

## Article

# Research on Retinal Vessel Segmentation Algorithm Based on a Modified U-Shaped Network

Xialan He <sup>1</sup>, Ting Wang <sup>1,\*</sup> and Wankou Yang <sup>2</sup>

<sup>1</sup> College of Information Science and Technology & College of Artificial Intelligence, Nanjing Forestry University, Nanjing 210037, China; hexialan@njfu.edu.cn

<sup>2</sup> College of Automation, Southeast University, Nanjing 210096, China; wkyang@seu.edu.cn

\* Correspondence: wangting10402@njfu.edu.cn; Tel.: +86-138-1454-5620

**Abstract:** Due to the limitations of traditional retinal blood vessel segmentation algorithms in feature extraction, vessel breakage often occurs at the end. To address this issue, a retinal vessel segmentation algorithm based on a modified U-shaped network is proposed in this paper. This algorithm can extract multi-scale vascular features and perform segmentation in an end-to-end manner. First, in order to improve the low contrast of the original image, pre-processing methods are employed. Second, a multi-scale residual convolution module is employed to extract image features of different granularities, while residual learning improves feature utilization efficiency and reduces information loss. In addition, a selective kernel unit is incorporated into the skip connections to obtain multi-scale features with varying receptive field sizes achieved through soft attention. Subsequently, to further extract vascular features and improve processing speed, a residual attention module is constructed at the decoder stage. Finally, a weighted joint loss function is implemented to address the imbalance between positive and negative samples. The experimental results on the DRIVE, STARE, and CHASE\_DB1 datasets demonstrate that MU-Net exhibits better sensitivity and a higher Matthew's correlation coefficient (0.8197, 0.8051; STARE: 0.8264, 0.7987; CHASE\_DB1: 0.8313, 0.7960) compared to several state-of-the-art methods.

**Keywords:** deep learning; retinal vessel segmentation; multi-scale information; selective kernel; attention mechanisms



**Citation:** He, X.; Wang, T.; Yang, W. Research on Retinal Vessel Segmentation Algorithm Based on a Modified U-Shaped Network. *Appl. Sci.* **2024**, *14*, 465. <https://doi.org/10.3390/app14010465>

Academic Editor: Jan Egger

Received: 5 December 2023

Revised: 22 December 2023

Accepted: 30 December 2023

Published: 4 January 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

Retinal vessel segmentation has great significance in the early diagnosis of eye-related diseases [1]. The presence of vessel leakage, obstruction, and swelling in the retina of the eye indicates that it is highly probable to be a complication of diabetes resulting from elevated blood sugar levels. Retinal vessel segmentation plays an important role in clinical medicine, which can extract vascular features from retinal fundus images for analyzing and detecting diseases. However, compared with other medical images, retinal blood vessels have intricate structures and smaller areas of interest, making manual segmentation time-consuming and limited by experience [2]. Therefore, it is essential to develop automatic retinal vessel segmentation methods, which not only assist doctors in diagnosis but also enable large-scale analysis of fundus images.

Currently, retinal vessel segmentation methods can be categorized into two groups: unsupervised methods [3] and supervised methods [4]. The unsupervised method does not need manual annotation as a reference and only realizes the segmentation according to the characteristics of the retinal vessels themselves, including clustering-based methods, level sets, model-based algorithms, etc. Kande et al. [5] used a matching filter to enhance image contrast and spatially weighted fuzzy C-means clustering to obtain segmentation results. Mardani et al. [6] designed a combination segmentation model based on density spatial clustering and morphological reconstruction to improve the performance of vessel segmentation. Zhao et al. [7] constructed an active contour model to learn different types of

regional feature information. The aforementioned unsupervised segmentation methods are straightforward and have made some advancements in retinal vascular image segmentation. However, owing to the unique nature of medical images and the intricacy of fundus images themselves, the extracted vascular information obtained using unsupervised methods is relatively coarse, with limited resistance to noise and segmentation accuracy. Consequently, it fails to fulfill the actual requirements of clinical assistance. As a result, researchers are increasingly focusing on and utilizing supervised methods in their studies.

The core of the supervised method is to use prior knowledge to construct algorithms and discover hidden patterns in the data, so that the algorithm can automatically extract retinal vessel information and achieve vessel segmentation. The biggest difference between traditional machine learning and deep learning lies in the feature extraction process: the former requires manual construction of image features, while the latter can automatically extract and select features. Ricci et al. [8] built feature vectors based on the width of blood vessels and pixel grayscale values, and used support vector machines for classification. Marin et al. [9] constructed a 7-D vector that consisted of gray-level and moment invariants-based features to represent each pixel, and employed a neural network architecture to classify pixels in retinal fundus images. Although these traditional machine learning methods perform well in some cases, handmade features rely too heavily on prior knowledge and therefore fail in many complex datasets [10].

Semantic segmentation is a fundamental topic in computer vision, aiming to assign semantic labels to each pixel in an image [11,12]. With the improvement of computer hardware, especially the GPU technology, deep learning algorithms have shown tremendous advantages in the field of image segmentation. Many network structures designed for semantic segmentation tasks based on deep convolutional neural networks have demonstrated excellent performance. These structures possess rich representation capabilities. Fully convolutional networks (FCN) play a crucial role in the development of semantic segmentation. FCN utilizes an encoder–decoder structure and employs a fully convolutional integrated class network to extract features across the entire backbone. Another significant contribution of semantic segmentation is the incorporation of skip connections, which aggregate low-level features into high-level features, allowing for the recovery of reduced details. Building upon FCN and skip connections, U-Net [13] introduced a U-shaped encoder–decoder structure, which further enhanced and extended the FCN architecture. Another specific architecture, Deeplabv3+ [14], designed to solve image segmentation problems, has also demonstrated excellent performance. It utilizes a spatial pyramid pooling module and an encoder–decoder structure to achieve semantic segmentation. The spatial pyramid pooling module captures contextual information by pooling features of different resolutions, while the encoder–decoder structure aids in obtaining clear object boundaries. However, Deeplabv3+ usually has higher computational complexity and a larger number of parameters than U-Net.

Inspired by the symmetrical structure and effective skip connections of U-Net, many variant methods based on U-shaped structures have been applied to medical image segmentation. Wu et al. [15] realized that the accuracy of segmentation can be improved by embedding the inception-residual block into the U-shaped symmetric encoder–decoder architecture. Dense dilated convolutions [16] and deformable convolutions [17,18] have also been introduced into the U-Net architecture to capture multi-scale local context information, which can improve the effectiveness of retinal vessel segmentation. Dilated convolution layers [19] can be applied to expand the receptive field and aggregate multi-scale features from different kernels, while selective kernels (SK) [20] evaluate information combinations from multiple kernels and select effective spatial scales. Phan et al. [21] placed selective kernels at skip connections to locate receptive fields related to lesion size, which can improve the utilization of multi-scale features in dermoscopy images. In recent years, attention mechanisms have gained significant note in various tasks, including attention gate networks [22], SENet [23,24], spatial attention mechanisms [25], etc. These methods have been introduced to efficiently capture important objects, locations, and channel information,

thereby enhancing the representation ability. Li et al. [26] developed an attention module based on the U-Net architecture to incorporate global information during the feature fusion process. Wang et al. [27] proposed Att U-Net (Attention U-Net), which utilizes an attention-gating mechanism at skip connections to minimize the impact of background pixels on segmentation. To enhance the network's segmentation ability for retinal blood vessels, Fu et al. [28] adopted DANet (dual attention networks), which leverages both spatial attention and channel attention to capture contextual features. The integration of the convolutional neural network and the transformer model also has shown promising results in image segmentation. Chen et al. [29] constructed a joint CNN-transformer structure as the encoder, utilizing the U-Net architecture. They also incorporated a cascade upsampling operation in the decoder to improve the accuracy of position information. In the context of medical images, Li et al. [30] designed the group transformer module and constructed the GT U-Net network. In addition, Liang et al. [31] proposed an adaptive feature fusion cascade transformer retinal segmentation algorithm to address the issue of incomplete segmentation of small blood vessels.

Inspired by the above, a modified U-shaped network called MU-Net, which focuses on extracting multi-scale features is proposed in this study. The encoding stage of MU-Net includes a multi-scale residual convolution module (MRCM), which is designed to incorporate hidden feature information from different scales. MRCM performs average pooling operations on the horizontal and vertical to generate attention maps, thereby achieving long-range context information aggregation. By integrating spatial information from two dimensions and combining local features with global contextual information, MRCM efficiently models the correlation between different pixels. A selective kernel unit (SKU) is utilized at the skip connections to capture multi-scale blood vessels, resulting in a feature map with adaptive receptive field size. This approach allows for the adaptive aggregation of multi-scale features within a single layer, eliminating the need for stacking multiple layers. SKU facilitates networks to control multi-scale information flow by learning global contextual information. In response to the gradient vanishing phenomenon caused by the increase in the number of parameters and the training difficulty of encoder–decoder merging branches in the network, a residual attention module (RAM) was designed in the decoding stage to efficiently suppress noise and redundant information.

