Retinal Vessel Segmentation Algorithm Based on U-NET Convolutional Neural Network

Yun-Hao Zhang, Jie-Sheng Wang*, Zhi-Hao Zhang

Abstract—**The segmentation of retinal vessels a crucial role in the accurate visualization, early intervention, and surgical planning for ophthalmic disorders. There are some problems in the process of retinal vascular imaging, such as noise, low contrast, imbalance of vascular background pixel ratio and distortion of capillary cutting. The retinal blood vessel images underwent a series of preprocessing steps to optimize the performance of image segmentation. These steps included converting the images to grayscale, normalizing the data, applying restricted contrast adaptive histogram equalization, performing gamma correction, and then normalizing the data again. The subsequent analysis utilized four segmentation algorithms based on the U-Net model, namely the U-Net segmentation algorithm, Res-UNet segmentation algorithm, DU-Net segmentation algorithm, and Sa-UNet segmentation algorithm, were selected to segment the retinal vessel images. The fundus images from the DRIVE public database were utilized to conduct simulation experiments in order to validate the efficacy of the adopted algorithms. The sensitivity, specificity, accuracy and AUC of Sa-UNet segmentation algorithm were 0.8573, 0.9835, 0.9905 and 0.9755, respectively.**

Index Terms—**Retinal vessel; Image segmentation; U-Net; Performance comparison**

I. Introduction

The morphological information of retinal blood vessels is indispensable for the early diagnosis of ocular diseases. In addition to the detection of ocular diseases, diabetic retinopathy can also be timely identified. Retinal blood vessels are the sole non-invasive observable blood vessels in the human body, and their detection and analysis hold significant application value for predicting and diagnosing aforementioned conditions.[1]. If not detected in time and treated effectively, it will lead to deterioration of the disease, blindness or even more serious consequences. Therefore, accurate segmentation of retinal fundus early detection of vascular abnormalities, playing a pivotal role in the prevention and treatment of various ophthalmic diseases. With the incessant advancement of science and technology, the deep learning technology has been continuously improved. Image segmentation technology is also gradually promoted in the medical field. Image segmentation algorithms based on deep learning can help doctors process and analyze complex fundus images, and further improve the speed and accuracy of computer image segmentation. Its optimization methods have attracted more and more attention [2].

Retinal blood vessels are of irreplaceable significance for the medical field. In order to achieve better performance, scholars at home and abroad have been studying retinal segmentation for many years. Computerization of the fundus was first proposed in 1982, and Chaudhuri published the first paper on retinal segmentation in 1989 [3]. Retinal vessel segmentation methods are commonly classified into two categories by scholars, namely unsupervised learning methods and supervised learning methods. The latter can be further divided into those involving manual feature extraction and those involving automatic feature extractional Neural Networks (CNNs) as dominant architectures [4]. Since CNN needs a large number of data sets, the number of medical image data is small, and the training model cannot be optimized through a large number of training. On the basis of full CNN, Ronneberger et al. proposed a U-shaped code structure to make the expansion path and contraction path symmetric [5].The network utilizes jump connections to establish links between the up-sampled output and the output of the sub-module at the same resolution in the encoder, which serves as input for the subsequent submodule in the decoder. Iglovikov et al. have enhanced U-Net's encoder by replacing it with VGG11 and proposed TernausNet. [6].

Drawing inspiration from the principles of residual connection and dense connection, Xiao et al. incorporated a form with residual connection and a form with dense connection into each sub-module of UNet, respectively. [7]. Res-UNet and Dense-UNet are proposed, where Res-UNet is used for retinal segmentation. The incorporation of the output from a layer within a sub-module into the subsequent layers' input, along with the derivation of a layer's input from the amalgamation of preceding layers' outputs, characterizes a dense connection. MultiResUNet employs a MutiRes module in conjunction with UNet [8]. The MutiRes module enhances residual connections by concatenating three 3x3 convolutions to generate a fused feature map, which is subsequently added to the input feature map through a 1x1 convolution operation. Deng et al. [9] combined the primary and secondary paths with the channel attention mechanism and used it to adaptively predict the potential key features to form a dual-path convolutional neural network DPCA-Net, which improved

Manuscript received July 1, 2023; revised October 18, 2023. This work was supported by the Basic Scientific Research Project of Institution of Higher Learning of Liaoning Province (Grant No. LJKZ0293), and Postgraduate Education Reform Project of Liaoning Province (Grant No. LNYJG2022137).

Yun-Hao Zhang is a postgraduate student of School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, 114051, P. R. China (e-mail: a18524334102@163.com).

Jie-Sheng Wang is a professor of School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, 114051, P. R. China (Corresponding author, phone: 86-0412-2538246; fax: 86-0412-2538244; e-mail: wang_jiesheng@126.com).

Zhi-Hao Zhang is a postgraduate student of School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, 114051, P. R. China (e-mail: 527351380@qq.com).

the accuracy of segmentation. Atli et al. [10] proposed a segmented deep learning model, Sine-Net, which firstly used up-sampling to capture thin vessel features, and then used down-sampling vessel features to capture thin vessel features and thick vessel features respectively, and added residuals to the model to assume deeper contextual information.

The retinal segmentation method based on deep learning is conclusively demonstrated to be superior and more reliable compared to the traditional segmentation approach. In this study, four segmentation algorithms based on U-Net, Res-UNet, DU-Net and Sa-UNet were selected to segment retinal vessel images. The fundus images from the DRIVE public database were utilized for conducting simulation experiments, and the evaluation metrics were compared against the reference images to derive the final conclusion.

II. Retinal Vessel Image Segmentation Algorithm Based on U-NET Convolutional Neural Network

The segmentation of retinal vessels plays a pivotal role in the field of medicine. Manual segmentation is laborious and difficult for clinical application. With the first paper on retinal vessel segmentation proposed in the late 1980s, and the emergence of AlexNet network in 2012, this technology has gradually improved from traditional segmentation scheme to deep learning. This section focuses on convolutional neural network (CNN), and introduces U-Net model, Res-UNet model, DU-UNet model and SA-UNet model in details. It lays the foundation for subsequent image segmentation.

A. Convolutional Neural Network

The Convolutional Neural Networks (CNN) is a deep learning model that has gained significant traction in the field of image processing. The CNN architecture primarily comprises an input layer, convolutional layers, ReLU activation layers, pooling layers, and fully connected layers. The complete convolutional neural network is formed by superimposing all layers.

