



Article

A Multi-Scale Attention Fusion Network for Retinal Vessel Segmentation

Shubin Wang , Yuanyuan Chen * and Zhang Yi

Intelligent Interdisciplinary Research Center, College of Computer Science, Sichuan University, Chengdu 610065, China; wangshubin@stu.scu.edu.cn (S.W.); zhangyi@scu.edu.cn (Z.Y.)

* Correspondence: chenyuanyuan@scu.edu.cn

Abstract: The structure and function of retinal vessels play a crucial role in diagnosing and treating various ocular and systemic diseases. Therefore, the accurate segmentation of retinal vessels is of paramount importance to assist a clinical diagnosis. U-Net has been highly praised for its outstanding performance in the field of medical image segmentation. However, with the increase in network depth, multiple pooling operations may lead to the problem of crucial information loss. Additionally, handling the insufficient processing of local context features caused by skip connections can affect the accurate segmentation of retinal vessels. To address these problems, we proposed a novel model for retinal vessel segmentation. The proposed model is implemented based on the U-Net architecture, with the addition of two blocks, namely, an MsFE block and MsAF block, between the encoder and decoder at each layer of the U-Net backbone. The MsFE block extracts low-level features from different scales, while the MsAF block performs feature fusion across various scales. Finally, the output of the MsAF block replaces the skip connection in the U-Net backbone. Experimental evaluations on the DRIVE dataset, CHASE_DB1 dataset, and STARE dataset demonstrated that MsAF-UNet exhibited excellent segmentation performance compared with the state-of-the-art methods.

Keywords: deep neural networks; attention mechanism; retinal vessel segmentation



Citation: Wang, S.; Chen, Y.; Yi, Z. A Multi-Scale Attention Fusion Network for Retinal Vessel Segmentation. *Appl. Sci.* **2024**, *14*, 2955. <https://doi.org/10.3390/app14072955>

Academic Editor: Thomas Lindner

Received: 28 February 2024

Revised: 27 March 2024

Accepted: 30 March 2024

Published: 31 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The retinal vascular system is a crucial component of the visual system, playing a key role in maintaining intraocular homeostasis and ensuring visual function. It serves as a vital actor in sustaining the normal functionality of visual tissues by regulating blood flow and adjusting vascular tension, ensuring an ample supply of blood and oxygen to visual tissues. Additionally, the retinal vascular system is involved in the regulation of intraocular pressure, which is essential for maintaining the morphology and structure of the eyeball [1,2].

Under normal circumstances, the retinal vascular network exhibits a highly organized distribution, including major vessels, such as the central artery and central vein, along with various branches and capillaries. This structured arrangement ensures a stable blood supply to the retina, establishing optimal conditions for the transmission and processing of visual signals. Nevertheless, in pathological conditions, alterations to this structure may occur, leading to the formation of vascular abnormalities. These abnormalities can result in issues such as insufficient nutrient supply and hypoxia, ultimately affecting the normal functioning of the visual system [2,3].

In the field of ophthalmology, the structure and function of retinal vessels play a crucial role in diagnosing and treating various ocular and systemic diseases. The diameter, reflectivity, curvature, and branching characteristics of retinal blood vessels are crucial indicators for various retinal and systemic diseases [4–6]. A quantitative analysis of retinal vessels can assist ophthalmologists in detecting and diagnosing the early stages of certain severe conditions [7,8].

In recent years, methods utilizing artificial intelligence (AI) technology for medical image segmentation have garnered widespread attention [9]. In medical image analysis, AI technology has become a potent tool, assisting doctors in diagnosing diseases quickly and accurately. Retinal vessel segmentation is a crucial task in medical image analysis. The manual segmentation of retinal vessels is a laborious and time-consuming task prone to inter-observer variability. With the aid of AI technology, particularly machine learning and deep learning algorithms, automated retinal vessel segmentation can be achieved, aiding ophthalmologists in better understanding and analyzing the morphology and structure of retinal vessels, thereby providing superior clinical decision support for ophthalmologists.

The advancements in medical image segmentation have primarily been driven by deep learning techniques. The well-known CNN architecture U-Net [10] has demonstrated excellent performance in medical image segmentation. However, with the increase in network depth, multiple pooling operations may lead to the problem of crucial information loss, and the handling of insufficient local contextual features caused by skip connections can affect the accurate segmentation of retinal vessels. To address these problems, we proposed a multi-scale attention fusion network (MsAF-Net) for retinal vessel segmentation. Our main contributions are as follows:

- (1) We propose a multi-scale feature extraction (MsFE) block to capture diverse scale information from low-level features, providing the network with richer contextual information.
- (2) We propose a multi-scale attention fusion (MsAF) block, which combines channel attention from low-level features and spatial attention from high-level features, enabling the network to comprehensively understand the content of the image.
- (3) Combining the MsFE block and MsAF block, we propose a novel model for retinal vessel segmentation. Experimental results on three datasets demonstrated that our proposed model exhibited strong competitiveness compared with other state-of-the-art methods.

2. Related Works

2.1. Multi-Scale Feature Extraction

Multi-scale feature extraction is a common technique used in various computer vision tasks. A popular method for multi-scale feature extraction involves using filters with different sizes or receptive fields for image convolution. One notable example of this is the Inception v1, v2, and v3 modules proposed by Szegedy et al. [11,12].

For the task of retinal vessel segmentation, Yang et al. [13] proposed a segmentation method based on U-Net that utilizes the inception module to replace the convolution operation in the encoder. Compared with traditional convolution operations, the inception module can extract features at multiple scales. Experimental results on two datasets demonstrated the superior performance of this method, showing its competitiveness. Shi et al. [14] proposed a novel segmentation method named MD-Net. MD-Net adopts a strategy of dense connections and multi-scale feature extraction, enabling the network to simultaneously focus on both local details and global features of the image, thereby better capturing the structure and morphology of retinal vessels. Additionally, this method enhances the network performance by effectively utilizing residual learning and multi-scale receptive field design. Experimental validation on multiple datasets demonstrated the performance of MD-Net, with results indicating high segmentation accuracy and performance metrics across different datasets, thus confirming its superiority and effectiveness in retinal vessel segmentation tasks.

2.2. Attention Mechanism

Vaswani et al. [15] introduced the Transformer architecture, allowing the model to attend to different parts of a sequence without being constrained by the sequence length, thus enhancing the model's ability to handle long-range dependencies. Subsequently, attention mechanisms have found widespread application in computer vision tasks [16–24].

