An attention directed generative adversarial network for retinal vessel segmentation

Zhenxiang He, Nianzu Lv, Wei Sun

Zhenxiang He, Nianzu Lv, Wei Sun, "An attention directed generative adversarial network for retinal vessel segmentation," Proc. SPIE 12506, Third International Conference on Computer Science and Communication Technology (ICCSCT 2022), 125063B (28 December 2022); doi: 10.1117/12.2661840

SPIE.

Event: International Conference on Computer Science and Communication Technology (ICCSCT 2022), 2022, Beijing, China

An attention directed generative adversarial network for retinal vessel segmentation

Zhenxiang He[a], Nianzu Lv[*b], Wei Sun[b]
[a]School of Intelligent Science and Technology, Tianfu College of Southwest University of Finance and Economics, Mianyang 621000, Sichuan, China; [b]School of Information Engineering, Xinjiang Institute of Technology, Aksu 13558, Xinjiang, China

ABSTRACT

Observing the structure of retinal blood vessels can help doctors diagnose the disease of patients, so the accurate segmentation of retinal blood vessels has important research significance. However, there are many problems in retinal vessel segmentation, such as complex and small vascular structure and low image contrast, which lead to low segmentation accuracy. To solve the above problems, this paper proposes an attention-directed adversarial network for retinal vascular segmentation. The purpose is to guide the network to learn useful information for vascular segmentation and ignore useless redundant information. The attention directed generative adversarial networks consist of generator and discriminator. The generator uses U-Net architecture and combines the high-low feature attention modules. The high-low level feature attention modules act on the high-level and low-level feature maps so that the model can strengthen the high-level and low-level spatial features respectively, eliminate redundant information, and guide the model to pay more attention to the vascular foreground information. The batch normalization layer in the generator is also removed to avoid the impact of unstable batch statistics on the segmentation results when the generator is trained in small batches. The discriminator consists of a stack of residual modules, which together with the generator form a conditional generation adversarial network. The experimental results show that the Se, Sp, Acc and AUC of this paper's method tested on the DRIVE dataset are 82.88%, 97.45%, 95.59%, and 97.86%, respectively, and all the indexes are better than the current mainstream retinal vessel segmentation algorithms.

Keywords: Retinal vessel segmentation, attention, generative adversarial networks, U-Net, ResNet

1. INTRODUCTION

Retinal blood vessels can characterize human diseases. Studies have shown that the degree of retinal vessel deformation is inextricably linked to the severity of ophthalmic diseases such as glaucoma, cataract and diabetic retinopathy. Detection of retinal vascular changes can effectively prevent and diagnose ophthalmic diseases. It is not only time-consuming and labor-consuming to observe the changes of retinal vessels manually, but also greatly affected by human subjective factors. Therefore, in order to assist physicians in the rapid diagnosis of relevant diseases, precise segmentation of fundus images using image processing techniques is required. However, the following four factors make it difficult to segment tiny blood vessels[1]: first, the shape and width of retinal vessels are different and cannot be represented by simple patterns; second, the local intensity, resolution and contrast of the image affect the detection effect; third, it may be affected by other structures such as light and lesion area; Fourth, the fundus image with low contrast and noise makes this segmentation task very challenging. Therefore, retinal vessel segmentation has always been hot and difficult research at home and abroad.

From the relevant research in recent years, the algorithms used to solve fundus image segmentation tasks can be broadly classified into the following five categories: vascular tracking-based, matched filtering, morphological processing, deformation modeling, and machine learning-based methods. Vessel tracking-based methods can obtain local information about blood vessels, but they are easily affected by vessel branches or intersections, resulting in the loss of tracking targets. Zhang et al.[2] used the PCA method to locate the optic disc and identify the initial point for tracking, and selected the edge candidates of the vessel on the semi-ellipse, considering a combination of three types of vessels. Odstrcilik et al.[3] combined the minimum threshold error and matched filtering together, which is a simple and effective method. In 2014, Karthika et al.[4] extended the wavelet algorithm to enhance the retinal blood vessel image and finally