(1) Convolutional Layer

The convolutional layer is the core layer of the CNN, which generates the computational amount in the network and contains multiple convolution kernels. By calculating the dot product between the area of the input image and the weight matrix of the filter, the result is processed by nonlinear activation function, and the output map is finally obtained. The deeper the network is, the more edges and feature detection was obtained. The input was selected as 4*4 and the convolution kernel as 3*3 for the calculation. The convolution calculation form and results are illustrated in Fig. 1, while the disassembled calculation process is presented in Fig. 2. In the fist step shown in Fig. 2(a), the 3x3 network positioned in the upper left corner of the convolution kernel is selected for performing dot product calculations with the convolution kernel and the first output result is obtained by accumulating the dot multiplication data. Slide one step to the right in Fig. 2(b) to get the second output. Slide the step down in Fig. 2(c) to get the third output result. Fig. 2(d) takes the 2*2 input network in the lower right corner as the last output result and serves as the final output.

(2) Pooling Layer

The conventional approach involves the periodic insertion of a pooling layer between successive convolutional layers, gradually reducing the spatial dimensions of the data volume. This reduction facilitates parameter count minimization and computational resource conservation within the network, effectively mitigating overfitting. However, it does not affect network depth. The most common pooling operations are Average pooling and Max pooling, in which the effect of maximum pooling is generally better than that of average pooling. This paper describes the maximum pooling in details. The maximum pooling method retains the most salient features of the feature map. As shown in Fig. 3, the maximum pooling is resolved. In Fig. 3(a), within the 2*2 pooling layer, the maximum value is 6, and the output result is 6. The step size for rightward movement is 2, as illustrated in Figure 3(b), while the maximum output value reaches 8. Similarly, the output results of Fig. 3(c) and Fig. 3(d) are 9 and 7 respectively. The whole output result is shown in Fig. 3(d).

(3) Activation Layer

The main function of activation function in activation layer is to further deepen the stacking of network layers and enhance the expression of network speed.

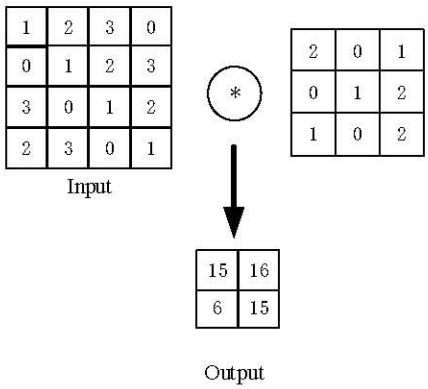


Fig. 1 Results of convolution calculation.

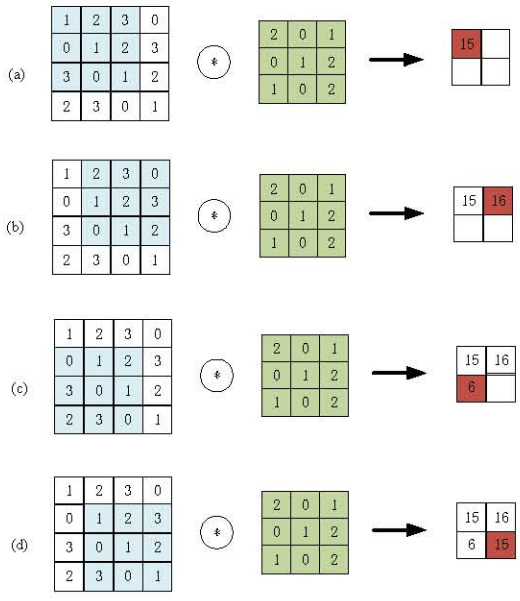


Fig. 2 Dismantling calculation process.

Commonly used activation functions are Sigmoid, ReLU, LeakyReLU, and Tanh functions. In the neural network architecture, ReLU activation functions are generally used instead of traditional Sigmoid or Tanh activation functions. As shown in Fig. 4, The Rectified Linear Unit (ReLU) activation function assigns a value of 0 to each negative input pixel, while it preserves the same value for each positive input pixel.

B. U-Net Model

U-Net has emerged as a prominent network among various methodologies employed for medical image segmentation in recent years. Since CNN cannot accurately segment the fine parts of the object, full convolutional neural network (FCN) is introduced to realize image pixel level segmentation. The advancement of FCN has led to the persistence of a substantial number of channels in U-Net during up-sampling, thereby facilitating the propagation of contextual information within the network at a deeper resolution. The middle connection part is changed to jump connection, and the encoder (down-sampling) and decoder (up-sampling) symmetry forms a U-shaped network. The U-Net architecture diagram is depicted in Fig. 5 [5].

According to the structural diagram, the left side of the model constitutes the encoder, while the right side constitutes the decoder. The encoder consists of four submodules, each of which contains two convolutional layers, and the maximum pooling method is used to realize subsampling. The resolution of the input image is 572 x 572 pixels, while the resolution of modules 1 to 5 is gradually reduced to 572 x 572, 284 x 284, 140 x 140, 68 x 68 and 32 x 32, respectively. As valid mode is employed for convolution, the resolution of each subsequent sub-module is halved compared to its preceding sub-module's resolution. Finally, a 1x1 convolution kernel is utilized in the last layer to reduce channel dimensions to a predetermined number. The decoder comprises of four sub-modules that progressively enhance the resolution of the output image through a step-by-step upsampling operation until it matches the same resolution as the input image. Consequently, due to the utilization of convolution operations, the effective mode results in a smaller actual output size compared to that of the input image. The network also employs jump connections to establish links between the up-sampling results and the outputs of sub-modules in the encoder that possess identical resolutions, serving as inputs to lower sub-modules within the decoder.

