U-Net with Attention Mechanism for Retinal Vessel Segmentation

Ze Si, Dongmei Fu(✉), and Jiahao Li

School of Automation and Electrical Engineering,
University of Science and Technology Beijing, No.30 Xueyuan Road,
Haidian District, Beijing, China
fdm2003@163.com

Abstract. Retinal vessel is the only vessel which can be observed directly, retinal vessel analysis is a crucial method for the screening and diagnosis of related diseases. In this paper, we propose a retinal vessel segmentation method based on U-Net and attention mechanism. Fully convolutional network (FCN) like U-Net have excellent performance on segmentation tasks, but there are problems for it to build long-range dependencies amoung different part of images because convolution layers extract features in local area, local feature based methods can lead to mistake in some segmentation scenes. In this paper, attention mechanism is used to solve this problem, and a new attention module is proposed, with two different attention module, long-range dependencies in different part of the image can be built efficiently. The proposed method was evaluated on DRIVE dataset, experiment result demonstrate that proposed method have better performance than the state-of-art methods.

Keywords: Segmentation · Deep learning · U-Net · Retinal vessel · Attention mechanism

1 Introduction

Retinal vessels are branches of the cerebral blood vessels, main function of retinal vessels is to provide nutrition to the retina. Retinal vessels at bottom of the eye are the only non-invasive parts of the vascular system, features of retinal vessels like width, angle and branch morphology can be used as a basis for diagnosis of vascular-related diseases. Ophthalmological blindness diseases such as glaucoma and diabetic retinopathy, etc. can be directly observed from retinal vasculopathy. So, retinal vessels segmentation is indispensable in clinical diagnosis. Figure 1 shows a retinal image and retinal vessel segmented by expert. Manual segmentation of retinal vessels is time consuming, due to the growing

Supported by Beijing Engineering Research Center of Industrial Spectrum Imaging, School of Automation and Electrical Engineering, University of Science and Technology Beijing (No.BG0150).

Y. Zhao et al. (Eds.): ICIG 2019, LNCS 11902, pp. 668–677, 2019.
https://doi.org/10.1007/978-3-030-34110-7_56

number of patients with retinal vasculopathy and lacking of trained specialists, an effective and precise method for retinal vessel segmentation is meaningful in clinical.

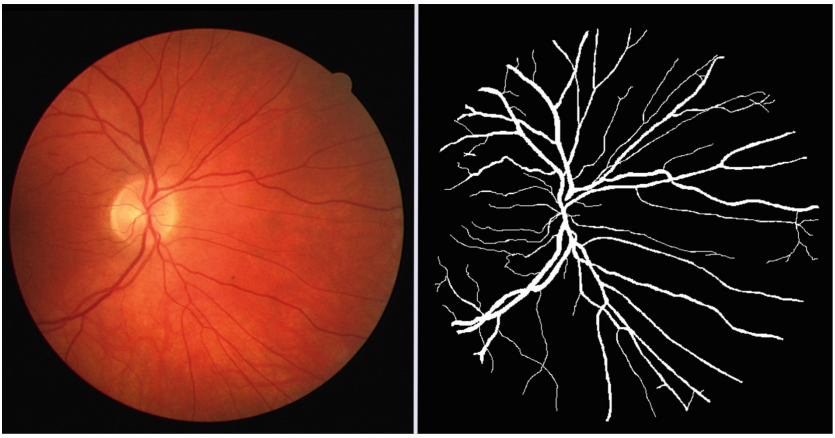

Fig. 1. Retinal images and vessel area's pixel level label in DRIVE database.

Reasearch of retinal vessel segmentation method has attract researchers in cv field. Deep learning method showed great performance in semantic segmentation task these years, some researches have tried related method on retinal vessle segmentation, for example, Hu et al. [1] proposed a method based on FCN for retinal vessel segmentation which perform state-of-the-art. Although deep learning method like FCN have great performance on related tasks, there are still some limitations for normal FCN model on retinal vessel segmentation task. Convolution operator has a local receptive field, and long-range dependencies in different regions of the image can be processed only after passing through many convolutional layers. Therefore, small scale neural network will not be able to learn and represent long-range dependencies across different image regions. Even for deep model, there's still some problems prevent the model from learning long-range dependencies. For example, optimization algorithms may have trouble discovering parameter values that carefully coordinate multiple layers to capture these dependencies, and these parameterizations may be statistically brittle and prone to failure when applied to previously unseen inputs [2].

