Retinal Vessel Segmentation Using Multi-scale Generative Adversarial Network with Class Activation Mapping

Minqiang Yang, Yinru Ye, Kai Ye, Xiping Hu(✉), and Bin Hu(✉)

Lanzhou University, Lanzhou 730099, Gansu, China
{yangmq,yeyr18,kye18,huxp,bh}@lzu.edu.cn

Abstract. Retinal vessel segmentation plays a significant role in the accurate diagnosis of retinal diseases. However, existing methods commonly omit micro-vessels in retinal images and generate some false-positive vessels. To alleviate this issue, we propose a multi-scale generative adversarial network with class activation mapping to achieve efficient segmentation. For the problem of small amount of data, we introduce a novel data augmentation method, which can generate multiple samples by cutting pixels from other samples. This method increases the diversity of samples and improve the robustness of the model. We compare our method with previous models with several metrics, and experiments show the superiority and effectiveness of our model.

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

1 Introduction

Retinal vessel segmentation has been a longstanding topic in medical image. In the treatment and evaluation of retinal diseases, such as macular degeneration, retinitis pigmentosa [26], the segmentation of retinal vessels in the fundus image is essential. However, traditional medical image processing mainly relies on the personal experience of experts to manually analyze and process images. This manual processing method is inefficient and subjectively affects the analysis results. Therefore, it is imperative for the computer to automatically process the retinal vessel segmentation image.

The methods of retinal segmentation consists of supervised learning and unsupervised learning. Supervised learning requires training through the original fundus images and the images of blood vessels manually labeled by experts. Common supervised learning methods applied to retinal segmentation include support vector machine (SVM) [16], random forest (RF) [30], and multilayer perceptron (MLP) [21]. Unsupervised methods do not require manual label to train the model, such as vessel tracking [36,38], template matching [5,12,31],

X. Gao et al. (Eds.): MobiHealth 2021, LNICST 440, pp. 95–105, 2022.
https://doi.org/10.1007/978-3-031-06368-8_7

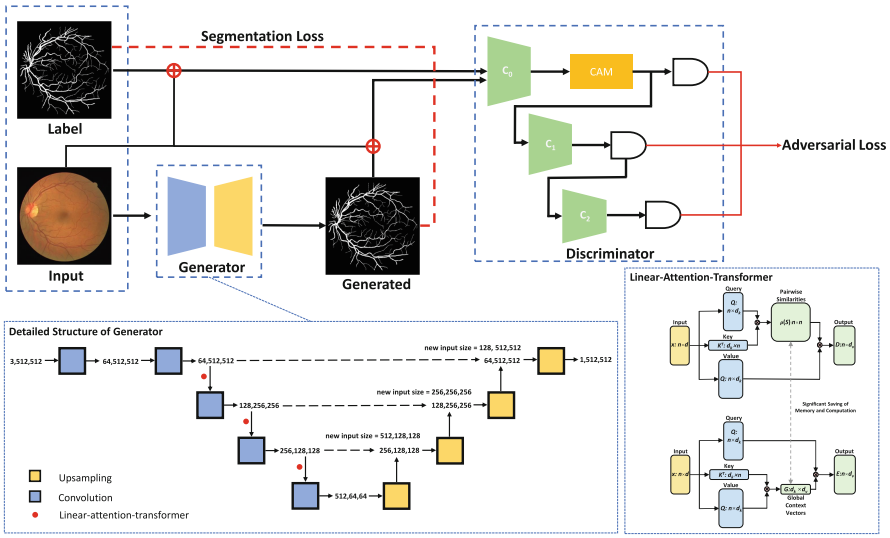

Fig. 1. Structure of our model.

multi-scale analysis [4,15,37], region growth [18]. Olaf et al. proposed U-net methods [14] which has greatly promoted the field of image segmentation, and has achieved great success in this field. However, although the image generated by the existing method has a good value on the metrics, the generated image is full of fragmented blood vessels, which is different from the real blood vessels. For doctors, this is not conducive to auxiliary diagnosis of medical diseases. But it is very challenging to generate images that are almost the same as the real blood vessel distribution, we propose a method that combines generative adversarial network (GAN) and end-to-end network generation. Inspired by locality-aware attention mechanism [22], we add class activation mapping (CAM) to network.