* 1797664658@qq.com

Third International Conference on Computer Science and Communication Technology (ICCSCT 2022)
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12506, 125063B
 · 0277-786X · doi: 10.1117/12.2661840

segmented the enhanced image to improve the segmentation accuracy from the data side. The morphological model-based approach describes the vessel boundaries with continuous curves, and Al-Diri et al.[5] proposed a new retinal vessel segmentation and measurement algorithm, which uses a Ribbon of Twins (RoTs), active contour model, to efficiently extract retinal vessels while maintaining the uniformity of vessel width. The machine learning based approach to dichotomizing vascular pixels and background pixels is mainly divided into unsupervised and supervised algorithms, where unsupervised algorithms do not require labeling of pixel points and supervised algorithms require labeling of pixel points as vascular or background. Salazar-Gonzalez et al.[6] used adaptive histogram equalization as well as robust distance transform to preprocess the image and then combined vector flow of graph cut algorithm for vessel segmentation. Wang et al.[7] combined convolutional neural networks and random forests to constitute a supervised learning algorithm based on features and integrated learning to improve the performance of vessel segmentation. Ronneberger et al.[8] proposed U-Net for cell segmentation, however, for the retinal vessel segmentation task, it still could not segment the fine vessels. Although the above algorithms improved the accuracy of retinal vessel segmentation, they still could not segment the fine vessels well.

Therefore, to address the challenges and shortcomings of existing retinal vessel segmentation algorithms, we combine attention mechanisms and adversarial networks to propose a new segmentation method. First, our proposed high and low level feature attention module, which acts on both high and low level feature maps, enables the network to focus on the fusion of high and low level features, eliminate irrelevant features and learn relevant features, which can better segment the fine retinal vessels; Second, we segment the retinal vessels using a conditional generative framework, combining the attention module and the U-net to form a generator for the conditional generative adversarial network, using the residual network as the discriminator, we further improve the segmentation ability of the generator and the discriminator by adversarial training between the generator and the discriminator; finally, the batch normalization layer in the generator is removed to avoid the problem of unstable batch statistics caused by the BN layer when training the segmentation network in small batches. In this paper, we also preprocess the original fundus images with contrast-limited adaptive histogram equalization and gamma correction to mitigate the effects of factors such as low contrast of illumination, foreground, and background, and alleviate the model training pressure.

2. METHOD

2.1 Attentional conditional generation adversarial network

Generative Adversarial Networks (GANs)[9] are derived from game theory ideas and consist of two parts: the generator model (G) and the discriminator (D) model. Among them, the generator generates images by inputting random noise signals; the discriminator is responsible for determining whether the input image is from the generator-generated image or the real image, and both are continuously iterated and optimized until the discriminator cannot discern the source of the input image. The objective function of GAN is as follows:

$$\min_{G}\max_{D} V(D,G)=\mathrm{E}_{x\sim p_{data}(x)}[\log(D(x))]+\mathrm{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{1}$$

where E denotes expectation, z denotes random noise, $p_z(z)$ denotes z obeying Gaussian distribution, x denotes real data, $x\sim p_{data}(x)$ denotes obeying real data distribution.

However, generative adversarial networks are too free to be constrained; Conditional Generative Adversarial Nets (CGAN), on the other hand, constrain generative adversarial networks with some additional information as a condition to guide data generation. The additional information can be image category labels or segmentation labels. The general framework of the method in this paper is shown in Figure 1. The objective function of CGAN is as follows:

$$L_{CGAN}=\mathrm{E}_{x,y\sim p_{data}(x,y)}[\log(D(x,y))]+\mathrm{E}_{x\sim p_{data}(x)}[\log(1-D(x,G(x)))] \tag{2}$$