Researchers have proposed many methods for retinal vessel segmentation that combine attention mechanisms with the U-Net backbone. Dong et al. [25] proposed a cascaded U-Net framework to progressively extract features at different hierarchical levels from images. They introduced a residual attention block, incorporating an attention mechanism to enhance the network's focus on crucial image regions, thereby improving the precision of vessel segmentation. The experimental results on the DRIVE dataset and CHASE_DB1 dataset demonstrated its outstanding segmentation performance. Guo et al. [26] introduced SA-UNet for retinal vessel segmentation. SA-UNet replaces the original convolutional blocks in the U-Net framework by incorporating DropBlock [27] and batch normalization. Additionally, a spatial attention module is integrated between the encoder and decoder. The segmentation performance of SA-UNet is excellent on the DRIVE dataset and CHASE_DB1 dataset.

However, despite the good performance exhibited by U-Net and its variants in the field of medical image segmentation, increasing the network depth may lead to issues such as the loss of crucial information due to multiple pooling operations and the inadequate handling of local contextual features caused by skip connections. These issues can have an impact on the accurate segmentation of retinal vessels.

3. Methodology

Figure 1 illustrates the framework of the proposed model. The model is implemented based on the U-Net architecture and incorporates two blocks: the MsFE block and the MsAF block. These blocks are inserted between the encoder and decoder at each layer of the U-Net backbone. The MsFE block extracts low-level features from different scales, while the MsAF block performs feature fusion across different scales. Ultimately, the output of the MsAF block replaces the skip connection in the original U-Net.

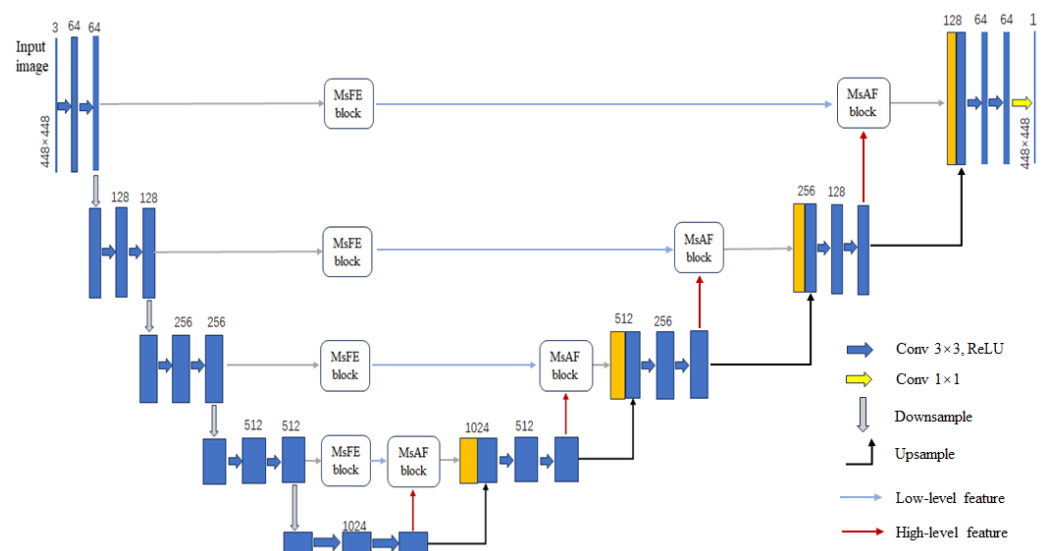


Figure 1. The framework of the proposed model.

3.1. MsFE Block

Multi-scale features have the capability to capture semantic information at different scales, providing richer contextual information. Therefore, inspired by [11,28], we employed the MsFE block to extract multi-scale information. As illustrated in Figure 2, the MsFE block consists of four parallel branches and a residual connection. The four parallel branches include convolutional operations of 1×1 , 3×3 , and 5×5 , as well as a 3×3 max-pooling operation. After concatenating the outputs of the four parallel branches, a 1×1 convolution and sigmoid activation are applied.

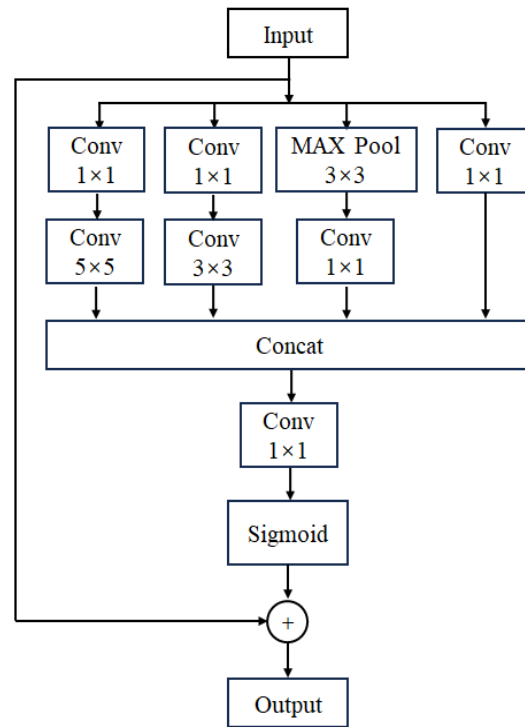


Figure 2. Illustration of the MsFE block. The MsFE block extracts features at different scales by using convolutional kernels of different sizes.

3.2. MsAF Block

The channel attention mechanism focuses on adjusting the weights of different channels in the network's feature maps to enhance useful features and suppress those that are less relevant to the current task. On the other hand, the spatial attention mechanism is concerned with how the network prioritizes different spatial positions in the image, allowing for a selective emphasis on crucial areas. Inspired by [16,29], we simultaneously introduced both channel attention and spatial attention mechanisms, designing a novel MsAF block. As illustrated in Figure 3, we applied channel attention to low-level features and spatial attention to high-level features. The final step involved merging the features extracted from each attention mechanism, enabling the network to comprehensively understand the content of the image. We describe the detailed operation below.