The existing models also have a problem that data collection and labeling of retinal blood vessels are time-consuming and labor-intensive and the amount of data is scarce compared to most tasks. Inspired by data augmentation method proposed by Haocong Rao et al. [23] and Phil Wang [29], we introduce a new data augmentation method called image stitching. this method cut a part of the area but randomly fill in the pixels of other data in the training set. The cropped and filled data can be flipped at a certain angle. Because the location, rotation angle and size of the grafting area are different, only one sample can produce infinite samples. This method increases the diversity of samples and reduces the occurrence of over-fitting.

Finally, we analyze the superiority of our model from a qualitative and quantitative perspective. For quantitative evaluation, we compared the evaluation indicators with other models and showed the changes of the indicators during the training process. For the qualitative point of view, we visualized the attention mechanism and the real classification result.

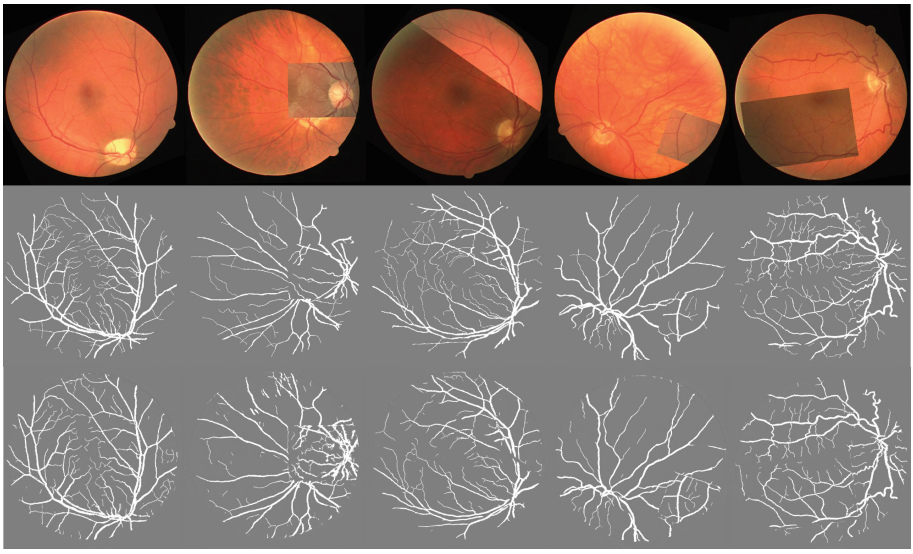

Fig. 2. Image stitching. The top row are eye fundus images after image stitching; the middle row are manual label; the bottom row are results obtained by our method.

2 Related Work

2.1 Image Segmentation

Image segmentation refers to finding the boundary of the region of interest (ROI) in the image, so that the pixels inside and outside the boundary have similar characteristics (intensity, texture, etc.). Medical image segmentation is the basis of other medical image processing. Medical image segmentation is usually used in the following use cases: liver segmentation for computed tomography (CT) images, breast-lesion segmentation, and retinal vessel segmentation. Kumar, S. S. proposed a method which can achieve automatic segmentation of liver and lesion from CT images needed for computer-aided diagnosis of liver [13]. Moi Hoon Yap used convolutional neural networks to detect breast ultrasound lesions automatically [35]. Tiejun Yang finded that SUD-GAN has a good performance in the field of retinal vessel segmentation [19].

