An improved U-Net method with High-resolution Feature Maps for Retinal Blood Vessel Segmentation

Chang-lei Dongye*1, Yanyan Ma2
1 Computer Department, Shandong University of Science and Technology, Qingdao, China
2 Computer Department, Shandong University of Science and Technology, Qingdao, China
*Corresponding author’s e-mail:dycl@sdust.edu.cn

Abstract. The accurate segmentation of retinal blood vessels is of great significance for the diagnosis of diseases such as diabetes, hypertension, microaneurysms and arteriosclerosis. However, manual segmentation of retinal blood vessels is time-consuming and laborious. This paper proposes a convolutional network structure based on U-Net for retinal vessel segmentation. First, a new convolution block, which makes full use of shallow high-resolution feature maps to minimize the information loss caused by downsampling, is added to the network. Second, the network was downscaled. Particularly, this network application conducts downsampling twice to reduce the complexity of the network and the number of parameters during training. In addition, we retain the original short connection, which merges the feature information of the shallow and deep networks. Therefore, this network can capture the details of blood vessels more effectively. We tested the work on the DRIVE data set and evaluated the accuracy, sensitivity, specificity and AUC, which were 0.9552, 0.7603, 0.9839, and 0.9789, respectively. A comprehensive comparison between the proposed algorithm and the existing algorithms shows that the various indicators of the algorithm perform well.

1. Introduction
Eyes are the human body's visual organs. They are located in the orbital cavities and have a complex structure[1]. The retinal blood vessels in the eyes are the only deep vascular system in the human body that can be directly observed without damage. Chronic and easily neglected diseases in the human body, such as hypertension, chronic kidney disease, diabetes, and atherosclerosis, are closely related to the structure of retinal blood vessels[2]. Therefore, the analysis and accurate segmentation of retinal images play vital roles in the early screening and prevention of vascular diseases. However, the segmentation of retinal blood vessels needs to be performed by professionals, which not only requires high professional skills, but it also consumes time and energy. Besides, it is easy for the process to be affected by the subjective factors of professionals; thus, different effects can be obtained for the same image. Therefore, with the rapid development of computer technology, deep learning has been applied to the segmentation of retinal blood vessels to assist medical diagnosis and treatment and prevent the occurrence of eye diseases.

In recent years, many researchers have made outstanding contributions to retinal segmentation. Their work can be divided into two categories: unsupervised learning and supervised learning.

Regarding the unsupervised learning method, A.K. Shukla et al. [3] propose a fractional filter-based efficient algorithm for retinal blood vessel segmentation. This method has high computational efficiency...
but relatively low sensitivity, and it can make some erroneous detections in pathological images. W. Wang et al. [4] proposed a retinal vascular automatic unsupervised segmentation method based on modified morphological transformation and fractal dimension, which has strong robustness. E. Emary et al. [5] use the fuzzy C-means objective function, which has accurate anti-noise and pathological performance. L. M. Liang et al. [6] use the level set blood vessel segmentation method incorporating the fusion region energy fitting information and shape prior to constructing a retinal blood vessel segmentation model. X. Wang et al. [7] proposed a new active contour model combining local and global information that can detect blood vessels in retinal angiography. However, unsupervised methods lack supervision information, and most of them rely on hand-crafted features to identify blood vessels, resulting in problems such as the insufficient segmentation of microvessels and the missegmentation of lesion areas.

The supervised learning method mainly achieves blood vessel segmentation through two stages of feature extraction and pixel segmentation. Q. Li et al. [8] use a cross-modal learning method that reconstructs the segmentation task into a cross-modal data conversion problem. The network can automatically learn blood vessel features during the training process. R. Biyani et al. [9] proposed a blood vessel segmentation algorithm based on an improved fuzzy min-max neural network supervised classifier. The algorithm is essentially a super classification method that achieves high performance. C. Tian et al. [10] proposed a retinal blood vessel segmentation method based on a multipath convolutional neural network that can effectively suppress noise and ensure the continuity of blood vessel segmentation. W. Xiancheng et al. [11] propose a method based on the U-Net neural network segmentation method that combines the features extracted by the shallow network and the features extracted by the deep network to achieve a good segmentation effect. The supervised learning method trains a model by examining the mapping function between the input and the output and the standard image segmented manually by the expert. However, some currently existing supervised learning methods still cannot identify small blood vessels well, and some networks have too many output parameters and overfitting problems. In order to solve this problem, this article incorporated the U-Net network, added a new convolution block on the basis of the original model, and built an end-to-end network model suitable for retinal vessel segmentation. The experiment proved that all performance indicators performed well.