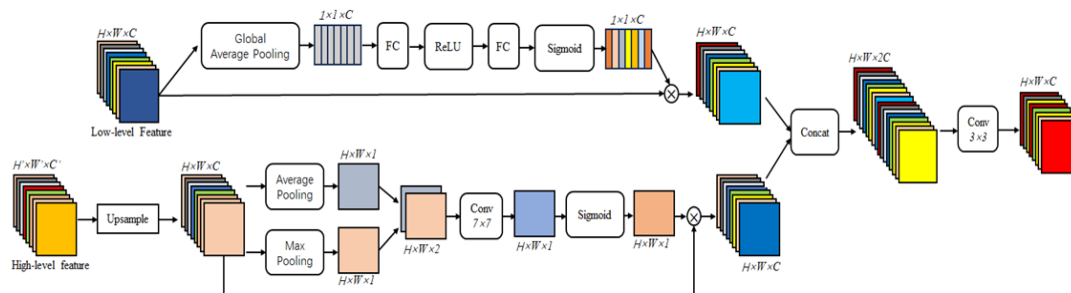


Figure 3. Illustration of the MsAF block. The MsAF block fuses channel attention from low-level features and spatial attention from high-level features.

First, we defined the low-level features as X_L and the high-level features after the upsample operation as X_H . For the low-level features X_L , global average pooling was applied to compress the features. Subsequently, two fully connected (FC) layers and an activation function were utilized to obtain channel-wise dependencies Y'_L . This can be expressed as

$$Y'_L = \theta(f_2(\delta(f_1(AvgPool(X_L))))), \quad (1)$$

where f_1 and f_2 denote the fully connected layers, θ denotes a sigmoid function, and δ denotes a ReLU function. Then, the channel attention map can be represented as

$$Y_L = X_L \cdot Y'_L, \quad (2)$$

where $Y_L \in R^{H \times W \times C}$.

For the high-level features X_H , we initially performed two pooling operations, concatenated the resulting two feature maps, and then utilized a 7×7 convolution to generate the spatial attention map Y'_H . This can be expressed as

$$Y'_H = \theta(f^{7 \times 7}([AvgPool(X_H); MaxPool(X_H)])), \quad (3)$$

where θ denotes a sigmoid function. Then, the final spatial attention map Y_H can be represented as

$$Y_H = X_H \cdot Y'_H, \quad (4)$$

where $Y_H \in R^{H \times W \times C}$.

We concatenated the obtained channel attention map X_L and spatial attention map X_H , performed a 3×3 convolution operation, and generated a multi-scale attention fusion feature map Y . This can be represented as follows:

$$Y = f^{3 \times 3}([X_L; X_H]), \quad (5)$$

where $Y \in R^{H \times W \times C}$, and $f^{3 \times 3}$ represents the convolution operation.

4. Experiments and Results

4.1. Dataset

The experiments were conducted using the DRIVE dataset [30], the CHASE_DB1 dataset [31], and the STARE dataset [32]. Table 1 displays the specific information of each dataset. Due to the small sizes of the three datasets, which may lead to overfitting, we utilized horizontal flips, vertical flips, rotations, addition of Gaussian noise, adjustment of brightness, and other methods to augment the data. After the augmentation, the training samples for the DRIVE dataset and CHASE_DB1 dataset reached 800 images each, and for the STARE dataset, there were 400 training samples.

Table 1. The specific information of each database.

Dataset	Total Images	Training Set	Testing Set
DRIVE dataset	40	20	20
CHASE_DB1 dataset	28	20	8
STARE dataset	20	10	10

4.2. Experimental Setup

All experiments employed the Adam optimizer, training was stopped after 200 epochs, and the input image size was uniformly resized to $448 \times 448 \times 3$. The Dice loss was used as the loss function, with an initial learning rate of 0.001. Evaluation metrics comprised the sensitivity (SE), F1-score (F1), specificity (SP), accuracy (ACC), and the area under the receiver operating characteristic curve (AUC).

4.3. Results

4.3.1. Comparison with Baseline Models

We conducted experiments on U-Net [10], Unet++ [33], and attention U-Net [34] as baseline models.

Table 2 displays the sensitivity, F1-score, specificity, accuracy, and AUC values of the proposed model and the baseline model on the DRIVE dataset. It can be observed that compared with the baseline models, although the specificity value of Unet++ was higher than that of the proposed model, the proposed model achieved the highest values in terms of sensitivity, F1-score, accuracy, and AUC.

Table 2. The segmentation results on the DRIVE dataset.

Model	SE	F1	SP	ACC	AUC
U-Net [10]	0.8452	0.8287	0.9798	0.9521	0.9738
Unet++ [33]	0.8510	0.8319	0.9865	0.9542	0.9741
Attention U-Net [34]	0.8427	0.8294	0.9807	0.9534	0.9726
The proposed model	0.8611	0.8383	0.9851	0.9629	0.9760

Tables 3 and 4 display the sensitivity, F1-score, specificity, accuracy, and AUC values of the proposed model and the baseline model on the CHASE_DB1 dataset and STARE dataset, respectively. It can be observed that compared with the baseline models, all evaluation metrics of the proposed model achieved the highest values on the CHASE_DB1 dataset and the STARE dataset.

The ROC curves for the proposed model and the baseline models on three datasets are depicted in Figure 4. It is evident that the proposed model achieved higher ROC values on all three datasets compared with the baseline models, indicating superior segmentation performance of the proposed model over the baselines.

