

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

SANet: a self-adaptive network for hyperreflective foci segmentation in retinal OCT images

Yao, Chenpu, Zhu, Weifang, Wang, Meng, Zhu, Liangjiu, Huang, Haifan, et al.

Chenpu Yao, Weifang Zhu, Meng Wang, Liangjiu Zhu, Haifan Huang, Haoyu Chen, Xinjian Chen, "SANet: a self-adaptive network for hyperreflective foci segmentation in retinal OCT images," Proc. SPIE 11596, Medical Imaging 2021: Image Processing, 115962Y (15 February 2021); doi: 10.1117/12.2580699

SPIE.

Event: SPIE Medical Imaging, 2021, Online Only

SANet: A self-adaptive network for hyperreflective foci segmentation in retinal OCT images

Chenpu Yao^{a,#}, Weifang Zhu^{a,#}, Meng Wang^a, Liangjiu Zhu^a, Haifan Huang^c, Haoyu Chen^c, Xinjian Chen^{a,b,*}

^a School of Electronics and Information Engineering, Soochow University, Suzhou, Jiangsu Province, 215006, China

^b State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou 215123, China

^c Shantou University and the Chinese University of Hong Kong, Joint Shantou International Eye Center, Shantou, China

ABSTRACT

Diabetic retinopathy (DR) is the most common chronic complication of diabetes and the first blinding eye disease in the working population. Hard exudates (HE) is an obvious symptom of diabetic retinopathy, which has high reflectivity to light and appears as hyperreflective foci (HRF) in optical coherence tomography (OCT) images. Based on the research and improvement of U-Net, this paper proposes a self-adaptive network (SANet) for HRF segmentation. There are two main improvements in the proposed SANet: (1) In order to simplify the learning process and enhance the gradient propagation, the ordinary convolution block in the encoder structure is replaced by a dual residual module (DRM). (2) The novel self-adaptive module (SAM) is embedded in the deep layer of the model, which enables the network to integrate local features and global dependencies adaptively, and makes it adapt to the irregular shape of HRF. The dataset consists of 112 2D OCT B-scan images, which were verified by four-fold cross validation. The mean and standard deviation of Dice similarity coefficient, Jaccard index, Sensitivity and Precision are $73.69 \pm 0.72\%$, $59.17 \pm 1.00\%$, $74.57 \pm 1.16\%$ and $75.54 \pm 1.35\%$, respectively. The experimental results show that the proposed method can segment HRF successfully and the performance is better than the original U-Net.

KEYWORDS: Diabetic retinopathy, hard exudates, hyperreflective foci, self-adaptive module, dual residual module

1. INTRODUCTION

Diabetic retinopathy (DR) generally has no obvious symptoms in the early stage, but when the patient has symptoms such as decreased vision, the optimal treatment time has been missed^[1]. The main symptom of DR, hard exudates (HE), appears as hyperreflective foci (HRF) in optical coherence tomography (OCT) images. As is shown in Fig.1, the irregular shapes, blurred boundaries and heterogeneous scale of HRF bring great challenges to the accurate HRF segmentation.

There are some researches focused on the detection of HE in retinal fundus images, such as support vector machine (SVM) based method^[2] and thresholding based method^[3]. Lammer et al proposed thresholding based

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

method to segment HE in polarization-sensitive optical coherence tomography (PS-OCT) images^[4]. Xie et al. has proposed a self-adaption threshold method to segment HRF in SD-OCT images^[5]. But there are very few deep learning studies which focus on the HRF segmentation in retinal OCT images. Varga et al. proposed a deep neural network based method to segment HRF in retinal OCT images, which mainly focused on the OCT image pre-processing^[6]. In order to overcome the shortcomings of the data imbalance problem and the influence of variant shapes and scale of HRF, we propose a self-adaptive network (SANet) based on U-Net, which can optimize HRF segmentation performance without significantly increasing the complexity of the neural network and make the network more adaptable.



Fig. 1. Examples of HRF in retinal OCT image. (a) An original B-scan of retinal OCT image with HRF (areas enclosed by orange rectangles). (b) The corresponding ground truth of HRF.