The U-Net network can be trained with a limited number of models, and data augmentation techniques can be incorporated into the network. This enables effective resolution of the challenge posed by small training datasets in the medical domain that necessitate extensive data detection. The network achieves high segmentation accuracy by segmenting each pixel individually. During the training process, self-regulation of the learning rate is achieved through utilization of a high Momentum value (0.99). To account for varying pixel probabilities within an image, weights are assigned to pixels to enhance their distinctiveness, which can be defined as:

$$w(x) = w_c(x) + w_0 * \exp\left(-\frac{(d_1(x)+d_2(x)^2)}{2\sigma^2}\right) \quad (1)$$

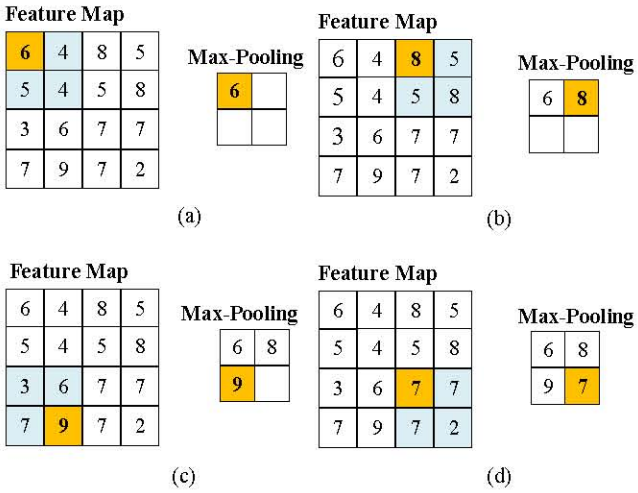


Fig. 3 Maximum pooling method.

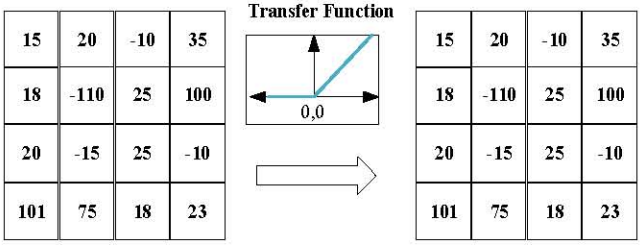


Fig.4 ReLU activation function.

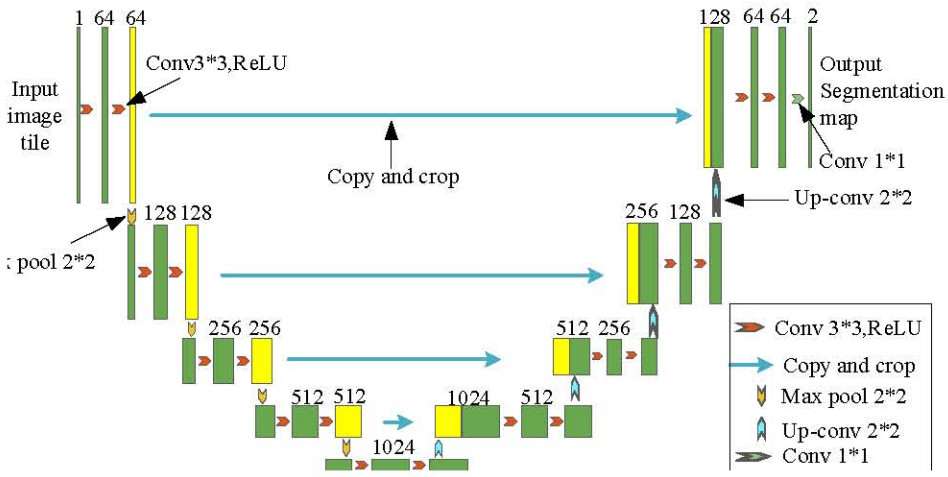


Fig. 5 U-Net structure.

where, w_c represents the balanced category frequency weight map; d_1 represents the distance from the nearest cell boundary; d_2 represents the distance from the boundary of the second near cell; In general, w_0 =10 and σ≈5 are set. The network weights are initialized by following a Gaussian distribution, where the standard deviation is determined as 2 divided by the total number of incoming nodes (N).

C. Res-UNet Model

In 2018, the Res-UNet model was released, which makes important improvements to U-Net[7]. It is difficult for the human eye to distinguish small blood vessels, the optic disc area is bright, the contrast is low, the thin-walled blood vessel detection and the boundary is not clear due to reflection are improved. Firstly, Contrast Limited Adaptive Histgram Equalization (CLAHE) [11] was used for prepossessing to increase the image contrast, adjust the fundus image to grayscale image, and adjust the size to 512*512. The gray-scale image was standardized by the following formula.

$$I = \frac{I - \mu}{\sigma} \tag{2}$$

where, μ and σ represent the mean and standard deviation of the gray-scale image I.

Due to the small number of samples, it was not enough to support the training demand. 500 overlapping 64*64 patches after image enhancement were randomly extracted from the training image to expand the number of training, and then the segmented images were obtained. The overall architecture of this network is roughly the same as that of U-Net, and a new scheme is added on the basis of the original weighted attention mechanism and jump connection. Each sub-module of U-Net is replaced by a form with residual connection and dense connection respectively.

(1) Attention unit

The circular region (ROI) and dark background were extracted, and the circular template ROI mask M was binarized as weighted attention by using Eq. (3). However, this weighted attention mechanism only focuses on the target model area of the ROI and ignores the noise background.

$$\zeta(x,y) = \begin{cases} F(x,y)*1.0 & (x,y)\in M \\ F(x,y)*0.0 & (x,y)\notin M \end{cases} \tag{3}$$

where, $F(x,y)$ denotes the feature at (x,y).

(2) Skip connection unit

The skip connection is added to the network according to the idea of Res-Net[12]. This operation increases the depth of the network, and according to the explanation in Res-Net, it can effectively mitigate over-fitting issues and enhance the model's accuracy.

(3) Loss function

Binary cross-entropy is used as the loss function of the training model, which is shown in Eq. (4).

$$L(p,q) = -\frac{1}{n}\sum_{k=1}^{n} q_k \log p_k + (1-q_k)\log(1-p_k) \tag{4}$$

where, n represents the number of pixels per image, and the background values corresponding to the predicted pixel values are denoted by p and q.

D. DU-Net Model

Basedo on local features of blood vessels, retinal vessel segmentation was performed in an end-to-end manner. Inspired by U-Net[5] and Deformable Convolutional Network[13], the deformable convolutional network is integrated into the proposed network. The DU-Net utilizes an up-sampling operator for enhancing output resolution, with the objective of extracting contextual information and achieving accurate positioning through the synergistic integration of low-level and high-level feature maps [14]. In addition, by adaptively adjusting the receptive field according to the distinctive features of each vessels, DU-Net proficiently captures retinal vessels with diverse shapes and sizes. The network exhibits a U-shaped architecture, flanked by encoders and decoders, wherein the original convolutional layers are substituted with deformable convolutional blocks. Training the novel model synergistically integrates low-level and high-level features, while adapting receptive fields and sampling locations to accommodate the size and shape of blood vessels for precise segmentation. The DU-Net, an exquisite extension of the U-Net framework, showcases its prowess by employing deformable convolutional blocks encoding and decoding units. The illustrious network architecture is depicted in Fig. 6 [14].