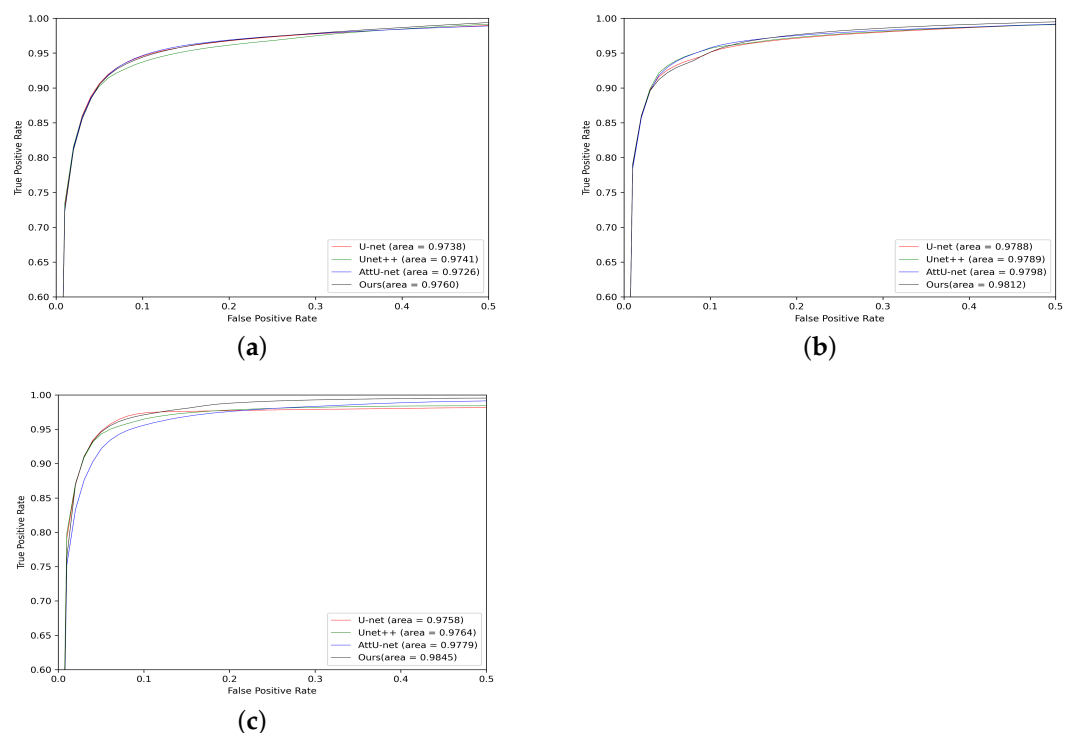


Figure 4. The ROC curves of the proposed model and baseline models. (a) The ROC curves on the DRIVE dataset. (b) The ROC curves on the CHASE_DB1 dataset. (c) The ROC curves on the STARE dataset.

Table 3. The segmentation results on the CHASE_DB1 dataset.

Model	SE	F1	SP	ACC	AUC
U-Net [10]	0.8459	0.8322	0.9805	0.9613	0.9788
Unet++ [33]	0.8443	0.8325	0.9823	0.9615	0.9789
Attention U-Net [34]	0.8517	0.8384	0.9825	0.9608	0.9798
The proposed model	0.8630	0.8435	0.9850	0.9721	0.9812

Table 4. The segmentation results on the STARE dataset.

Model	SE	F1	SP	ACC	AUC
U-Net [10]	0.8320	0.8217	0.9814	0.9644	0.9758
Unet++ [33]	0.8425	0.8253	0.9829	0.9650	0.9764
Attention U-Net [34]	0.8349	0.8236	0.9817	0.9637	0.9779
The proposed model	0.8614	0.8511	0.9866	0.9754	0.9845

4.3.2. Comparison with State-of-the-Art Segmentation Methods

The comparison of the performance between the proposed model and state-of-the-art methods on the DRIVE dataset is shown in Table 5. Compared with the other advanced models, MsAF achieved the highest sensitivity value of 0.8611 and the highest F1-score value of 0.8383 on the DRIVE dataset, surpassing the highest values obtained by the other advanced models by 0.14% and 0.26%, respectively. The differences in specificity, accuracy, and AUC compared with the highest values obtained by the other advanced models were 0.19%, 0.72%, and 1.15%, respectively.

Table 5. Comparison results on the DRIVE dataset.

Model	SE	F1	SP	ACC	AUC
Jin et al. [35]	0.7894	0.8203	0.9870	0.9697	0.9856
Guo et al. [36]	0.7891	0.8249	0.9804	0.9561	0.9806
Wang et al. [37]	0.7940	0.8270	0.9816	0.9567	0.9772
Li et al. [38]	0.7791	0.8218	0.9831	0.9574	0.9813
Zhang et al. [39]	0.8215	0.8267	0.9845	0.9701	0.9867
Wu et al. [40]	0.8520	0.8297	-	0.9555	0.9814
Li et al. [41]	0.8324	-	0.9757	0.9574	0.9820
Wang et al. [42]	0.8071	0.8251	0.9782	0.9565	0.9801
Gegundez-Arias et al. [43]	0.8597	-	0.9690	0.9547	0.9837
Lin et al. [44]	0.8361	0.8287	0.9740	0.9563	0.9799
Guo et al. [26]	0.8212	0.8263	0.9840	0.9698	0.9864
Li et al. [45]	0.8291	0.8302	0.9852	0.9622	0.9859
Wei et al. [46]	0.8302	0.8018	0.9826	0.9581	0.9821
Shen et al. [47]	0.8056	0.8357	0.9854	0.9680	0.9875
The proposed model	0.8611	0.8383	0.9851	0.9629	0.9760

The comparison of the performance between the proposed model and state-of-the-art methods on the CHASE_DB1 dataset is shown in Table 6. Compared with the other advanced models, MsAF achieved the highest sensitivity value of 0.8630 and the highest F1-score value of 0.8435 on the DRIVE dataset, surpassing the highest values obtained by the other advanced models by 0.57% and 0.85%, respectively. The differences in specificity, accuracy, and AUC compared with the highest values obtained by the other advanced models were 0.46%, 0.34%, and 0.94%, respectively.

Table 6. Comparison results on the CHASE_DB1 dataset.

Model	SE	F1	SP	ACC	AUC
Jin et al. [35]	0.8229	0.7853	0.9821	0.9724	0.9863
Guo et al. [36]	0.8155	0.7883	0.9752	0.9610	0.9804
Wang et al. [37]	0.7888	0.7983	0.9801	0.9627	0.9840
Li et al. [38]	0.7970	0.8073	0.9823	0.9655	0.9851
Wu et al. [40]	0.7996	0.8031	-	0.9642	0.9823
Gegundez-Arias et al. [43]	0.8044	-	0.9698	0.9663	0.9880
Lin et al. [44]	0.8448	0.8332	0.9795	0.9668	0.9861
Guo et al. [26]	0.8573	0.8153	0.9835	0.9755	0.9905
Li et al. [45]	0.7856	0.8355	0.9896	0.9660	0.9876
Wei et al. [46]	0.8196	0.8138	0.9824	0.9678	0.9872
Shen et al. [47]	0.8250	0.8350	0.9875	0.9734	0.9906
The proposed model	0.8630	0.8435	0.9850	0.9721	0.9812

The comparison of the performance between the proposed model and state-of-the-art methods on the CHASE_DB1 dataset is shown in Table 7. Compared with the other advanced models, MsAF achieved the highest sensitivity value of 0.8630 and the highest F1-score value of 0.8435 on the DRIVE dataset, surpassing the highest values obtained by the other advanced models by 0.48% and 0.20%, respectively. The differences in specificity, accuracy, and AUC compared with the highest values obtained by the other advanced models were 0.91%, 0.05%, and 0.81%, respectively.

Table 7. Comparison results on the STARE dataset.

