PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Fully automated segmentation of hyperreflective foci in OCT images using a U-shape network

Zhu, Liangjiu, Zhu, Weifang, Feng, Shuanglang, Chen, Xinjian

Liangjiu Zhu, Weifang Zhu, Shuanglang Feng, Xinjian Chen, "Fully automated segmentation of hyperreflective foci in OCT images using a U-shape network," Proc. SPIE 11313, Medical Imaging 2020: Image Processing, 1131308 (10 March 2020); doi: 10.1117/12.2548085

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

Fully automated segmentation of hyperreflective foci in OCT images using a U-shape network

Liangjiu Zhu [1,#] , Weifang Zhu [1,2#] , Shuanglang Feng [1] , Xinjian Chen [1,3,*]

[1]School of Electronics and Information Engineering, Soochow University, Suzhou 215006, China
[2]Collaborative Innovation Center of IoT Industrialization and Intelligent Production, Minjiang University, Fuzhou 350108, China
[3]State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou 215123, China

ABSTRACT

Diabetic retinopathy (DR), a highly specific vascular complication caused by diabetes, has been found a major cause of blindness in the world. Early screening of DR is crucial for prevention of vision loss. Hard exudates (HEs) is one of the main manifestations of DR, which is characterized by hyper-reflective foci (HF) in retinal optical coherence tomography(OCT) images. In this paper, a fully automated method based on U-shape network is proposed to segment HF in retinal OCT images. Compared with the original U-Net, there are two main improvements in the proposed network:(1) The ordinary 3×3 convolution is replaced by multi-scale convolution based on dilated convolution, which can achieve adaptive receptive fields of the images. (2) In order to ignore irrelevant information and focus on key information in the channels, the channel attention module is embedded in the model. A dataset consisting of 112 2D OCT B-scan images was used to evaluate the proposed U-shape network for HF segmentation with 4-fold cross validation. The mean and standard deviation of Dice similarity coefficient, recall and precision are **73.26±2.03%**, **75.71±1.98%** and **74.28±2.67%,** respectively**.** The experimental results show the effectiveness of the proposed method.

KEYWORDS: Hard exudations, hyper-reflective foci, U-Net, multi-scale convolution, channel attention

1. INTRODUCTION

Many studies have focused on hard exudates (HEs) segmentation in retinal fundus images, such as k-means based segmentation[1], thresholding based segmentation[2]. However, fundus image only provides two-dimensional information about the HEs. Because of the density and opacity, HEs are also distinguishable in retinal optical coherence tomography (OCT) images as hyper-reflective foci (HF), which can provide the three-dimensional information of HF such as shape, extent and spatial distribution. There is a strong positive correlation between HEs and HF in OCT images (Matthews correlation coefficient between HEs and HF is 0.94[3]). So automated segmentation of HF in OCT images is of great importance.

As is shown in Fig.1, the main characteristic of HF is that their intensity values are larger than the surrounding area. The main challenges for HF segmentation include the various sizes, irregular shapes and blurred boundaries. Moreover, the volume proportion between background and HF is very large, which causes the common and difficult class imbalance problem in medical image segmentation.

To the best knowledge of us, only reference [3] proposed a deep learning based method to segment HF in OCT images, which combines GoogleNet[4] and ResNet[5]. In this paper, we propose an improved U-shape network for HF segmentation in OCT images, which achieves good performances. We use a four-layer U-Net[6] model as the baseline. To deal with the problem of the varying sizes of objects, we propose multi-scale convolution modules (MSCM) and insert them into the encoder path, which can help the network achieve adaptive receptive fields. In order to guide the model to pay more attention to useful information in the channels, we add channel attention module (CAM) after each max-pooling and the result shows a great improvement compared with the baseline. Last but not least, we adopt data augmentation strategies, including random flips, random rotations, and Gaussian noise addition during training process, to increase the generalization of the model.