There's less semantic information in retinal images compared with images in other complex tasks, in some cases, it is difficult to distinguish pixels belong to the vessels and some retinal tissues by local features solely. Therefore, it is necessary to make use of global information and establish long-range dependencies in different regions of the image.

U-Net is one of the most common segmentation model for medical image segmentation. U-Net is an encoder-decoder model with skip connection structure, encoder extract features, decoder maps the low-resolution features to the

high-resolution space, skip connection structure fuse multi-features to enhance the details of the segmentation. Based on the improved U-Net, in this paper, attention modules are added to encoders and decoders to build long-range dependencies by global information.

Improved U-Net with attention modules in this paper is an end-to-end method. Different from previous deep learning method, proposed method established position-wise and channel-wise long-range dependencies. By the use of global information, proposed method performed better than the state-of-the-art methods.

2 Related Work

2.1 Segmentation Model

In the early years, fully connected neural networks was the first ANN used for segmentation task, features around each pixel are extracted and fully connected neural network work as a classifier to classify all the pixels. With the development of neural networks, convolutional neural network (CNN) replaced fully connected model in segmentation problem. Patch centered on the pixel is feed into the model to extract features, and a fully connected layer is trained as a classifier. Application of CNN avoids feature extraction process, but repeated storage and redundant convolution computation caused by overlapping image patches makes it time consuming and inefficiency [3]. Long et al. [4] proposed fully convolutional network (FCN), which is more precise and efficient, from then on, almost all semantic segmentation studies have adopted this basic structure. Deeplab series model [5–7] showed impressive results in the semantic segmentation tasks by the use of dilated convolution, conditional random field, ASPP and other techniques. U-net [8] is widely used in medical image segmentation, with dense encoder-decoder and skip-connection structure, it have great advantage in clinical image segmentation.

2.2 Attention Mechanism

Attention mechanism was proposed by Benigio et al. [9]. Similar to human attention, attention mechanism intended to screen the high value information that is most useful for current task in the overall information received. Vaswani et al. [10] proposed a method for establishing global dependencies of input information using self-attention mechanism and applied this in machine translation. Attention mechanism was widely used in the field of NLP in the early years. In recent years, attention mechanism has attracted researchers in CV field. Wang et al. used attention module to establish the temporal and spatial dependencies of video sequences. This method had greatly improved the video classification performance [11]. Zhang et al. introduced self-attention mechanism in GAN to generate consistent scenes using complementary features of images [2].

Retinal vessel segmentation methods based on FCN in the published works improve the segmentation performance by feature fusion, but during the feature

extraction process and mapping process, global features are not used and relationship between different features is not considered. So in this paper, a retinal vessel segmentation method with attention mechanism is proposed.

3 Method

3.1 Model Structure

The basic segmentation network in this paper is an improved U-net. For an input image, the model outputs a probability map which indicate the probability of all the pixels belong to the vessel area. A module called channel attention module is proposed, and it is added after each skip-connection structure, in second basic block we introduced a position attention module based similar to the non-local model proposed by Wang et al. [11]. Figure 2 shows the structure of proposed model.

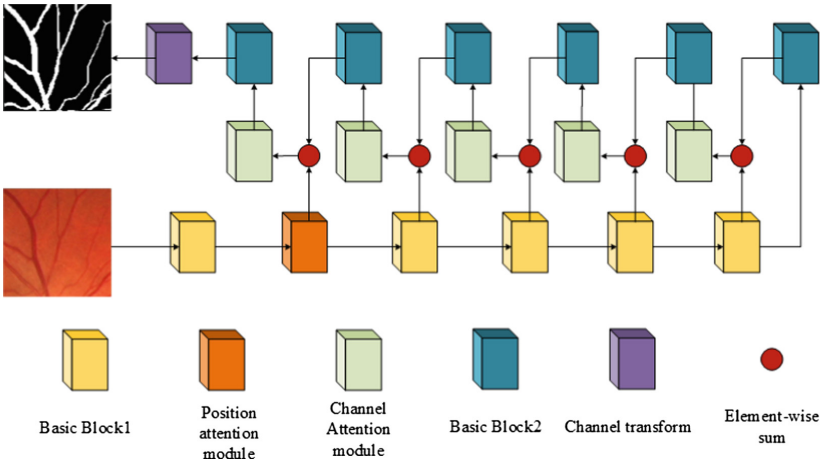

Fig. 2. Structure of proposed model.