The deformable convolutional block's detailed design is presented within the specified region. In the U-Net architecture, it consists of a convolutional encoder on the left and a decoder on the right. Complex details and changes of retinal blood vessels are captured using deformable convolutional blocks at various stages of encoding and decoding, which enable efficient acquisition of local, dense and adaptable receptive field information. Each deformable convolution block contains a convolution offset layer, a convolution layer, a normalization layer and an activation layer, which constitute the core of the deformable convolution. During the decoding phase, another standard convolution layer after the merge is also inserted to adjust the number of filters in the convolution offset layer. With this structure, DU-Net is able to learn key features and generate detailed retinal vessel segmentation results.

The original goal of introducing deformable convolutional neural networks is to determine whether pixels in an image belong to blood vessels. This task can be viewed as a pixel-level classification problem, where the classification of pixels is determined based on information about their surrounding neighborhoods. The neighborhood refers to a small image block defined at the center of the target pixel. To classify the target pixel, we extract the pixel values from a small image block centered on that pixel, obtaining highly localized information. In order to maximize local feature capture while reducing computational complexity, we set the size of these small image blocks to 29 × 29 pixels. Fig. 7 illustrates the structure of a deformable convolutional neural network.

E. SA-UNet Model

Based on U-Net network architecture, a lightweight spatial attention network (SA-UNet) is proposed by introducing the improvement of spatial attention module. [15]. This method does not require a large number of labeled training samples and can effectively leverage the available annotated data in an enhanced manner. This module generates attention maps across spatial dimensions and applies them to the input feature maps for adaptive feature refinement. Additionally, in order to address the issue of overfitting, this approach incorporates a structured Dropout convolutional block as a replacement for the original convolutional block in U-Net. To maintain network simplicity, SA-UNet introduces a structured Dropout convolutional block that integrates DropBlock and Batch Normalization (BN) to substitute U-Net's initial convolutional block. The specific network structure is shown in Fig. 8[15], where the structured Dropout convolution blocks and 2*2 maximum convolutions are included in each step of the encoder. After each fast convolutional layer, DropBlock, batch normalization, and a ReLU activation function are incorporated.

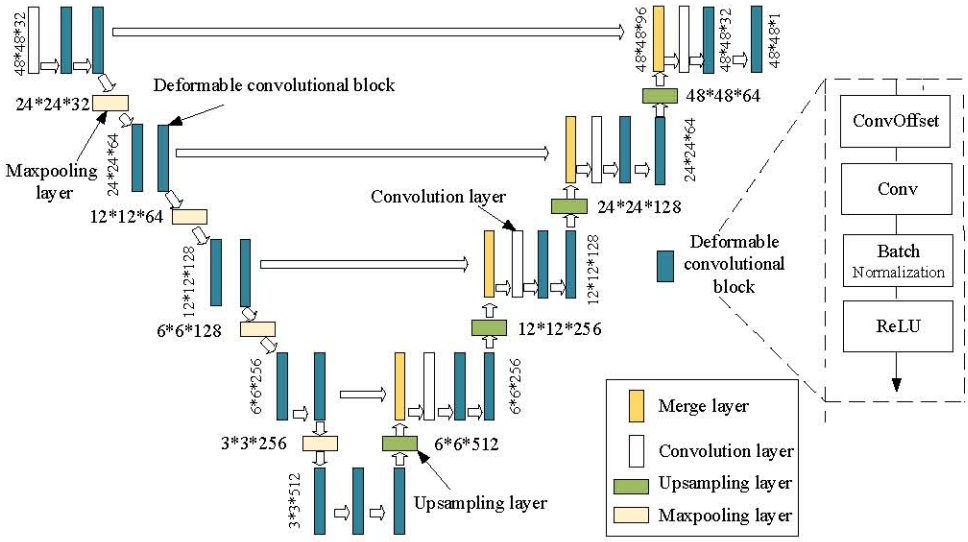


Fig. 6 DU-Net network architecture diagram.

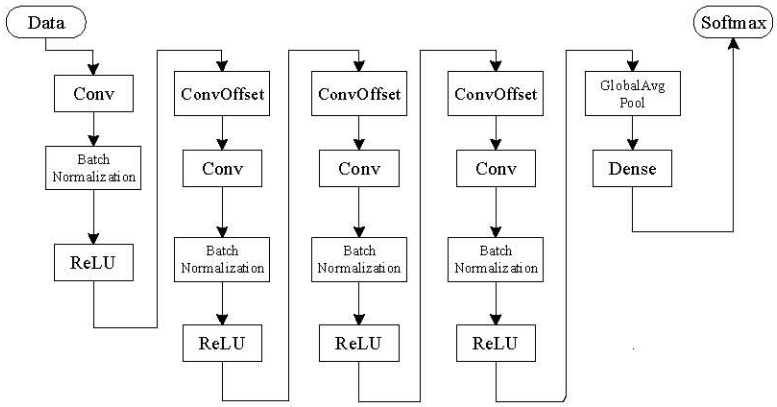


Fig. 7 Structure of the deformed convolutional network.

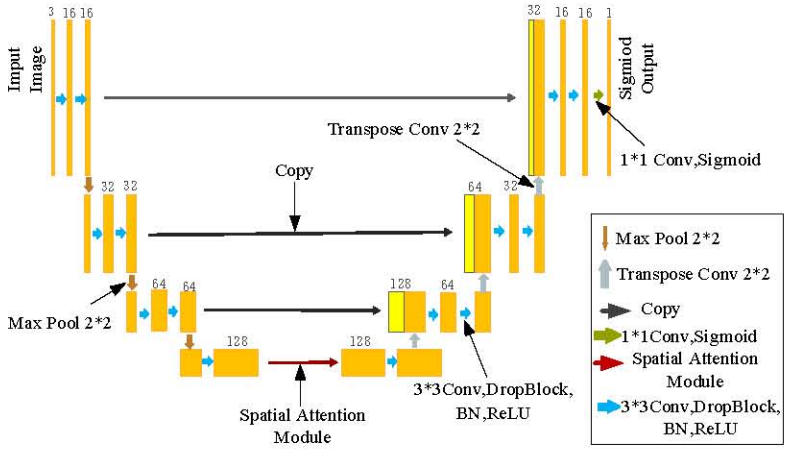


Fig. 8 Network architecture of SA-UNet model.

