International Core Journal of Engineering
Volume 6 Issue 6, 2020
ISSN: 2414-1895
DOI: 10.6919/ICJE.202006_6(6).0016

Multi-path Retinal Vessel Segmentation Algorithm Based on Improved U-shaped Network

Shuguang Zhang [1, a], Shuling Chen [1, b]

[1]School of Electronics and Information Engineering, AnHui Jianzhu University, HeFei 230601, China.

[a]zsg@ahjzu.edu.cn, [b]ShuLing_C@163.com

Abstract

In view of the complexity of the fundus retinal vascular morphology and the difficulty of segmentation of microvessels, this paper proposes a multi-path dual U-shaped retinal vascular segmentation algorithm based on the U-Net model that can be adaptively adjusted based on the vascular morphology and structure. The algorithm uses jump connections to build a convolutional neural network with two symmetrical encoder-decoder structures to obtain more connection paths and fuse high-level semantic information with low-level feature information. The network coding part uses a deformable convolution structure to adaptively adjust the receptive field to capture retinal blood vessels of various shapes and sizes. At the same time, a residual module that shares parameters is proposed to make fuller use of the detailed feature information of each layer. The experimental results show that the average accuracy on the public data set of DRIVE (Digital Retial Images for Vessel Extraction) is 95.58%, the sensitivity is 78.56%, and the specificity is 98.39%, which can better segment the fundus retinal microvessels.

Keywords

Retinal blood vessel, Image segmentation, Convolutional neural network, U-shaped network.

1. Introduction

Color fundus images are a common basis for the inspection and diagnosis of ophthalmic diseases [1]. It is also the only part of the human body where blood vessels can be directly observed. By observing changes in the angle, width, and shape of retinal blood vessels, it is possible to judge vascular-related diseases, such as glaucoma, age-related macular degeneration, diabetes, hypertension, and coronary heart disease [2]. In the fundus-retinal computer-assisted diagnosis system, the veins and states of blood vessels are an important direction for assisting diagnosis. In current fundus retinal computer-aided diagnosis systems, whether based on unsupervised methods or supervised methods, they rely heavily on the gold standard features of manual markers to characterize the differences between blood vessels and backgrounds between blood vessels and background. The diagnosis segmentation algorithm is to accurately and efficiently identify the characteristics of blood vessels, and then to meet the needs of assisted diagnosis of fundus diseases.

2. Related work

The main research content of retinal fundus image segmentation includes the following five directions [3]: matching filtering direction, mathematical morphology direction, blood vessel tracking direction,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202006_6(6).0016

deformation model direction, and deep learning direction. Among them, the accuracy of blood vessel segmentation based on the deep learning direction is the highest .

Traditional deep learning methods include principal component analysis, clustering, or fusion of artificially produced features as the prior information of the data, so as to obtain separable vascular features and combine with the classifier to obtain the final segmentation result. Reference [4] uses two orthogonal line detectors with a fixed length but a change in direction to calculate the average gray value on the blood vessel segment, and sends the extracted feature vector to the support vector machine for blood vessel segmentation. Reference [5]] Using filters, morphological operators and linear detection operators to extract retinal vascular morphological features as prior information of the classifier. In [6], the image segmentation problem was transformed into a variable model of the energy minimization problem. By applying different regular terms in the level set algorithm, the adaptability to arbitrary changes in vascular morphological structure characteristics was achieved, and different energy terms were achieved. Characterize information such as different image intensities, shapes, and texture features. Reference [7] uses the vascular direction information, combined with decision tree and ensemble learning for pixel classification.