*Corresponding author: E-mail:xjchen@suda.edu.cn, # indicates these authors contributed equally to this work

Medical Imaging 2020: Image Processing, edited by Ivana Išgum, Bennett A. Landman,
Proc. of SPIE Vol. 11313, 1131308 · © 2020 SPIE · CCC
code: 1605-7422/20/$21 · doi: 10.1117/12.2548085

(a) (b)

Fig.1. Example of HF in retinal OCT image. (a) An original B-scan of retinal OCT image with HF (areas enclosed by a green rectangle). (b) The corresponding ground truth of HF.

2. METHODS

In this section the proposed method is introduced including the five parts: overall structure of the network, multi-scale convolution module, channel attention module, data augmentation strategy and loss function.

2.1 Overall structure of the network

U-Net[6], which consists of a contracting path to capture context and a symmetric expanding path that enables precise localization, is widely used in medical image segmentation. In this paper, an improved U-shape network is proposed for HF segmentation in retinal OCT images. Fig.2 (a) shows the overall structure of the proposed method. As can be seen from Fig.1, the sizes of HF vary greatly, which requires the segmentation network to self-regularize the different receptive fields. To achieve this goal, improved multi-scale convolution modules (MSCM) are designed and inserted into the encoder path of the network, which is shown in Fig.2 (b). Moreover, after multiple feature extractions in the encoder path, there will be a lot of irrelevant information in channels. In order to guide the model to focus on useful channel information, two channel attention modules (CAM) are added to the bottom two max-pooling layers, which is shown in Fig.2 (c).

2.2 Multi-scale convolution module with dilation convolution

Considering the varying sizes of the HF, we adopt multi-scale convolution module to self-regularize the receptive fields of the network. The receptive fields of different branches of this structure are different, so both small targets and big ones can be well segmented. However, the original multi-scale structure used one 3×3 convolution, one 5×5 convolution and one 7×7 convolution. As is known to all, the parameters of 7×7 convolution and 5×5 convolution grow exponentially compared with that of 3×3 parameters. So in this paper, we propose an improved multi-scale convolution module based on dilated convolution[7] with dilation rates of 1, 2, and 3 to achieve the same receptive fields, which is shown in Fig.2 (b). By this way, we can obtain multi-scale receptive fields and reduce network parameters effectively.

2.3 Channel attention module

Channel attention module (CAM) was first proposed by Hu et al.[8]. CAM was originally used for classification tasks, which won first place in ILSVRC[9] 2017. The key idea of CAM is to discard redundant information and guide the model to focus on the useful channel information. We adopt this idea and redesign CAM in our segmentation model as shown in Fig.2 (c). After a global average-pooling of the input feature map, a $c\times1\times1$(c represents the channel numbers of the input) vector is got, in which each 1×1 point represents the global information of the corresponding channel feature map. Then it passes two fully connected layers in order to fit more nonlinear functions. After a sigmoid activation function, we get a c×1×1 vector with element values between 0 and 1 as a channel weight. Finally, the initial input feature map is multiplied by this channel weight vector to get the weighted channel information. Experiments show that CAM has better effect on high-level semantic information, so we insert two CAMs in the bottom two encoders.

