



Retinal Vessel Segmentation in Medical Diagnosis using Multi-scale Attention Generative Adversarial Networks

Minqiang Yang¹ · Yinru Ye¹ · Kai Ye¹ · Wei Zhou¹ · Xiping Hu¹ · Bin Hu¹

Accepted: 29 July 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

With the advancement of medical technology, the demand for efficient and precise medical diagnosis is growing. Retinal vessel segmentation using artificial intelligence techniques is vital for medical diagnosis of degenerative retinal diseases. In this paper, we introduce a multi-scale generative adversarial network with class activation mapping, which can effectively improve the efficiency and accuracy of vessel segmentation using artificial intelligence. The task of vessel segmentation is better achieved due to our proposed architecture, which incorporates an attention mechanism and a multi-scale discrimination. It not only strengthens the ability to locate and segment fine retinal vessels, but also enables the model to have the ability to discriminate different receptive fields. To tackle the instability problem caused by unsupervised learning of generative adversarial networks, we introduce a supervised segmentation loss to improve model stability and convergence speed. And we propose a data augmentation method by reconstructing and combining fundus images to make the model obtain better generalization ability. We compare our method with previous models by several metrics and perform ablation study on each component of the model, demonstrating the superiority and effectiveness of the model.

Keywords Retinal vessel segmentation · Multi-scale generative adversarial network · Class activation mapping · Data augmentation

1 Introduction

The combination of medical diagnosis and artificial intelligence (AI) technology has become an indispensable

part of the medical field [1–3]. Funduscopic examination is an imperative basis for the treatment and evaluation of retinal diseases [4]. Relevant experiments have shown that in the case of hypertension and diabetes, the retinal vessels may have related conditions such as tortuosity, irregularity and narrowing [5]. So in funduscopic examination, precise retinal vessel segmentation in fundus images plays a crucial role in assisting decision-making in the assessment of retinal diseases. However, traditional medical image processing mainly relies on the personal experience of ophthalmologists to manually process and analyze the fundus images. This manual processing method is time-consuming and prone to subject bias, which ultimately affects the analysis results and treatment effects. Therefore, it is desirable to use AI technology to automatically segment retinal vessels.

Retinal vessel segmentation has been a long-standing and important task in medical image processing. With the widespread applications of deep learning [6], many researchers used generative adversarial networks (GAN) [7, 8] to make many attempts to complete the task of image segmentation and achieved considerable results. Combined with the privacy protection of edge computing [9], disease

✉ Xiping Hu
huxp@lzu.edu.cn

Bin Hu
hb@lzu.edu.cn

Minqiang Yang
yangmq@lzu.edu.cn

Yinru Ye
yeirl8@lzu.edu.cn

Kai Ye
kye18@lzu.edu.cn

Wei Zhou
zhouw21@lzu.edu.cn

¹ School of Information Science and Engineering, Lanzhou University, Tianshui South Road, Lanzhou, 730000, Gansu, China

diagnosis based on retinal vessels has a broad application space.

In medical diagnosis, the shape of tortuous retinal vessels is often related to retinopathy [10], so the vessel distribution and the shape of retinal vessel also have impacts on diagnosis and treatment. In the retinal vessel segmentation task, the boundary consists of the distribution, shape and direction of the retinal vessel. Many previous models can indeed have good results on the metrics in this task, but in fact the segmented retinal vessels will appear fragmented, irregular and other shapes, which influence the diagnosis.

The existing models also meet problems that data collection and data labeling are time-consuming and labor-intensive, and the small amount of data causes overfitting. Therefore, a data augmentation method should be applied.

Based on previous methods [11–14], we propose a novel architecture GAN to achieve retinal vessel segmentation. Inspired by the locality-aware attention mechanism [15, 16], we add class activation mapping (CAM) into the network. Moreover, we combine segmentation loss and adversarial loss to jointly train the model, which can make the generated retinal vessel segmentation images have accurate segmentation boundaries. Also, inspired by previous data augmentation methods [17–19], we introduce a novel data augmentation method - image stitching. It erratically crops a part of the area of one sample in the training set then randomly fills in the pixels of another sample. The cropped data can be flipped at any angle. Due to the different positions, rotation angles and sizes of the filled regions, only using two samples can generate unlimited training samples. This method increases the diversity of samples and avoids model overfitting.

Finally, we analyze the superiority of our model with qualitative, quantitative experiment and ablation study. For quantitative evaluation, we compare the evaluation metrics with other models and show the changes of the metrics during the training process. For the qualitative point of view, we visualized the attention mechanism and the real classification result. For the ablation study, we verified the effectiveness of each component embedded in the model by comparing the impact of the combination of different components.

The main contributions of our work are:

1. We propose a multi-scale GAN model with CAM to restore vessels, which are more in line with the real vessel distribution.
2. We introduce a novel data augmentation methods called image stitching to solve the problem of small retinal image dataset, whose distribution of vessels is less affected by processing such as rotation and flipping.
3. We utilize the supervised segmentation loss to improve the training stability and convergence speed.

2 Related Work

2.1 Image Segmentation

Image segmentation refers to finding the boundary of the target region in the image, so that the pixels inside and outside the boundary have different characters (intensity, texture, etc.). It is an significant support for computer-aided systems in the process of diagnosis and treatment. Medical image segmentation is usually used in the following use cases: breast-lesion segmentation [20], liver segmentation on computed tomography (CT) scans [21, 22], retinal vessel segmentation [23, 24]. Moi Hoon Yap accomplished the task of breast ultrasound lesions recognition through convolutional neural networks (CNN) [25]. Lianfen Huang achieved automatic segmentation of liver and lesion from CT images needed for computer-aided diagnosis of liver by a single-block linear detection algorithm [26].

Many breakthroughs have been achieved in the field of biomedical image segmentation through AI and machine learning methods. Dasgupta used implicit advantages of combining convolutional neural networks and structured prediction to achieve strong performance of segmentation task [27]. E Schonfeld introduced U-Net which performed great ability of segmentation [24]. U-Net used the encoder to downsampling the input and performed upsampling through the decoder to match the input resolution with the output resolution for precise localization [18]. Skipping connections to route data between the resolutions of the two modules improves the ability of network to accurately segment details.