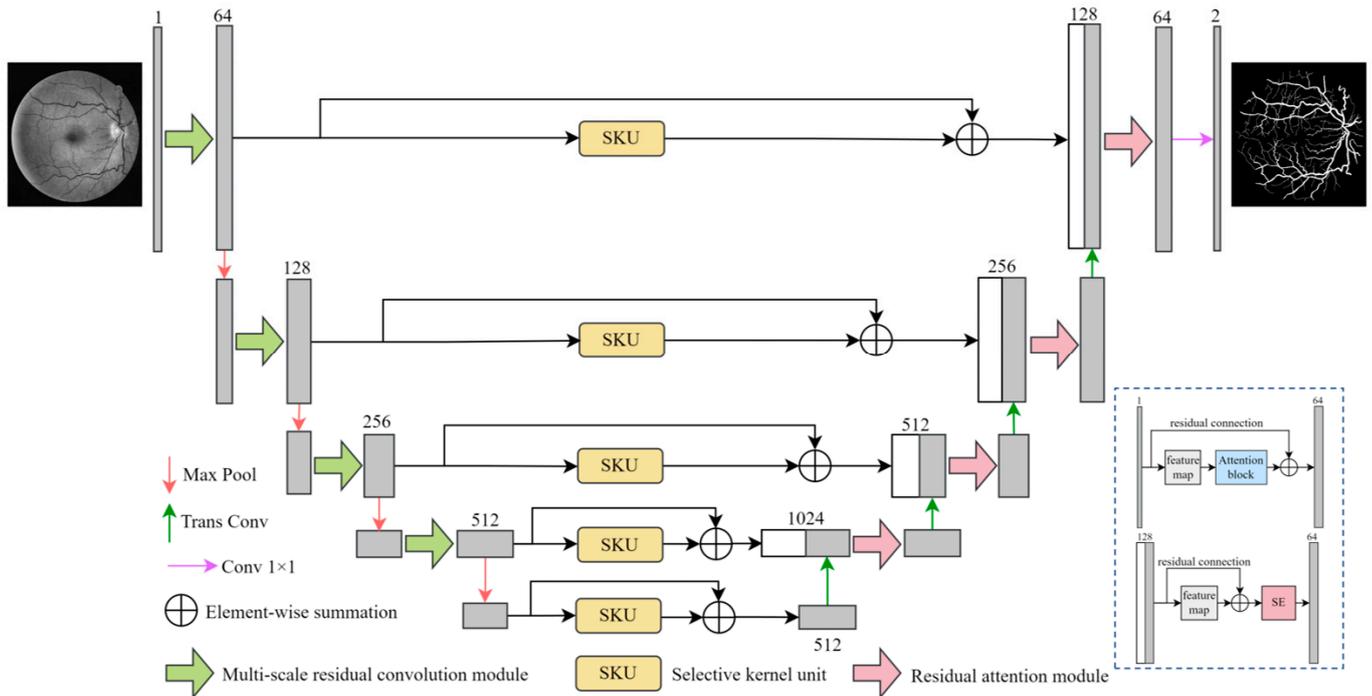
## 2. Methods

This section details the MU-Net model for automatically segmenting vessel structures of retinal fundus images. The overall framework of the method is first presented. Then, the design of each module and the loss function of this network are explained in detail.

### 2.1. Network Architecture

The architecture adopts a symmetric four-order downsampling and upsampling network structure, as shown in Figure 1. The encoder utilizes MRCM for multi-scale feature extraction, along with a downsampling max-pooling operation using a pooling size of  $2 \times 2$  and stride = 2. With each downsampling step, the number of feature channels is doubled. For the first-level MRCM, a convolution which is of quantity 64 with a kernel size of  $3 \times 3$  (stride = 1, padding = 1) is performed. Then, three sets of convolution operations are applied to the feature map ( $C \times H \times W$ ) at different scales:  $1 \times 1$  (stride = 1, padding = 0),  $3 \times 3$  (stride = 1, padding = 1), and  $5 \times 5$  (stride = 1, padding = 2). Additionally, a  $1 \times 1$  convolution operation is employed to adjust the number of channels. In the MRCM, the quantity of filters in each level is equal to the number of feature channels in that level. All other parameters remain consistent with the first level. In the improved decoder section, a channel attention mechanism is added to the concatenated feature maps, while maintaining the original U-Net decoder's structural framework. The decoder consists of a RAM and a  $2 \times 2$  transposed convolution upsampling, which reduces feature channels by half and gradually reconstructs a segmented image with the same spatial size as the input. Instead of using simple bilinear interpolation, trainable transposed convolutions are used

for upsampling, enabling adaptive recovery of detailed feature information. In Figure 1, the embedding of the SKU in the skip connections of the U-shaped network is depicted using yellow boxes, while a residual structure is employed in this framework. Then, the concatenated feature maps are processed by RAM. In the last layer of the MU-Net, a  $1 \times 1$  convolution operation and sigmoid activation function are utilized to achieve the final segmentation graph. In Figure 1, graphical representations of adding attention mechanisms and residual connections to the encoder–decoder section of the U-Net framework are displayed in the bottom right corner. The implementation process is explained in detail in Sections 2.2 and 2.4.



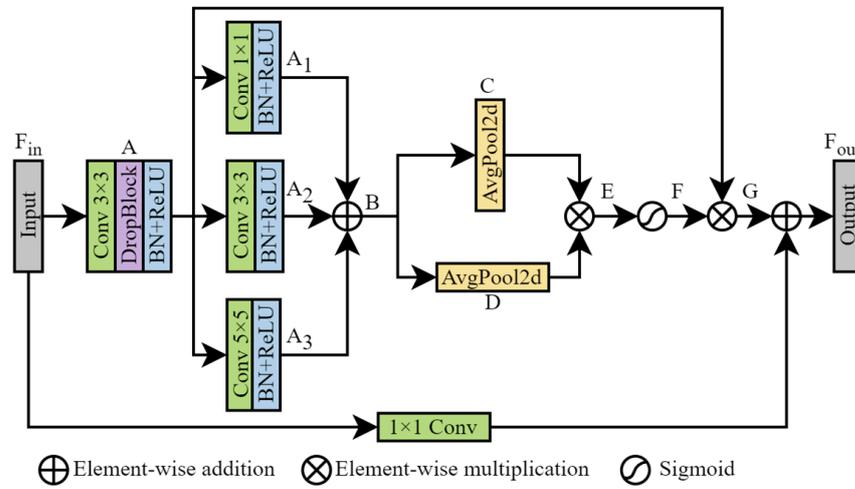
**Figure 1.** Architecture of MU-Net.

## 2.2. Multi-Scale Residual Convolution Module

The architecture of inception networks enhances the utilization of computing resources by increasing both their depth and width, while maintaining the same computing budget. It achieves this by incorporating filters of different sizes on the same layer to process feature information from multiple scales. The features are then aggregated in the next layer to extract fused features from multiple scales in the subsequent inception module [32]. Additionally, the use of residual connections has been proven to be advantageous in training deep learning models [33]. Taking inspiration from the inception module and residual connections, a multi-scale residual convolution module was designed in the encoder section, as depicted in Figure 2. This module consists of convolutional layers, batch normalization (BN), DropBlock, corrected linear units (ReLU), and Softmax layers.

To extract blood vascular feature information comprehensively, three types of convolution blocks with different scales are utilized to achieve multi-scale extraction of retinal blood vascular features. Different convolution kernels have varying receptive fields. Convolution kernels with smaller receptive fields are designed to capture small targets and local detailed information. On the other hand, convolution kernels with larger receptive fields can provide larger targets and global semantic information. Inspired by the successful application of DropBlock in recent computer vision research [34,35], this study employed DropBlock as a regularization technique in the network at MRCM. DropBlock is a structured version of DropOut, with the key distinction that it removes contiguous regions in a layer's feature map instead of randomly discarding independent units. By eliminating

features in continuous regions, DropBlock forces the network to rely on evidence from other regions that are relevant to the data.