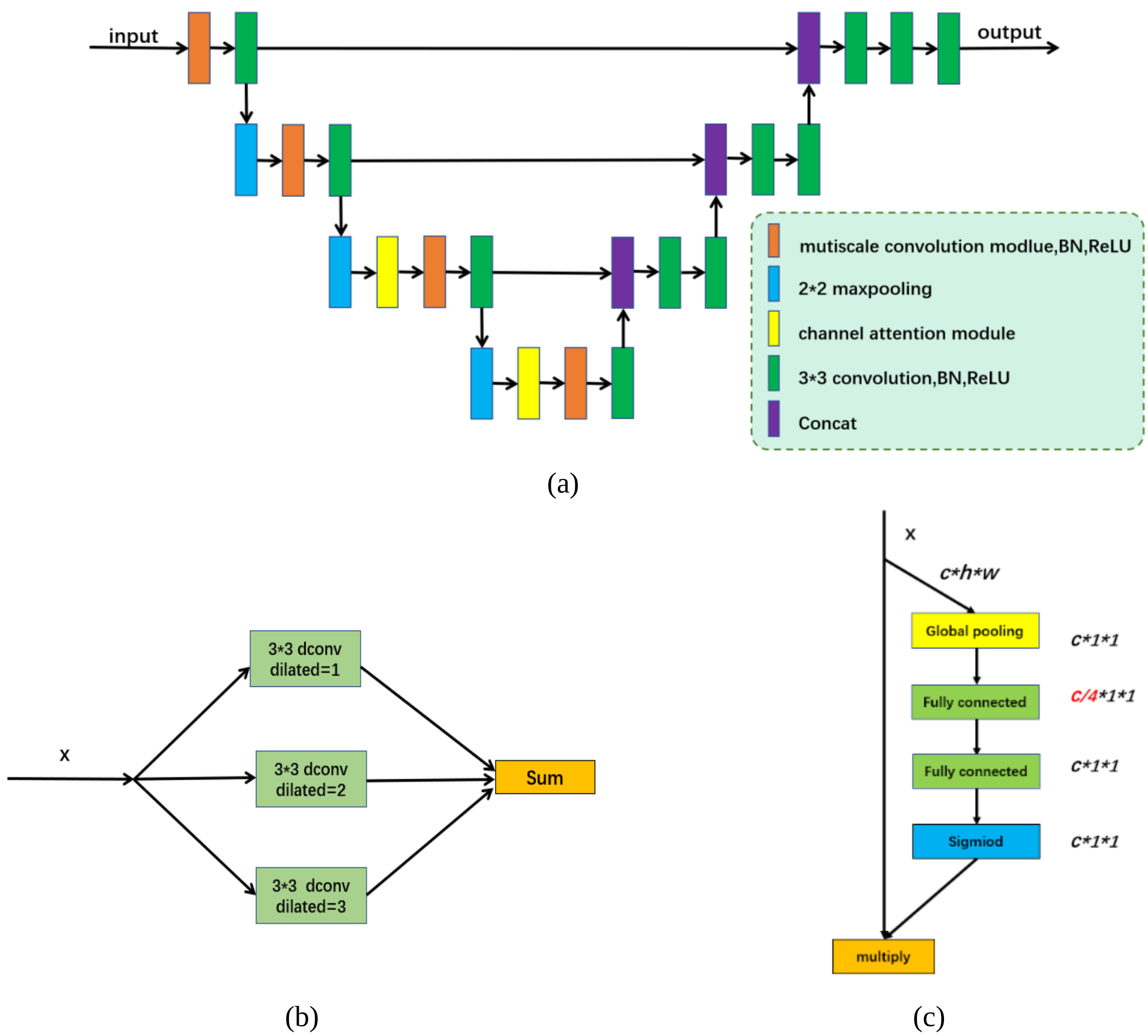

Fig.2 (a) Overall structure of the network. (b) MSCM: multi-scale convolution module with dilation convolution. (c) CAM: channel attention module.

2.4 Data augmentation strategy

To increase the generalization of the model, we adopt online augmentation strategy including left and right flipping, up and down flipping, random rotation and additive Gaussian noise addition. For each round of training, 2-4 of these augmentation methods are used.

2.5 Loss function

According to our statistics, the foreground (HF) only accounts for 0.3% of the whole image, which means the segmentation of HF is a typical data imbalance problem. To solve the data imbalance problem, a combination of binary cross-entropy loss and Dice loss is adopted as the loss function, which is described as:

$$L(Y,\overline{Y}) = -\frac{1}{N}\sum_{b=1}^{N}\left(\frac{1}{2}\cdot Y_b \cdot \log \overline{Y_b} + \frac{2\cdot Y_b \cdot \overline{Y_b}}{Y_b + \overline{Y_b}}\right) \tag{1}$$

where $\overline{Y}_b$ and Y_b denote the flatten predicted probabilities and the flatten ground truths of b^{th} image respectively, and N indicates the batch size.