Model	SE	F1	SP	ACC	AUC
Jin et al. [35]	0.7428	0.8079	0.9920	0.9729	0.9868
Li et al. [38]	0.7715	0.8146	0.9886	0.9701	0.9881
Li et al. [41]	0.8189	-	0.9887	0.9759	0.9912
Gegundez-Arias et al. [43]	0.8441	-	0.9764	0.9754	0.9926
Lin et al. [44]	0.8566	0.8491	0.9819	0.9681	0.9874
Li et al. [45]	0.7616	0.8022	0.9957	0.9653	0.9889
Wei et al. [46]	0.8341	0.8012	0.9865	0.9672	0.9876
Shen et al. [47]	0.8140	0.8000	0.9817	0.9686	0.9832
The proposed model	0.8614	0.8511	0.9866	0.9754	0.9845

4.3.3. Qualitative Analysis

In order to visually observe the segmentation results more intuitively, this work introduced qualitative analysis for performance visualization. A sample image was selected from each respective test set, and the corresponding segmentation results are presented in Figures 5–7. As demonstrated in these figures, the proposed model exhibited satisfactory segmentation performance, showcasing its ability to effectively detect vessels in retinal images.

Examining the red box in Figure 5, it is evident that the proposed model accurately identified vessels in the retinal image, whereas the baseline model mistakenly identified a branching vessel at the same location, resulting in false positives. In the green box of Figure 5, it is apparent that all models failed to recognize a small vessel, leading to false negatives.

In Figure 6, the proposed model accurately identified vessels in the retinal image. In the red box of Figure 6, the baseline model erroneously identified a branching vessel, resulting in false positives. In the green box of Figure 6, Attention U-Net accurately recognized this vessel, while U-Net and Unet++ only identified a portion, resulting in false negatives.

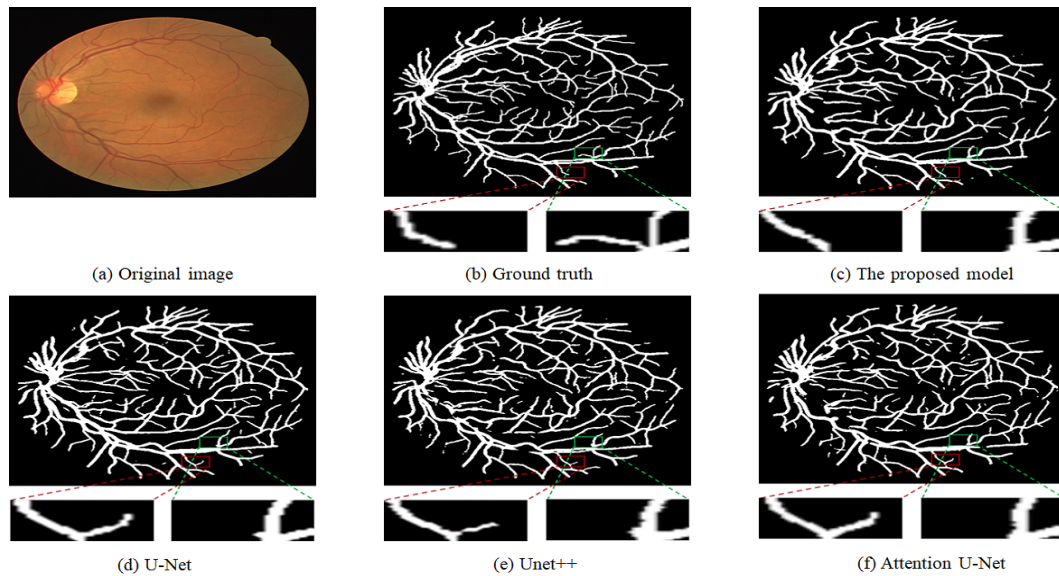


Figure 5. The segmentation result on the DRIVE dataset. The red boxes in subfigures (c–f) represent false positives, while the green boxes represent false negatives.

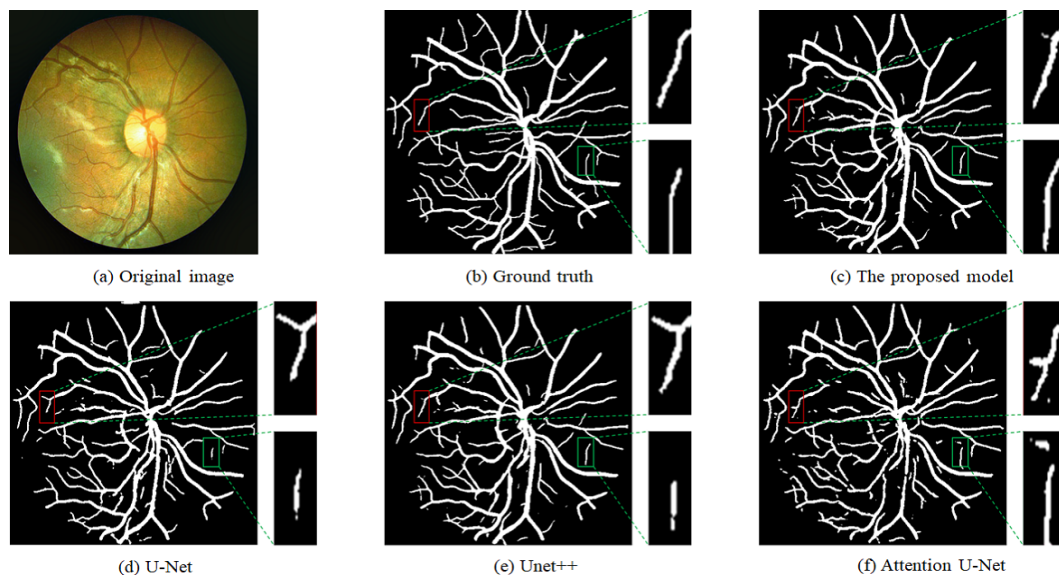


Figure 6. The segmentation result on the CHASE_DB1 dataset. The red boxes in subfigures (c–f) represent false positives, while the green boxes represent false negatives.

Examining the green box in Figure 7, it is observed that the proposed model accurately identified this vessel more completely. Unet++ recognized only a small portion, while U-Net and Attention U-Net almost failed to identify this vessel, resulting in false negatives. In the red box of the ground truth in Figure 7, representing a completely background area, all models recognized an incomplete vessel, leading to false positives.