2.2 Generative Adversarial Network

There are two main types of popular generative models, variational autoencoder (VAE) [28] and generative adversarial network (GAN) [29]. VAE has an excellent effect on online scheduling [30]. Compared with VAE, GAN has better results in image generation. GAN, which can achieve the improvement of image quality [31] and image colorization [32], has been applied in many fields, such as image segmentation [7], artwork generation [33], and video generation [34]. There are several approaches for GAN to improve the authenticity of generated retinal vessel segmentation images from different perspectives. The first approach is to increase the size of the training process through more complex computations (e.g. UNet-GAN [18] adopted the same batch size and truncation technique as BIGGAN [35]). The second way is to improve the training stability of network structure by using a combination of strided convolution and transposed convolution [36]. The third way is architectural modification (e.g. Style-GAN [37] achieved more realistic

face images by building a mapping network and adding style modules).

3 Our Approach

In this paper, we introduce a multi-scale attention GAN, which consists of a discriminator and a generator. And in order to solve the problem of small amount of dataset, we innovate a data augmentation method called image stitching.

3.1 Discriminator

In this paper, we propose a multi-scale attention discriminator. From the perspective of multi-scale discrimination, multi-scale discrimination enables the discriminator to better maximize the distance between the true and false distribution of the whole and small retinal blood vessels through different receptive fields. As shown in Fig. 1, the discriminator is mainly composed of two parts: CAM [38] and continuous classifiers C1, C2, C3. Spectral normalization is applied in downsampling of classifier [39]. The C1 representation can accept a 128×128 receptive field, and the C2 representation can accept a 256×256 receptive field. The input image, label and the generated segmentation image

are respectively sent to C0 as the input, and then are downsampled by C1 after passing through CAM. The same operation is performed in C2. For a pair of images (eye fundus images and retinal vessel segmentation image), C0, C1, and C2 are all trained to predict the true and false of the image. The true and false sample input of the discriminator is the image superimposed by the fundus image and the segmentation label and the generated segmentation image, respectively. And the discriminator will distinguish the samples combined with the fundus image and the generated segmentation map from the real distribution in the sample space. From the perspective of CAM, CAM will perform weighted linear superposition of different spatial positions on the multi-channel hidden vectors generated by C1, so as to obtain more important information in the input image. It enables the model to efficiently focus on the distribution of retinal vessels, thereby improving training efficiency.

3.2 Generator

The input and output of the generator are RGB retinal fundus images and single-channel retinal segmentation images, respectively. Our proposed model mainly consists of three parts, an encoder, an adaptive module, and a decoder. The encoder structure computes a multi-channel hidden vector through adaptive layer instance normalization

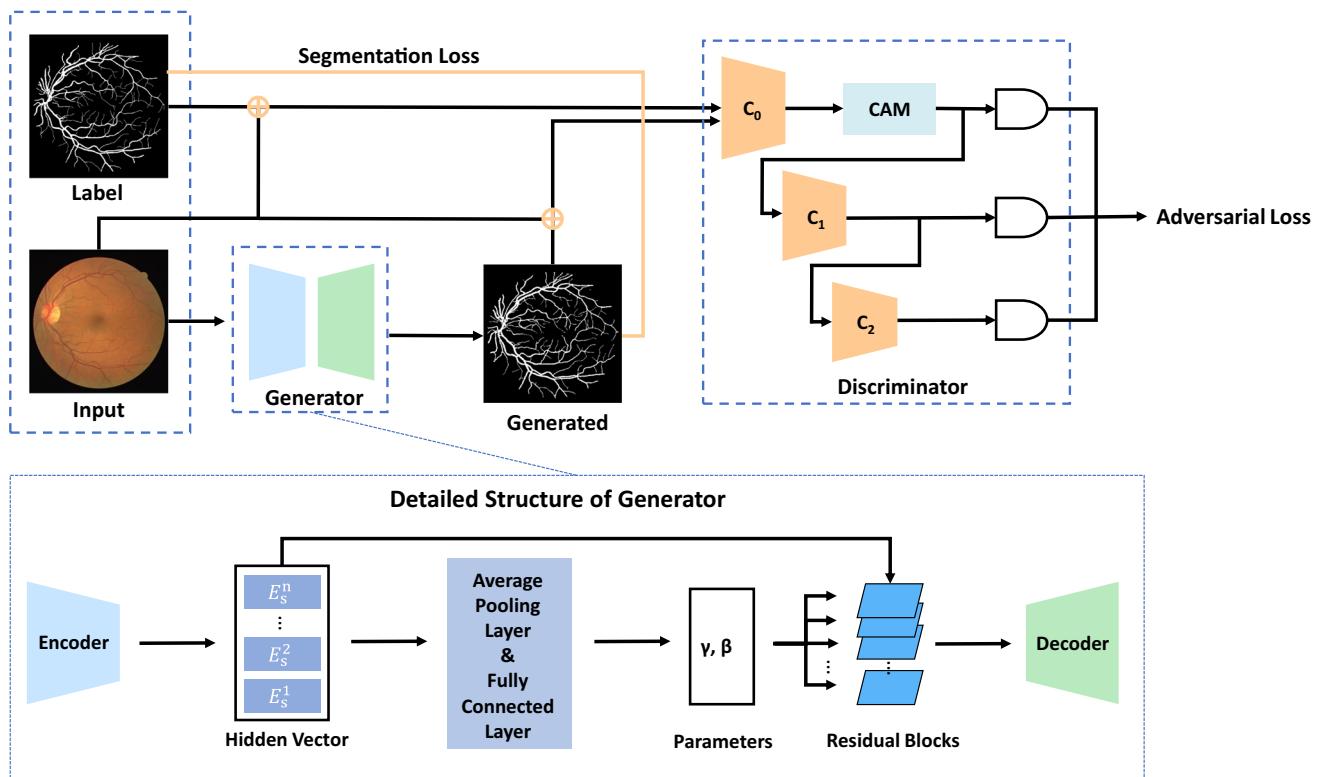


Fig. 1 Architecture of our model

and double-layer downsampling. The adaptive module uses the output of the encoder to calculate the parameters γ and β [40] through the average pooling layer and the fully connected layer. The decoder will combine the parameters of the adaptive module and the convolution normalization including the residual block to calculate, and finally output the segmented image through the upsampling composed of sub-pixel convolution. The above architecture can improve the expressive ability of the generator, so as to better retain the information to complete the retinal segmentation task.