3. RESULTS

3.1 Dataset

The dataset used in this paper contains 2D-images from OCT scans acquired by a Topcon Atlantis DRI-1 OCT scanner at 1050 nm, with 20μm lateral resolution and 8μm axial resolution. The image size is 512(width of B-scans) *992(height of B-scans). The dataset includes 112 retinal OCT B-scan images with ground truth from 28 different patients (4 2D B-scan images are selected from each patient randomly), which are labeled under the supervision of a senior ophthalmologist.

3.2 Parameter settings

The proposed network is realized on the Pytorch[10] framework. In the training process, the SGD[11] algorithm with an initial learning rate of 0.01, momentum of 0.9 and weight decay of 0.0001 is used to optimize the network. The batch size is set to 2 and the number of epochs is 60. We use four-fold cross validation to evaluate the performance of the proposed method. Each fold includes 28 images from 7 different patients.

3.3 Ablation experiment

To evaluate the performance of our method, comparison experiments between the FCN[12] network, the original U-Net with increased channels (Wide U-Net) and our improved U-shape network have been performed. More concretely, we have increased original U-Net's channels by a quarter to compose Wide U-Net[13], whose parameters are similar to our method in order to prove that the improvement of our method is not simply due to the increase of network parameters. We also do ablation experiments to prove the effectiveness of our proposed MSCM and CAM modules.

3.4 Results

Fig.3 shows an example of segmentation results by these 4 methods. The FCN can only distinguish the approximate location of the lesion area and there are many false positives and false negatives. The baseline (original U-Net) generates many false positives. Compared with the baseline, our method can remove more false positives in that we adopt the channel attention module which can discard redundant information and guide the model to focus on the useful channel information. Besides, the multi-scale convolution module guides the model to focus on a wider range of receptive fields. So our method can get a better result.

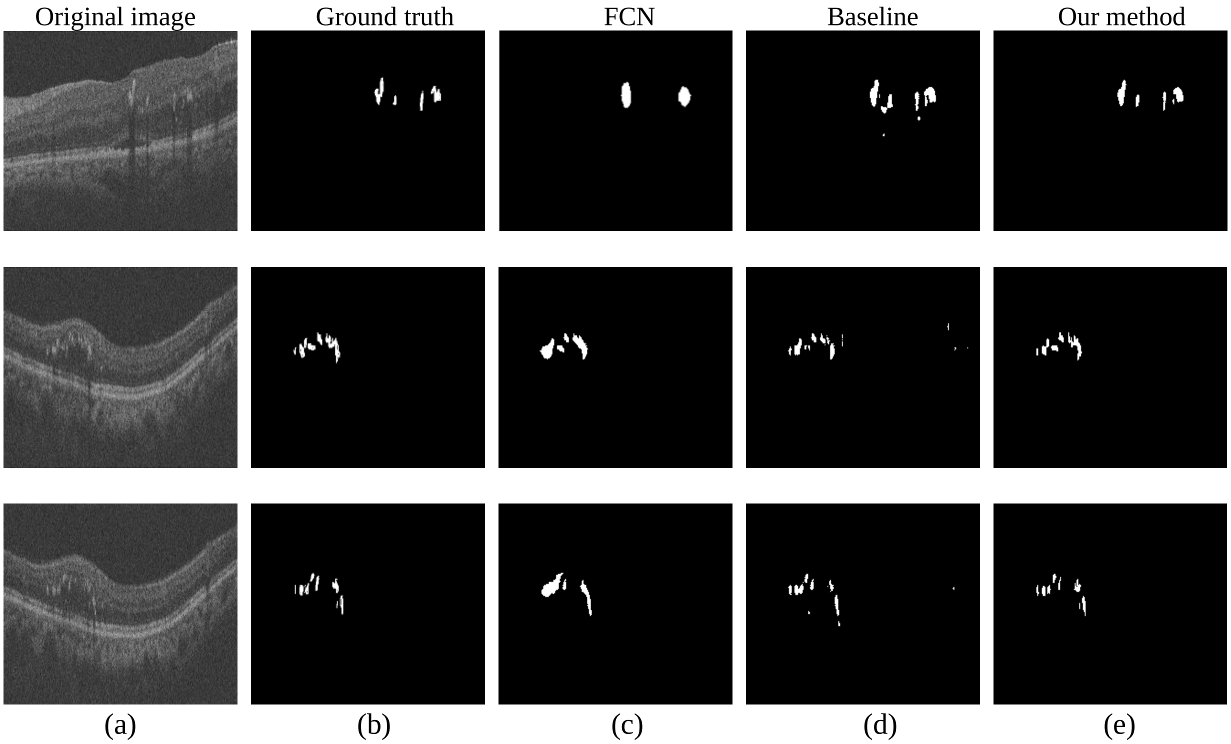

Fig.3 Qualitative comparison between FCN, Baseline and our method.