The retinal image segmentation algorithm based on deep learning is different from the traditional machine learning algorithm's dependence on the prior information of image data, and has a strong data representation ability. By training a large number of standard maps marked by experts to automatically learn retinal vascular features in different scenarios, the generalization ability is stronger. Literature [8] proposed the DCNN method, treating the fundus vessel segmentation task as a binary classification task, and considering several variants of the method, combined with structured prediction, multi-pixel classification can be achieved at the same time, although it has a good effect on the detection of microvessels , But the amount of training parameters required is too large. [9] proposed a multi-level CNN, which takes the original image and the image after maximum pooling as multi-level input, and adds a spatial-dropout layer to the multi-level network. This model has the ability to classify other types of images. Potential, but the multi-level brings about tens of millions of calculations, causing calculation difficulties. Reference [10] proposed a recursive CNN and a recursive residual CNN based on the self-encoding architecture, using the functions of the residual network and RCNN. Although it shows better performance in segmentation tasks with the same number of network parameters, small blood vessels The segmentation effect is not ideal. Reference [11] embeds dilated convolution and DenseNet into a U-shaped network, which improves the network feature reuse ability and the overall receptive field of the network, and solves the problem of insufficient microvascular segmentation. There are some problems such as microvascular segmentation and rupture of the blood vessels around the optic disc and the ends of the main vessels.

Aiming at the problems of low accuracy of retinal microvascular segmentation and excessive calculation, this paper optimizes the classic encoder-decoder UNet network [12], and proposes a convolutional neural network with a double u-shaped structure. This network It has the following advantages: First, the shared weight residual module is introduced. On the one hand, the residual learning can make full use of the feature information of each layer; on the other hand, the two convolution block weights in this module are shared, reducing training parameters and enhancing the network. Secondly, using deformable convolution [13], deformable convolution is used for tasks with geometric deformation. In the retinal vascular segmentation task, complex vascular structures can be detected to further improve the generalization ability of the model; finally It is the overall network architecture. The network in this paper is expanded from the existing U-Net network structure. The model in this paper has more information flow paths, which only adds a small amount of calculation and a small model complexity. In addition, this paper preprocesses the original image to enhance the image contrast, which further improves the accuracy of the network segmentation of retinal blood vessel images in this paper.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202006_6(6).0016

3. Network Structure

3.1 Shared Weight Residual Module

When receiving the Traditional CNN extracts image features through convolutional and pooling layers, and layer-by-layer forward and backward gradient propagation to optimize parameters. The classic ResNets [11] builds a residual block based on traditional CNN, that is, every two layers of convolution in multi-layer convolution adds a connection to form a residual block, which solves the problem of network degradation and reduces the error rate. The residual block is very easy to learn the identity function, and it will not affect the performance of the network. Therefore, this article adds the classic residual block in ResNets to the model.

The DDU-Net (Deformal Double U-net) model proposed in this paper introduces more branch structures, and the model needs more training parameters, making the network training process more difficult. To solve this problem, this paper proposes a weight sharing residual block module, as shown in Fig.1. Different from the standard residual block in the ResNets network, the residual module proposed in this paper has two convolutional layers sharing the same parameters, and a drop out layer is added between the two variable convolutional layers to avoid overfitting, effectively improve the robustness and generalization of the model.

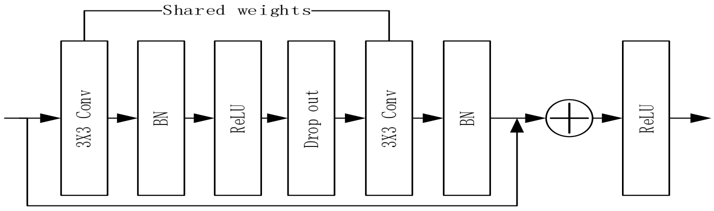

Figure 1. Schematic diagram of the shared weight residuals module

3.2 Deformable Convolution

Fundus retinal arteriovenous vessels have large differences in shape and size, and the retinal optic disc, central depression area and borders will interfere with the segmentation task. It is difficult to sample the specifications of ordinary convolution to adapt to the geometric deformation of the fundus image. In this paper, a deformable convolution method is used. This method adds an offset variable to the position of each sampling point in the convolution kernel. Through these variables, the convolution kernel can randomly sample near the current position, as shown in Fig.2.