**Figure 2.** Multi-scale residual convolution module.

Input  $F_{in}^l \in R^{B \times C' \times H \times W}$  is first fed into a  $3 \times 3$  convolution layer for feature extraction, resulting in a new feature map  $A^l \in R^{B \times C \times H \times W}$ . Then, feature map  $A^l$  is passed through three parallel convolution layers with kernels of  $1 \times 1$ ,  $3 \times 3$ , and  $5 \times 5$ , generating three new feature maps  $A_1^l$ ,  $A_2^l$ , and  $A_3^l$ . To preserve multi-scale feature information, feature maps  $B^l \in R^{B \times C \times H \times W}$  are obtained by element-wise addition, which is calculated as follows:

$$B^l = f_1(A^l) \oplus f_2(A^l) \oplus f_3(A^l) \tag{1}$$

where  $f_1$ ,  $f_2$ , and  $f_3$  represents the  $1 \times 1$ ,  $3 \times 3$ , and  $5 \times 5$  convolution operations, respectively, and  $\oplus$  denotes the element-wise addition.

In order to capture the entire scene relationship through the attention of a spatial attention mechanism, adaptive average pooling operations are performed to  $B^l$  on the H and W dimensions respectively. The resulting outputs  $C^l \in R^{B \times C \times 1 \times W}$  and  $D^l \in R^{B \times C \times H \times 1}$  are fused adopting matrix multiplication to generate a spatial attention feature map  $E$  using sigmoid. The feature map  $F^l$  is calculated as follows:

$$F^l = \sigma(\text{pool}_c(B^l) \otimes \text{pool}_r(B^l)) \tag{2}$$

where  $\sigma$  is a sigmoid activation,  $\text{pool}_c$  and  $\text{pool}_r$  present the average pooling of columns and rows, respectively, and  $\otimes$  denotes the element-wise multiplication.

Next, the elements of  $A^l$  and  $F^l$  are multiplied to obtain the spatial weight map  $G^l \in R^{B \times C \times H \times W}$  for each pixel. This operation can selectively aggregate contexts based on the attention map, and contain abundant contextual information. Then, the spatial weight map  $G$  is added to the input image  $F_{in}^l$  through a residual connection. Finally, the output of MRCM is as follows:

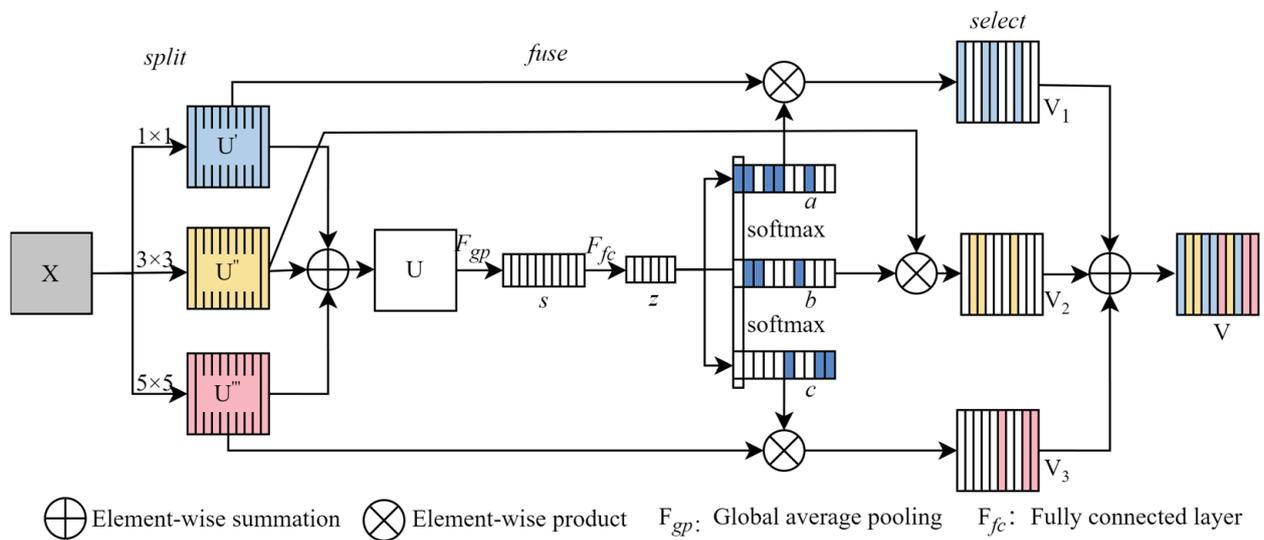
$$F_{out}^l = F_{in}^l + G^l \tag{3}$$

The long-range contextual information captured by MRCM in both H and W dimensions is beneficial for semantic segmentation. The convolutional module in the traditional U-Net is replaced by MRCM, which enhances the information on the tiny features of blood vessels and then efficiently extracts the small blood vascular features.

### 2.3. Selective Kernel Unit

U-Net utilizes ordinary skip connections, but often fails to effectively utilize the multi-scale information output from the encoder. Retinal fundus images commonly exhibit blood vessels with varying sizes and shapes. The receptive field sizes of feature representations

should vary in order to capture multi-scale spatial semantic information efficiently [14,36]. However, standard convolutions generate feature representations with fixed receptive fields and kernel parameters when sliding over feature maps. This can result in intra-class inconsistency as the feature representation of pixels belonging to the same category may differ in different areas [37]. For instance, in the task of retinal blood vessel segmentation, blood vessels may have varying widths, colors, and textures in different areas, or the same blood vessel may exhibit bends or bifurcations. If the receptive field of a standard convolution is not large enough to cover the entire blood vessel, it may only capture part of its features, leading to inaccurate segmentation results. In a study conducted by Wang et al. [38], three distinct cross-sectional strength distribution maps of blood vessels were defined. Building on this research and drawing inspiration from SK, we introduced SKU, with a detailed structure shown in Figure 3. The SKU utilizes three kernels of different sizes to generate multi-scale information. Subsequently, a gated softmax operation is employed to fuse the information obtained from the multi-scale convolution kernels.



**Figure 3.** Selective kernel unit.

In the split part, the input feature map  $X \in R^{B \times C \times H \times W}$  undergoes three transformations  $F_1 : X \rightarrow U' \in R^{B \times C \times H \times W}$ ,  $F_2 : X \rightarrow U'' \in R^{B \times C \times H \times W}$  and  $F_3 : X \rightarrow U''' \in R^{B \times C \times H \times W}$  with kernel sizes of  $1 \times 1$ ,  $3 \times 3$ , and  $5 \times 5$ . Note that  $F_1$ ,  $F_2$ , and  $F_3$  are composed of convolutions, BN, and ReLU functions in sequence.

In the fuse part, a gating mechanism is employed to selectively filter the output of the previous layer, allowing each branch to carry a distinct flow of information into the next neuron. First, the outputs of different branches are fused by element-wise addition:

$$U = U' \oplus U'' \oplus U''' \tag{4}$$

Then,  $U$  is subjected to global average pooling operations, which aim to obtain channel-wise feature information as  $s \in R^C$ . Here,  $C$  is the number of feature channels. The  $s$  is calculated by reducing  $U$  across the spatial dimensions  $H \times W$ :

$$s = F_{gp}(U) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U(i, j) \tag{5}$$

where  $H$  and  $W$  represent the height and width directions, respectively. In addition, a fully connected layer is applied to reduce the dimensionality and obtain a compact feature

$z \in R^{t \times 1}$ , which is utilized to guide precise and adaptive selection. The calculation of  $z$  is as follows:

$$z = F_{fc}(s) = \delta(BN(M \times s)) \tag{6}$$

where  $\delta$  is the ReLU function,  $BN$  denotes the batch normalization,  $M \in R^{t \times C}$  denotes the learnable parameters in the fully connected layer. To study the impact of  $t$  on the efficiency of the model, a reduction ratio  $r$  is used to control its value:

$$t = \max(C/r, L) \tag{7}$$

where  $L$  denotes the minimal value of  $t$  ( $L = 32$  is a typical setting in this experiments) [20].