Figures 8–10 display the differential images of segmentation results on a different dataset each. Through the analysis of the differential images, we can distinctly observe the distribution of false positives and false negatives across different models and datasets. Upon comparing the differential images of different models, as depicted in Figures 8–10, it is evident that the proposed model exhibited lower false positives and false negatives compared with the baseline model across all three datasets. Furthermore, contrasting the differential images of different datasets revealed significant variations in the quantities of false positives and false negatives. In the segmentation of the DRIVE dataset and CHASE_DB1 dataset, missegmentation of major vessels led to a higher occurrence of false

positives and false negatives. Conversely, in the segmentation of the STARE dataset, major vessels were accurately segmented, with only minor vessels exhibiting no segmentation, resulting in lower false positives and false negatives. After an analysis, we attributed these differences to variations in image quality and annotation accuracy between the datasets.

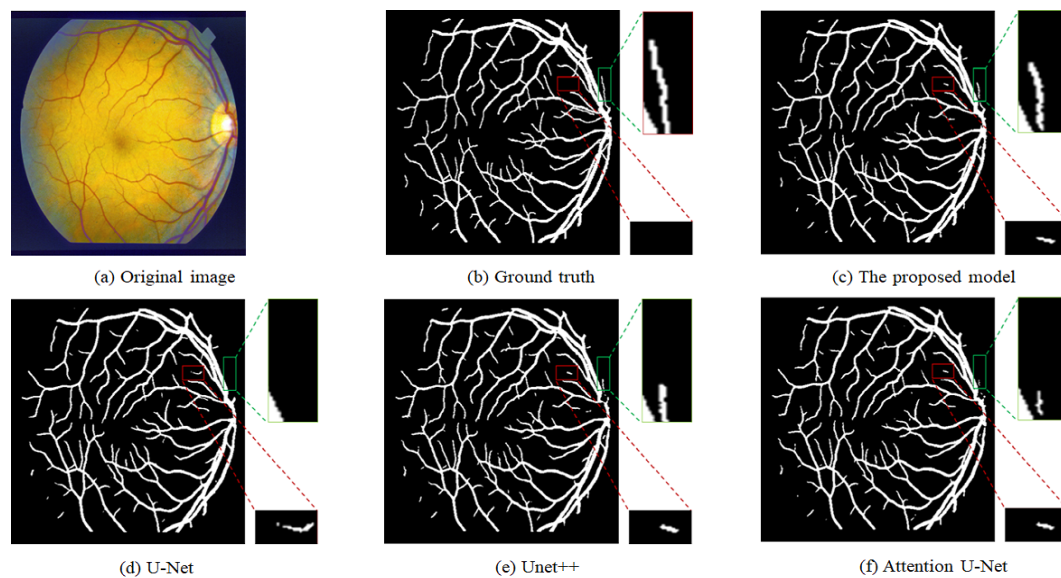


Figure 7. The segmentation result on the STARE dataset. The red boxes in subfigures (c–f) represent false positives, while the green boxes represent false negatives.

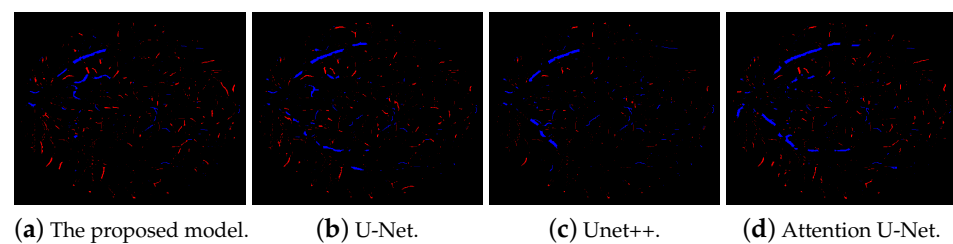


Figure 8. The differential images from the DRIVE dataset.

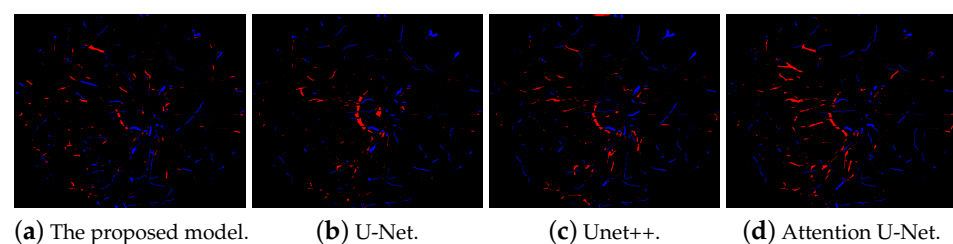


Figure 9. The differential images from the CHASE_DB1 dataset.

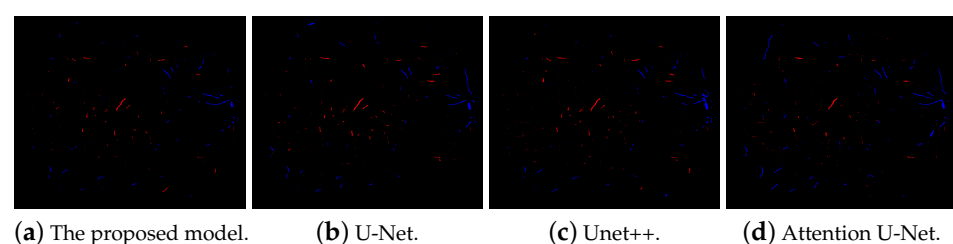


Figure 10. The differential images from the STARE dataset.

4.4. Discussion

Due to the complex morphology of retinal vessels, which includes branching, bending, and irregular shapes, and the small proportion of vessels leading to severe class imbalance, retinal vessel segmentation poses significant challenges. Additionally, retinal fundus images are often affected by noise, resulting in a low image contrast. Therefore, the task of retinal vessel segmentation is highly challenging.

In retinal vessel segmentation tasks, false positives and false negatives can lead to different consequences. First, false positives may result in incorrect diagnoses or unnecessary treatments. If non-vessel regions are erroneously labeled as blood vessels, clinicians may mistakenly believe that abnormalities exist and proceed with unnecessary further examinations or therapies. On the other hand, false negatives may lead to overlooking lesions or abnormalities in the retina, such as vessel occlusion or abnormal vessel morphology. If crucial vessel regions are erroneously excluded, clinicians may miss important clues for diagnosing diseases, leading to delayed treatment or inadequate therapy. Therefore, reducing the occurrences of false positives and false negatives is crucial in retinal vessel segmentation tasks.