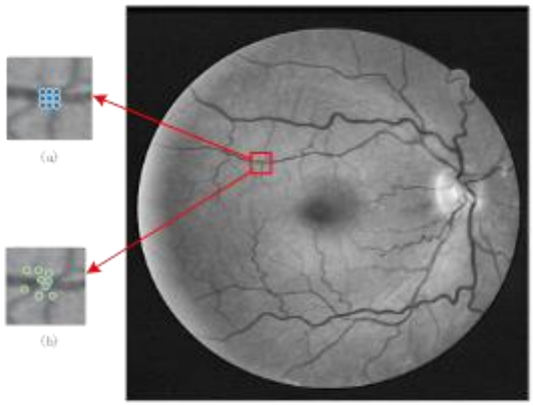

Figure 2. shows the sampling method of the ordinary convolution kernel and deformable convolution with a size of 3x3. (a) is 9 sampling points (blue) for ordinary convolution, and (b) is deformable convolution (green). After the coordinates of the sampling points are added to the displacement vector, the deformable convolution can be scaled.

Take 3x3 convolution as an example. For each output $y(p_0)$, ordinary convolution needs to sample 9 positions from x. These 9 positions are diffused from the center position (0,0) to the surrounding

ISSN: 2414-1895 DOI: 10.6919/ICJE.202006_6(6).0016

$x(p_0)$, and the upper left corner of the 1) The lower right corner of the representation, the position of other coordinate points is similar, as shown in the following formula (1).

$$\mathcal{R} = \{(-1,-1),(-1,0),\dots,(0,1),(1,1)\} \tag{1}$$

$p_0$traverses all positions in $\mathcal{R}$, as shown in the following formula

$$y(p_0) = \sum_{p_n \in \mathcal{R}} \omega(p_n) \cdot x(p_0 + p_n) \tag{2}$$

Deformable convolutions are added to the scope of ordinary convolutions, and learnable parameters ∇p_n are added to allow the sampling points to diffuse into a non-mesh shape. And set the weight of each sampling point to increase the degree of freedom, for unwanted sampling points can be set to 0. Equation 3 represents convolution deformable operation.

$$y(p_0) = \sum_{p \in \mathcal{R}} \omega(p_n) \cdot x(p_0 + p_n + \nabla p_n) \tag{3}$$

∇p_n represents the position offset of the sampling point k, which is obtained by applying different convolutional layers to the feature map of the same input. The deformable convolution layer has the same offset as the current convolution layer, and the output is $2\nabla p_n$ channel s, where ∇p_n channel is the learned offset. Because ∇p_n usually not an integer, bilinear interpolation is used to determine the offset sample value. Fig.3 specifically describes the specific process of deformable convolution.

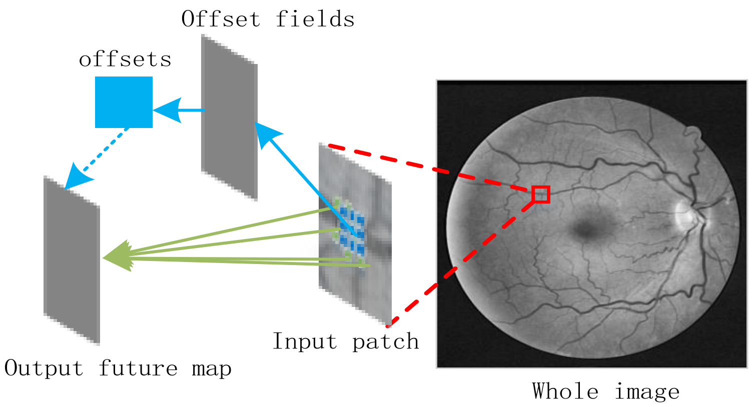

Figure 3. The specific process of deformable convolution

Offset field is learned from Input patch and features. The number of $2p_n$ channels is p_n for 2D offsets, and the convolution kernel and offSets are learned simultaneously. By learning the offset, the generalization ability of the scale transformation and scale transformation of the model can be improved. The deformable convolution structure is shown in Fig. 4:

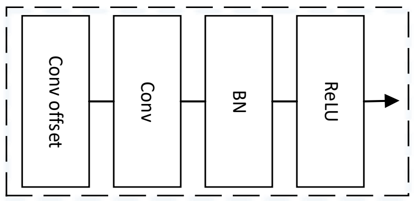

Figure 4. Deformable convoluutional block (DCB)

3.3 Multi-branch Encoder Decoder Structure

Although U-Net can achieve good segmentation results in the field of medical image segmentation, the number of U-Net information flow paths is limited. For this reason, this paper proposes a multi-branch encoder-decoder deformable volume based on U-Net. Product neural network (DDU-Net). Fig. 5 shows the structure of the DDU-Net network. The shape can be seen as the superposition of two U-shaped networks. The contraction path and expansion path of each U-shaped network include three structural blocks and are symmetrical to each other. The two U-shaped networks are symmetrical to each other. The downsampling uses a deformable convolution with a step size of 2 and a shared weight residual block, and uses a transposed convolution with a step size of 2 and a

ISSN: 2414-1895
DOI: 10.6919/ICJE.202006_6(6).0016

shared weight residual block. The last layer of the network maps multi-channel features to a feature map with n channels. In this paper, n = 2 represents vascular pixels and non-vascular pixels. Finally, the resulting feature map is put into the Softmax classifier $C_{softmax}$, and the probability value of each pixel is output. The posterior probability of the m-th pixel O_m output is expressed as follows:

$$P(O_m) = C_{softmax}(a_m^{out}) = \frac{\exp(a_m^{out})}{\sum_i \exp(a_j^{out})} \tag{4}$$

a_m^{out} represents the m-th activation output, a_j^{out} represents the j-th activation output.

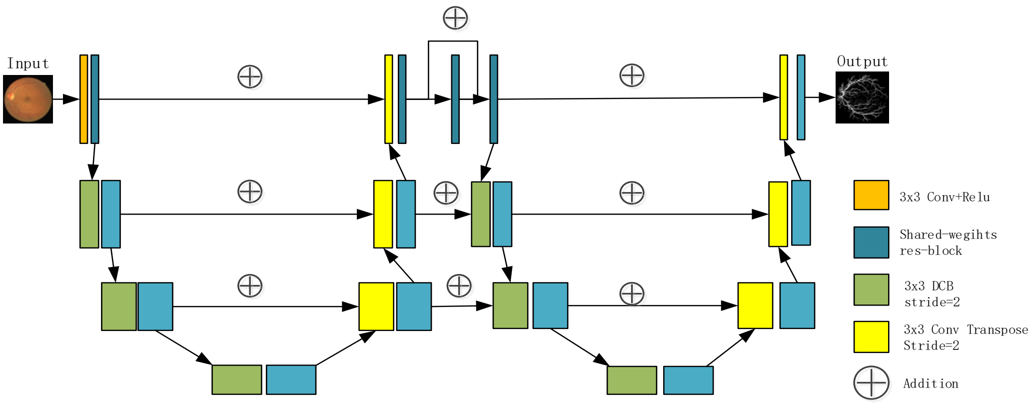

Figure 5. DDU-Net Network Structure,

4. Experimental Results and Analysis

The hardware configuration used in this experiment is Intel Core i7-8700K, GTX 1070Ti, 8G video memory. On Ubuntu 16.04 system, the Pytorch machine learning library is used to train and test the network model in this paper.

4.1 Image Preprocessing and Image Enhancement

This paper uses the public fundus image segmentation dataset DRIVE [12] to experiment the segmentation performance of DDU-Net network. The DRIVE data set contains 40 fundus images, 20 for the training set, and the remaining 20 for the test set. Each image consists of 565x584 pixels, corresponding to the segmented and mask images manually labeled by experts. The fundus retina color image consists of three channels: red, green, and blue. The contrast of the green channel image is high, and the structure of blood vessels can be clearly observed. Therefore, this paper uses the green channel of the fundus image for subsequent blood vessel segmentation. In order to further improve the effect of image segmentation, the contrast-limited histogram equalization method (CLAHE) is used to solve the problem of uneven illumination caused by fundus image collection. The Gamma correction method is used as follows:

$$V_{out} = AV_{in}^{\gamma} \tag{5}$$

V_n represents the input image,A is a constant coefficient, γ is an adjustable parameter,V_{out} is the output image, In order to better observe the details at low resolution of the image, Set γ value to 1.2, Further enhance the contrast of the image. The pre-processed image is shown in Fig.6:

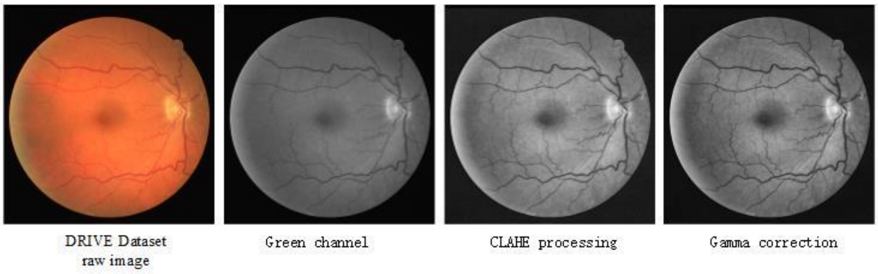

Figure 6. Fundus image preprocessing,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202006_6(6).0016

Because the selected DRIVE data set is small and easily overfitting, this paper randomly selects 190,000 48x48 pixel size patches from the training set as a new training set Fig.7 shows 48x48 size patches and corresponding ground truth blocks. To find the best model, 10% of the new training set is selected as the validation set, which is used to adjust the parameters of the model. Finally, use the test set to evaluate the performance of the network.

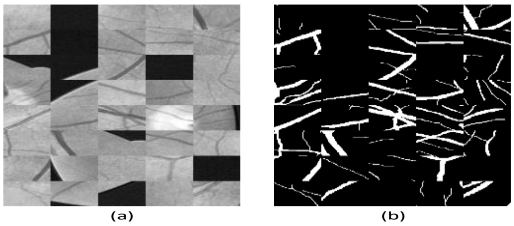

Figure 7. 48x48patches extracted from the training set

4.2 Model Training and Parameter Settings

In this paper, the DDU-Net model is trained to predict, on a pixel-by-pixel basis, whether each pixel is foreground (vascular) or background (non-vascular). The common loss function for general classification tasks is the cross-entropy loss function, but for fundus images, objects such as the optic disc and retinal blood vessels will only occupy a small part of the entire Fig., using the cross-entropy loss function is not the best. This paper uses the Dice coefficient loss function [13, 14] to replace the common cross-entropy loss function. The Dice coefficient is a set of similarity metric functions. It is often used in the field of medical image segmentation to evaluate the quality of segmentation results. The specific definition is as follows:

$$\mathrm{L_{loss}} = L_{dice} = 1 - \sum_k^K 2\omega \sum_i^N \frac{P(k,i)g(k,i)}{(\sum_i^N p^2_{(k,i)} + \sum_i^N g^2_{(k,i)})} \tag{6}$$

N is the number of pixels, k is the category number, $\mathrm{p(k,i)} \in [0,1]$ and $\mathrm{g(k,i)} \in \{0,1\}$ respectively represent the predicted probability and the category represented by the ground truth label, $\sum_{\mathrm{k}} \omega_k$ are the category weights. In addition, this paper uses a commonly used optimization algorithm, the stochastic gradient descent method, to optimize the parameters in the network. The initial weight parameters are initialized using a random parameter initialization method that conforms to the normal distribution. The initial learning rate is set to 0.20, the drop out parameter is set to 0.25, and the training batch-size to 32 and epoch to 150.