3.3 Image Stitching

In the field of medical image process, data samples are often scarce due to difficult acquisition and time-consuming data collection. For neural networks, a diverse and sufficient dataset can help the training of the network model more efficient. To this end, we propose a data augmentation model called image stitching to augment fundus image samples, which can increase the diversity of samples and the robustness of the model. Firstly, it randomly selects two different samples and their label images in the dataset, and then randomly selects two points in the image. The StartPoint is in the upper left corner and the EndPoint is in the lower right corner. Then the corresponding pixels in one set of images and labels are cropped based on the two points, embedded in the other set of images, and finally resizing, rotating, and mirroring are performed probabilistically. Details of these processing steps can be found in Algorithm 1. In Fig. 2, the first row is the data after image stitching, the second row is the image after retinal vessel segmentation, and the third row is the label after image stitching. Our method can generate a large number of

samples, which is actually a regularization, which alleviates the overfitting of the model.

Data: All original images I and tags of all original images L

Result: The stitched image $StitchedImage$ and the corresponding label $StitchedLabel$

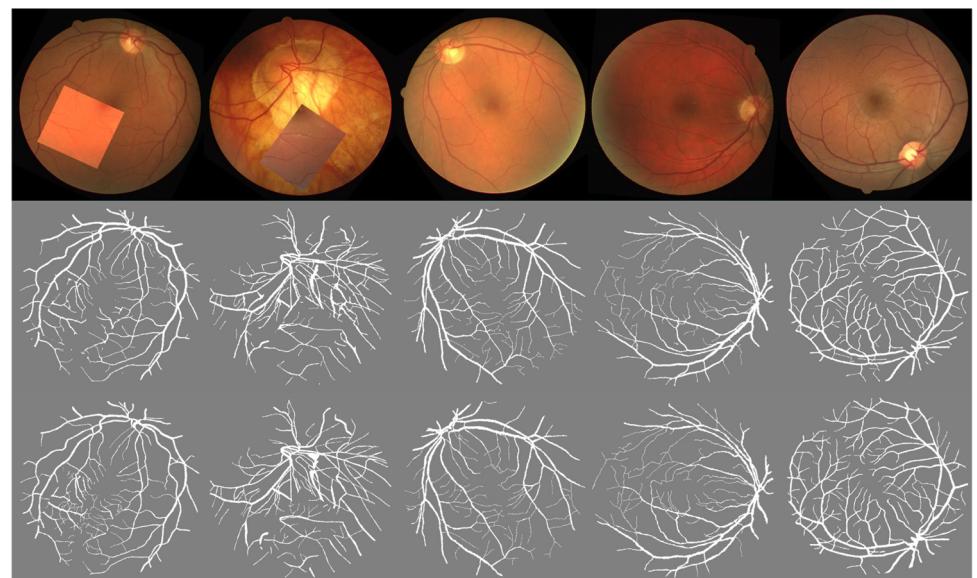
```

1 OriginalImage0, OriginalLabel0 ←
  RandomChoice ( $I, L$ ) ;
2 OriginalImage1, OriginalLabel1 ←
  RandomChoice ( $I, L$ ) ;
3 StartPoint ← RandomChoice ( $Image.Height$ ,
   $Image.Width$ ) ; EndPoint ←
  RandomChoice ( $Image.Height$ ,  $Image.Width$ ) ;
4 StartPoint, EndPoint ← Order (StartPoint, EndPoint) ;
5 OriginalImage0 ← Crop (OriginalImage0, StartPoint,
  EndPoint) ;
6 OriginalLabel0 ← Crop (OriginalLabel0, StartPoint,
  EndPoint) ;
7 StitchedImage ← Paste (OriginalImage1,
  OriginalImage0, StartPoint) ;
8 StitchedLabel ← Paste (OriginalLabel1,
  OriginalLabel0, StartPoint) ;
9 Spin ← RandomChoice(0,360);
10 Flip ← RandomBinary(0,1);
11 StitchedImage ← Rotate (StitchedImage, Spin) ;
12 StitchedLabel ← Rotate (StitchedImage, Spin) ;
13 if Flip == True then
14   | StitchedImage ← Transpose (StitchedImage) ;
15   | StitchedLabel ← Transpose (StitchedImage) ;
16 end

```

Algorithm 1 Image stitching.

Fig. 2 Image stitching: The top row are eye fundus images after image stitching; the middle row are results obtained by our method; the bottom row are manual label



3.4 Loss Functions

Our model mainly contains two kinds of loss functions in the training process, one is the adversarial loss, and the other is the end-to-end segmentation loss. The specific details are as follows:

Adversarial Loss The adversarial loss aims to guide the discriminator to distinguish between two parts of the input, one is the overlay of the generated vessel segmentation image and the fundus image, and the other is the overlay of the ground truth and the fundus image. The former is generally regarded as the source domain and the latter as the target domain, which have different internal distribution. With the iterative process going on, it attempts to make the distribution probability of the output of the generator approach the distribution of the target domain continuously.

$$\min_G \max_D L_{adv}^{x \rightarrow y} = \mathbb{E}_{y \sim \mathcal{Y}} [(D(y + x))^2] + \mathbb{E}_{x \sim \mathcal{X}} [(1 - D(G(x) + x))^2]. \quad (1)$$