In the select part, soft attention is applied between channels to select information of different sizes. This selection process is guided by the compact feature information  $z$ , and softmax operation is applied channel-wise:

$$a_i = \frac{e^{A_i z}}{e^{A_i z} + e^{B_i z} + e^{C_i z}}, b_i = \frac{e^{B_i z}}{e^{A_i z} + e^{B_i z} + e^{C_i z}}, c_i = \frac{e^{C_i z}}{e^{A_i z} + e^{B_i z} + e^{C_i z}} \tag{8}$$

where  $A, B, C \in R^{C \times t}$ , and  $a, b, c$  denote the soft attention vector for  $U', U''$  and  $U'''$ , respectively. Note that  $a_i$  is the  $i^{th}$  element of  $a$ , and  $A_i \in R^{1 \times t}$  is the  $i^{th}$  row of  $A$ , so as  $b_i, B_i, c_i$  and  $C_i$ . Finally, the attention weights are applied to various kernels to obtain the feature map  $V$ , where  $V = [V_1, V_2, \dots, V_C], V_i \in R^{B \times C \times H \times W}$ .

$$V_i = a_i \times U'_i + b_i \times U''_i + c_i \times U'''_i, a_i + b_i + c_i = 1 \tag{9}$$

#### 2.4. Residual Attention Module

After concatenating the output of the encoder with the feature map generated by SKU and the upsampled feature map, the resulting information contains valuable key information across channels. This key information can greatly enhance the model’s ability to identify lesion areas [39]. To effectively utilize this key information and address the challenges arising from the increased parameter count and training difficulties when merging multiple encoder–decoder branches in the network, this study introduces a RAM for decoding. The structure of RAM is composed of two  $3 \times 3$  convolutions, BN, ReLU,  $1 \times 1$  convolutional, and SE composition, as depicted in Figure 4. The  $1 \times 1$  convolution is responsible for adjusting the channel dimensions. Squeezing operators compress global spatial information into a channel descriptor, while the excitation operator maps the input descriptor to a set of channel weights. As mentioned in the previous section, extracting microvascular features is highly challenging. The advanced features of each channel graph can be considered as responses to specific categories, such as thin blood vessels, thick blood vessels, and other noise. Introducing the SE mechanism in the latter half of the RAM enables dynamic modeling of the interdependence between feature channels. By increasing the weight of feature channels in the vessel category and suppressing the weight of feature channels in the others category, small vascular features can be more easily extracted.

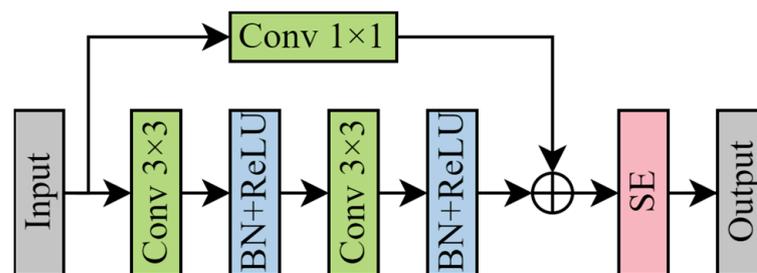


Figure 4. Residual attention module.

### 2.5. Weighted Joint Loss Function

Considering the imbalanced distribution of background and vessel pixels in retinal fundus images, a weighted joint loss function composed of weighted binary cross-entropy loss and dice loss is used to train the network and it is defined as follows:

$$L = L_{WBCE} + \alpha L_{Dice} \quad (10)$$

$$L_{WBCE} = -\frac{1}{N} \sum_{i=1}^N (\beta y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \quad (11)$$

$$L_{Dice} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (12)$$

To address the imbalanced problem, a combination of the weighted binary cross-entropy loss for pixel-wise classification and the dice loss for crossing regions is employed. In Equation (10), the hyperparameter  $\alpha$  is introduced to balance these two losses and set it to 0.9 in the experiment. Equation (11) involves the number of pixels  $N$  in each training image and the ground truth  $y_i$  and predicted probability  $p_i$  of pixel  $i$ , where  $\beta$  is a balance factor used to measure the false positive rate and false negative rate. When  $\beta < 1$ , the false positive rate decreases while the false negative rate increases. Contrarily, when  $\beta > 1$ , the false negative rate decreases while the false positive rate increases. In the context of retinal vascular segmentation, false positives may result in overtreatment, while false negatives may lead to missed diagnosis or inadequate treatment, increasing the risk of disease deterioration and visual loss. Therefore, it is crucial to minimize the false negative rate. In this study  $\beta = 2$ . In Equation (12),  $X$  represents the magnitude of the segmentation result, and  $Y$  represents the magnitude of the ground truth.  $|X \cap Y|$  represents the common elements between  $X$  and  $Y$ . This formula is a loss function widely employed for image segmentation tasks.

## 3. Experiment and Results

In this section, all experiments were performed on the DRIVE, START, and CHASE datasets. The pre-processing methods are first presented, and various evaluation metrics are adopted to quantify segmentation results. Then the experimental setup and network training process are described. Next, the results of the ablation experiments are given to verify the effectiveness of the proposed method. In the end, this study is compared with other neural network methods mentioned in the literature.

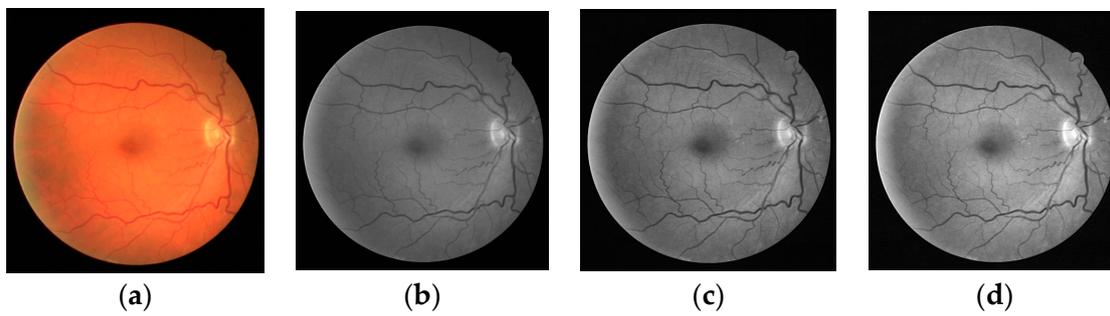
### 3.1. Datasets and Pre-Processing

Three publicly available retinal fundus image datasets: DRIVE, STARE, and CHASE\_DB1 were applied to evaluate our method. The DRIVE dataset contains 40 color fundus images with a resolution of  $565 \times 584$  pixels from the Netherlands diabetes retinopathy screening project [40], where 20 images are performed for training and 20 images for testing. To assist ophthalmologists in diagnosing eye diseases, 20 images with a resolution of  $700 \times 605$  pixels are selected from the STARE dataset [41]. As the official does not provide a clear division of training and testing sets, the leave-one-out method is applied for cross-validation. Specifically, each time 19 images are selected for training, and the remaining one is for testing. The above operation is repeated and finally the average of the 20 trained model results is calculated. The CHASE\_DB1 dataset contains 28 images with a resolution of  $999 \times 960$  pixels, corresponding to two images per patient of 14 children in the UK Children's Hearing and Health Research Project [42], where the first 20 images are employed to form the training set and the last eight images to form the test set.

Due to the optical effect during photography, fundus images often exhibit low contrast between blood vessels and the background, making it challenging to distinguish between them effectively. To address this issue, three preprocessing methods are utilized in this study.

- In this study, the fundus images are separated into RGB three-channel feature maps. It was found that the green channel exhibited moderate brightness and better contrast between blood vessels and the background, in comparison to the other channels [43,44]. Consequently, the fundus vascular images with green channels were selected as the post-processing images.
- The contrast-limited adaptive histogram equalization (CLAHE) improves the contrast between blood vessels and the background while minimizing noise amplification, resulting in smoother processed fundus images [45].
- The gamma transformation is a nonlinear method commonly employed for image enhancement [46]. It can enhance the brightness of the darker areas in an image, thereby improving its visibility. In this study, the gamma value is set to 1.2.

Figure 5 illustrates the original retinal image and its corresponding preprocessed image. The enhanced image exhibits richer details and higher contrast compared to the original image, thereby facilitating the subsequent segmentation steps.



**Figure 5.** Examples of the image pre-processing: (a) original image, (b) green channel image, (c) CLAHE, and (d) gamma correction.

To expand the limited number of training samples, two data augmentation operations are employed, including random horizontal and vertical flipping with a 50% probability, and random rotation with a 40% probability (angle ranges from  $-10$  to  $10$  degrees). These operations aim to improve the generalization ability of the model.