4.3 Performance Evaluation Standards

In order to quantitatively compare the results of blood vessel segmentation and the results of ground truth manually labeled by experts, this paper uses three general evaluation indicators, including accuracy (Ac), sensitivity (Sn), and specificity (Sp).

$$\mathrm{Ac} = \frac{\mathrm{TP+TN}}{TP+TN+FP+FN} \tag{7}$$

$$\mathrm{S_n} = \frac{TP}{TP+FN} \tag{8}$$

$$\mathrm{S_p} = \frac{TN}{TN+FP} \tag{9}$$

In the formula TF(true positive), the number of correctly segmented blood vessel pixels, TN(true negative), the number of correctly segmented background pixels, FP(false negative), the number of incorrectly segmented background pixels, and (false positive), indicating Number of wrongly segmented blood vessel pixels. In addition, in order to evaluate the performance of different neural networks, the receiver operating cha2.4 Results analysis

This paper uses four performance evaluation criteria to evaluate the performance of the model. Fig. 7 shows the effect of retinal blood vessel segmentation on the DRIVE test data set. Fig.8 (a) is the original pre-processed image; Fig.8 (b) is the ground truth manually labeled by experts; By comparing the segmented image of model training with the ground truth labeled by the experts, we can observe that DDU-Net has a good segmentation effect.racteristic (ROC) curve is used to measure the accuracy

International Core Journal of Engineering
Volume 6 Issue 6, 2020
ISSN: 2414-1895
DOI: 10.6919/ICJE.202006_6(6).0016

of retinal vascular segmentation, and the algorithm's segmentation effect is judged by calculating the sum of the area under the ROC curve (AUC).

(a)

(b)

(c)

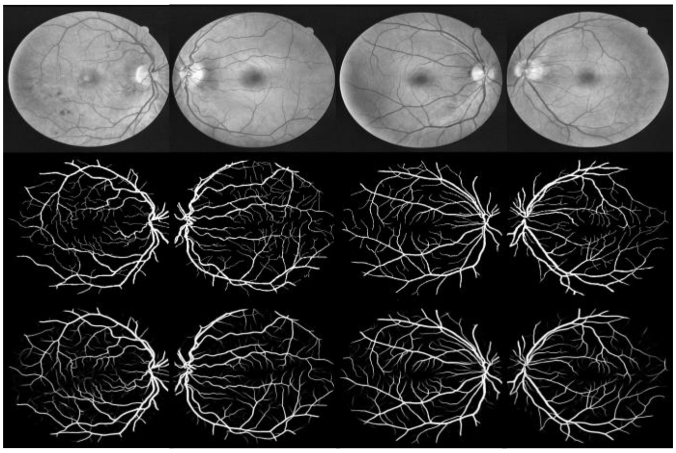

Figure 8. DDU-Net network segmentation effect on the DRIVE test data set

The ROC curve of the algorithm proposed in this paper on the DRIVE database is shown in Fig.9. The AUC value is 97.93%.

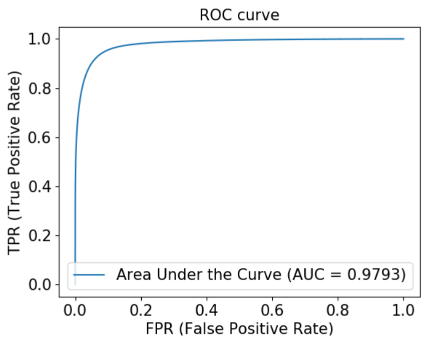

Figure 9. ROC curve of the method in this paper

In order to better highlight the vascular segmentation performance of the algorithm model proposed in this paper, Table 1 gives the accuracy, sensitivity, specificity, and AUC values of different deep learning fundus vascular segmentation algorithms on the DRIVE database. Are the best value.

Table 1. Retinal vessel segmentation results on the DRIVE dataset

Method		Ac	Sn	Sp	AUC
Ref.[8]		0.9535	0.7811	0.9807	0.9765
Ref.[9]		0.9533	0.7464	0.9836	0.9752
Ref.[10]		0.9556	0.7751	0.9816	0.9782
Our method		0.9559	0.7854	0.9810	0.9793

As can be seen from Table 2, the accuracy, sensitivity, specificity, and AUC value of the proposed algorithm on the DRIVE database are higher than the above algorithms, which shows that the algorithm in this paper can correctly segment vascular pixels and non-vascular pixels. In this paper, the training time of the DRIVE database is about 7.5 hours, and the test is about 9 seconds, which can reach the level of computer-aided medical diagnosis.