Segmentation Loss In addition to unsupervised adversarial loss, we design a supervised segmentation loss that aims to accelerate and stabilize the convergence and iteration of unsupervised GAN with the efficiency and accuracy of supervised learning. The fundus dataset has the property of imbalance, i.e. most pixels do not belong to blood vessels, therefore, it is a better choice to use a loss function that can significantly increase the cost of model classification errors. Because the loss function used to balance the samples may not be robust to the learned features based on a small number of samples. In summary, we choose mean square error (MSE) as the segmentation loss to improve the segmentation effect.

$$\min_G L_{seg}^{x \rightarrow y} = \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{n^2} \sum_{i,j}^n (x_{i,j} - G(x)_{i,j})^2 \right]. \quad (2)$$

4 Experiments

4.1 Dataset

In this article, we consider the Drive dataset [41] for retinal vessel segmentation. It is a standard dataset for retinal vessel extraction, and is also used to evaluate model by a large number of literature. So the dataset is quite representative and reliable. Drive dataset contains 40 fundus images with resolution of 565*584 pixels, which the training set and the test set are equally separated, each with 20 samples. These samples are from screening of patients with diabetic retinopathy. For each image, there is a manual segmentation

performed by an ophthalmological expert to distinguish the region of the retinal vessel as a label for each fundus image. When we revisit this data, we realize the difficulty of data collection of retinal fundus images. So we came up with an image augmentation technique called image stitching. It rotates, and crops two different images to form an image with a resolution of 512*512 pixels, and performs the same operation on the label.

4.2 Evaluation Metrics

For quantitative evaluation, we choose six metrics to verify the performance of our model: accuracy (ACC), sensitivity, specificity and area under the curve of receiver operating characteristics (AUC-ROC), F1-score (F1s), precision. As shown in Fig. 3, we enumerate the five samples in the test set after segmentation. In the third row of Fig. 3, the green retinal vessel represents the correctly predicted retinal vessel (TP), the red vessel represents the real retinal vessel that was not predicted (FN), and the blue vessel represents the non-existent retinal vessel generated by the generator (FP). Figure 3 shows our model can clearly identify the vessels, and there are very few unidentified and incorrectly identified vessels. It proves the superiority of the model.

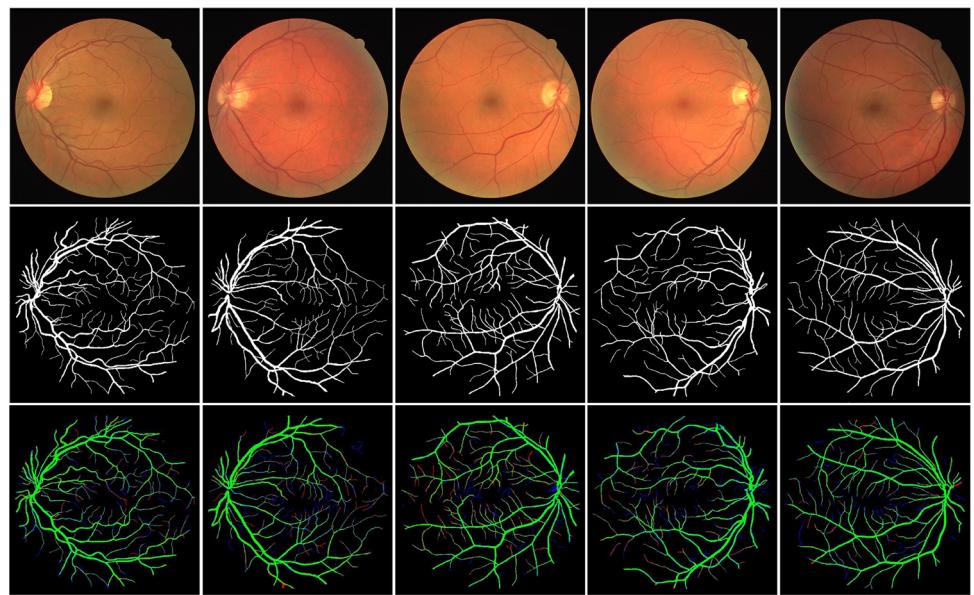
4.3 Comparisons

As shown in Table 1, we utilize six different metrics as the basis for the quantitative analysis of 9 models. It can be clearly observed from the evaluation metrics that our model has a good effect on overall metrics. The recall of our model is as high as 80.40% based on the specificity of 97.57%, combined with the previous description of the metrics in Section 4.2, which means that our model can fully learn the distribution of retinal vessels, especially fine vessels, so that the generated segmentation images are more accurate and precise. Compared with other models, which may focus more on a single metric, our model can complete the task of vessel segmentation while ensuring that the overall metrics are better, especially sensitivity and accuracy. By observing Table 1, it can be found that only our model can still have an accuracy of 96.03% under the premise of ensuring a high recall, which shows that our model can efficiently complete the task of blood vessel segmentation and obtain more realistic segmentation images.

4.4 Ablation Study

In Table 2, we compare the effectiveness of each component in the GAN model on the DRIVE dataset by ablation study, and analyze the models by six metrics. We can clearly find from the Table 2 that when each component is not added, the ACC is only 0.8769 but the sensitivity reaches 0.8512.

Fig. 3 Segmentation results of an example in the DRIVE dataset: The first row are eye fundus images; the second row are manual label; the third row are visualization of classification effects



At first glance it seems that this model can achieve good metrics. But when we observe the visualization of the output in the fourth column of Fig. 4, we clearly find that the vessel segmentation images generated by the model in this case are far from the real retinal vessel distribution. The model classifies most of the pixels as vessels, which can improve the sensitivity, but the actual application effect is not good.

After the model is separately embedded in CAM, multi-scale discrimination and image stitching, we can observe that all metrics except sensitivity are significantly improved. It means that the GAN model embedded in each component can segment the retinal vessels more accurately, and has a stronger ability to perceive vessel boundaries. In terms of image performance, the images generated by the model embedded with a single component are more in line with the real retinal vessel distribution than the model without any component embedded. From the retinal vessel distribution

marked by the red box in Fig. 4, it can be clearly observed that the model can basically identify the main vessels, but it is still insensitive to the fine vessels and the generated vessels have discontinuous features.