2. Related technologies

2.1. Full convolutional network

A deep convolutional neural network is divided into two parts. The first part is composed of a series of convolutional layers and pooling layers, and the latter part is added the fully connected layer. This requires the input image size to be fixed. However, there is no fully connected layer in the fully convolutional network [12], so the input image does not need to have a fixed dimension. In addition, after the last convolution pooling operation, the feature map is upsampled to restore the size of the input image so that each pixel in the image can be predicted.

2.2. U-Net architecture

Net [13] is an encoder-decoder architecture. The encoder corresponds to the downsampling part, which is composed of a series of convolution and pooling layers. It can extract the characteristics of the input image and capture the global information of the image. The decoder, which is composed of several convolution and upsampling layers, corresponds to the upsampling part and can expand the feature map to every pixel. In addition, U-Net combines the effective information of the shallow layer and deep layer networks with short connections. The function of the shallow layer neural network is to improve the training accuracy and resolution of the feature map while the deep layer neural network can extract the complex features of the image. The above networks can be combined to obtain more segmentation details. The output layer in the network uses the softmax activation function to classify each pixel in the final feature map. The U-Net architecture is shown in figure1.
3. Research methods

3.1. Image preprocessing

The collection of images is usually affected by external factors, such as different light intensities, shooting angles and collection equipment, which can easily cause image noise, resulting in a low contrast for blood vessels and a fundus background. In order to reduce the noise in the image and improve the contrast, the image needs to be preprocessed. In this experiment, the following treatments were mainly performed, and the comparison of the effects is shown in figure 2. As figure 2 shows, the contrast between the blood vessels in the image and the background after processing is significantly improved, which plays a very important role in the accurate segmentation of blood vessels.

- The color fundus image consists of three channels: red, green and blue. The overall contrast of the green channel is relatively high, that is, the blood vessels are relatively obvious. Therefore, the green channel image serves as the basis for the subsequent segmentation of blood vessels.
- The fundus image is standardized. The method achieves centralization by removing the mean and obtains data with a mean value of 0 and a variance of 1 that obey a standard normal distribution.
- Contrast-limited adaptive histogram equalization (CLAHE) is used, and contrast limiting must be used for each small area of the fundus image. The contrast between the retinal blood vessels and the background is enhanced without amplifying the noise so that the structure and characteristics of the blood vessels in the fundus image can be better seen.
- Gamma correction of the image [14] can enhance the darker blood vessel parts in the fundus image while having no effect on the brighter part and smoothly expand the details of the darker part of the image.
3.2. Data enhancement

The DRIVE (digital retinal images for vessel extraction) [15] data set was used in this experiment. This data set has only 20 training set images and 20 test set images with a resolution of 565 × 584, and each image corresponds to the results of manual segmentation by professionals. Owing to the small number of images in this data set, it is necessary to enhance the data set and expand the data set.

In this experiment, each image is divided into 10,000 patches, and the resolution of each data block is 32 × 32. Therefore, the data set can be amplified into 200000 patches. The first 90% of the patches are used as the training set, and the remaining 10% of the patches are used as the validation set.

4. Method and architecture

An improved architecture based on U-Net is proposed in this paper. The architecture is composed of an input layer, a hidden layer and an output layer, as shown in Figure 3. The hidden layer is divided into two parts, downsampling and upsampling, but a new convolution block is added to the upsampling part. This architecture has the following advantages:

- In the original U-Net architecture, there are many upsampling and downsampling operations, which can easily cause overfitting problems. The improved network structure in the paper only uses upsampling and downsampling twice, thus retaining the short connection between the contraction path and the expansion path. Since we divide the original image into several image blocks, if more upsampling and downsampling operations are used, the position information and detailed information of the blood vessels will be seriously lost.

- In order to make full use of the low complexity features extracted by the shallow neural network, this algorithm performs upsampling after the first downsampling to obtain a high-resolution feature map. A short connection is made between the high-resolution feature map and the feature map obtained after the first convolution. After a series of convolutions and the last upsampled feature map, a short link is made. By adding high-resolution feature maps to the network, the information loss caused by downsampling is reduced. Combined with the original short link, it can more effectively capture the details and position information of blood vessels and improve the segmentation efficiency of blood vessels.