5. Conclusion

Correct segmentation of retinal blood vessels of the fundus is of great significance to assist doctors in diagnosis. Aiming at the difficulty of subtle blood vessel segmentation in fundus images, this paper

ISSN: 2414-1895 DOI: 10.6919/ICJE.202006_6(6).0016

proposes a dual U-shaped network structure model based on the classic UNet network, which fuses shared residuals and deformable convolutions. First, the green channel color information is used in the pre-processing stage, and the image is further processed using CLAHE processing and GAMMA transformation, and the image is enhanced by operations such as rotation and flip. The pre-processed image is then sent to the dual U-shaped network structure designed in this paper. The U-shaped structure can extract the local features of the image in an end-to-end manner. Enhance the robustness of the model. The model designed in this paper is suitable for small-scale data sets and is used for fundus retinal blood vessel segmentation, which can accurately segment vascular and non-vascular pixels. In addition, the model is also applicable to the segmentation of medical image data sets.

Acknowledgments

CERNET Innovation Project (NGII20190602).

References

[1] Abràmoff M D, Garvin M K, Sonka M. Retinal Imaging and Image Analysis[J]. IEEE Rev Biomed Eng, 2010, 3:169-208.

[2] Kirbas C , Quek F . A review of vessel extraction techniques and algorithms[J]. ACM Computing Surveys, 2004, 36(2):81-121.

[3] Chengyi Zhu, Beiyi Zou, Yao Xiang, et al. Research progress of retinal vessel segmentation in color fundus images [J]. Journal of Computer-Aided Design & Computer Graphics, 2015, 27(11):2046-2057.

[4] Ricci E , Perfetti R . Retinal Blood Vessel Segmentation Using Line Operators and Support Vector Classification[J]. IEEE Transactions on Medical Imaging, 2007, 26(10):1357-1365.

[5] C. Zhu, J. Cui, B. Zou. Retinal Vessel Segmentation Based on Multiple Feature Fusion and Random Forest[J]. Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao/Journal of Computer-Aided Design and Computer Graphics, 2017, 29(4):584-592.

[6] Mikael Rousson, Nikos Paragios. Prior Knowledge, Level Set Representations & Visual Grouping[J]. International Journal of Computer Vision, 76(3):231-243.

[7] An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation[J]. IEEE Transactions on Biomedical Engineering, 2012, 59(9):2538-2548.

[8] Liskowski P , Krawiec K . Segmenting Retinal Blood Vessels with Deep Neural Networks[J]. IEEE Transactions on Medical Imaging, 2016:1-1.

[9] Ngo L , Han J H . Multi-level deep neural network for efficient segmentation of blood vessels in fundus images[J]. Electronics Letters, 2017, 53(16):1096-1098.

[10] Alom M Z , Hasan M , Yakopcic C , et al. Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation[J]. 2018.

[11] Liming Liang, Xiaoqi Sheng, Kai Guo, et al. Improved U-Net Fundus Retinal Vessel Segmentation [J / OL], Application Research of Computers,2019,37（4）[2019-02-16]，http://doi.org/10.19734/j,issn.1001-3695.2018.09.0775

[12] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [C]// International Conference on Medical Image Computing & Computer-assisted Intervention. 2015.

[13] Dai J, Qi H, Xiong Y, et al. Deformable Convolutional Networks[J]. 2017.

[14] He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[J]. 2015.

[15] Staal J , Abramoff, M.D, Niemeijer M , et al. Ridge-based vessel segmentation in color images of the retina[J]. IEEE Transactions on Medical Imaging, 2004, 23(4):501-509.

[16] Laina I, Rupprecht C, Belagiannis V, et al. Deeper Depth Prediction with Fully Convolutional Residual Networks[C]// 2016.

[17] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in 3D Vision (3DV), 2016 Fourth International Conference on. IEEE, 2016, pp. 565–571.