By embedding the CAM and improving the multi-scale discriminant receptive field of the discriminator in the GAN architecture, we can find that sensitivity and AUC-ROC have been improved to a certain extent, especially the sensitivity has increased from 0.7435 to 0.8208. This is because a more detailed discriminative receptive field enables the discriminator to better distinguish the local capillary distribution features, which in turn enables the generator to synthesize a retinal vessel distribution that is more in line with the input image.

After analyzing the metrics and visualization of multiple models, we realized that the combination of CAM, multi-scale discrimination and data augmentation can improve

Table 1 Comparison with baselines

Methods	DRIVE					
	ACC	Sensitivity	Specificity	AUC-ROC	F1s	Precision
B-COSFIRE filter [42]	0.9442	0.7655	0.9704	0.9614	—	—
YT Zhao et al [43]	0.9540	0.7420	0.9820	0.8620	—	—
DCNN [44]	0.9453	0.7426	0.9735	0.9516	—	—
Decision Tree [45]	0.9480	0.7406	0.9807	0.9747	—	—
Multi-scale CNN [46]	0.9521	0.7779	0.9780	0.9782	—	—
U-Net [47]	0.9531	0.7537	0.9820	0.9755	0.8142	—
R2U-Net [47]	0.9556	0.7792	0.9813	0.9784	0.8171	—
U-GAN [8]	0.9615	0.7798	0.9820	—	—	—
Proposed	0.9603	0.8040	0.9757	0.9500	0.7807	0.7627

Value of metric is higher, model is better

Table 2 Ablation study

Components			DRIVE					
CAM	Multi-scale	Image Stitching	ACC	Sensitivity	Specificity	AUC - ROC	F1s	Precision
✗	✗	✗	0.8769	0.8512	0.8795	0.9039	0.5553	0.4162
✓	✗	✗	0.9614	0.7435	0.9829	0.9337	0.7721	0.8083
✗	✓	✗	0.9624	0.7838	0.9801	0.9497	0.7857	0.7925
✗	✗	✓	0.9510	0.8621	0.9600	0.9376	0.7561	0.6783
✓	✓	✗	0.9578	0.8208	0.9713	0.9439	0.7736	0.7553
✓	✗	✓	0.9464	0.7971	0.9611	0.9089	0.7718	0.7438
✗	✓	✓	0.9594	0.7936	0.9758	0.9508	0.7744	0.7618
✓	✓	✓	0.9603	0.8040	0.9757	0.9500	0.7807	0.7627

the segmentation ability of the model and synthesize more accurate retinal vessel distribution picture. Overall, the results obtained by combining all components outperformed all other combination.

4.5 Result

As the number of iterations increases, the right part of Fig. 5 shows how the corresponding metrics of model change during training. Overall, we can observe an upward trend in

metrics as the number of iterations increases. In addition, in the right part of Fig. 5, we also show the visualization of the retinal vessel segmentation at 1k, 10k, and 20k iterations, which more directly show how our method accomplishes the task. Combining the metrics and generated segmentation figure at 1k, it is not difficult to find that the model in the early stage of training will tend to predict pixels as retinal vessel. Although this way can make the recall reach a high level, it does not actually restore the distribution of real retinal vessels, which predict some fragmented, irregular

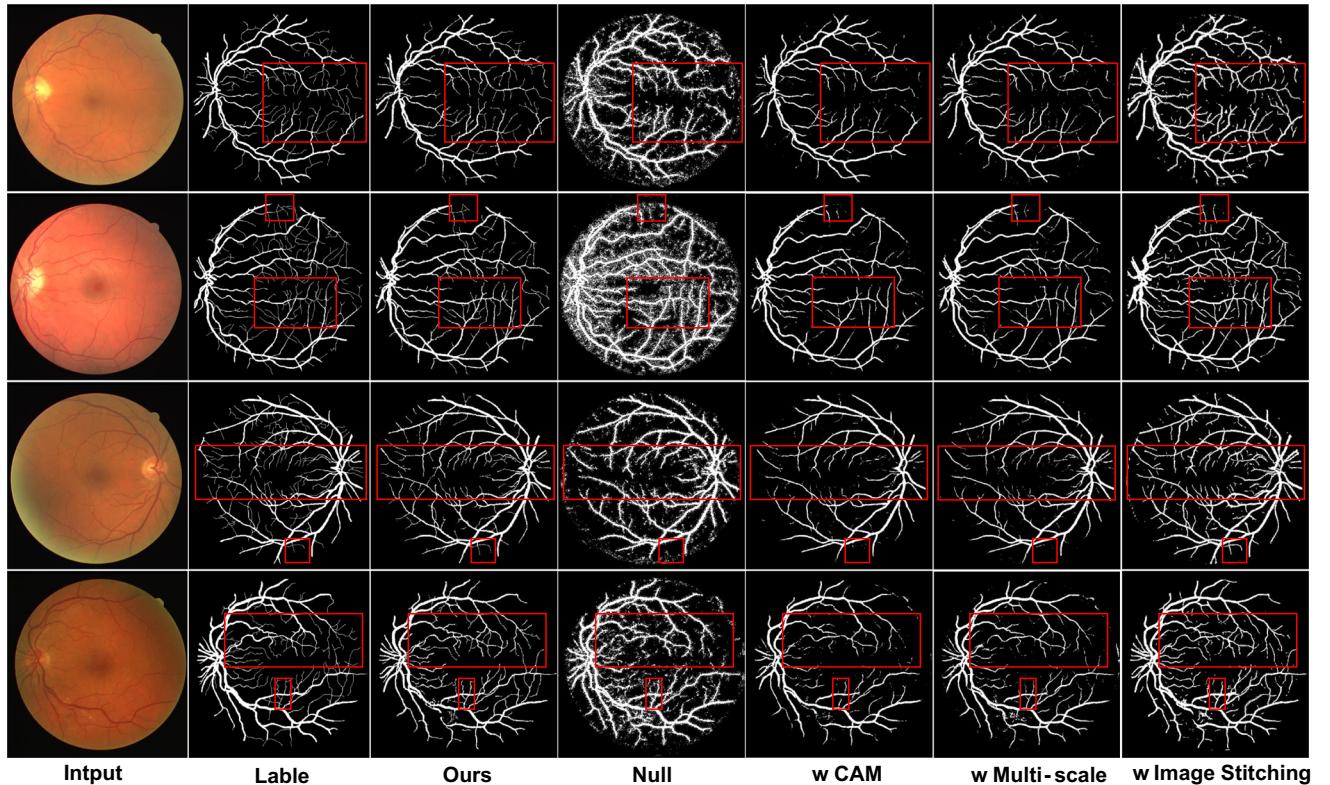
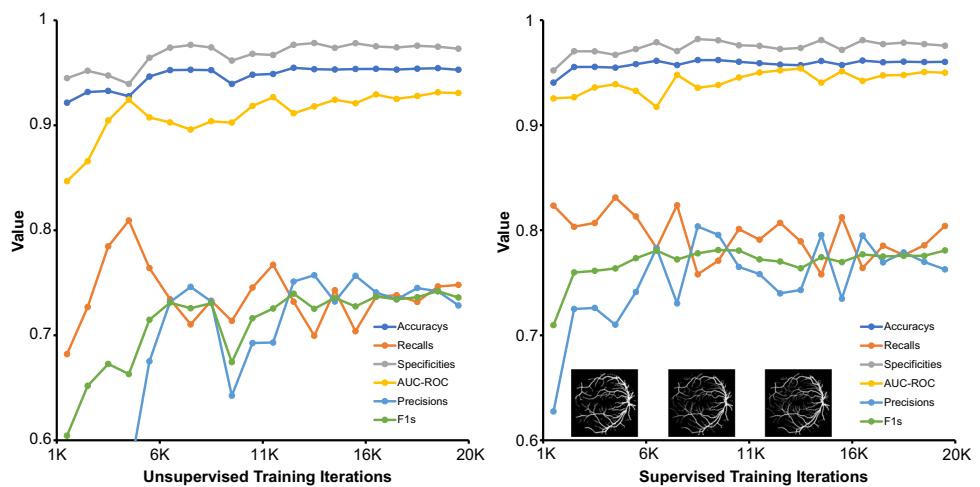