The proposed MsFE block extracts low-level features from different scales, and the MsAF block integrates channel attention from low-level features and spatial attention from high-level features, enabling the model to comprehend the image content more comprehensively. As observed in Figures 5–7, the proposed model accurately identified small vessels in the images, demonstrating lower false negatives compared with the baseline network. Meanwhile, the analysis of the differential images in the previous section for Figures 8–10 demonstrated that the false positives and false negatives of the proposed model were lower than those of the baseline model across all three datasets. All of these indicate that the fusion of features from different scales contributed to the improvement of accuracy in retinal vessel segmentation.

After a comprehensive analysis of the data in Tables 5–7, it is evident that in comparison with other state-of-the-art models, the proposed model achieved the highest values in sensitivity and F1-score on the three datasets. Although the proposed model did not attain the highest values in specificity, accuracy, and AUC, the differences compared with the other advanced models were minimal. Sensitivity measures a model's ability to identify true positive samples, while the F1-score provides a balanced evaluation of the model's classification performance on both positive and negative samples, which is particularly useful in scenarios with class imbalance. In the context of retinal vessel segmentation tasks, correctly identifying vessel pixels in the image is crucial. The small proportion of vessel pixels in retinal images led to a severe class imbalance issue. Therefore, given that the proposed model attained the highest values in sensitivity and F1-score on the DRIVE dataset, CHASE_DB1 dataset, and STARE dataset, it demonstrated excellent segmentation performance.

Overall, the segmentation performance of the proposed model was satisfactory. However, there were still some segmentation errors due to the unique morphology of the vessels and factors such as a low image contrast. Additionally, the limited size of the datasets may pose challenges to the generalization capability of the model.

5. Conclusions

In this paper, we propose a novel model for retinal vessel segmentation built upon the U-Net backbone architecture. For each layer of the U-Net backbone, two additional blocks, namely, the MsFE block and the MsAF block, were incorporated between the encoder and decoder. The MsFE block extracts low-level features from different scales, while the MsAF block performs attention fusion across various scales. Experimental evaluations were conducted on the DRIVE dataset, CHASE_DB1 dataset, and STARE dataset, demonstrating that MsAF-UNet exhibited competitive performance.

In future work, our focus will be on investigating multi-scale attention fusion mechanisms to further enhance the segmentation performance of retinal vessels.

Author Contributions: Conceptualization, S.W., Y.C. and Z.Y.; methodology, S.W., Y.C. and Z.Y.; software, S.W., Y.C. and Z.Y.; validation, S.W.; formal analysis, S.W., Y.C. and Z.Y.; investigation, S.W., Y.C. and Z.Y.; resources, Y.C. and Z.Y.; data curation, S.W.; writing—original draft, S.W.; writing—review and editing, Y.C. and Z.Y.; visualization, S.W.; supervision, Y.C. and Z.Y.; project administration, Y.C. and Z.Y.; funding acquisition, Y.C. and Z.Y. All authors read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant no. 61702349) and National Major Science and Technology Projects of China (grant no. 2018AAA0100201).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are openly available in the DRIVE dataset at <https://drive.grand-challenge.org/> accessed on 28 February 2024, the CHASE_DB1 dataset at <https://www.kaggle.com/datasets/rashasrhanalharthi/chase-db1/> accessed on 28 February 2024, and the STARE dataset at <https://cecas.clemson.edu/~ahoover/stare/probing/index.html> accessed on 28 February 2024.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Tso, M.O.; Jampol, L.M. Pathophysiology of hypertensive retinopathy. *Ophthalmology* **1982**, *89*, 1132–1145. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Wong, T.Y.; Mitchell, P. Hypertensive retinopathy. *N. Engl. J. Med.* **2004**, *351*, 2310–2317. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Cunha-Vaz, J.; De Abreu, J.F.; Campos, A. Early breakdown of the blood-retinal barrier in diabetes. *Br. J. Ophthalmol.* **1975**, *59*, 649–656. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Mitchell, P.; Leung, H.; Wang, J.J.; Rochtchina, E.; Lee, A.J.; Wong, T.Y.; Klein, R. Retinal vessel diameter and open-angle glaucoma: The Blue Mountains Eye Study. *Ophthalmology* **2005**, *112*, 245–250. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Kipli, K.; Hoque, M.E.; Lim, L.T.; Mahmood, M.H.; Sahari, S.K.; Sapawi, R.; Rajae, N.; Joseph, A. A review on the extraction of quantitative retinal microvascular image feature. *Comput. Math. Methods Med.* **2018**, *2018*, 4019538. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Cheung, C.Y.I.; Zheng, Y.; Hsu, W.; Lee, M.L.; Lau, Q.P.; Mitchell, P.; Wang, J.J.; Klein, R.; Wong, T.Y. Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors. *Ophthalmology* **2011**, *118*, 812–818. [\[CrossRef\]](#) [\[PubMed\]](#)
7. MacGillivray, T.; Trucco, E.; Cameron, J.; Dhillion, B.; Houston, J.; Van Beek, E. Retinal imaging as a source of biomarkers for diagnosis, characterization and prognosis of chronic illness or long-term conditions. *Br. J. Radiol.* **2014**, *87*, 20130832. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Irshad, S.; Akram, M.U. Classification of retinal vessels into arteries and veins for detection of hypertensive retinopathy. In Proceedings of the 2014 Cairo International Biomedical Engineering Conference (CIBEC), Giza, Egypt, 11–13 December 2014; pp. 133–136.
9. Kufel, J.; Bargiel-Łączek, K.; Kocot, S.; Koźlik, M.; Bartnikowska, W.; Janik, M.; Czogalik, Ł.; Dudek, P.; Magiera, M.; Lis, A.; et al. What is machine learning, artificial neural networks and deep learning?—Examples of practical applications in medicine. *Diagnostics* **2023**, *13*, 2582. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015*; Proceedings, Part III 18; Springer: Cham, Switzerland, 2015; pp. 234–241.
11. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–15 June 2015; pp. 1–9.
12. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
13. Yang, D.; Liu, G.; Ren, M.; Xu, B.; Wang, J. A multi-scale feature fusion method based on u-net for retinal vessel segmentation. *Entropy* **2020**, *22*, 811. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Shi, Z.; Wang, T.; Huang, Z.; Xie, F.; Liu, Z.; Wang, B.; Xu, J. MD-Net: A multi-scale dense network for retinal vessel segmentation. *Biomed. Signal Process. Control* **2021**, *70*, 102977. [\[CrossRef\]](#)
15. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
16. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