Figure 3. Network structure (Above each layer of the convolution block is the number of channels of the network. The resolution of each layer of the input image is at the lower left of the convolution block.)
The downsampling part is composed of alternating convolutional layers, dropout layers, and pooling layers. During the training process, the path can be contracted to extract the global information of the image. The upsampling part consists of upsampling, convolutional layers, and Batch Normalization (BN) layers alternately, and this part can expand the path and extract the information of each pixel of the image. The pooling layer adopts maximum pooling, and the image block becomes \( \frac{1}{4} \) of the original after passing through the pooling layer. The model selected is the most common ReLU activation function [16].

5. Experimental results and analysis

5.1. Performance evaluation indexes

In the experiment, in order to better judge the segmentation effect of the model, it is necessary to compare the segmentation result with the ground truths manually marked by experts. This experiment uses a confusion matrix to represent the pixel segmentation results, as shown in Table 1.

| GT vascular Points | GT background points |
|--------------------|----------------------|
| Predict vascular points | True positive (TP) | False positive (FP) |
| Predict background points | True negative (TN) | False negative (FN) |

In Table 1, true positives (TP) are the blood vessel points are correctly classified in the predicted image. False positives (FP) are the background points that are falsely classified as blood vessel points. True negatives (TN) are the blood vessel points that are classified as background points. False negatives (FN) are the background points in the predicted image that are correctly segmented [17].

The basic statistical results in Table 1 can be extended to multiple evaluation indicators: Accuracy, Sensitivity, and Specificity, which are defined in Table 2.

| Index | Formula | Meaning |
|-------|---------|---------|
| Acc   | \( \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \) | The proportion of correct judgment results with respect to the total observations |
| Sen   | \( \text{Sensitivity} = \frac{TP}{TP + FN} \) | It measures the proportion of actual positives that are correctly classified as such |
| Spe   | \( \text{Specificity} = \frac{TN}{TN + FP} \) | It measures the proportion of actual negatives that are correctly identified as such |

The ROC (receiver operating characteristic) curve is an important curve used to evaluate the quality of a classifier. It uses the FPR (False Positive Rate) as the abscissa. In addition, its uses the TPR (True Positive Rate) as the ordinate. The AUC value is the area under the ROC curve. The closer the AUC value is to 1, the better the segmentation performance of the model.

5.2. Analysis of the experimental results

The experimental hardware environment in this article includes an Intel(R) Xeon(R) CPU E5-2690 v4 @ 2.60 GHz, a GTX1080 GPU used to accelerate image processing, and 16 GB of memory. The convolutional neural network is implemented under the open source framework of TensorFlow 1.14.0.

In the training phase, a stochastic gradient descent optimizer (SGD) is used. The initial learning rate is set to 0.01, the epoch size is set to 32, and the N_epochs is set to 100. In addition, the number of learning parameters of this network is 0.65 M while the number of learning parameters of the classic U-
Net is 32.9 M. In addition, the corresponding running time (the time required for one iteration) has decreased to a quarter of the original running time. In the network output, first, the $1 \times 1$ convolution kernel is used to obtain the feature map. Then, the softmax activation function is used to predict each pixel and obtain the probability of whether the point is a blood vessel point or a background point. Finally, an appropriate threshold of 0.5 is set. If the probability is greater than 0.5, the point is classified a blood vessel point; otherwise, it is classified as a background point.

Figure 4 shows the segmentation results at the end of the blood vessel with low contrast. It is difficult to identify these microvessels at the intersection. The results show that the method used in this experiment is not affected by factors such as low contrast and changes in the blood vessel shape. Under the influence of various unfavorable factors, it still shows good segmentation performance and has good robustness.

![Original image Local Original image GT image Our method](image)

5.3. Comparsion with existing methods

In order to analyze the effectiveness of the method adopted in this paper, the quantitative results of this experiment were compared against those of existing methods, as shown in Table 3.

| Method            | Year | Sen   | Spe   | Acc   | AUC   |
|-------------------|------|-------|-------|-------|-------|
| Q. Li[8]          | 2016 | 0.7569| 0.9816| 0.9527| 0.9738|
| C. Zhu[18]        | 2017 | 0.7447| 0.9838| 0.9606| 0.9632|
| M. J. Garcia[21]  | 2018 | 0.7300| 0.9200| 0.9600| 0.9720|
| D.Cortinovis[11]  | 2018 | 0.7537| 0.9843| 0.9549| 0.9777|
| C. Wang[19]       | 2019 | 0.7986| 0.9736| 0.9511| 0.9740|
| Z.Tingyue[20]     | 2019 | 0.8053| 0.9767| 0.9546| 0.9771|
| Proposed method   | 2020 | 0.7603| 0.9839| 0.9552| 0.9789|