Fig. 4 Segmentation results of each component: The first column are fundus images; the second column are labels; the third row are the results of our model; the fourth column are the results without CAM,

multi-scale and image stitching; the fifth column are results with CAM; the sixth column are the results with multi-scale discriminator; the seventh column are results with image stitching

Fig. 5 Iteration performance: The left part is the performance of the metrics in iterations using only unsupervised learning. The right part is the performance of the metrics in the iterative process after combining unsupervised and supervised learning



shapes. When the model is trained to 10k iterations, it can basically segment the main vessel distribution. But it still cannot segment some fine vessel well. When the training reaches 20k iterations, the segmented images that are almost the same as the label and the model accomplish task of accurate and fine retinal vessel segmentation.

In the left part of Fig. 5, we perform unsupervised learning training without using our proposed segmentation

loss. Firstly, by comparing the metrics, we find that adding the segmentation loss has significantly improved the final effect of the model. Also, comparing the training process of the two, the metric of the model combined with the segmentation loss approaches a stable value significantly faster in the iterative process than the metric of the model with only the adversarial loss, which means that our proposed final model converges faster. Finally, when we

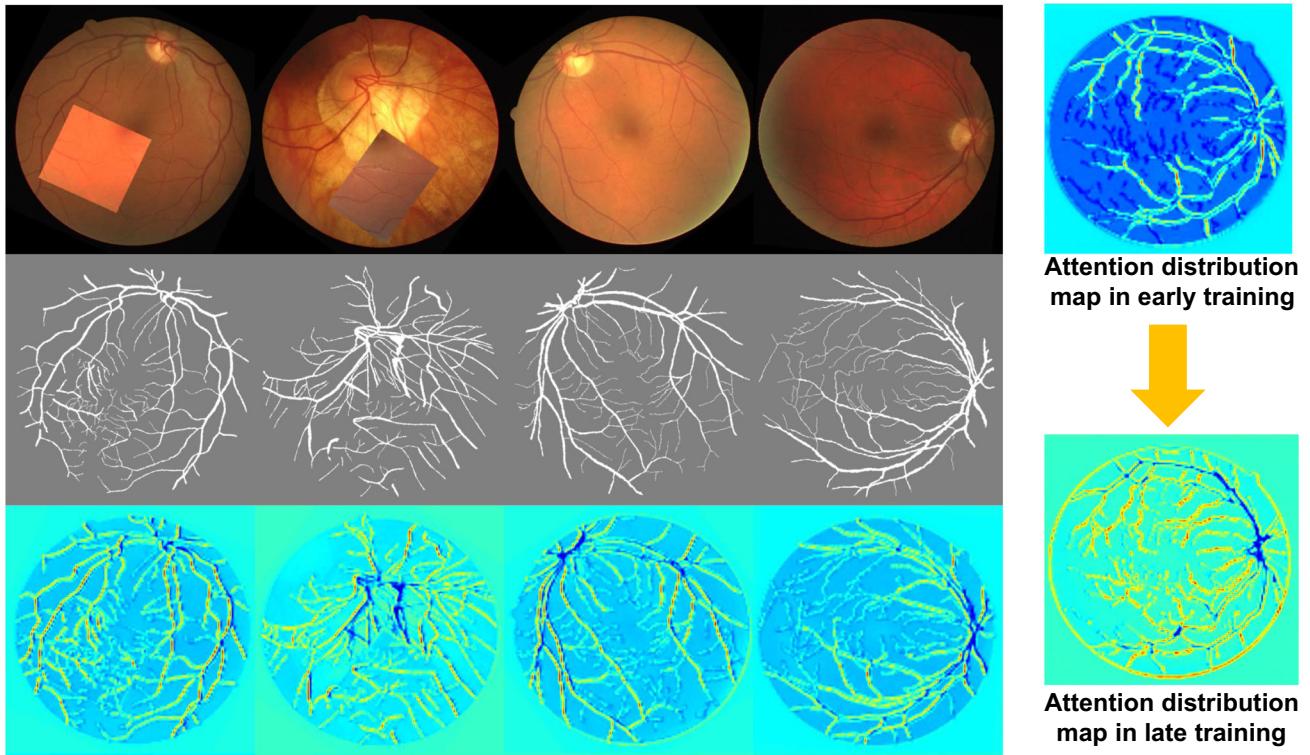


Fig. 6 Class activation mapping: Left part: The first row are fundus images; the second row are manual labels; the third row are fundus images with CAM. Right part: The picture above is the attention distribution map in early training, the picture below is the attention distribution map in late training

revisit the unsupervised method, we can see that this method is still not stable, and adding a segmentation loss can make the training stable.