17. Zhao, H.; Zhang, Y.; Liu, S.; Shi, J.; Loy, C.C.; Lin, D.; Jia, J. Psanet: Point-wise spatial attention network for scene parsing. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 267–283.
18. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.
19. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-attention generative adversarial networks. In Proceedings of the International Conference on Machine Learning. PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 7354–7363.
20. 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, Republic of Korea, 27 October–2 November 2019; pp. 9167–9176.
21. Rundo, L.; Han, C.; Nagano, Y.; Zhang, J.; Hataya, R.; Militello, C.; Tangherloni, A.; Nobile, M.S.; Ferretti, C.; Besozzi, D.; et al. USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. *Neurocomputing* **2019**, *365*, 31–43. [[CrossRef](#)]
22. Wang, Y.; Dou, H.; Hu, X.; Zhu, L.; Yang, X.; Xu, M.; Qin, J.; Heng, P.A.; Wang, T.; Ni, D. Deep attentive features for prostate segmentation in 3D transrectal ultrasound. *IEEE Trans. Med. Imaging* **2019**, *38*, 2768–2778. [[CrossRef](#)] [[PubMed](#)]
23. Zhou, Y.; Huang, W.; Dong, P.; Xia, Y.; Wang, S. D-UNet: A dimension-fusion U shape network for chronic stroke lesion segmentation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *18*, 940–950. [[CrossRef](#)] [[PubMed](#)]
24. Ni, J.; Wu, J.; Tong, J.; Chen, Z.; Zhao, J. GC-Net: Global context network for medical image segmentation. *Comput. Methods Programs Biomed.* **2020**, *190*, 105121. [[CrossRef](#)] [[PubMed](#)]
25. Dong, F.; Wu, D.; Guo, C.; Zhang, S.; Yang, B.; Gong, X. CRAUNet: A cascaded residual attention U-Net for retinal vessel segmentation. *Comput. Biol. Med.* **2022**, *147*, 105651. [[CrossRef](#)] [[PubMed](#)]
26. Guo, C.; Szemenyei, M.; Yi, Y.; Wang, W.; Chen, B.; Fan, C. Sa-unet: Spatial attention u-net for retinal vessel segmentation. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 1236–1242.
27. Ghiasi, G.; Lin, T.Y.; Le, Q.V. Dropblock: A regularization method for convolutional networks. *Adv. Neural Inf. Process. Syst.* **2018**, *31*.
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
29. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
30. Staal, J.; Abramoff, M.D.; Niemeijer, M.; Viergever, M.A.; Van Ginneken, B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **2004**, *23*, 501–509. [[CrossRef](#)] [[PubMed](#)]
31. Fraz, M.M.; Remagnino, P.; Hoppe, A.; Uyyanonvara, B.; Rudnicka, A.R.; Owen, C.G.; Barman, S.A. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 2538–2548. [[CrossRef](#)] [[PubMed](#)]
32. Hoover, A.; 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. [[CrossRef](#)] [[PubMed](#)]
33. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, 20 September 2018*; Proceedings 4; Springer: Cham, Switzerland, 2018; pp. 3–11.
34. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
35. 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. [[CrossRef](#)]
36. Guo, S.; Wang, K.; Kang, H.; Zhang, Y.; Gao, Y.; Li, T. BTS-DSN: Deeply supervised neural network with short connections for retinal vessel segmentation. *Int. J. Med. Inform.* **2019**, *126*, 105–113. [[CrossRef](#)] [[PubMed](#)]
37. Wang, B.; Qiu, S.; He, H. Dual encoding u-net for retinal vessel segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd International Conference, Shenzhen, China, 13–17 October 2019*; Proceedings, Part I 22; Springer: Cham, Switzerland, 2019; pp. 84–92.
38. 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.
39. Zhang, M.; Yu, F.; Zhao, J.; Zhang, L.; Li, Q. BEFD: Boundary enhancement and feature denoising for vessel segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference, Lima, Peru, 4–8 October 2020*; Proceedings, Part V 23; Springer: Cham, Switzerland, 2020; pp. 775–785.
40. Wu, Y.; Xia, Y.; Song, Y.; Zhang, Y.; Cai, W. NFN+: A novel network followed network for retinal vessel segmentation. *Neural Netw.* **2020**, *126*, 153–162. [[CrossRef](#)]
41. Li, D.; Rahardja, S. BSEResU-Net: An attention-based before-activation residual U-Net for retinal vessel segmentation. *Comput. Methods Programs Biomed.* **2021**, *205*, 106070. [[CrossRef](#)] [[PubMed](#)]

42. Wang, B.; Wang, S.; Qiu, S.; Wei, W.; Wang, H.; He, H. CSU-Net: A context spatial U-Net for accurate blood vessel segmentation in fundus images. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 1128–1138. [[CrossRef](#)] [[PubMed](#)]
43. Gegundez-Arias, M.E.; Marin-Santos, D.; Perez-Borrero, I.; Vasallo-Vazquez, M.J. A new deep learning method for blood vessel segmentation in retinal images based on convolutional kernels and modified U-Net model. *Comput. Methods Programs Biomed.* **2021**, *205*, 106081. [[CrossRef](#)]
44. Lin, Z.; Huang, J.; Chen, Y.; Zhang, X.; Zhao, W.; Li, Y.; Lu, L.; Zhan, M.; Jiang, X.; Liang, X. A high resolution representation network with multi-path scale for retinal vessel segmentation. *Comput. Methods Programs Biomed.* **2021**, *208*, 106206. [[CrossRef](#)] [[PubMed](#)]
45. Li, J.; Gao, G.; Yang, L.; Liu, Y. GDF-Net: A multi-task symmetrical network for retinal vessel segmentation. *Biomed. Signal Process. Control* **2023**, *81*, 104426. [[CrossRef](#)]
46. Wei, X.; Yang, K.; Bzdok, D.; Li, Y. Orientation and Context Entangled Network for Retinal Vessel Segmentation. *Expert Syst. Appl.* **2023**, *217*, 119443. [[CrossRef](#)]
47. Shen, X.; Xu, J.; Jia, H.; Fan, P.; Dong, F.; Yu, B.; Ren, S. Self-attentional microvessel segmentation via squeeze-excitation transformer Unet. *Comput. Med. Imaging Graph.* **2022**, *97*, 102055. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.