2.2 Generative Adversarial Network

GAN [8] has been applied in many fields, such as vessel generation [34], artwork generation [25], and video generation [28]. It can achieve the improvement of image quality [27], image coloring [39]. There are several approaches for GAN to improve the authenticity of generated vessel images from different perspectives. The first one is large-scale training relies on complex calculations(e.g. BIGGAN [3] generated realistic images by increasing batchsize and

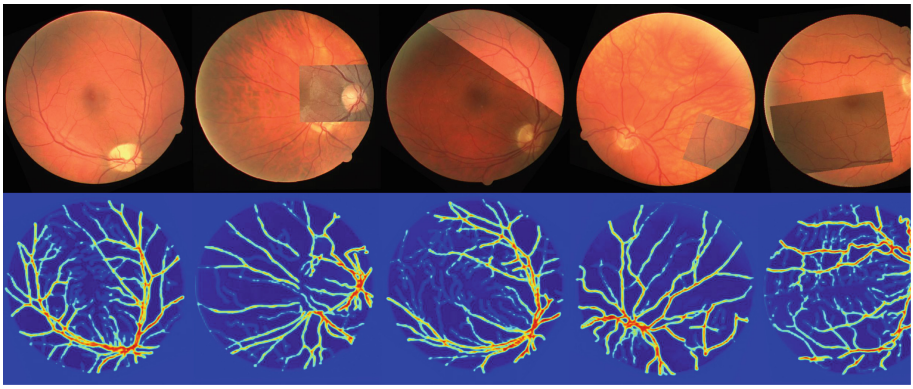

Fig. 3. Class activation mapping. The first row are eye fundus images; the second row are eye fundus images with class activation mapping.

truncation techniques). The second approach is to enhance the training stability (e.g. DCGAN [20] proposed a network structure consisting of stride convolution and transposed convolution to improve training stability). The third one is architectural modifications (e.g. AMCGC-LSTM [33] put spatial attention module into the network).

3 Our Approach

3.1 Discriminator

Compared with the traditional model [10], we use a multi-scale discriminator, which can improve the perceptual ability of the discriminator. It can better distinguish the fine blood vessels and the overall blood vessel distribution. Compared with the previous methods of applying discriminators of different scales to images of different sizes, we present a more direct method to superimpose the input image and segmented images and transmit them to the discriminator. The discriminator passes the input features of different layers to the corresponding size classifier for distinction. As shown in Fig. 1, the discriminator has two components: class activation mapping [41] and classifier. The C1 representation can accept a 128×128 receptive field, and the C2 representation can accept a 256×256 receptive field. The input image and label and the generated segmentation image are respectively used as the input of C0, and then are down-sampled by C1 after passing through CAM [41]. The same is true for C2. For a pair of images (eye fundus images and blood vessel segmentation image), C0, C1, and C2 are all trained to predict the true and false of the image. In addition to the multi-scale effect that is conducive to segmentation, we also use an attention mechanism to allow the model to learn features in the image more efficiently. This attention mechanism was first proposed by CAM. The idea of CAM is to superimpose the weighted linear sum of these visual patterns in different spatial

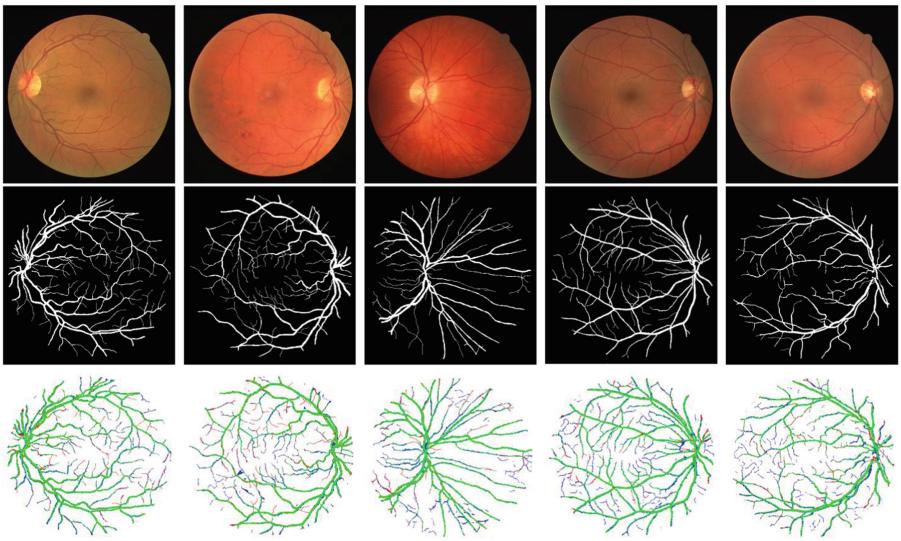

Fig. 4. Segmentation results of an example in the DRIVE dataset. The first row are eye fundus images; the second row are munual lable; the third row are visualization of real classification.

positions and get important information in the image on the input. It enables the model to efficiently focus on the distribution of blood vessels, thereby improving training efficiency.