To quantitatively evaluate the performance of the model, Dice similarity coefficient (DSC), precision and recall are used as the evaluation metrics. As is vividly shown in Table 1, compared with the Baseline, our improved U-shape network achieves remarkable improvements in DSC, Recall and Precision. The result of Wide U-Net is not good, which indicates that simply increasing the parameters can not achieve good performance. The performance of FCN in all of the evaluation metrics is bad which proves the importance of skip connections in U-shape network .The results of ablation experiments show that the proposed CAM and MSCM are all effective.

Table 1: The performance of segmentation with different evaluation metrics (units:%)

Architecture	Params	DSC	Recall	Precision
FCN	40.53M	48.59±1.59	58.22±1.87	49.73±2.32
Baseline	1.84 M	71.12±2.11	74.67±2.37	71.52±3.24
Wide U-Net	2.87M	72.01±2.68	75.64±2.63	70.47±2.89
Baseline +CAM	1.85M	72.78±2.09	74.84±2.25	73.95±2.74
Baseline +MSCM	2.58M	72.95±1.95	74.77±2.19	74.08±2.84
Our Methods	2.59M	**73.26±2.03**	**75.71±1.98**	**74.28±2.67**

4.CONCLUSIONS

In this paper, we propose an improved U-shape network to automatically segment HF in retinal OCT images. In the proposed network, we introduce MSCM based on dilated convolution to obtain multi-scale receptive fields. Moreover, we introduce CAM to discard redundant information and guide the model to focus on the useful channel information. We also adopt online augmentation strategy[14] to increase the generalization of the model. Binary cross-entropy loss and Dice coefficient loss are effectively combined to constraint the model. As a result, our methods achieve the best result in DSC, recall and precision, compared with FCN and original U-Net. The experimental results demonstrate the effectiveness and practicability of the method, which may provide great help to the ophthalmologists in diagnosis of HEs.

5.ACKNOWLEDGEMENTS

This work was supported in part by the National Key R&D Program of China under Grant 2018YFA0701700, and in part by the National Basic Research Program of China (973 Program) under Grant 2014CB748600, the National Natural Science Foundation of China (NSFC) under Grant 61622114, 81401472, and in part by Collaborative Innovation Center of IoT Industrialization and Intelligent Production, Minjiang University under Grant IIC1702.

6. REFERENCE

[1] Wei Lin. (2008). Automated detection of retinal hard exudation based on K-nearest neighbor map region (Doctoral dissertation, School of Biomedical Engineering, SHANG HAI JIAO TONG UNIVERSITY).

[2] Walter, T., Klein, J. C., Massin, P., & Erginay, A. (2002). A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina. IEEE transactions on medical imaging, 21(10), 1236-1243.

[3] Chengcheng Yu. (2018). Auto segmentation and relationship analysis of SD-OCT diabetic retinal image lesion.

[4] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[6] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[7] Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.

[8] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141).

[9] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Berg, A. C. (2015). Imagenet large scale visual recognition challenge. International journal of computer vision, 115(3), 211-252.

[10] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., ... & Lerer, A. (2017). Automatic differentiation in pytorch.

[11] Bottou, L. (2012). Stochastic gradient descent tricks. In Neural networks: Tricks of the trade (pp. 421-436). Springer, Berlin, Heidelberg.

[12] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

[13] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (pp. 3-11). Springer, Cham.

[14] Van Dyk, D. A., & Meng, X. L. (2001). The art of data augmentation. Journal of Computational and Graphical Statistics, 10(1), 1-50.