Basic block1 in Fig. 2 consists of three residual blocks [12] and a convolutional layer with 3*3 kernel and 2 strides, basic block2 consists of three residual blocks and a deconvolutional layer. Compared with U-net in previous work, max-pooling is replaced by convolutional layer with 2 strides to reduce loss of information. Batch normalization layers [10] are used to reduce the risk of overfitting and simplify the training process. To reduce computing cost, concats in feature fusion stages are replaced by summation.There are two kinds of attention moudles in proposed model, position attention moudle and channel attention moudle, position attention moudle same as non-local moudle proposed by Wang [11] is used to build long-range dependencies of different regions of the feature map, channel attention moudle is used to build long-range dependencies of different channels of the feature map.

3.2 Channel Attention Mechanism

Figure 3 shows the basic idea of attention mechanism, each point of the feature map can be treated as a Query, points around the Query point are treated as Key, each S is calculated as the similarity of corresponding Key and Query, weights A is normalization of S weighted sum of Value is the attention value. In general, Key and Value is the same, A is Query element's encode vector, which contains Query's relationship with both local and global features. So attention module can capture both local and long-range dependencies.

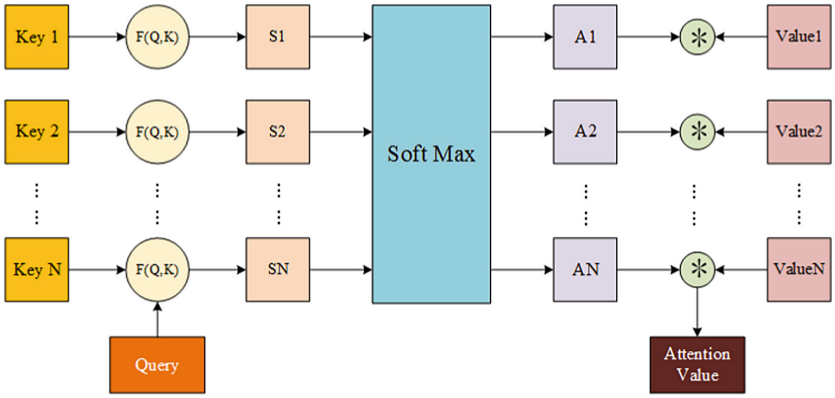

Fig. 3. Attention mechanism

For image semantic segmentation task, each channel of layer's output can be regarded as a response of certain semantic feature. Establishing dependencies between different semantic features is meaningful that interdependent features can be used to improve the feature's representation of specific semantics. Based on this motivation, channel attention mechanism is proposed in this paper.

Analogy the position attention module like non-local module, each channel of channel attention module's output is a weighted sum of all the input feature map's channels as shown in Fig. 4.

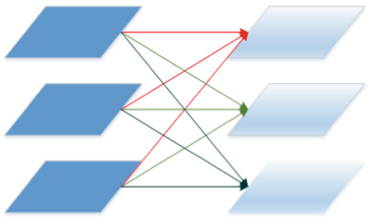

Fig. 4. Channel attention mechanism schematic diagram.