In Fig. 6, we show the attention distribution maps produced by CAM. The attention distribution map reflects the attention of model paid to each area of image. The closer the color is to red, the more attention model pays to this area. On the contrary, the closer color is to blue, the less attention is paid to this area. From the perspective of spatial dimension, the model focuses more on retinal vessels, while other elements not related to vessel receive less attention. In terms of iterations, the model will focus on the distribution of the main retinal vessels in the early stage. Because the pixels of the early main vessels account for a large proportion and the loss caused by training also accounts for the majority, the model will give priority to learning the characteristics of the main vessels. In the later stage of model training, the generator can stably generate main vessels with high-accuracy, and the model will focus on fine retinal vessels. The above allows the model to segment retinal vessels more efficiently, thereby synthesizing higher-quality retinal vessels.

5 Conclusion

In this paper, we aim to achieve better retinal vessel segmentation to assist medical diagnosis. By combining CAM and multi-scale discrimination, our proposed generative adversarial network architecture can segment vessel that are more in line with the distribution of real vessel. The analyses of various experiments support our two hypotheses that the attention distribution map obtained from the CAM can guide the discriminator to pay more attention to the vessels in the segmentation image that have large deviations from the fundus image, and the multi-scale discriminator can discriminate in different receptive fields so that the generator can segment finer vessels. Furthermore, we found that supervised segmentation loss can improve training stability and convergence speed. Through ablation study and comparative test, we demonstrate the effectiveness and superiority of each component in the proposed architecture in the task of retinal vessel segmentation.

6 Statements and Declarations

The core idea of our manuscript was submitted to the 10th EAI International Conference on Wireless Mobile Communication and Healthcare and won the best paper award, which is recommended to the Journal of Mobile Networks and Applications. And this manuscript is an extension of the previous one.

The extensions are as follows:

1. The generator part of multi-scale attention generative adversarial network is modified. Refer to Section 3.2 for details.
2. The update of the loss function. Compared with the previous focal loss, we replaced it with the supervised MSE loss. Refer to Section 3.4 for details.
3. We explore the ablation experiments of each component embedded in the model and visually compare the images generated by the model after embedding each component. Refer to Table 2, Fig. 4 and Section 4.4 for details.
4. We expanded the metrics used to evaluate the model from four to six in quantitative experiments, ablation experiment and iteration performance. Refer to Tables 1 and 2, Fig. 5 for details.
5. For iteration performance, the number of iterations of the model is extended from the 10k to 20k, and the convergence performance of the model is more fully evaluated. We also visualized the generated images of the model during the iterations. And we compared the models before and after adding supervised learning loss, and compared the gap of convergence speed between the two. Refer to Fig. 5 and Section 4.5 for details.
6. For CAM, we visualize the results of CAM, and show the changes of the attention distribution map in early and late training. Refer to Fig. 6 and Section 4.5 for details.

Acknowledgements This work was supported in part by the National Key Research and Development Program of China (Grant No. 2019YFA0706200), in part by the National Natural Science Foundation of China (Grant No.61632014, No.61627808), in part by the Natural Science Foundation of Gansu Province, China (Grant No. 22JR5RA488). Supported by Supercomputing Center of Lanzhou University.

References

1. Ning Z, Dong P, Wang X, Hu X, Guo L, Hu B, Guo Y, Qiu T, Kwok RY (2020) Mobile edge computing enabled 5G health monitoring for internet of medical things: A decentralized game theoretic approach. *IEEE J Sel Areas Commun* 39(2):463–478
2. Sun K, Xin Y, Ma Y, Lou M, Qi Y, Zhu J ASU-NET: U-shape adaptive scale network for mass segmentation in mammograms. *J Intell Fuzzy Syst* (Preprint), 1–16
3. Yang M, Ma Y, Liu Z, Cai H, Hu X, Hu B (2021) Undisturbed mental state assessment in the 5G era: a case study of depression detection based on facial expressions. *IEEE Wirel Commun* 28(3):46–53
4. Schneiderman H (1990) The funduscopic examination. *Clinical Methods: The History, Physical, and Laboratory Examinations*
5. Amemiya T, Bhutto I (2001) Retinal vascular changes and systemic diseases: corrosion cast demonstration. *Ital J Anat Embryol Arch Ital Anat Embriol* 106(2 Suppl 1):237–244