where x represents the color fundus image, y represents the pixel-level label corresponding to the color fundus image, $D(x,y)$ represents the probability that the discriminator predicts (x,y) to be a true sample, and $D(x,G(x))$ represents the probability that the discriminator predicts $(x,G(x))$ to be a false sample. The goal of the generator is to minimize its loss function to segment the vessels accurately, and the goal of the discriminator is to maximize its loss function so that it cannot discriminate the generator output, so CGAN optimizes the following function:

$$G^* = \arg\min_G [\max_D [\mathrm{E}_{x,y\sim p_{data}(x,y)}[\log D(x,y)] + \mathrm{E}_{x\sim p_{data}(x)}[\log(1-D(x,G(x)))]]] \quad (3)$$

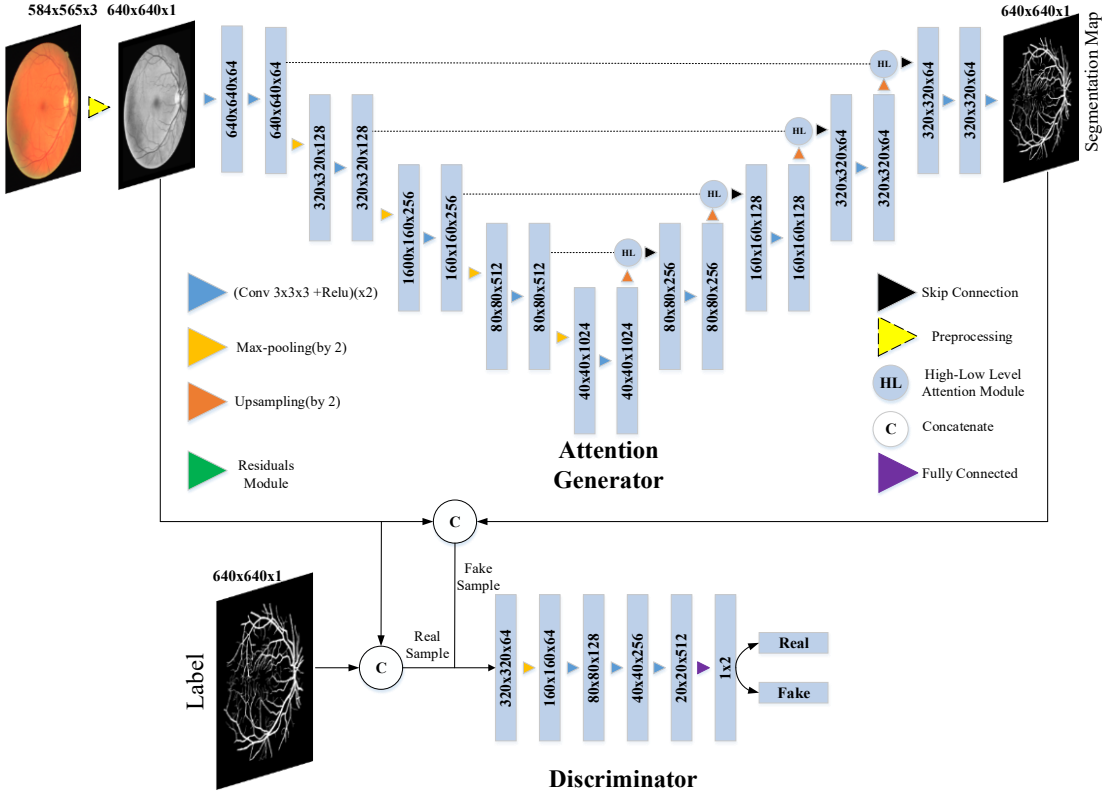

Figure 1. Attention condition generation adversarial network framework.