The most important part of convolutional attention module is the introduction of spatial attention module, which is used for detection and classification. To generate a spatial attention graph, spatial attention (SA) constructs an effective feature descriptor by applying maximum pooling and average pooling operations along the channel axis on the feature graph, respectively. Subsequently, these descriptors are concatenated together to form a comprehensive representation, as illustrated in Fig. 9. [15].

After maximum pooling and average pooling, the output features $F \in R^{H*W*C}$ and $F_{ab}^{s} \in R^{H*W*1}$ $F_{mb}^{s} \in R^{H*W*1}$ are obtained. Then the above convolutional layer and Sigmod activation function are used to generate the spatial attention map $F_{ab}^{s} \in R^{H*W*1}$. In conclusion, the spatial attention formula is expressed as follows.

$$\begin{aligned} F^{s} &= F \cdot M^{s}(F) \\ &= F \cdot \sigma(f^{7\times7}([MaxPool(F); AvgPool(F)])) \\ &= F \cdot \sigma(f^{7\times7}([F_{mp}^{s}; F_{ap}^{s}])) \end{aligned} \tag{5}$$

where, $f^{7\times7}$ represents the convolution operation with a convolution kernel of 7x7.

III. Simulation and Result Analysis

A. Public Datasets

The retinal blood vessels were trained and tested by using the fundus images in the DRIVE dataset. The DRIVE dataset was screened and established by the leadership team of Niemeijer in the Netherlands[16]. The dataset contains 40 images, 20 color images for training and 20 color images for testing. All images included in the dataset were collected and selected by more than 400 people who participated in the experiments. Participants were between 25 and 90 years of age, which includes the majority of patients at risk for disease. The image size was 564*584 and was taken at 45°. Among the 40 images, there were 33 healthy fundus images without eye diseases and 7 fundus images with early diabetic lesions. Each fundus image was segmented by two experts. There were two vascular structure maps and one mask. For manual segmentation results, the first expert marking was 12.7% and the second expert marking was 12.3%. The image segmentation results are used as the standard to test the segmentation algorithm performance, so this paper uses the segmentation results of the first expert as the standard segmentation image.

B. Evaluation Indexes of Retinal Vessel Image Segmentation

The purpose of retinal vessel segmentation is to segment every pixel in the image. In order to test the segmentation performance, the confusion matrix evaluation model is used to test the validity. The following four performance measures were employed in this study: sensitivity, specificity, accuracy and area under the curve (AUC) under the Receiver Operating Characteristic (ROC). Their expressions are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{6}$$

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

$$Sensitivity = \frac{TP + TN}{TP + FN + TN + FP} \tag{8}$$

where, TP is the number of vessels in correct segmentation; TN is the number of background in correct segmentation; FN represents the number of pixels that misclassified blood vessels into background; FP represents the number of pixels that missegmented the background into vessels.

The working characteristic curve is crucial for assessing the overall performance of binary vessel segmentation, as it illustrates the relationship between false positive rate (FPR) and true positive rate (TPR). A larger area under the ROC curve (AUC) indicates superior performance of the segmentation algorithm and greater robustness. Herein lie the mathematical definitions of these metrics.

$$TPR = \frac{TP}{TP + FN} \tag{9}$$

$$FPR = \frac{TP}{FP + TN} \tag{10}$$

C. Image Preprocessing and Image Enhancement

(1) Image Preprocessing

Due to the manual shooting process, the fundus images in the dataset exhibited uneven illumination and low contrast in both vascular pixels and background pixels. The detection of small blood vessel features is improved to optimize the segmentation performance of the neural network, we implemented a preprocessing step on the fundus image prior to segmentation.

Gray transformation. The data values of the red, green and blue channels of the color fundus image were converted to *Gray* level by the gray level conversion strategy shown in Eq. (11), In this context, Gray denotes the grayscale value, while R, G and B respectively denote, green and blue color channels.

$$\text{Gray} = \text{R} * 0.299 + \text{G} * 0.587 + \text{B} * 0.114 \tag{11}$$

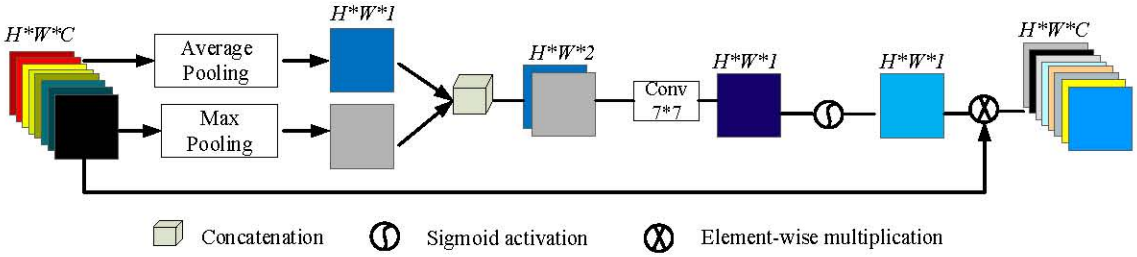


Fig. 9 Spatial attention module.

Data standardization. The mean and standard deviation of fundus images, after being transformed into grayscale, are normalized to ensure that the data exhibits characteristics consistent with a normal distribution. This process of standardization can be mathematically represented by the following formula:

$$X_s = \frac{X_g - \mu}{\sigma} \tag{12}$$

where, X_g is the grayscale image sample data, μ is the sample data mean, σ is the data standard deviation, and X_s is the normalized image data.

Because the contrast is still very low after grayscale processing, the complete vascular information in the image cannot be extracted. Contrast Limited Adaptive Histogram Equalization (CLAHE) [11] can be used to reallocate the brightness and change the image contrast, which can significantly enhance the local contrast and image details. CLAHE is used to normalize the image contrast enhancement data, which can not only effectively eliminate the amplification problem of noise, but also enhance the contrast and the distinction between blood vessel pixels and background pixels, and obtain complete blood vessel information.

Gamma correction. Gamma correction is a nonlinear operation that can enhance the fundus image and change the brightness of the dark areas without affecting the areas with strong brightness. The formula is described as follows:

$$f(I) = I^{\gamma} \tag{13}$$

where, I is the input image data, γ is the correction parameter, and $f(I)$ is the output image data.