3.2 Generator

The construction of the generator is inspired by previous work. In the previous work [2], the auto-encoder network constitutes a highly efficient method of dense prediction. Especially U-net can excel in many complex image segmentation tasks. In these image classification networks, basically follow a pattern. The encoder gradually downsamples the input to learn the global features of the input image, and the decoder gradually upsamples to make the output resolution consistent with the input. The two are connected by layer jumps. Data transmission in the same resolution layer can further improve the ability of the network to accurately segment. As shown in Fig. 1, we reconstructed the generator through the residual network, and added linear attention transformation [11] on the original basis. Linear attention transformation can discard information that is irrelevant to blood vessels in the process of downsampling.

3.3 Image Stitching

Since there are often problems in medicine such as small sample size and difficulty in collecting, we propose a data augmentation model to alleviate this

problem. Image stitching is a data augmentation method that can increase sample diversity and model robustness. It cuts one picture, embeds it in another picture, and then performs size change, rotation, and mirroring. The details of these processing steps can be seen in Algorithm 1. As shown in Fig. 2, the first row is the data after image stitching, the second row is the real label after image stitching, and the third row is the image after model segmentation. It can be seen that our method can generate a large number of samples, thereby reducing over-simulation.

Algorithm 1: Image stitching

Data: All original images I and tags of all original images L
Result: The stitched image *StitchedImage* and the corresponding label *StitchedLabel*

```
OriginalImage0, OriginalLabel0 ← RandomChoice(I, L);
OriginalImage1, OriginalLabel1 ← RandomChoice(I, L);
StartPoint ← RandomChoice(Image.Height, Image.Width);
EndPoint ← RandomChoice(Image.Height, Image.Width);
StartPoint, EndPoint ← Order(StartPoint, EndPoint);
OriginalImage0 ← Crop(OriginalImage0, StartPoint, EndPoint);
OriginalLabel0 ← Crop(OriginalLabel0, StartPoint, EndPoint);
StitchedImage ← Paste(OriginalImage1, OriginalImage0, StartPoint);
StitchedLabel ← Paste(OriginalLabel1, OriginalLabel0, StartPoint);
Spin ← RandomChoice(360);
Flip ← RandomChoice(1);
StitchedImage ← Rotate(StitchedImage, Spin);
StitchedLabel ← Rotate(StitchedLabel, Spin);
if Flip == True then
    StitchedImage ← Transpose(StitchedImage);
    StitchedLabel ← Transpose(StitchedLabel);
end
```

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 is aim to guide the discriminator to distinguish the generated vessel image and ground truth. 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, it attempts to and make the distribution probability of the output of the generator approach the distribution of the target domain continuously.

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

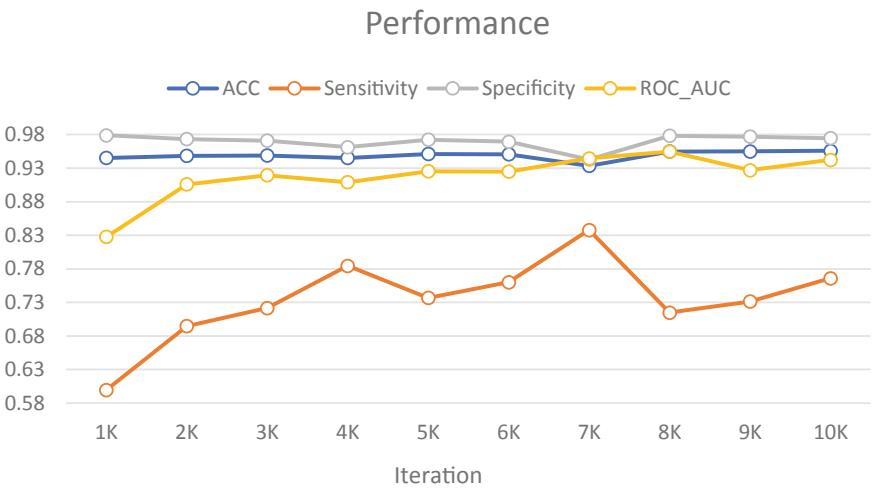

Fig. 5. Iteration performance.