2.2 Attention generators and discriminators

In the U-Net, the high-resolution feature maps have more detailed features, which help the network to obtain foreground location information; the low-resolution feature maps have more semantic features, which help the network to obtain foreground category information. The U-Net decoder recovers image details by up-sampling, but this leads to edge blurring and loss of location detail information; therefore, the U-Net performs joint channel processing of low-level features and high-level features in a direct hop-connected manner, which helps to supplement location detail features and improve segmentation accuracy. However, the low-level feature maps lack semantic information and contain a lot of useless background information, and the high-level feature maps contain many useless features after upsampling operations. To solve this problem, this paper introduces the attention mechanism[10,11] and designs an attention module to add the high and low-layer feature maps to increase the rich detail information and semantic information of the image feature maps, which is used to guide the network to select the features that help retinal vessel segmentation; and the weighted features are integrated by hopping connections to achieve the suppression of useless features in the low layer feature maps and high layer feature maps. The weighted features are integrated by hopping connections to achieve the purpose of suppressing useless features in the low-level feature maps and the high-level feature maps, thus improving the retinal vessel segmentation performance.

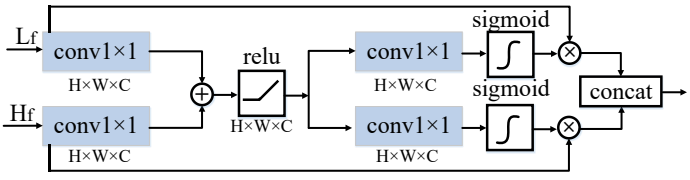

Figure 2. High-Low level (HL) attention module.

Figure 2 shows the attention module designed in this paper. L_f is the low-level feature map, H_f is the high-level feature map, and conv1x1 denotes the convolution of 1×1. Firstly, L_f does the convolution operation of 1×1 to get the feature map with constant size and number of channels C. Meanwhile, H_f is processed with the same, and the processed feature maps H_f L_f are added together for feature fusion, and then the intermediate feature map is obtained by the ReLu function; then it is divided into two paths, which are convolved by 1×1 and sigmoid function, respectively, to learn the probability map, and then with the corresponding feature map. The important pixels are given higher weights and the irrelevant pixels are given smaller weights, to suppress irrelevant features while enhancing useful features; finally, the feature maps obtained from the two paths are channel spliced. The low-level feature map attention probability map can be described as follow:

$$V=\sigma_1[W_lL_f+b_l+W_hH_f+b_h] \tag{4}$$

$$A_l=\sigma_2[W_{l1}V+b_{l1}] \tag{5}$$

where σ_1 is the ReLu function, σ_2 is the sigmoid function, W_l, W_h, W_{l1} is the 1×1 convolution kernel parameter, and b_l, b_h, b_{l1} is the bias. The high-level feature map attention probability map formula is similar to the above formula. The semantic dependencies can be further captured using the convolution of 1×1. Normalization of the attention[11,12] coefficients using the softmax activation function produces sparse activation values, so σ_2 uses the sigmoid activation function, resulting in a probability map for each feature map. The attention generator network is shown in Figure 3.