6. Ning Z, Sun S, Wang X, Guo L, Wang G, Gao X, Kwok RY (2021) Intelligent resource allocation in mobile blockchain for privacy and security transactions: a deep reinforcement learning based approach. *Sci China Inf Sci* 64(6):1–16
7. Yang T, Wu T, Li L, Zhu C (2020) SUD-GAN: deep convolution generative adversarial network combined with short connection and dense block for retinal vessel segmentation. *J Digit Imaging* 33(4):946–957
8. Wu C, Zou Y, Yang Z (2019) U-GAN: generative adversarial networks with U-Net for retinal vessel segmentation. In: 2019 14th international conference on computer science & education (ICCSE). IEEE, pp 642–646
9. Ning Z, Sun S, Wang X, Guo L, Guo S, Hu X, Hu B, Kwok R (2021) Blockchain-enabled intelligent transportation systems: a distributed crowdsensing framework. *IEEE Transactions on Mobile Computing*
10. McClintic BR, McClintic JI, Bisognano JD, Block RC (2010) The relationship between retinal microvascular abnormalities and coronary heart disease: a review. *Am J Med* 123(4):374–1
11. Durugkar I, Gemp I, Mahadevan S (2016) Generative multi-adversarial networks. arXiv:[1611.01673](https://arxiv.org/abs/1611.01673)
12. Iizuka S, Simo-Serra E, Ishikawa H (2017) Globally and locally consistent image completion. *ACM Trans Graph (ToG)* 36(4):1–14
13. Demir U, Unal G (2018) Patch-based image inpainting with generative adversarial networks. arXiv:[1803.07422](https://arxiv.org/abs/1803.07422)
14. Wang T-C, Liu M-Y, Zhu J-Y, Tao A, Kautz J, Catanzaro B (2018) High-resolution image synthesis and semantic manipulation with conditional GANs. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8798–8807
15. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision, pp 618–626
16. Mao C, Yao L, Luo Y (2022) ImageGCN: Multi-relational image graph convolutional networks for disease identification with chest x-rays. *IEEE Transactions on Medical Imaging*
17. Rao H, Xu S, Hu X, Cheng J, Hu B (2021) Augmented skeleton based contrastive action learning with momentum LSTM for unsupervised action recognition. *Inform Sci* 569:90–109
18. Schonfeld E, Schiele B, Khoreva A (2020) A U-Net based discriminator for generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 8207–8216
19. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. *J Big Data* 6(1):1–48
20. Pham DL, Xu C, Prince JL (2000) Current methods in medical image segmentation. *Annu Rev Biomed Eng* 2(1):315–337
21. Campadelli P, Casiraghi E, Esposito A (2009) Liver segmentation from computed tomography scans: a survey and a new algorithm. *Artif Intell Med* 45(2-3):185–196
22. Dou Q, Chen H, Jin Y, Yu L, Qin J, Heng P-A (2016) 3D deeply supervised network for automatic liver segmentation from CT volumes. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 149–157
23. Vlachos M, Dermatas E (2010) Multi-scale retinal vessel segmentation using line tracking. *Comput Med Imaging Graph* 34(3):213–227
24. Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 234–241
25. Yap MH, Pons G, Martí J, Ganau S, Sentís M, Zwiggelaar R, Davison AK, Martí R (2017) Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J Biomed Health Inform* 22(4):1218–1226
26. Huang L, Weng M, Shuai H, Huang Y, Sun J, Gao F (2016) Automatic liver segmentation from CT images using single-block linear detection. *BioMed Res Int* :2016
27. Dasgupta A, Singh S (2017) A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. In: 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017). IEEE, pp 248–251
28. Kingma DP, Welling M (2013) Auto-encoding variational Bayes. arXiv:[1312.6114](https://arxiv.org/abs/1312.6114)
29. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville AC, Bengio Y (2014) Generative adversarial nets. In: NeurIPS, pp 2672–2680
30. Wang X, Ning Z, Guo S, Wen M, Poor V (2021) Minimizing the age-of-critical-information: An imitation learning-based scheduling approach under partial observations. *IEEE Trans Mob Comput* :1–1. <https://doi.org/10.1109/TMC.2021.3053136>
31. Tong T, Li G, Liu X, Gao Q (2017) Image super-resolution using dense skip connections. In: ICCV, pp 4809–4817
32. Zhang R, Isola P, Efros AA (2016) Colorful image colorization. In: ECCV, vol 9907, pp 649–666
33. Tan WR, Chan CS, Aguirre HE, Tanaka K (2017) ArtGAN: Artwork synthesis with conditional categorical GANs. In: ICIP, pp 3760–3764
34. Tulyakov S, Liu M, Yang X, Kautz J (2018) MoCoGAN: Decomposing motion and content for video generation. In: CVPR, pp 1526–1535
35. Brock A, Donahue J, Simonyan K (2019) Large scale GAN training for high fidelity natural image synthesis. In: ICLR
36. Radford A, Metz L, Chintala S (2016) Unsupervised representation learning with deep convolutional generative adversarial networks. In: ICLR
37. Karras T, Laine S, Aila T (2019) A style-based generator architecture for generative adversarial networks. In: IEEE conference on computer vision and pattern recognition, CVPR 2019, Long Beach, CA, USA, June 16–20, 2019, pp 4401–4410
38. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2016) Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2921–2929
39. Miyato T, Kataoka T, Koyama M, Yoshida Y (2018) Spectral normalization for generative adversarial networks. arXiv:[1802.05957](https://arxiv.org/abs/1802.05957)
40. Kim J, Kim M, Kang H, Lee K (2019) U-GAT-IT: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. arXiv:[1907.10830](https://arxiv.org/abs/1907.10830)
41. Staal JJ, Abramoff MD, Niemeijer M, Viergever MA, van Ginneken B (2004) Ridge based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging* 23(4):501–509
42. Azzopardi G, Strisciuglio N, Vento M, Petkov N (2015) Trainable cosfire filters for vessel delineation with application to retinal images. *Med Image Anal* 19(1):46–57
43. Zhao Y, Rada L, Chen K, Harding SP, Zheng Y (2015) Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images. *IEEE Trans Med Imaging* 34(9):1797–1807
44. Chen Y (2017) A labeling-free approach to supervising deep neural networks for retinal blood vessel segmentation. arXiv:[1704.07502](https://arxiv.org/abs/1704.07502)
45. Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, Barman SA (2012) An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans Biomed Eng* 59(9):2538–2548
46. Hu K, Zhang Z, Niu X, Zhang Y, Cao C, Xiao F, Gao X (2018) Retinal vessel segmentation of color fundus images using

- multiscale convolutional neural network with an improved cross-entropy loss function. Neurocomputing 309:179–191
47. Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK (2018) Recurrent residual convolutional neural network based on u-net (R2U-Net) for medical image segmentation. arXiv:[1802.06955](https://arxiv.org/abs/1802.06955)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.