The tables shows that the AUC values of the existing algorithms for vascular segmentation on the DRIVE data set are all lower than those in this paper, indicating that this algorithm has strong robustness on this data set. Z.Tingyue et al.[20] provide a blood vessel segmentation method based on a fully convolutional network. Its sensitivity is 0.045 higher than that of the algorithm in this paper, but the algorithm in this paper has obvious advantages in other respects. W.Xiancheng [11] uses the traditional U-Net neural network with two upsampling operations and two downsampling operations, and its
specificity is approximately 0.04% better than that of our method. Its advantage is almost negligible. Our method provides great improvements, which illustrate the importance of using high-resolution feature maps after adding convolution blocks. C.Zhu et al.[18] provide a retinal segmentation method based on the combination of multifeature fusion and a random forest. Its accuracy is 0.0054 higher than that of the proposed algorithm, and its advantage is almost negligible. The specificity is also the same as that of the algorithm in this paper, but the AUC of the algorithm in this paper is higher than that of C.Zhu et al.[18]. These results show that the classification ability of the proposed method is stronger. In conclusion, the proposed method constructs an end-to-end network using the two processes of feature extraction and pixel classification in supervised learning. The algorithm reduces the number of training parameters, improves the training effect, and accurately segments retinal blood vessels to reduce the omission and misjudgment of blood vessel pixels. Therefore, it can be comprehensively judged that the method presented in this paper has the better performance in all aspects.

In order to intuitively compare the segmentation performance of retinal vessels, the results of this experiment are compared with those of the three network models of the original U-Net, the dense U-Net[22] and N4-Fields[23], as shown in Figure 5. Dense U-Net adds a dense mechanism on the basis of the U-Net network. N4-Fields is an image transformation method using the nearest neighbors of a neural network. The segmentation details are enlarged and compared, and it is concluded that the model proposed in this paper can accurately segment main blood vessels and small blood vessels. Meanwhile, the continuity of the segmented blood vessels is relatively good. This demonstrates that the proposed method has significantly improved the segmentation effect because of adding the convolution block.

![Figure 5. The segmentation results of different models.](image-url)
6. Conclusion
The paper proposes an improved convolutional network structure based on U-Net network, with application in retinal vessel segmentation. The proposed structure adds new convolutional blocks to make use of shallow high-resolution feature maps. It reduces the information loss caused by downsampling. Moreover, the algorithm in this paper keeps the original short connection and more effectively captures the details of blood vessels. Finally, the proposed algorithm in this paper only uses two downsampling, which reduces the network complexity and improves the training performance.

We compare our model with the U-Net model and other representative methods on the widely used DRIVE benchmark data set. The experimental results show that our method has better effects in terms of the specificity and AUC. In the future, we plan to focus on the segmentation of blood vessels in lesion areas. It is a challenging task to accurately segment tiny and indistinct vessels around lesions because of the abnormal and noisy regions in lesion areas. In addition, we plan to improve the robustness of the model by applying the model to other data sets. We expect larger datasets and more accurate manual annotation to further improve the performance of our method.