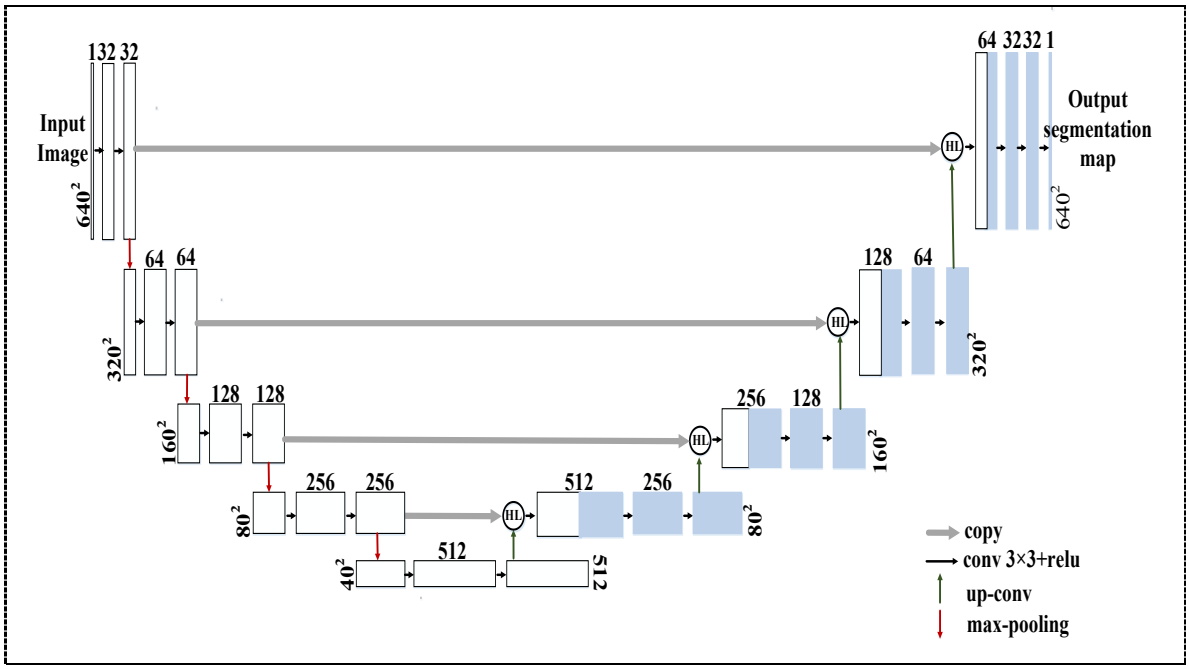

Figure 3. Attention generator network.

The discriminator input consists of the pre-processed fundus image, the generator output, and the image pixel-level labels, where the pre-processed fundus image and the generator output constitute the fake samples, and the pre-processed fundus image and the pixel-level labels constitute the real samples. In this paper, the method uses the residual network as the discriminator, and the residual module helps to solve the problem of network gradient disappearance and network degradation, which can train a deeper network while ensuring the good performance of the network, and the discriminator network structure is shown in Table 1.

Table 1. Discriminator network structure.

Output size	Operation	Number of operations
$320\times 320\times 64$	conv_7×7	1
$160\times 160\times 64$	maxpool_3×3	1
$160\times 160\times 64$	basicblock	2
$80\times 80\times 128$	basicblock	2
$40\times 40\times 256$	basicblock	2
$20\times 20\times 512$	basicblock	2
$1\times 1\times 512$	avgpool_20×20	1
1×2	fc	1

Here conv_7×7 represents convolution, the kernel size is 7, and the step size is 2. Maxpool_3×3 indicates maximum pooling with size 3 and step size 2. Avgpool_20×20 is the average pooling, size 20. *Fc* is the fully connected layer, and basicblock is the residual module, the specific structure is shown in Figure 4.

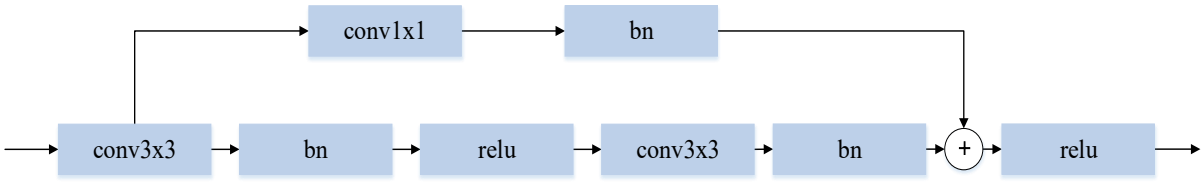

Figure 4. Residual module (basicblock).

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1 Data introduction

We use the DRIVE[12] dataset to verify the performance of the algorithm. The data set was published by Staal et al. in 2004. It is specially used for retinal blood vessel segmentation research. The data is color images obtained by screening patients with diabetic retinopathy. Screening population aged 25-90 years, with a resolution of 565×584 per image. The tag binary mask is marked by two groups of experts, and the mask corresponds to the image, where pixels 1 and 0 represent the blood vessel and background, respectively.