Data normalization is performed, whereby the image data is standardized and the pixel values of the image are adjusted within the range of [0, 1]. The advantages of data normalization are that it makes data processing easier, allows us to control the training gradient and train our model faster. The normalization formula is described in Eq. (14).

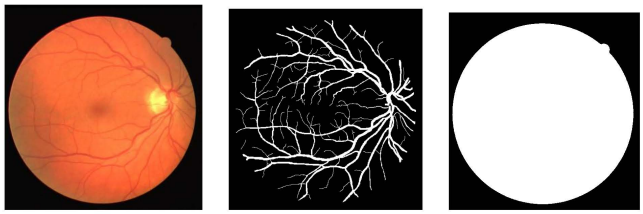
$$X_{out} = \frac{X_{in} - \min}{\max - \min} \tag{14}$$

where, X_{in} represents the pixel value of the input image, min denotes the minimum value of the data, max signifies the maximum value of the data, and X_{out} indicates the pixel value of the normalized image data.

The overall pre-processed image results are shown in Fig. 11. After pre-processing, it can be seen that the contrast of the image is significantly enhanced, and the vascular information is more prominent, which lays the foundation for the next segmentation.

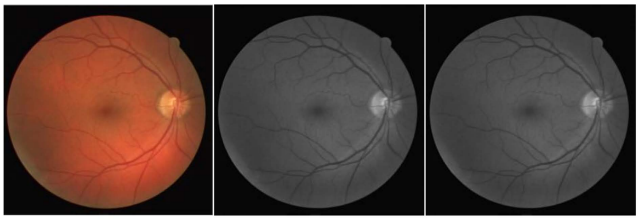
(2) Image Enhancement

Because there are too few fundus images in the public dataset, it cannot meet the needs of image training. Although the problem of image over-fitting can be solved by shallow training or optimizing the loss function, it cannot solve the fundamental problem. In view of too few images, this problem can be solved by image sectioning. The adoption of random slicing enables the augmentation of image samples, which is crucial for deep learning due to its requirement for a large training dataset. This method offers the advantage of generating a proportional number of slice samples regardless of factors such as original image size, slice step, or slice size. Consequently, it effectively expands the training data and enhances the performance of deep learning models. The number can be set manually, and a large number of picture data can be cut out. In this paper, in order to prevent the loss of details by amplifying the data, the number of up and down sampling is reduced so as to reduce the occurrence of over-fitting problem. The image block size used in this paper is 48*48, the total number of images collected was 25,000. Fig. 12 shows the randomly extracted images.

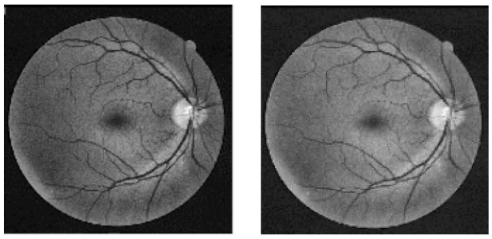


(a) Original image (b) Standard segmented (c) Mask diagram Image

Fig. 10 Images in the DRIVE dataset.

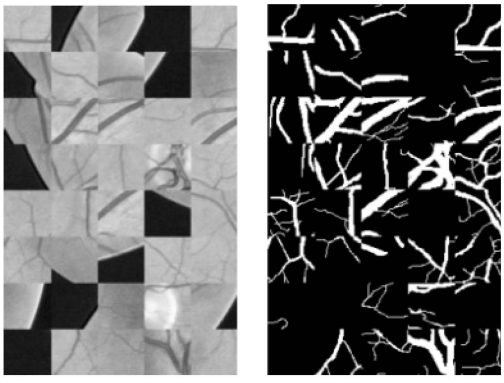


(a) Original Image (b) Grey Processing (c) Standardization



(d) CLAHE (e) Gamma Correction

Fig. 11 Overall preprocessed images.



(a) Crop the Image Blocks (b) Standard segmented image

Fig. 12 Random extracted images.

D. Experimental Results and Analysis

In a controlled experimental setting, we trained and tested the dataset using four distinct segmentation algorithms, namely U-Net model, Res-UNet model, DU-Net model, and SA-UNet model. Subsequently, we conducted a comparative analysis of the generated segmentation images produced by these diverse algorithms while evaluating their respective performance indicators.

(1) Segmentation Results

We evaluated four distinct retinal segmentation algorithms using the DRIVE dataset and presented the corresponding experimental results in FIG. 13, 14, 15, and 16 respectively. Each algorithm performed two datasets in each DRIVE database.

(2) Comparative Experimental Analysis

In order to compare the segmentation effect of different segmentation methods more intuitively, the images in the DRIVE dataset were selected and segmented by the above methods. The U-Net model, Res-UNet model, DU-Net model and SA-Net model were compared with the manual segmentation method. The second column is the standard image of manual segmentation, the third column is segmented by U-Net model, the fourth column is segmented by Res-UNet model, the fifth column is segmented by DU-Net model, and the sixth column is segmented by SA-UNet model. The segmentation results have been specifically presented in Fig. 17.

(3) Comparison of Segmentation Performance Indexes under Different Algorithms

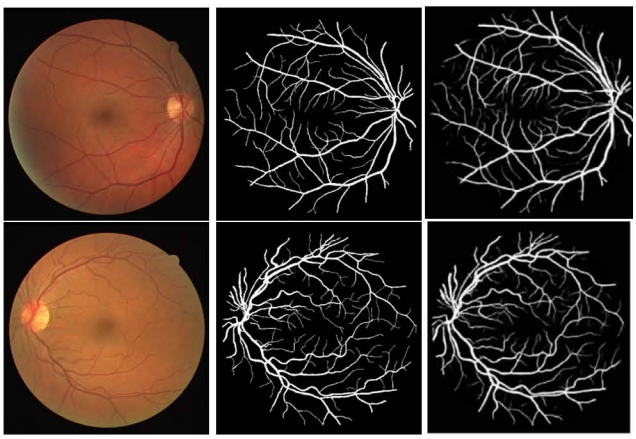
In order to reflect the differences in performance of the five algorithms more intuitively, the effectiveness and feasibility of the methods are summarized. Based on the DRIVE dataset, the segmentation was performed by the methods described above. The accuracy, specificity, sensitivity, and AUC of the segmentation results were analyzed and are presented in Table I. Based on the DRIVE dataset, a comparison of the four algorithms can be observed from the data in Table I. For the segmentation accuracy and sensitivity, SA-UNet model was the highest, reaching 0.9905 and 0.8573. The highest specificity was the Res-UNet model (0.9892). The highest AUC was 0.9856 for DU-Net model. From the comprehensive analysis, SA-UNet model is better. The advantages and disadvantages of various algorithms are elaborated in details below.