Channel attention mechanism proposed in this paper indicated as (5)

$$y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) x_j \tag{1}$$

x_i is the i^{th} channel of input feature map, y_i is the corresponding output channel. x_i and y_i are transformed into column vectors, $\{x_i, y_i\} \in R^N$, N is the size of each channel, f is the correlation measurement (6)

$$f(x_i, x_j) = e^{E((x_i - E(x_i))(x_j - E(x_j)))} \tag{2}$$

$E((x_i - E(x_i))(x_j - E(x_j)))$ is covariance of x_i and x_j, $E(x_i)$ is replaced by the mean value of x_i, f is expressed as (7) in this way.

$$f(x_i, x_j) = e^{\frac{(x_i - \bar{x_i})' (x_j - \bar{x_j})}{N}} \tag{3}$$

C is normalization coefficient like it in non-local module [11].

$$C(x) = \sum_{\forall j} f(x_i, x_j) \tag{4}$$

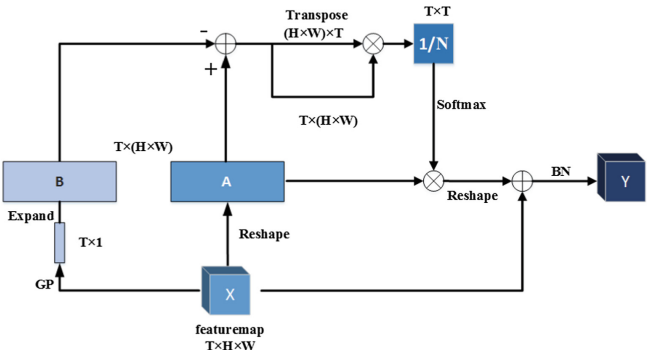

Fig. 5. Channel attention module.

The implementation process of channel attention module is shown in Fig. 5. Input feature map $X \in R^{T \times H \times W}$ is reshaped to $A \in R^{T \times N}$. The i^{th} row of A corresponds to the vector expansion from the i^{th} channel of original feature map. Expected value of the i^{th} channel is approximated by the mean value of it. A global pooling layer is used to calculate the mean value of each channel of original feature map X, the output of global pooling layer is expanded to $B \in R^{T \times N}$. The result of element-wise subtraction between A and B preform matrix multiplication with the transpose of itself and then each element is multiplied by the factor $1/N$ to get the covariance matrix for all the channels of input

feature map. Finally, a SoftMax layer is applied to get the channel attention map $CM \in R^{T \times T}$, each element in CM can be expressed as (9)

$$CM_{ij} = \frac{e^{\frac{\sum_{k=1}^{N}(A-B)_{ik}(A-B)'_{kj}}{N}}}{\sum_{w=1}^{T} e^{\frac{\sum_{k=1}^{N}(A-B)_{ik}(A-B)'_{kw}}{N}}} \tag{5}$$

Here CM_{ij} represents the correlation of the j^{th} channel and the i^{th} channel of input feature map. Result of a matrix multiplication between CM and A is reshaped and add back the input feature map to obtain Y.

3.3 Data Processing Method

During the training of deep model, a large number of training samples with labels are required. For retinal vessel segmentation task, training samples with labels is limited, data augmentation is necessary.

Medical images generally have large size, it is a common method to cut image into patches. In this way, all the patch can be treated as training samples. In this paper, attention mechanism is used to build position-wise and channel-wise long-range dependencies, we need sufficient information in a single patch, therefore, patches with 256 * 256 size are used, in this paper, all the images and labels are flip and rotated (30° each time), then sliding window with 256 * 256 size and 128 strides is used to cut the images into patches.

4 Experiments and Results

4.1 Materials

To demonstrate the performance of proposed method, we evaluated our method on public dataset DRIVE. DRIVE is consisted of 40 retinal images, in which 20 images in both training set and testing set, All the images have the same size of 565 * 584. To get more and larger patches, in the training stage, all the images are resized to 1695 * 1752, in testing stage, the segmentation result will be resized back to 565 * 584.

4.2 Result Comparison

Commonly used evaluation metrics for retinal vessel segmentation task are accuracy (ACC), sensitivity (SE), specificity (SP), and AUC. Figure 6 shows several segmentation results generated by proposed method, from which we can see proposed method have a good performance on both thick vessels and tiny vessels. To prove the validity of proposed method, comparation of proposed method with other methods in this years on DRIVE is shown in Table 1.