3.2 Fundus image data expansion and pre-processing

Due to the small amount of data in the training set in the DRIVE dataset, direct use for training would lead to poor neural network training results, so the images were rotated and horizontally flipped, and the training set images were increased to 4800 images, of which 4560 images were used to train the network and the remaining 240 images were used for the validation set. To reduce the effects of lighting, noise, and low foreground and background contrast factors and further improve the network segmentation accuracy, the fundus images are preprocessed. The preprocessing steps are as follows: Step 1, the color fundus images are converted to grayscale images; Step 2, contrast limited adaptive histogram equalization (CLAHE)[13] is applied to the grayscale images to enhance the background and foreground contrasts while suppressing the presence of noise; Step 3, gamma correction is applied to reduce the effects of uneven illumination and centerline reflection phenomena; Step 4, the mean and variance of each grayscale image are calculated and image normalization is performed to enable the network to better learn the image distribution and speed up the network training; Step 5, the image is subjected to a complementary 0 operation to make the image size 640×640. The results after Steps 1-3 are shown in Figure 5.

(a) Original image (b) Grayscale images (c) after CLAHE and gamma correction

Figure 5. Fundus image preprocessing example.

3.3 Evaluation indicators

Retinal segmentation is the division of pixel points into vascular and non-vascular categories. To better evaluate the segmentation accuracy of retinal vessels in color fundus images, we evaluated the algorithm performance using sensitivity (se), specificity (sp), accuracy (acc) and area under the ROC curve. Sensitivity indicates the proportion of correctly segmented vessel points to all gold standard vessel points. Specificity indicates the proportion of correctly segmented background points to all gold standard background points. Accuracy represents the ratio of the correct segmentation of vascular points and background points to the whole image pixels. AUC refers to the area under the ROC curve, the closer the area is to 1, the better the retinal segmentation effect. The calculation formula is as follows:

$$Se = \frac{TP}{TP + FN} \tag{7}$$

$$Sp = \frac{TN}{TN + FP} \tag{8}$$

$$Acc = \frac{TP + TN}{TP + FN + TN + FP} \tag{9}$$

where TP, TN, FP, and FN denote the number of pixels whose blood vessels are correctly segmented, the number of pixels whose background is correctly segmented, the number of pixels whose blood vessels are incorrectly segmented, and the number of pixels whose background is incorrectly segmented, respectively.

3.4 Training process

We are experimenting with a computing platform with a GPU model NVIDIA GTX1080Ti. We build the network with keras under the tensorflow deep learning framework. First train the discriminator, then train the generator, both alternately. We use Adam optimizer, parameter $\beta_1 = 0.5$, parameter $\beta_2 = 0.999$. We set the learning rate to 2e-4, the training batch size to 3, and the training epoch = 70, and use the validation set to find the best model in the training process. Finally, the retinal vessel segmentation network proposed in this paper is tested on the test set.

3.5 Experimental results

Table 2 shows the comparison results between our algorithm and other algorithms under four evaluation indexes. The results of our method in the four evaluation indexes were Se = 82.88%, Sp = 97.45%, Acc = 95.59% and AUC = 97.86%, where Se, Acc and AUC were higher than the current advanced methods and improved the retinal vascular segmentation effect.

Figure 6 is the ROC curve and PR (Precision Recall) curve of the proposed method. Figure 7 is the segmentation effect comparison of different algorithms. The red rectangular box is the local segmentation of retinal blood vessels. By comparing the experimental results in References[14,19], we know that our method has high vascular connectivity, is effective in segmenting larger vessels, can accurately segment the main vessel branches, and has good results in segmenting some fine vessels, which can effectively alleviate the problem of vessel segmentation breakage.