2. METHODS

2.1 Overall structure of the network

U-Net^[7] can perform multi-scale prediction and deep supervision, which is widely used in medical image segmentation. This paper proposes a SANet based on U-Net for the segmentation of HRF in retinal OCT images, whose structure is shown in Fig.2.

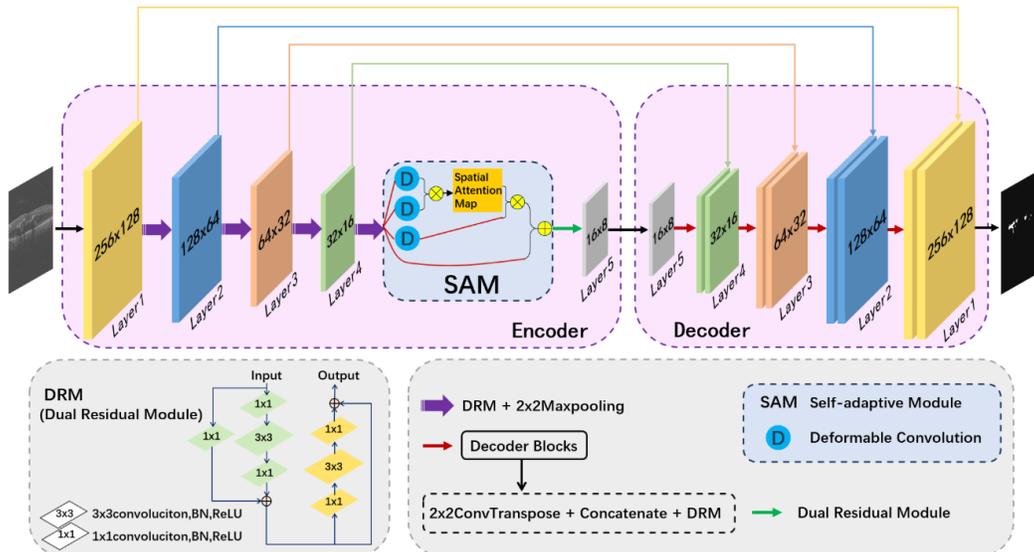


Fig. 2. The structure of self-adaptive net (SANet)

In order to simplify the learning process and enhance the gradient propagation, each ordinary convolution in the encoder structure is replaced by a dual residual module (DRM), which greatly improves the learning efficiency when the depth of the model is increasing.

2.2 Self-adaptive module

A self-adaptive module (SAM) is embedded into the top of the encoder path, as shown in Fig.3. In SAM, spatial attention mechanism^{[8][9]} is introduced, which can selectively aggregate the features of each location by weighted sum of all positions. The attention mechanism captures the spatial dependence between any two positions of the feature map, that is, any two positions with similar characteristics can improve each other. This enables the network to adaptively solve the problems such as uneven distribution of HRF and blurred boundaries.

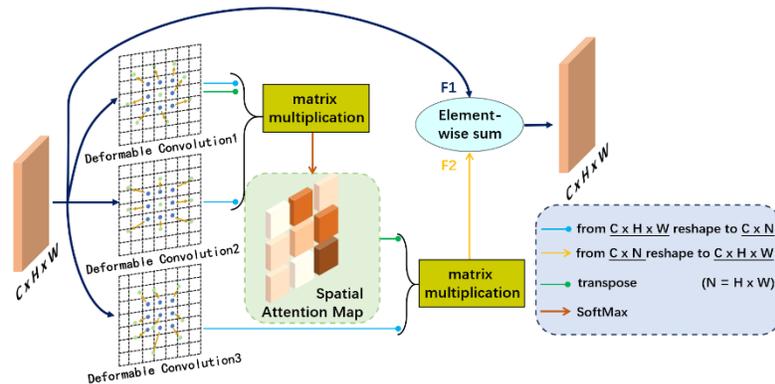


Fig. 3. Self-adaptive Module (SAM)

In addition, to further improve the performance, three deformable convolutions (shown as Fig. 4)^{[10][11]} are embedded in SAM.