Segmentation Loss. If you only apply the adversarial loss of unsupervised learning, the efficiency is actually low, so we have an additional end-to-end segmentation loss for supervised learning. Because the dataset has unbalanced characteristics and the traditional loss functions will not be able to learn well with features of few samples and difficult samples, we apply focal loss as the segmentation loss.

$$\min_{G} L_{seg}^{x \to y} = \mathbb{E}_{y \sim \mathcal{Y}}[L_{ft}(G(y))]. \tag{2}$$

To prevent an extreme imbalance between background and retinal vessels classes, we apply focal loss to adapt the weight of positive (vessels) and negative (background) samples. For weighting factor $\alpha \in [1,0]$, 1- α refers to as inverse class frequency to balance the significance of positive and negative samples.

$$L_{fl} = \begin{cases} -\alpha(1-\hat{y})^{\gamma} \log \hat{y}, y = 1 \\ -(1-\alpha)\hat{y}^{\gamma} \log(1-\hat{y}), y = 0 \end{cases} \tag{3}$$

4 Experiments

4.1 Dataset

In this article, we use the DRVIE dataset [24] for retinal vessel segmentation. The size of the images and the corresponding labels in the data set is 565×584. For data augmentation, all images will be processed by image stitching, rotation, mirroring and other operations, and finally will be cropped to a size of 512×512. The data set is divided into two parts for training set and testing set respectively, with a sample size ratio of 20:20. The metrics and results of our model in this article have been iteratively trained 10,000 times.

4.2 Evaluation Metrics

For quantitative evaluation, we choose four evaluation metrics: accuracy(ACC), area under curve of receiver operating characteristic (AUC-ROC), sensitivity and specificity. ACC and AUC-ROC are very authoritative and intuitive indicators in classification algorithms, which can evaluate the effect of segmentation. In the field of medical images, sensitivity and specificity are very important, they can well evaluate the superiority of image segmentation algorithms. Accuracy is a metric to calculate the ratio between pixels which are classfied correctly and total pixels in the dataset. Sensitivity, also called recall, is a metric to measures the proportion of all predicted positive samples to all positive samples. Specificity indicates the proportion of the true negative samples among all the predicted negative samples.

Table 1. Comparison with baselines. Value of metric is higher, model is better.

Methods	DRIVE			
	ACC	Sensitivity	Specificity	AUC - ROC
Odstrcilik et al. [17]	0.9341	0.7847	0.9512	0.9569
Azzopardi et al. [1]	0.9614	0.7655	0.9704	0.9614
YT Zhao et al. [40]	0.9540	0.7420	0.9820	0.8620
C Wu et al. [32]	0.9514	0.7696	0.9780	0.8909
G Azzopardi et al. [1]	0.9442	0.7655	0.9704	0.9614
Y Chen [6]	0.9453	0.7426	0.9735	0.9516
MM Fraz et al. [7]	0.9480	0.7406	0.9807	0.9747
K Hu et al. [9]	0.9521	0.7779	0.9780	0.9782
Ours	0.9559	0.7659	0.9746	0.9424

4.3 Comparisons

As shown in Table 1, we use quantitative analysis to evaluate our model. It can be clearly seen that our model has a good effect on various metrics. The compared model may be more focused on a certain metric, and we can complete the segmentation task while ensuring that the overall metrics are better. Especially for the sensitivity and specificity, from the evaluation results it attests our model can ensure the segmentation of real blood vessels and can be helpful for the auxiliary diagnosis of medical diseases in practical sense.

