. SA-Net: A scale-attention network for medical image segmentation. *PLoS ONE* **2021**, *16*, e0247388.
37. Tian, C.; Fang, T.; Fan, Y.; Wu, W. Multi-path convolutional neural network in fundus segmentation of blood vessels. *Biocybern. Biomed. Eng.* **2020**, *40*, 583–595.
38. Miao, Y.; Cheng, Y. Automatic extraction of retinal blood vessel based on matched filtering and local entropy thresholding. In Proceedings of the 2015 8th International Conference on Biomedical Engineering and Informatics (BMEI), Shenyang, China, 14–16 October 2015; pp. 62–67.
39. Marín, D.; Aquino, A.; Gegúndez-Arias, M.E.; Bravo, J.M. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Trans. Med. Imaging* **2010**, *30*, 146–158.
40. Aslani, S.; Sarnel, H. A new supervised retinal vessel segmentation method based on robust hybrid features. *Biomed. Signal Process. Control* **2016**, *30*, 1–12.
41. Li, L.; Verma, M.; Nakashima, Y.; Nagahara, H.; Kawasaki, R. Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 3656–3665.
42. Atli, İ.; Gedik, O.S. Sine-Net: A fully convolutional deep learning architecture for retinal blood vessel segmentation. *Eng. Sci. Technol. Int. J.* **2021**, *24*, 271–283.
43. Gu, R.; Wang, G.; Song, T.; Huang, R.; Aertsen, M.; Deprest, J. Ourselin, S.; Vercauteren, T.; Zhang, S. CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation. *IEEE Trans. Med. Imaging* **2020**, *40*, 699–711.
44. Mo, J.; Zhang, L. Multi-level deep supervised networks for retinal vessel segmentation. *Int. J. Comput. Assist. Radiol. Surg.* **2017**, *12*, 2181–2193.
45. Yan, Z.; Yang, X.; Cheng, K.T. A three-stage deep learning model for accurate retinal vessel segmentation. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 1427–1436.
46. Jin, Q.; Meng, Z.; Pham, T.D.; Chen, Q.; Wei, L.; Su, R. DUNet: A deformable network for retinal vessel segmentation. *Knowl.-Based Syst.* **2019**, *178*, 149–162.
47. Wang, D.; Haytham, A.; Pottenburgh, J.; Saeedi, O.; Tao, Y. Hard attention net for automatic retinal vessel segmentation. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 3384–3396.