Table 2. Experimental results of different algorithms on DRIVE dataset (%).

Method	Se	Sp	Acc	AUC
Orlando[14]	78.50	96.70	-	-
Liskowski[15]	77.63	97.68	94.95	97.20
Li[16]	75.69	98.16	95.27	97.38
Orlando[17]	78.97	96.84	-	-
Dasgupta[18]	76.91	98.01	95.33	97.44
Yan[19]	76.53	**98.18**	95.42	97.52
Alom[20]	77.92	98.13	95.56	97.84
Khoong[21]	82.81	97.23	95.39	97.81
Ours	**82.88**	97.48	**95.59**	**97.86**

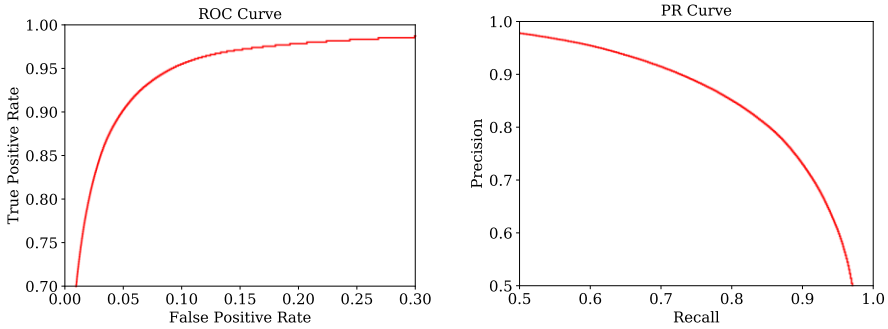

Figure 6. ROC curve and PR (Precision Recall) curve.

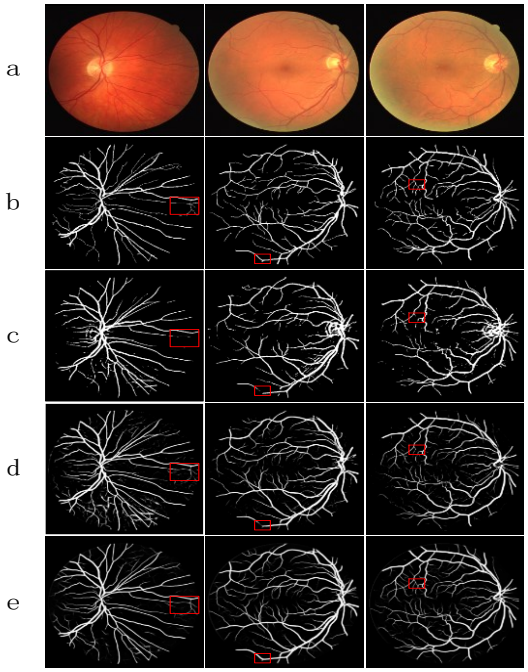

Figure 7. Example of the segmentation effect of the method in this paper on the DRIVE data set. (a): the original color fundus image; (b): the gold standard image; (c): the segmentation result of Reference[14]; (d): Reference[19] segmentation result graph; (e): the segmentation result graph of this method.

4. CONCLUSION

Retinal vascular segmentation can help doctors observe the status of blood vessels, which is important for ophthalmic disease prevention and diagnosis. Many fundus retinal vessel segmentation algorithms have poor segmentation accuracy due to the low quality of color fundus images and the small size of blood vessels. Therefore, we propose conditional generation adversarial network based on attention mechanism for enhancing fundus retinal segmentation. We designed an attention module to give higher weight to useful pixels and lower weight to useless pixels. Combining the attention module with U-Net as a generator to make the network focus on relevant features and suppress irrelevant features can effectively improve vessel segmentation accuracy. The residual network is used as a discriminator to form an attention-conditional generative adversarial network, which further optimizes the generator and strengthens the generator's ability to segment tiny vessels through adversarial learning. Compared with current advanced algorithms, the Se, Sp, Acc, and AUC of segmentation of this paper's algorithm are improved, and it also has some reference significance for segmenting other tissue structures.