4.4 Result

In Fig. 5, we can observe the change of each metric as the number of iterations increases. As shown in Fig. 3, we use the CAM, which makes blood vessels be recognized by neural networks, to make the model more efficient to segment blood vessels. Red means that the model pays more attention to this area, and

blue means that the model pays less attention to this area. Through the image generated by the attention mechanism in the bottom row, we can clearly see that the attention of model is placed on the blood vessels. And for other irrelevant factors such as background information, our model are not paid much attention to, which allows the generator able to generate better images. For the second row of Fig. 4, it's easily to observe the predicted result of our model. The green blood vessel represents the correctly predicted blood vessel (TP), the red vessel represents the real blood vessel that was not predicted (FN), and the blue vessel represents the non-existent blood vessel generated by the generator (FP). It means our model can clearly identify the main vessels, and there are very few unidentified and incorrectly identified vessels. It proves the superiority of the model.

5 Conclusion

We propose an architecture for retinal vessel segmentation. It can distinguish between global and local perception domains through a multi-scale CAM discriminator, and the generator can also use linear attention to transform more efficient blood vessel segmentation. In addition, we also introduce a data augmentation method called image stitching, so that the model can accept a sufficient variety of images. The evaluation of vision and metrics can show that our model has excellent performance in retinal blood vessel segmentation. It proves that this method achieves fast and accurate blood vessel segmentation.

Acknowledgement. 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, No. 61802159), in part by Fundamental Research Funds for Central Universities (lzujbky-2019-26).

References

1. Azzopardi, G., Strisciuglio, N., Vento, M., Petkov, N.: Trainable COSFIRE filters for vessel delineation with application to retinal images. Med. Image Anal. **19**(1), 46–57 (2015)
2. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
3. Brock, A., Donahue, J., Simonyan, K.: Large scale GAN training for high fidelity natural image synthesis. In: ICLR (2019)
4. Budai, A., Bock, R., Maier, A., Hornegger, J., Michelson, G.: Robust vessel segmentation in fundus images. Int. J. Biomed. Imaging (2013)
5. Chaudhuri, S., Chatterjee, S., Katz, N., Nelson, M., Goldbaum, M.: Detection of blood vessels in retinal images using two-dimensional matched filters. IEEE Trans. Med. Imaging **8**(3), 263–269 (1989)
6. Chen, Y.: A labeling-free approach to supervising deep neural networks for retinal blood vessel segmentation. arXiv preprint arXiv:1704.07502 (2017)

7. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Trans. Biomed. Eng. **59**(9), 2538–2548 (2012)
8. Goodfellow, I.J., et al.: Generative adversarial nets. In: NeurIPS, pp. 2672–2680 (2014)
9. Hu, K., et al.: Retinal vessel segmentation of color fundus images using multi-scale convolutional neural network with an improved cross-entropy loss function. Neurocomputing **309**, 179–191 (2018)
10. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Trans. Graph. (ToG) **36**(4), 1–14 (2017)
11. Kitaev, N., Kaiser, L., Levskaya, A.: Reformer: the efficient transformer. In: ICLR (2020)
12. Kovács, G., Hajdu, A.: A self-calibrating approach for the segmentation of retinal vessels by template matching and contour reconstruction. Med. Image Anal. **29**, 24–46 (2016)
13. Kumar, S., Moni, R., Rajeesh, J.: Automatic liver and lesion segmentation: a primary step in diagnosis of liver diseases. SIViP **7**(1), 163–172 (2013). https://doi.org/10.1007/s11760-011-0223-y
14. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
15. Moghimirad, E., Rezatofighi, S.H., Soltanian-Zadeh, H.: Retinal vessel segmentation using a multi-scale medialness function. Comput. Biol. Med. **42**(1), 50–60 (2012)
16. Noh, K.J., Park, S.J., Lee, S.: Scale-space approximated convolutional neural networks for retinal vessel segmentation. Comput. Methods Programs Biomed. **178**, 237–246 (2019)
17. Owen, C.G., et al.: Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program. Invest. Ophthal. Vis. Sci. **50**(5), 2004–2010 (2009)
18. Palomera-Perez, M.A., Martinez-Perez, M.E., Benitez-Perez, H., Ortega-Arjona, J.L.: Parallel multiscale feature extraction and region growing: application in retinal blood vessel detection. IEEE Trans. Inf. Technol. Biomed. **14**(2), 500–506 (2009)
19. Park, K.B., Choi, S.H., Lee, J.Y.: M-GAN: retinal blood vessel segmentation by balancing losses through stacked deep fully convolutional networks. IEEE Access **8**, 146308–146322 (2020)
20. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: ICLR (2016)
21. Rahebi, J., Hardalaç, F.: Retinal blood vessel segmentation with neural network by using gray-level co-occurrence matrix-based features. J. Med. Syst. **38**(8), 1–12 (2014). https://doi.org/10.1007/s10916-014-0085-2
22. Rao, H., et al.: A self-supervised gait encoding approach with locality-awareness for 3d skeleton based person re-identification. IEEE Trans. Pattern Anal. Mach. Intell. (2021)
23. Rao, H., Xu, S., Hu, X., Cheng, J., Hu, B.: Augmented skeleton based contrastive action learning with momentum LSTM for unsupervised action recognition. Inf. Sci. **569**, 90–109 (2021)
24. Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M.A., Van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. IEEE Trans. Med. Imaging **23**(4), 501–509 (2004)