For medical image segmentation, the role of U-Net model is irreplaceable. It modifies the middle convolutional layer of the full convolutional network by adding jump structure between the encoder and decoder with symmetric structure and connecting them. Its advantages are that it supports a small number of training models and uses random slices to enhance the images, which greatly reduces the training time. By detecting the pixels, the segmentation is more accurate. High gradient descent is utilized in training so that the trained model is self-regulating at the learning rate. In order to compensate for the different probability of each image pixel, the model introduces weight division to make the pixel more representative. The drawback lies in the fact that incorporating effective convolution techniques poses challenges to model design in terms of complexity and generalizability. There is a loss of boundary before Feature Map merging, and the extraction of multi-scale feature information is not accurate enough.

Res-UNet makes important improvements to the U-Net model. Fig. 17 shows that the segmentation of small vessels at the end of blood vessels is more accurate. The problem of high luminance and low contrast in the optic disc region with unsatisfactory segmentation results is effectively solved. The problems of thin-walled vessels that are easy to break and unclear blood vessel boundaries caused by light reflection are improved. Non-vascular regions were less likely to be segmented as vascular regions. However, the effect of fundus image segmentation for larger lesions is quite different, and the lesion area is easy to be segmented as a blood vessel. DU-Net represents an advancement over U-Net by replacing the original convolution layer with a deformable convolution block and incorporating an up-sampling operator to enhance output resolution. The primary objective is to integrate low-level and high-level feature maps for extracting contextual information retinal segmentation. Furthermore, DU-Net demonstrates adaptability by on blood vessel size and shape, thereby effectively capturing diverse variations in retinal blood vessel morphology. In the decoding stage, another normal convolutional layer after merging was inserted to adjust the number of filters in the convolutional offset layer. As can be seen in the Table I, its overall performance has been greatly improved, whose AUC was 0.9856.

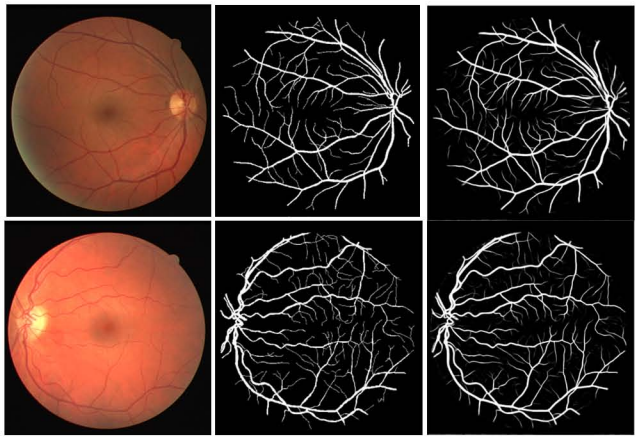
(a) Original image (b) Standard picture (c) Segmented results

Fig. 13 Segmentation results of U-Net model.



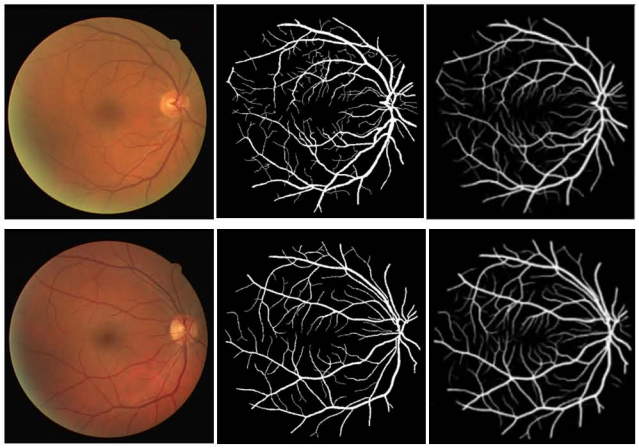
(a) Original image (b) Standard picture (c) Segmented results

Fig. 14 Segmentation results of the Res-UNet model.



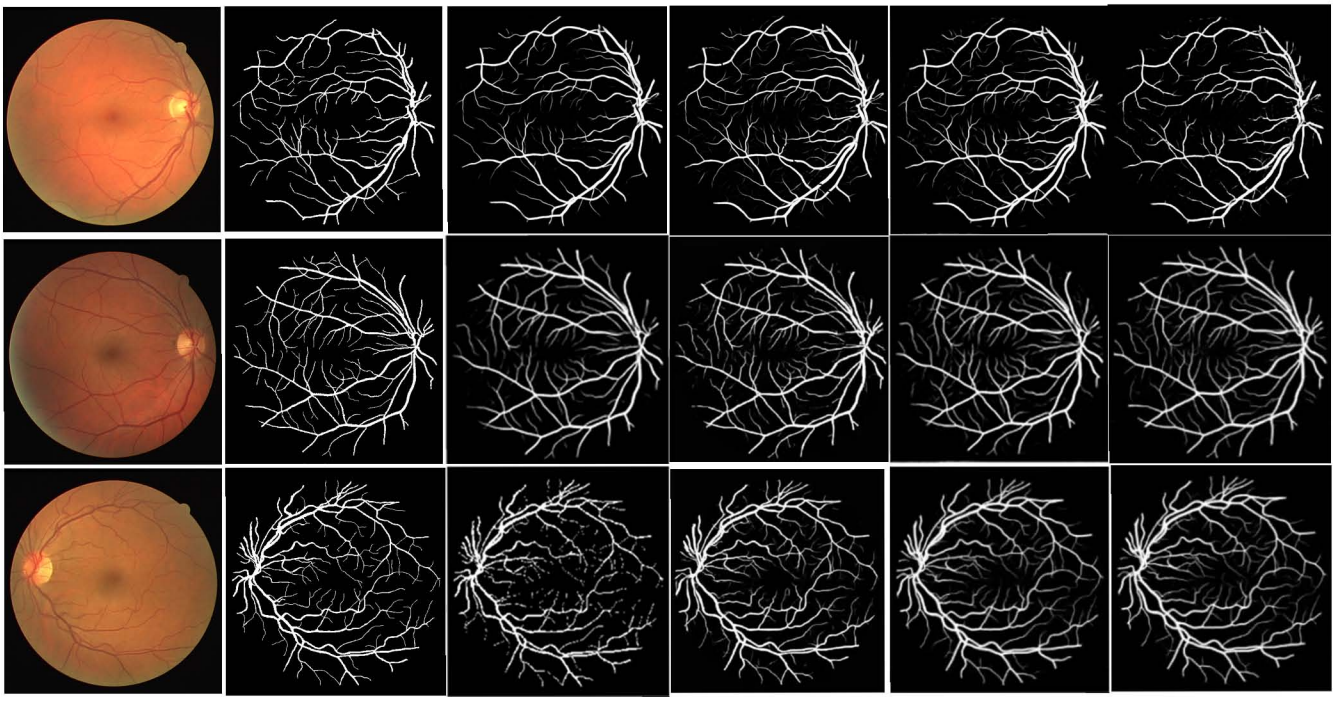
(a) Original image (b) Standard picture (c) Segmented results

Fig. 15 Segmentation results of DU-Net model.