REFERENCES

[1] Fraz, M. M., Barman, S. A., Remagnino, P., et al., "An approach to localize the retinal blood vessels using bit planes and centerline detection," Computer Methods and Programs in Biomedicine 108(2), 600-616 (2012).

[2] Zhang, J., Li, H., Nie, Q., et al., "A retinal vessel boundary tracking method based on Bayesian theory and multi-scale line detection," Computerized Medical Imaging and Graphics 38(6), 517-525 (2014).

[3] Odstrcilik, J., Kolar, R., Budai, A., et al., "Retinal vessel segmentation by improved matched filtering: Evaluation on a new high-resolution fundus image database," IET Image Processing 7(4), 373-383 (2013).

[4] Karthika, D. and Marimuthu, A., "Retinal image analysis using contourlet transform and multistructure elements morphology by reconstruction," IEEE World Cong. on Computing and Communication Technologies, 54-59 (2014).

[5] Al-Diri, B., Hunter, A. and Steel, D., "An active contour model for segmenting and measuring retinal vessels," IEEE Transactions on Medical Imaging 28(9), 1488-1497 (2009).

[6] Salazar-Gonzalez, A., Kaba, D., Li, Y., et al., "Segmentation of the blood vessels and optic disk in retinal images," IEEE Journal of Biomedical and Health Informatics 18(6), 1874-1886 (2014).

[7] Wang, S., Yin, Y., Cao, G., et al., "Hierarchical retinal blood vessel segmentation based on feature and ensemble learning," Neurocomputing 149, 708-717 (2015).

[8] Ronneberger, O., Fischer, P. and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," Inter. Conf. on Medical Image Computing and Computer-Assisted Intervention, 234-241 (2015).

[9] Salimans, T., Goodfellow, I., Zaremba, W., et al., "Improved techniques for training gans," Advances in Neural Information Processing Systems, 2234-2242 (2016).

[10] Anderson, P., He, X., Buehler, C., et al., "Bottom-up and top-down attention for image captioning and visual question answering," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 6077-6086 (2018).

[11] Jetley, S., Lord, N. A., Lee, N., et al., "Learn to pay attention," arXiv preprint arXiv:1804.02391, (2018).

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

[13] Reza, A. M., "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," Journal of VLSI signal processing systems for Signal, Image and Video Technology 38(1), 35-44 (2004).

[14] Orlando, J. I. and Blaschko, M., "Learning fully-connected CRFs for blood vessel segmentation in retinal images," Inter. Conf. on Medical image computing and Computer-Assisted Intervention, 634-641 (2014).

[15] Liskowski, P. and Krawiec, K., "Segmenting retinal blood vessels with deep neural networks," IEEE Transactions on Medical Imaging 35(11), 2369-2380 (2016).

[16] Li, Q., Feng, B., Xie, L. P., et al., "A cross-modality learning approach for vessel segmentation in retinal images," IEEE Transactions on Medical Imaging 35(1), 109-118 (2016).

[17] Orlando, J. I., Prokofyeva, E. and Blaschko, M. B., "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," IEEE transactions on Biomedical Engineering 64(1), 16-27 (2017).

[18] Dasgupta, A. and Singh, S., "A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation," 2017 IEEE 14th Inter. Symp. on Biomedical Imaging (ISBI 2017), 248-251 (2017).

[19] Yan, Z., Yang, X. and Cheng, K. T., "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," IEEE Transactions on Biomedical Engineering 65(9), 1912-1923 (2018).

[20] Alom, M. Z., Hasan, M., Yakopcic, C., et al., "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," arXiv preprint arXiv:1802.06955, (2018).

[21] Khoong, W. H., "BUSU-Net: An ensemble u-net framework for medical image segmentation," arXiv preprint arXiv:2003.01581, (2020).