25. Tan, W.R., Chan, C.S., Aguirre, H.E., Tanaka, K.: ArtGan: artwork synthesis with conditional categorical GANs. In: ICIP, pp. 3760–3764 (2017)
26. Teng, T., Lefley, M., Claremont, D.: Progress towards automated diabetic ocular screening: a review of image analysis and intelligent systems for diabetic retinopathy. Med. Biol. Eng. Compu. **40**(1), 2–13 (2002). https://doi.org/10.1007/BF02347689
27. Tong, T., Li, G., Liu, X., Gao, Q.: Image super-resolution using dense skip connections. In: ICCV, pp. 4809–4817 (2017)
28. Tulyakov, S., Liu, M., Yang, X., Kautz, J.: Mocogan: Decomposing motion and content for video generation. In: CVPR, pp. 1526–1535 (2018)
29. Wang, P.: Linear attention transformer (2020). https://github.com/lucidrains/linear-attention-transformer
30. Wang, S., Yin, Y., Cao, G., Wei, B., Zheng, Y., Yang, G.: Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. Neurocomputing **149**, 708–717 (2015)
31. Wang, Y., Ji, G., Lin, P., Trucco, E.: Retinal vessel segmentation using multiwavelet kernels and multiscale hierarchical decomposition. Pattern Recogn. **46**(8), 2117–2133 (2013)
32. Wu, C., Zou, Y., Yang, Z.: U-Gan: generative adversarial networks with u-net for retinal vessel segmentation. In: 2019 14th International Conference on Computer Science and Education (ICCSE), pp. 642–646. IEEE (2019)
33. Xu, S., et al.: Attention based multi-level co-occurrence graph convolutional LSTM for 3D action recognition. IEEE Internet Things J. (2020)
34. Yang, T., Wu, T., Li, L., Zhu, C.: 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 (2020). https://doi.org/10.1007/s10278-020-00339-9
35. Yap, M.H., et al.: Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J. Biomed. Health Inform. **22**(4), 1218–1226 (2017)
36. Yin, Y., Adel, M., Bourennane, S.: Retinal vessel segmentation using a probabilistic tracking method. Pattern Recogn. **45**(4), 1235–1244 (2012)
37. Yu, H., Barriga, S., Agurto, C., Zamora, G., Bauman, W., Soliz, P.: Fast vessel segmentation in retinal images using multi-scale enhancement and second-order local entropy. In: Medical imaging 2012: computer-aided diagnosis, vol. 8315, p. 83151B. International Society for Optics and Photonics (2012)
38. Zhang, J., Li, H., Nie, Q., Cheng, L.: A retinal vessel boundary tracking method based on Bayesian theory and multi-scale line detection. Comput. Med. Imaging Graph. **38**(6), 517–525 (2014)
39. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 649–666. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_40
40. Zhao, Y., Rada, L., Chen, K., Harding, S.P., Zheng, Y.: 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 (2015)
41. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2921–2929 (2016)