### 3.2. Evaluation Metrics

The segmentation performance of the MU-Net model is evaluated from four metrics: accuracy (*Acc*), sensitivity (*Sen*), specificity (*Spe*), Matthew's correlation coefficient (*Mcc*) [47], and area under the receiver operating characteristic curve (*AUC*). The calculation formulas for the first four evaluation metrics are as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$Sen = \frac{TP}{TP + FN} \quad (14)$$

$$Spe = \frac{TN}{TN + FP} \quad (15)$$

$$Mcc = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (16)$$

where *TP* is true positive, which means the number of correctly segmented vessel pixels. *TN* is true negative, which means the number of correctly segmented background pixels. *FP* is false positive, which means the number of pixels that recognize the background as blood vessels. *FN* is false negative, which means the number of pixels that recognize the blood vessels as background. Accuracy represents the percentage of correctly segmented pixels among all pixels. Sensitivity is exploited to measure the ability to correctly detect

vessel pixels. Specificity means the ability to recognize non-vessel pixels. The model's performance is evaluated using the *Mcc* to prevent data imbalance [48]. A perfect prediction is achieved when the *Mcc* value reaches 1. The receiver operating characteristic curve (ROC) takes the false positive rate as the horizontal axis and the true positive rate as the vertical axis. The closer the *AUC* value is to 1, the better the segmentation effect.

### 3.3. Implementation Details

The model is trained based on the training and testing sets described in Section 3.1. The experimental conditions for the training are provided in Table 1.

**Table 1.** Experimental conditions.

Experimental Environment	Details
Software	Pycharm
Programing language	Python 3.8
Operating system	Ubuntu 20.04
Deep learning framework	Pytorch 1.10.0
GPU	NVIDIA 3090

The hyperparameters are not determined by the network but are adjusted during the complete training stage, considering accuracy and necessary modifications. To enhance the network's performance, these settings need to be changed and optimized based on the training of retinal fundus images. The model's training input utilizes the original image's resolution size. In this study, the weights of each convolutional layer are initialized using the Kaiming method, and the network parameters are updated using the Adam optimizer. The initial learning rate is set to  $1 \times 10^{-4}$ . The study employs a poly-learning strategy with a power setting of 0.9, resulting in a linear decrease in the learning rate from  $10^{-4}$  to  $10^{-6}$  within 200 epochs. The momentum coefficients  $\beta_1$  and  $\beta_2$  are set to 0.9 and 0.999, respectively. The batch size is set to 2 and when the epoch reaches 200, the experimental results tend to converge. Taking into account the experimental period, this article selected 400 epochs. For the DropBlock setting, the size of the dropped block is set to 7, and the drop rate for all datasets is set to 0.15. For the hyperparameters  $\alpha$  of the weighted joint loss function, we aimed to achieve a maximum balance between the weighted binary cross entropy loss and dice loss through multiple experiments. Ultimately,  $\alpha$  is set to 0.9. The training parameters of the retinal vessel segmentation model are provided in Table 2.

**Table 2.** Training parameter settings.

Training Parameters	Details
Epochs	400
Batch size	2
Image size(pixels)	$565 \times 584/700 \times 605/999 \times 960$
Initial learning rate	0.0001
Optimization algorithm	Adam ( $\beta_1 = 0.9, \beta_2 = 0.999$ )

### 3.4. Ablation Experiment

To evaluate the effectiveness of the proposed modules, ablation experiments were conducted on three datasets, and the results are shown in Table 3. Based on the U-Net model, the parameters of the ablation experiment were set the same as in the method mentioned in Section 3.3.

**Table 3.** Results of ablation experiments on three datasets.

Methods	DRIVE				STARE				CHASE_DB1			
	Acc	Sen	Spe	AUC	Acc	Sen	Spe	AUC	Acc	Sen	Spe	AUC
U-Net	0.9546	0.7554	0.9836	0.9766	0.9622	0.7469	<b>0.9876</b>	0.9810	0.9589	0.7405	0.9810	0.9761
U-Net + MRCM	0.9683	0.7977	<b>0.9856</b>	0.9780	0.9637	0.7946	0.9829	0.9768	0.9676	0.8077	0.9828	0.9782
U-Net + SKU	0.9695	0.8042	0.9816	0.9828	0.9685	0.8104	0.9835	<b>0.9831</b>	0.9652	0.8142	0.9827	0.9796
U-Net + RAM	0.9676	0.7856	0.9842	0.9806	<b>0.9744</b>	0.7912	0.9866	0.9812	0.9647	0.8062	<b>0.9854</b>	0.9794
MU-Net	<b>0.9690</b>	<b>0.8197</b>	0.9833	<b>0.9853</b>	0.9693	<b>0.8264</b>	0.9821	0.9803	<b>0.9752</b>	<b>0.8313</b>	0.9849	<b>0.9860</b>

Note: Bold indicates the optimal value.

To further evaluate the performance of the proposed algorithm in retinal vessel segmentation, it was quantitatively compared with some representative algorithms. Table 4 shows the comparison results of our method and other methods.

**Table 4.** Performance scores of the different algorithms on the three datasets.

Methods	Year	DRIVE				STARE					
		Acc	Sen	Spe	AUC	Mcc	Acc	Sen	Spe	AUC	Mcc
U-Net [13]	2015	0.9546	0.7554	0.9836	0.9766	0.7857	0.9622	0.7469	0.9876	0.9810	0.7889
Zhou et al. [49]	2017	0.9469	0.8078	0.9674	-	0.7656	0.9585	0.8065	0.9761	-	0.7830
Deeplabv3+ [14]	2018	0.9526	0.7411	0.9694	0.9702	0.6729	0.9562	0.7402	0.9706	0.9764	0.6580
R2U-Net [50]	2018	0.9553	0.7735	0.9818	0.9784	-	0.9632	0.7944	0.9832	0.9819	-
HCF [51]	2019	0.9753	0.8176	0.9709	-	0.7659	<b>0.9751</b>	0.8239	0.9749	-	0.7818
SA-UNet [52]	2021	0.9583	0.7962	0.9781	0.9643	-	0.9642	0.8212	0.9722	0.9769	-
EEA Unet [53]	2021	0.9577	0.7918	0.9708	-	0.7115	0.9445	0.8021	0.9561	-	0.7115
Att U-Net [27]	2022	0.9531	0.7635	<b>0.9840</b>	0.9791	0.7963	0.9689	0.7695	0.9887	0.9815	<b>0.8021</b>
Zhang et al. [54]	2022	<b>0.9701</b>	0.7719	0.9799	0.8895	0.7399	0.9691	0.6912	<b>0.9911</b>	0.8391	0.7327
Liang et al. [1]	2023	0.9568	0.8054	0.9789	0.9807	-	0.9648	<b>0.8397</b>	0.9795	<b>0.9850</b>	-
MU-Net	2023	0.9690	<b>0.8197</b>	0.9833	<b>0.9853</b>	<b>0.8051</b>	0.9693	0.8264	0.9821	0.9803	0.7987

Methods	Year	CHASE_DB1				
		Acc	Sen	Spe	AUC	Mcc
U-Net [13]	2015	0.9589	0.7405	0.9810	0.9761	0.7416
Zhou et al. [49]	2017	0.9520	0.7553	0.9751	-	0.7398
Deeplabv3+ [14]	2018	0.9516	0.7319	0.9725	0.9662	0.6978
R2U-Net [50]	2018	0.9624	0.7405	0.9848	0.9813	-
HCF [51]	2019	0.9518	0.7559	0.9758	-	0.7379
SA-UNet [52]	2021	0.9672	0.8249	0.9822	0.9779	-
EEA Unet [53]	2021	0.9340	0.6457	0.9653	-	0.6508
Att U-Net [27]	2022	0.9604	0.7821	0.9854	0.9823	0.7900
Zhang et al. [54]	2022	0.9811	<b>0.8506</b>	<b>0.9981</b>	0.9142	0.7587
Liang et al. [1]	2023	0.9635	0.8240	0.9775	0.9836	-
MU-Net	2023	<b>0.9752</b>	0.8313	0.9849	<b>0.9860</b>	<b>0.7960</b>

Note: Bold indicates the optimal value.

### 3.5. Comparisons

#### 3.5.1. Quantitative Analysis

Based on the experimental data in Table 3, the following observations can be made:

- (1) When comparing U-Net on the DRIVE dataset with the introduction of MRCM and SKU, there are improvements in *Acc*, *Sen*, *Spe*, and *AUC* by 0.0137/0.0149, 0.0423/0.0488, 0.0020/−0.002, and 0.0014/0.0062 respectively. Similar performance improvements were observed in other datasets. These results indicate that the strategy of utilizing multi-scale residual modules and the selective kernel unit efficiently learns and locates multi-scale features of blood vessels.
- (2) The introduction of the residual attention module resulted in increased *Acc*, *Sen*, *Spe*, and *AUC* on the STARE dataset by 0.0122, 0.0443, −0.0010, and 0.0002 respectively. These findings suggest that the residual attention module is particularly effective in capturing microvessels.

- (3) Our MU-Net showed improvements in *Acc*, *Sen*, *Spe*, and *AUC* of 0.0163, 0.0908, 0.0039, and 0.0099 respectively when compared to the U-Net baseline on the CHASE\_DB1 dataset.