(a) Original image (b) Standard picture (c) Segmented results

Fig. 16 Segmentation results of SA-UNet model.



(a) Original image (b) Standard picture (c) U-Net (d) Res-UNet (e) DU-Net (f) SA-UNet

Fig. 17 Comparison of DRIVE dataset segmentation.

TABLE I. PERFORMANCE COMPARATIVE RESULTS UNDER VARIOUS METHODS

Method	ACC	Sn	Sp	AUC
U-Net	0.9639	0.8237	0.9818	0.9250
Res-UNet	0.9742	0.7912	0.9892	0.9823
DU-Net	0.9697	0.8325	0.9341	0.9856
SA-UNet	0.9905	0.8573	0.9835	0.9755

For the SA-UNet model, the spatial attention module is introduced to make it a lightweight network. Data augmentation is stronger by using a small number of training samples. The module conducts an inferred attention force calculation across spatial dimensions and subsequently applies element-wise multiplication with the input feature map to achieve adaptive feature refinement. To mitigate the risk of over-fitting, we employ a structured Dropout convolution block that integrates DropBlock and Batch Normalization techniques as a rapid replacement for the original convolution operation. As shown in the segmentation results of Fig. 17, the distinction between vessel regions and background is more obvious, and the detection of thin-walled vessels and edges is greatly improved. Unnecessary features are suppressed and the expressive power of the network is improved. Among the five segmentation methods, its comprehensive performance was the best.

IV. CONCLUSION

In this paper, four segmentation methods were used to segment retinal vessels. It can be seen that each segmentation algorithm has different segmentation results, as well as suitable segmentation range. Through continuous improvement, the stability and applicability of the supervised learning method for automatic feature extraction are much stronger than those of the supervised learning method for manual feature extraction. Through the continuous strengthening of U-Net network, the algorithm

can retain the integrity of blood vessels to the greatest extent, and maintain better segmentation effect for small vessels and thin-walled vessels. Among these four methods, the SA-UNet model had the better comprehensive performance, and the specificity, accuracy, sensitivity and AUC of the segmentation results were 0.9835, 0.9905, 0.8573 and 0.9755, respectively, whose performance is excellent.

REFERENCES

[1] X. Z. Mei, H. Jiang and J. Sun, "Retinal Vessel Image Segmentation Based on Dense Attention Network," *Computer Engineering*, vol. 46, no. 3, pp. 267-272, 2020.

[2] X. Li, Y. C. Jiang, M. L. Li and S. Yin, "Lightweight Attention Convolutional Neural Network for Retinal Vessel Image Segmentation," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1958-1967, 2020.

[3] S. Chaudhuri and S. Chatterjee, "Detection of Blood Vessels in Retinal Images Using Two-Dimensional Matched Filters", *IEEE Transactions on Medical Imaging*, vol. 8, no. 3, pp. 9-26, 1989.

[4] L. Zhang, C. Wu, X. Y. Fan, C. J. Gong, S. Y. Li and H. Liu, "Retinal Vessel Segmentation via Self-Adaptive Compensation Network", *Acta Optica Sinica*, vol. 43, no.14, pp. 1-18, 2023.

[5] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional Networks for Biomedical Image Segmentation", *Computers in Biology and Medicine*, vol. 136, pp. 234-241, 2015.

[6] Iglovikov, Vladimir and A. Shvets, "TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation", arXiv:1801.05746, 2018.

[7] X. Xiao, S. Lian, Z. M. Luo and S. Z. Li, "Weighted Res-UNet for High-Quality Retina Vessel Segmentation," *In 2018 9th International Conference on Information Technology in Medicine and Education (ITME),* pp. 327-331, IEEE, 2018.

[8] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation," *Neural Networks*, vol. 121, pp. 74-87, 2020.

[9] X. B. Deng, Q. Liu, X. Chen, K. C. He, M. L. Quan and Y. L. Liu, "Segmentation of Retinal Blood Vessels Based on Channel Attention Dual Path Architecture Network," *Chinese Journal of Medical Imaging Technology*, vol. 37, no. 10, pp. 1543-1547, 2021.

[10] B. Atil and O. S. Gedik, "A Fully Convolutional Deep Learning Architecture for Retinal Blood Vessel Segmentation", *Engineering Science and Technology,* vol. 24, no. 2, pp. 271-283, 2021.

[11] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization", *Graphics Gems*, vol. 5, pp. 474-485, 1994.

[12] J. Long, E. Shelhamer and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440, 2015.

[13] J. Dai, H. Z. Qi, Y, W, Xiong, Y. Li, G. D. Zhang, H. Hu and Y. C. Wei, "Deformable Convolutional Networks," *International Conference on Computer Vision*, vol. 89, pp. 764-773, 2017.

[14] Q. G. Jin, Z. P. Meng, T. D. Pham, Q. Chen, L. Y. Wei and R. Su, "DUNet: A Deformable Network for Retinal Vessel Segmentation," *Knowledge-Based Systems*, vol. 178, pp. 149-162, 2019.

[15] C. L. Guo, M. Szemenyei, Y. G. Yi, W. Wang, B. Chen and C. Q. Fan, "SA-UNet: Spatial Attention U-Net for Retinal Vessel Segmentation," *International Conference on Pattern Recognition*, IEEE, pp. 1236-1242, 2021.

[16] M. Niemeijer, J. J. Staal, B. V. Ginneken, M. Loog and M. D. Abràmoff, "Comparative Study of Retinal Vessel Segmentation Methods on a New Publicly Available Database", *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 5370, pp. 648-656, 2004.