References
[1] C. Oyster, “The human eye: Structure and function, chapter 15,” Sinauer Associates, Inc. Sunderland, Massachusetts, vol. 327, pp. 343–348, 1999.
[2] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, “Blood vessel segmentation methodologies in retinal images—a survey,” Computer methods and programs in biomedicine, vol. 108, no. 1, pp. 407–433, 2012.
[3] A. K. Shukla, R. K. Pandey, and R. B. Pachori, “A fractional filter based efficient algorithm for retinal blood vessel segmentation,” Biomedical Signal Processing and Control, vol. 59, p. 101883, 2020.
[4] W. Wang, W. Wang, and Z. Hu, “Retinal vessel segmentation approach based on corrected morphological transformation and fractal dimension,” IET Image Processing, vol. 13, no. 13, pp. 2538–2547, 2019.
[5] E. Emary, H. M. Zawbaa, A. E. Hassanien, G. Schaefer, and A. T. Azar, “Retinal vessel segmentation based on possibilistic fuzzy c-means clustering optimised with cuckoo search,” in 2014 International Joint Conference on Neural Networks (IJCNN). IEEE, 2014, pp. 1792–1796.
[6] L.-M. Liang, C.-L. Huang, F. Shi, J. Wu, H.-J. Jiang, and X. Chen, “Retinal vessel segmentation using level set combined with shape priori,” Chin. J. Comput, vol. 41, no. 7, pp. 1678–1692, 2018.
[7] X. Wang, Z. Jiang, W. Li, R. Zarei, G. Huang, A. Ulhaq, X. Yin, B. Zhang, P. Shi, M. Guo et al., “Active contours with local and global energy basedon fuzzy clustering and maximum a posteriori probability for retinal vessel detection,” Concurrency and Computation: Practice and Experience, vol. 32, no. 7, p. e5599, 2020.
[8] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, “A crossmodality learning approach for vessel segmentation in retinal images,” IEEE transactions on medical imaging, vol. 35, no. 1, pp. 109–118, 2015.
[9] R. Biyani, B. M. Patre, and U. V. Kulkarni, “Retinal vessel segmentation using enhanced fuzzy min-max neural network,” Multimedia Tools and Applications, vol. 78, no. 24, pp. 35 053–35 073, 2019.
[10] C. Tian, T. Fang, Y. Fan, and W. Wu, “Multi-path convolutional neural network in fundus segmentation of blood vessels,” Biocybernetics and Biomedical Engineering, 2020.
[11] W. Xiancheng, L. Wei, M. Bingyi, J. He, Z. Jiang, W. Xu, Z. Ji, G. Hong, and S. Zhaomeng, “Retina blood vessel segmentation using a u-net based convolutional neural network,” in Procedia Computer Science: International Conference on Data Science (ICDS 2018), Beijing, China, 2018, pp. 8–9.
[12] E. Shelhamer, J. Long, and T. Darrell, “Fully convolutional networks for semantic segmentation,”
IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 4, pp. 640–651, 2017.

[13] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[14] S. Rahman, M. M. Rahman, M. Abdullah-Al-Wadud, G. D. Al-Quaderi, and M. Shoyaib, “An adaptive gamma correction for image enhancement,” EURASIP Journal on Image and Video Processing, vol. 2016, no. 1, pp. 1–13, 2016.

[15] J. Staal, M. D. Abrámoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” IEEE transactions on medical imaging, vol. 23, no. 4, pp. 501–509, 2004.

[16] J. Kim, S. Kim, and M. Lee, “Convolutional neural network with biologically inspired on/off relu,” in International Conference on Neural Information Processing. Springer, 2015, pp. 316–323.

[17] C. Zhu, B. Zou, X. Yao, Q. Yan, Y. Liang, and J. Cui, “A survey of retinal vessel segmentation in fundus images,” J Comput Aided Des Comput Graph, vol. 27, no. 11, pp. 2046–57, 2015.

[18] C. Zhu, J. Cui, B. Zou, Y. Chen, and J. Wang, “Retinal vessel segmentation based on multiple feature fusion and random forest,” Journal of Computer Aided Design & Computer Graphics, vol. 29, no. 4, pp. 584–592, 2017.

[19] C. Wang, Z. Zhao, Q. Ren, Y. Xu, and Y. Yu, “Dense u-net based on patchbased learning for retinal vessel segmentation,” Entropy, vol. 21, no. 2, p. 168, 2019.

[20] Z. Tingyue, T. Chen, and L. Zhenkun, “Multi-scale retinal vessel segmentation based on fully convolutional neural network,” Acta Optica Sinica, vol. 39, no. 2, p. 0211002, 2019.

[21] M. J. García-Tarifa, F. J. Martínez-Murcia, and J. M. Górriz, “Retinal blood vessel segmentation by multichannel deep convolutional autoencoder,” in International Joint Conference SOCO’18-CISIS’18-ICEUTE’18: San Sebastián, Spain, June 6-8, 2018 Proceedings, vol. 771. Springer, 2018, p. 37.

[22] Y. Ganin and V. Lempitsky, “N4-fifields: Neural network nearest neighbor fifields for image transforms,” in Asian Conference on Computer Vision. Springer, 2014, pp. 536–551.

[23] Y. Zeng, X. Chen, Y. Zhang, L. Bai, and J. Han, “Dense-u-net: densely connected convolutional network for semantic segmentation with a small number of samples,” in Tenth International Conference on Graphics and Image Processing (ICGIP 2018), vol. 11069. International Society for Optics and Photonics, 2019, p. 110692B.