In the quantitative analysis, U-Net demonstrated good accuracy in retinal blood vessel segmentation. However, it had low sensitivity, indicating limitations in accurately extracting blood vessels. This is mainly attributed to the weak feature extraction ability of traditional convolution and U-Net susceptibility to image contrast and noise, which affects its precise blood vessel extraction, leading to vessel rupture and inadequate segmentation. To address these issues, we proposed a solution that enhances the extraction of multi-scale feature information, locates the receptive field matching the blood vessel scale at the skip connection, and utilizes channel attention to extract key information between channels. This improved U-Net encoder–decoder structure resulted in better prediction results compared to the U-Net baseline.

In addition, this study compared the performance of MU-Net with other CNN methods, including SA-UNet, EEA-UNet, and several other techniques, in retinal vessel segmentation. On the DRIVE dataset, MU-Net achieved the highest *Sen*, *AUC*, and *Mcc* values of 0.8197, 0.9853, and 0.8051, respectively. Similarly, on the CHASE\_DB1 dataset, MU-Net obtained the highest *Acc*, *AUC*, and *Mcc* values of 0.9752, 0.9860, and 0.7960, respectively. Although the STARE dataset did not yield the highest performance metric, MU-Net still demonstrated competitive results with the second highest *Acc*, *Sen*, and *Mcc* values compared to the best performing methods (HCF, Liang, and EEA-UNet, respectively). Overall, MU-Net outperforms these CNN methods in terms of quantitative performance.

Our proposed method also has some limitations. As shown in Tables 3 and 4, the calculated values of *Spe* for all methods, including MU-Net, are not dominant. In supervised training, finding the right balance between sensitivity and specificity can be challenging. Increasing the specificity of segmentation may lead to the identification of more non-vascular pixels as vascular pixels, resulting in reduced sensitivity. Therefore, we prioritize higher sensitivity by sacrificing a small portion of specificity. Our method achieved better sensitivity results compared to the method with the highest *Spe* value.

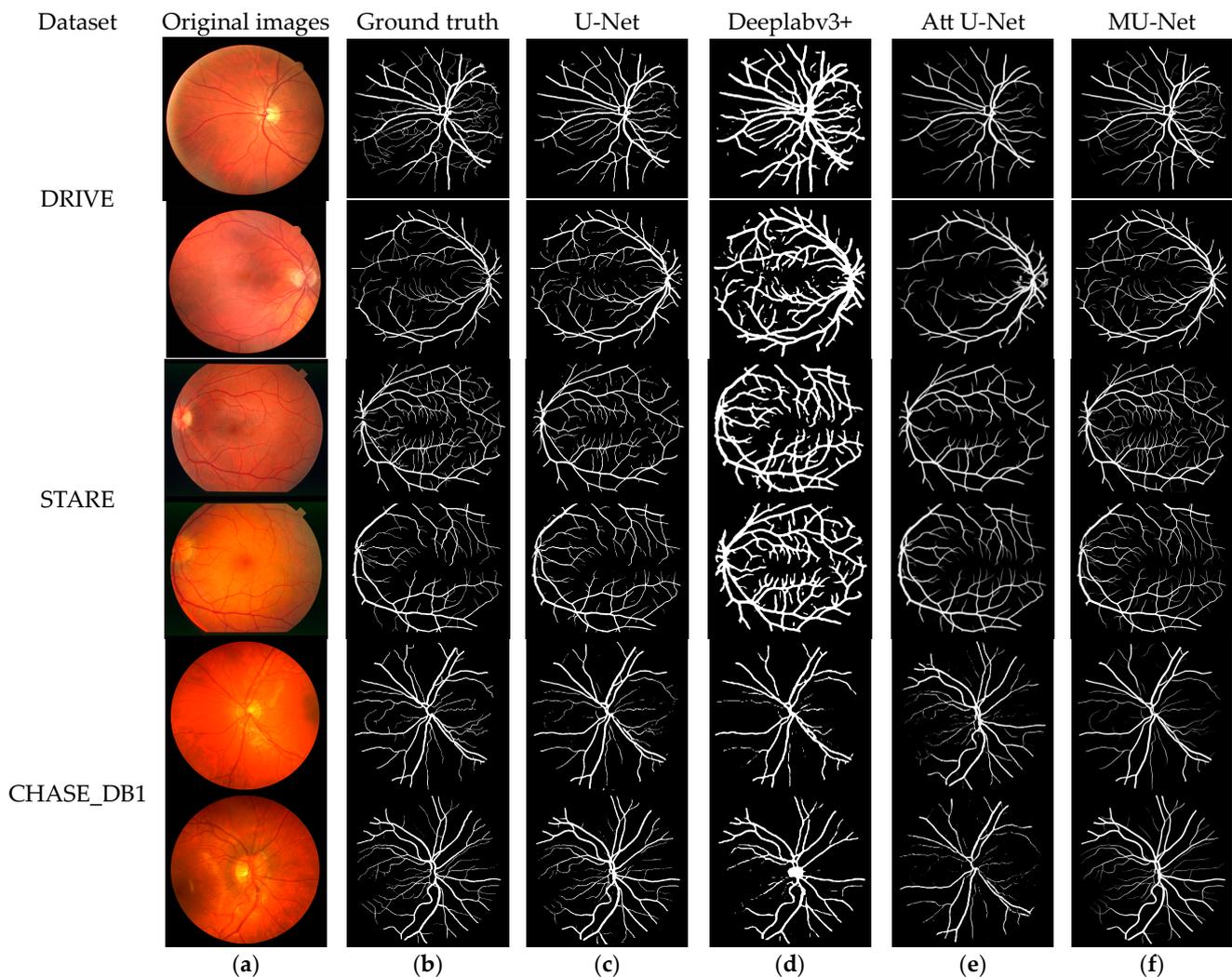
### 3.5.2. Qualitative Analysis

Figure 6 displays the segmentation results of U-Net, Deeplabv3+, Att U-Net, and our proposed MU-Net. Each dataset is represented by two rows of segmentation result images. In the DRIVE dataset, the first row consists of healthy fundus images, while the second row contains retinopathy images. To ensure consistency, both the U-Net and Att U-Net models were trained using the same parameter set and experimental environment.

To observe further the segmentation results of tiny blood vessels, we opted to enlarge and focus on specific details within the overall image which is shown in Figure 7. Local areas of interest are outlined using green and red rectangles.

Figures 6 and 7 show that compared with our proposed MU-Net, Deeplabv3+, and Att U-Net can only segment relatively large blood vessels, while tiny blood vessels are often lost at the ends or intersections of the blood vessel tree. This is evident in the bottom row of Figure 6. The Deeplabv3+ algorithm, for instance, exhibits vessel thickening, merging, or inadequate accuracy. As can be clearly seen from Figure 6, the segmentation results of Deeplabv3+ tend to contain more content, leading to a significant thickening of blood vessels. The possible reason for this is that Deeplabv3+ employs convolution with different dilation rates instead of downsampling operations. While this increases the receptive field of the network, dilation convolution involves discrete sampling on the feature map. Discrete sampling is more effective in obtaining feature information on large targets. However, for small targets, setting a high inflation rate may cause the learned features to lack correlation due to the large interval. This can result in incorrect feature information and affect the prediction of small objects, ultimately leading to unclear segmentation boundaries and judgment errors. Although the Att U-Net algorithm successfully segments microvessel ends in the STARE dataset, it shows insufficient segmentation and noticeable vessel rupture in the Drive and CHASE\_DB1 datasets, respectively. Analyzing retinal blood vessels,