As shown in Table 1 proposed method performs better than other methods on most of the evaluate index. Among these evaluate metrics, Se, Sp and Acc are

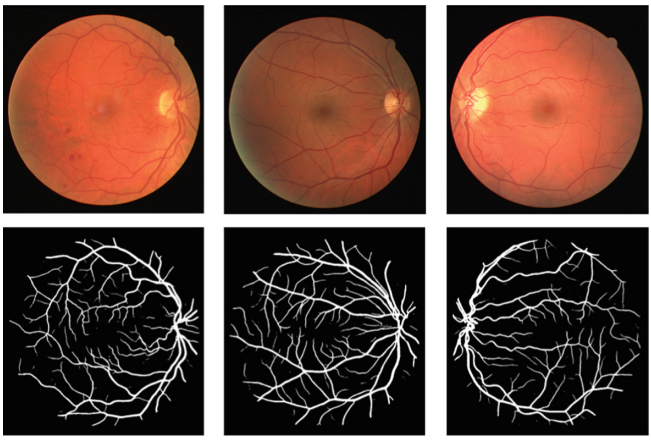

Fig. 6. Segmentation result

associated with select of threshold, Se and Sp can not measure the performance of model solely, Acc in class imbalance task is not an appropriate metric for model evaluation. Compared with other metrics, AUC not rely on the threshold, and it's a suitable metric to evaluate models for class imbalance tasks. As shown in Table 1, we got highest AUC which demonstrate the effectiveness of our proposed method.

Table 1. Performance comparision of proposed method on DRIVE

Method	Year	Se	Sp	Acc	AUC
Marin [13]	2011	0.7067	0.9801	0.9452	0.9588
Cheng [14]	2014	0.7252	0.9798	0.9474	0.9648
Roychowdhury [15]	2014	0.7250	0.9830	0.9520	0.9620
Wang [16]	2015	0.8173	0.9733	0.9533	0.9475
Azzopardi [17]	2015	0.7655	0.9704	0.9442	0.9614
Liskowski [18]	2016	0.7569	0.9816	0.9527	0.9738
Li [19]	2016	0.7569	0.9816	0.9527	0.9738
Dasgupta [20]	2017	0.7691	0.9801	0.9533	0.9744
Alom [21]	2018	0.7798	0.9813	0.9556	0.9784
Hu [1]	2018	0.7772	0.9793	0.9533	0.9759
Proposed	**2019**	**0.8156**	**0.9837**	**0.9687**	**0.9807**

5 Conclusion

In this paper, a U-Net with attention mechanism for retinal vessel segmentation is proposed. For the data imbalanced problem, a novel loss function called dice entropy loss function is used, that allowed the model to focus more on the vessel area. Comparative experiments show the efficiency of proposed method.

References

1. Kai, H., 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)
2. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-attention generative adversarial networks. arXiv preprint. arXiv:1805.08318 2018
3. Liu, Y., Dongmei, F., Huang, Z., Tong, H.: Optic disc segmentation in fundus images using adversarial training. IET Image Process. **13**(2), 375–381 (2018)
4. 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)
5. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062 (2014)
6. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 (2017)
7. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2018)
8. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
9. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
10. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
11. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7794–7803 (2018)
12. He, K., Zhang, X., Ren, S., Sun, J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
13. Marín, D., Aquino, S., Gegúndez-Arias, M.E., Bravo, J.M.: A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. IEEE Trans. Med. Imaging **30**(1), 146–158 (2011)
14. Cheng, E., Du, L., Wu, Y., Zhu, Y.J., Megalooikonomou, V., Ling, H.: Discriminative vessel segmentation in retinal images by fusing context-aware hybrid features. Mach. Vis. Appl. **25**(7), 1779–1792 (2014). https://doi.org/10.1007/s00138-014-0638-x

15. Roychowdhury, S., Koozekanani, D.D., Parhi, K.K.: Blood vessel segmentation of fundus images by major vessel extraction and subimage classification. IEEE J. Biomed. Health Inf. **19**(3), 1118–1128 (2014)
16. 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)
17. 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)
18. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. IEEE Trans. Med. Imaging **35**(11), 2369–2380 (2016)
19. Li, Q., Feng, B., Xie, L.P., Liang, P., Zhang, H., Wang, T.: A cross-modality learning approach for vessel segmentation in retinal images. IEEE Trans. Med. Imaging **35**(1), 109–118 (2016)
20. Dasgupta, A., Singh, S.: 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), pp. 248–251. IEEE (2017)
21. Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K.: Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955 (2018)