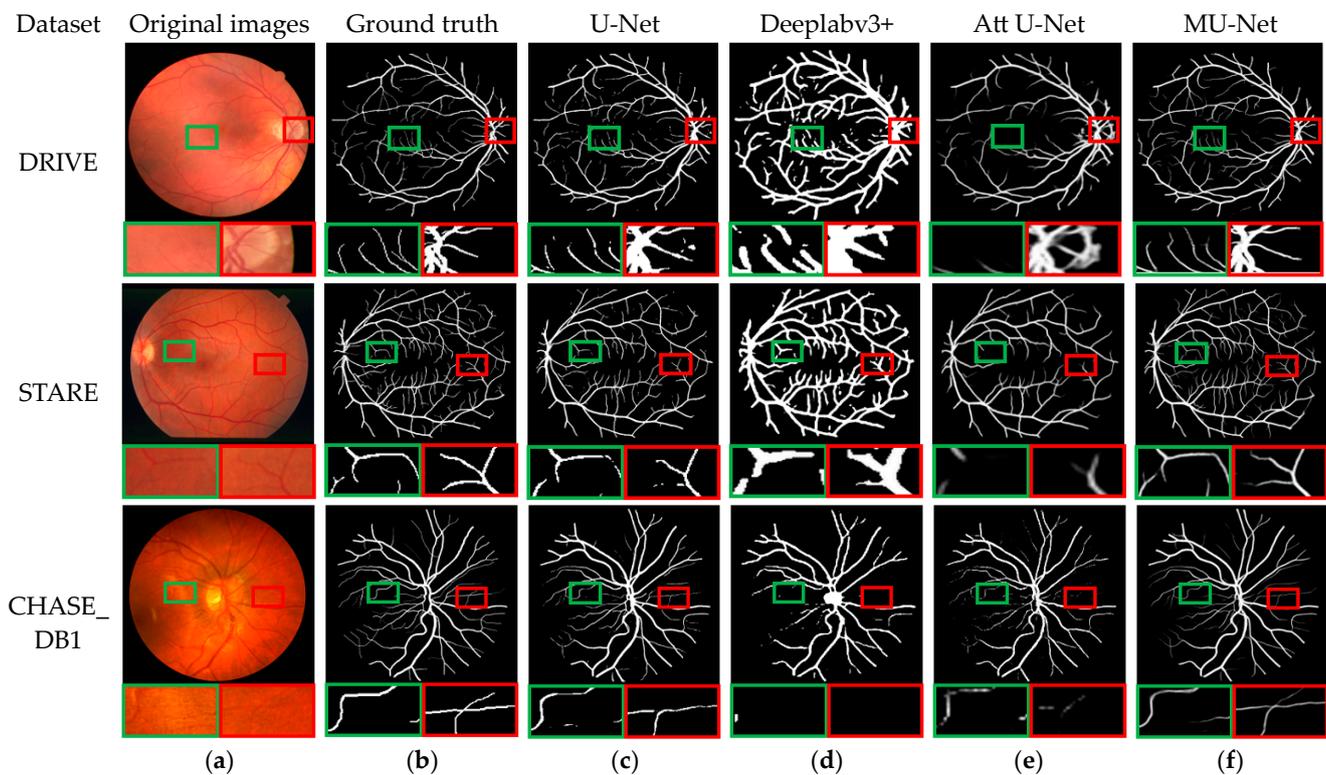
particularly microvessels, is crucial for diagnosing, monitoring, and planning treatment for eye diseases like macular degeneration, diabetic retinopathy, and retinal vein occlusion. These diseases can cause retinal ischemia due to vascular occlusion, leading to retinal neovascularization. In the advanced stages of these diseases, abnormal growth of new, small blood vessels can lead to vision impairment or even blindness. Therefore, the loss of small vessels in vessel segmentation may result in missed diagnoses and delayed treatment for these diseases.



**Figure 6.** Comparison of overall segmentation results: (a) original images, (b) ground truth, (c) U-Net segmentation results [13], (d) Deeplabv3+ segmentation results [14], (e) Att U-Net segmentation results [27], and (f) MU-Net segmentation results.

Further comparison is made between U-Net and MU-Net, which have relatively good performance. In comparison to the U-Net baseline, MU-Net focuses more on the connectivity of microvessels, resulting in clearer and well-connected vessel outlines at the ends. However, U-Net sometimes produces discontinuous prediction results for certain microvessels. Refer to the second row of Figure 7 for retinal blood vessel images containing numerous microvessels. Our proposed MU-Net integrates multi-scale features in the encoder stage and incorporates skip connections. Furthermore, it leverages the crucial information between channels in the decoder stage to enhance the segmentation results of fine blood vessels. This method not only provides accurate segmentation results for tiny blood vessels but also preserves the spatial morphology more effectively. As a result, the

approach proposed in this paper enhances the qualitative performance of segmentation by improving the segmentation quality.



**Figure 7.** Comparison of local segmentation results: (a) original images, (b) ground truth, (c) U-Net segmentation results [13], (d) Deeplabv3+ segmentation results [14], (e) Att U-Net segmentation results [27], and (f) MU-Net segmentation results. Note: Local areas of interest are outlined using green and red rectangles.

#### 4. Discussion

Retinal blood vessel segmentation is a complex task due to the inconsistent shape and size of blood vessels at different levels of image granularity, as well as their tendency to bend and intersect in various configurations. Conventional machine learning approaches often struggle to accurately detect and segment these blood vessels, particularly microvessels, leading to limited segmentation accuracy and sensitivity. Consequently, enhancing the algorithm's sensitivity to identify microvessels holds significant importance in facilitating doctors' diagnosis.

The experiments revealed that the baseline model has limitations in accurately extracting microvessels and tends to break at the ends of blood vessels, resulting in low sensitivity. To address these issues, this study proposes the use of a multi-scale residual convolution module in the encoder stage to enhance blood vessel features. Additionally, selective nuclear units are implemented at skip junctions to localize receptive fields that match the size of retinal vessels. Furthermore, a residual attention module is designed in the decoder stage to reduce noise interference and remove redundant information. The MU-Net efficiently segments microvessels and maintains proper connectivity at the end of blood vessels.

However, this study still has limitations and requires further optimization of the retinal microvessel segmentation model. On the one hand, we will continue to enhance the contrast between blood vessels and background by optimizing the preprocessing steps, which will ultimately improve the segmentation accuracy. At the same time, we have drawn inspiration from Wen et al. [55], who proposed a dual-threshold iterative algorithm to extract weak vascular pixels and enhance vascular connectivity. This work has provided

valuable insights for our future endeavors. Additionally, we plan to employ post-processing techniques to enhance the visual effects of the segmented image. On the other hand, we plan to incorporate multiple skip connections from Unet++ [56] to effectively fuse feature information at different scales and stages. This will enable more precise segmentation of microvessel structures. In our future work, the sensitivity of microvessel segmentation and good connectivity of blood vessel ends need to be further enhanced.

**Author Contributions:** Conceptualization, X.H. and T.W.; Methodology, X.H. and T.W.; writing—original draft, X.H.; writing—review and editing, T.W.; Supervision, T.W. and W.Y.; Funding acquisition, W.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Natural Science Foundation of China under Nos. 62276061 and 62006041.

**Data Availability Statement:** The data presented in this study is available on request from the corresponding authors, and the dataset was jointly completed by the team, so the data is not publicly available.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Liang, L.; Feng, J.; Peng, R.; Zeng, S. U-Shaped Retinal Vessel Segmentation Combining Multi-Label Loss and Dual Attention. *J. Comput.-Aided Des. Comput. Graph.* **2023**, *35*, 75–86.
- Roychowdhury, S.; Koozekanani, D.D.; Parhi, K.K. Iterative Vessel Segmentation of Fundus Images. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 1738–1749. [[CrossRef](#)] [[PubMed](#)]
- Sinaga, K.P.; Yang, M. Unsupervised K-Means Clustering Algorithm. *IEEE Access* **2020**, *8*, 80716–80727. [[CrossRef](#)]
- Yang, Y.; Wan, W.; Huang, S.; Zhong, X.; Kong, X. RADCU-Net: Residual attention and dual-supervision cascaded U-Net for retinal blood vessel segmentation. *Int. J. Mach. Learn. Cybern.* **2023**, *14*, 1605–1620. [[CrossRef](#)]
- Kande, G.B.; Savithri, T.S.; Subbaiah, P.V. Retinal Vessel Segmentation using Histogram Matching. In Proceedings of the 2008 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS 2008), Macao, China, 30 November–3 December 2008; Volumes 1–4, p. 129.
- Mardani, K.; Maghooli, K. Enhancing retinal blood vessel segmentation in medical images using combined segmentation modes extracted by DBSCAN and morphological reconstruction. *Biomed. Signal Process.* **2021**, *69*, 102837. [[CrossRef](#)]
- Zhao, Y.; Rada, L.; Chen, K.; Harding, S.P.; Zheng, Y. Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images. *IEEE Trans. Med. Imaging* **2015**, *34*, 1797–1807. [[CrossRef](#)] [[PubMed](#)]
- Ricci, E.; Perfetti, R. Retinal Blood Vessel Segmentation Using Line Operators and Support Vector Classification. *IEEE Trans. Med. Imaging* **2007**, *26*, 1357–1365. [[CrossRef](#)]
- Marin, D.; Aquino, A.; Gegundez-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* **2011**, *30*, 146–158. [[CrossRef](#)]
- Kaluri, R.; Ch, P.R. Optimized feature extraction for precise sign gesture recognition using self-improved genetic algorithm. *Int. J. Eng. Technol. Innov.* **2018**, *8*, 25–37.
- Zhou, B.; Zhao, H.; Puig, X.; Fidler, S.; Barriuso, A.; Torralba, A. Scene Parsing through ADE20K Dataset. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 5122–5130.
- Caesar, H.; Uijlings, J.; Ferrari, V. COCO-Stuff: Thing and Stuff Classes in Context. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018), Salt Lake City, UT, USA, 18–23 June 2018; pp. 1209–1218.
- Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Part III*; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241.
- Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Computer Vision—ECCV 2018, Part VII*; Springer: Cham, Switzerland, 2018; Volume 11211, pp. 833–851.
- Wu, Y.; Xia, Y.; Song, Y.; Zhang, D.; Liu, D.; Zhang, C.; Cai, W. Vessel-Net: Retinal Vessel Segmentation Under Multi-path Supervision. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019, Part I*; Springer: Cham, Switzerland, 2019; Volume 11764, pp. 264–272.
- Gu, Z.; Cheng, J.; Fu, H.; Zhou, K.; Hao, H.; Zhao, Y.; Zhang, T.; Gao, S.; Liu, J. CE-Net: Context Encoder Network for 2D Medical Image Segmentation. *IEEE Trans. Med. Imaging* **2019**, *38*, 2281–2292. [[CrossRef](#)]
- 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](#)]
- Deng, X.; Ye, J. A retinal blood vessel segmentation based on improved D-MNet and pulse-coupled neural network. *Biomed. Signal Process.* **2022**, *73*, 103467. [[CrossRef](#)]

19. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. In Proceedings of the International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2–4 May 2015.
20. Li, X.; Wang, W.; Hu, X.; Yang, J. Selective Kernel Networks. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019), Long Beach, CA, USA, 15–20 June 2019; pp. 510–519.
21. Phan, T.; Kim, S.H.; Yang, H.J.; Lee, G.S. Skin Lesion Segmentation by U-Net with Adaptive Skip Connection and Structural Awareness. *Appl. Sci.* **2021**, *11*, 4528. [[CrossRef](#)]
22. Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention gated networks: Learning to leverage salient regions in medical images. *Med. Image Anal.* **2019**, *53*, 197–207. [[CrossRef](#)]
23. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal.* **2020**, *42*, 2011–2023. [[CrossRef](#)]
24. Luo, K.; Wang, T.; Ye, F. U-Net segmentation model of brain tumor MR image based on attention mechanism and multi-view fusion. *J. Image Graph.* **2021**, *26*, 2208–2218.
25. Woo, S.; Park, J.; Lee, J.; 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.
26. Li, X.; Jiang, Y.; Li, M.; Yin, S. Lightweight Attention Convolutional Neural Network for Retinal Vessel Image Segmentation. *IEEE Trans. Ind. Inform.* **2021**, *17*, 1958–1967. [[CrossRef](#)]
27. Wang, S.; Li, L.; Zhuang, X. *AttU-NET: Attention U-Net for Brain Tumor Segmentation*; Springer: Cham, Switzerland, 2022; Volume 12963, pp. 302–311.
28. 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.
29. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.
30. Li, Y.; Wang, S.; Wang, J.; Zeng, G.; Liu, W.; Zhang, Q.; Jin, Q.; Wang, Y. GT U-Net: A U-Net Like Group Transformer Net-work for Tooth Root Segmentation. In *Machine Learning in Medical Imaging, MLMI 2021*; Springer: Cham, Switzerland, 2021; Volume 12966, pp. 386–395.
31. Liang, L.; Lu, B.; Long, P.; Yang, Y. Adaptive feature fusion cascade Transformer retinal vessel segmentation algorithm. *Opto-Electron. Eng.* **2023**, *50*, 230161.
32. Koshy, R.; Mahmood, A. Optimizing Deep CNN Architectures for Face Liveness Detection. *Entropy* **2019**, *21*, 423. [[CrossRef](#)]
33. Shafiq, M.; Gu, Z. Deep Residual Learning for Image Recognition: A Survey. *Appl. Sci.* **2022**, *12*, 8972. [[CrossRef](#)]
34. Ghiasi, G.; Lin, T.; Le, Q.V. DropBlock: A regularization method for convolutional networks. In Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, QC, Canada, 3–8 December 2018; p. 31.
35. Guo, C.; Szemenyei, M.; Pei, Y.; Yi, Y.; Zhou, W. SD-Unet: A Structured Dropout U-Net for Retinal Vessel Segmentation. In Proceedings of the 2019 IEEE 19th International Conference on Bioinformatics And Bioengineering (BIBE), Athens, Greece, 28–30 October 2019; pp. 439–444.
36. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 30TH IEEE Conference On Computer Vision And Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.
37. Li, K.; Qi, X.; Luo, Y.; Yao, Z.; Zhou, X.; Sun, M. Accurate Retinal Vessel Segmentation in Color Fundus Images via Fully Attention-Based Networks. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 2071–2081. [[CrossRef](#)] [[PubMed](#)]
38. Wang, X.; Jiang, X. Retinal vessel segmentation by a divide-and-conquer funnel-structured classification framework. *Signal Process.* **2019**, *165*, 104–114. [[CrossRef](#)]
39. Liang, L.; Zhou, L.; Yin, J.; Sheng, X. Fusion multi-scale transformer skin lesion segmentation algorithm. *J. Jilin Univ. (Eng. Technol. Ed.)* **2022**, 1–13.
40. 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](#)]
41. 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. [[CrossRef](#)] [[PubMed](#)]
42. 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. Ophthalmology Vis. Sci.* **2009**, *50*, 2004–2010. [[CrossRef](#)]
43. Yan, Z.; Yang, X.; Cheng, K. Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 1912–1923. [[CrossRef](#)]
44. Zhang, Y.; Chung, A.C.S. Deep Supervision with Additional Labels for Retinal Vessel Segmentation Task. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018, Part II*; Springer: Cham, Switzerland, 2018; Volume 11071, pp. 83–91.
45. Sidhu, R.K.; Sachdeva, J.; Katoch, D. Segmentation of retinal blood vessels by a novel hybrid technique- Principal Component Analysis (PCA) and Contrast Limited Adaptive Histogram Equalization (CLAHE). *Microvasc. Res.* **2023**, *148*, 104477. [[CrossRef](#)]
46. Wu, Z.; Cen, S. Image dehazing algorithm based on adaptive gamma correction estimation. *Chin. J. Liq. Cryst. Disp.* **2022**, *37*, 106–115. [[CrossRef](#)]
47. Le, N.; Ou, Y.Y. Incorporating efficient radial basis function networks and significant amino acid pairs for predicting GTP binding sites in transport proteins. *BMC Bioinform.* **2016**, *17*, 183–192. [[CrossRef](#)] [[PubMed](#)]

48. Le, N.; Nguyen, T.; Ou, Y. Identifying the molecular functions of electron transport proteins using radial basis function networks and biochemical properties. *J. Mol. Graph. Model.* **2017**, *73*, 166–178. [[CrossRef](#)] [[PubMed](#)]
49. Zhou, L.; Yu, Q.; Xu, X.; Gu, Y.; Yang, J. Improving dense conditional random field for retinal vessel segmentation by discriminative feature learning and thin-vessel enhancement. *Comput. Methods Programs Biomed.* **2017**, *148*, 13–25. [[CrossRef](#)] [[PubMed](#)]
50. 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.
51. Khowaja, S.A.; Khuwaja, P.; Ismaili, I.A. A framework for retinal vessel segmentation from fundus images using hybrid feature set and hierarchical classification. *Signal Image Video Process.* **2019**, *13*, 379–387. [[CrossRef](#)]
52. 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.
53. Sathananthavathi, V.; Indumathi, G. Encoder Enhanced Atrous (EEA) Unet architecture for Retinal Blood vessel segmentation. *Cogn. Syst. Res.* **2021**, *67*, 84–95.
54. Zhang, Y.; Fang, J.; Chen, Y.; Jia, L. Edge-aware U-net with gated convolution for retinal vessel segmentation. *Biomed. Signal Process.* **2022**, *73*, 103472. [[CrossRef](#)]
55. Liu, W.; Yang, H.; Tian, T.; Cao, Z.; Pan, X.; Xu, W.; Jin, Y.; Gao, F. Full-Resolution Network and Dual-Threshold Iteration for Retinal Vessel and Coronary Angiograph Segmentation. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 4623–4634. [[CrossRef](#)]
56. 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, Proceedings of the 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS, Granada, Spain, 20 September 2018*; Springer: Cham, Switzerland, 2018; pp. 3–11.

**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.