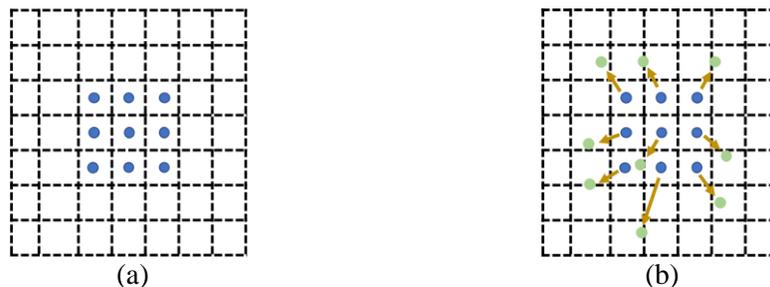


Fig. 4. Illustration of the sampling locations in a 3×3 standard convolution and a deformable convolution. (a) regular sampling grid (blue points) of standard convolution. (b) deformed sampling locations (light green points) with augmented offsets (dark yellow lines) in deformable convolution.

Compared with the standard convolution, the deformable convolution adds the 2D offset on the sampling position of the conventional grid. An offset variable is added to each sampling position in the convolution kernel, which can implement random sampling near the current position without being limited to the previous regular grid points. The module learns the offset in the task and does not require additional supervision, and the sampling grid can be deformed freely. Therefore, the information of the objects with different sizes perceived by the three deformable convolution layers can be dynamically fused by two matrix multiplication operations. Finally, the element-wise addition of F2 and the initial input F1 is used as the output of the SAM. These can improve the generalization ability of the network and adapt to the various shape and size of HRF.

2.3 Loss function

Since HRF accounts for a small proportion of the whole picture and the target area is diversified, we use the combination of the binary cross-entropy (BCE) loss and the Dice loss^[12] as the joint loss, which are defined as follows:

$$L_{Dice} = 1 - \frac{2 \sum_i^N p_i q_i + \sigma}{\sum_i^N p_i + \sum_i^N q_i + \sigma} \quad (1)$$

$$L_{BCE} = -\frac{1}{N} \sum_i^N (p_i \log(q_i) + (1 - p_i) \log(1 - q_i)) \quad (2)$$

$$L_{Total} = L_{BCE} + L_{Dice} \quad (3)$$

3. RESULTS

3.1 Datasets

The dataset used in this paper contains 28 retinal OCT volumes from 28 eyes acquired by a Topcon Atlantis DRI-1 OCT scanner at 1050 nm, with 20 μ m lateral resolution and 8 μ m axial resolution. The image size is 512 * 992. Each volume contains 128 B-scan images. Due to the similarity between slices, 4 B-scans from each volume are randomly selected. So there are totally 112 B-scans in our dataset and the corresponding ground truth are manually labeled under the supervision of a senior ophthalmologist.

In order to reduce the cost of computation, all OCT images are resized to 256 * 512. To improve the generalization ability of the model and prevent the model from over fitting, the data are augmented online, including vertical flip, horizontal flip, random rotation, affine transformation and adding Gaussian noise.

3.2 Implementation Details

The dataset is split randomly into 4 folds according to subjects. Four-fold cross validation strategy is used to evaluate the performance of the proposed method. In the training process, the stochastic gradient descent (SGD) 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 was set to 2, and the number of epochs was 60.

3.3 Evaluation Metrics

To evaluate the performance of the proposed SANet, four metrics including Dice similarity coefficient (Dice), Jaccard index (Jaccard), Sensitivity and Precision are adopted, which are defined as follows:

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (4)$$

$$Jaccard = \frac{TP}{TP + FP + FN} \quad (5)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

where TP denotes true positive, FP denotes false positive, FN denotes false negative and TN denotes true negative.

3.4 Results

Fig.5 shows examples of segmentation results of three networks. U-Net's segmentation results are slightly rough, and there are many false positives and false negatives. The results of the baseline (the convolution block in U-Net replaced by DRM) have improved, but there are still many false positives. Compared with the baseline, SANet can remove more false positives and get better results.

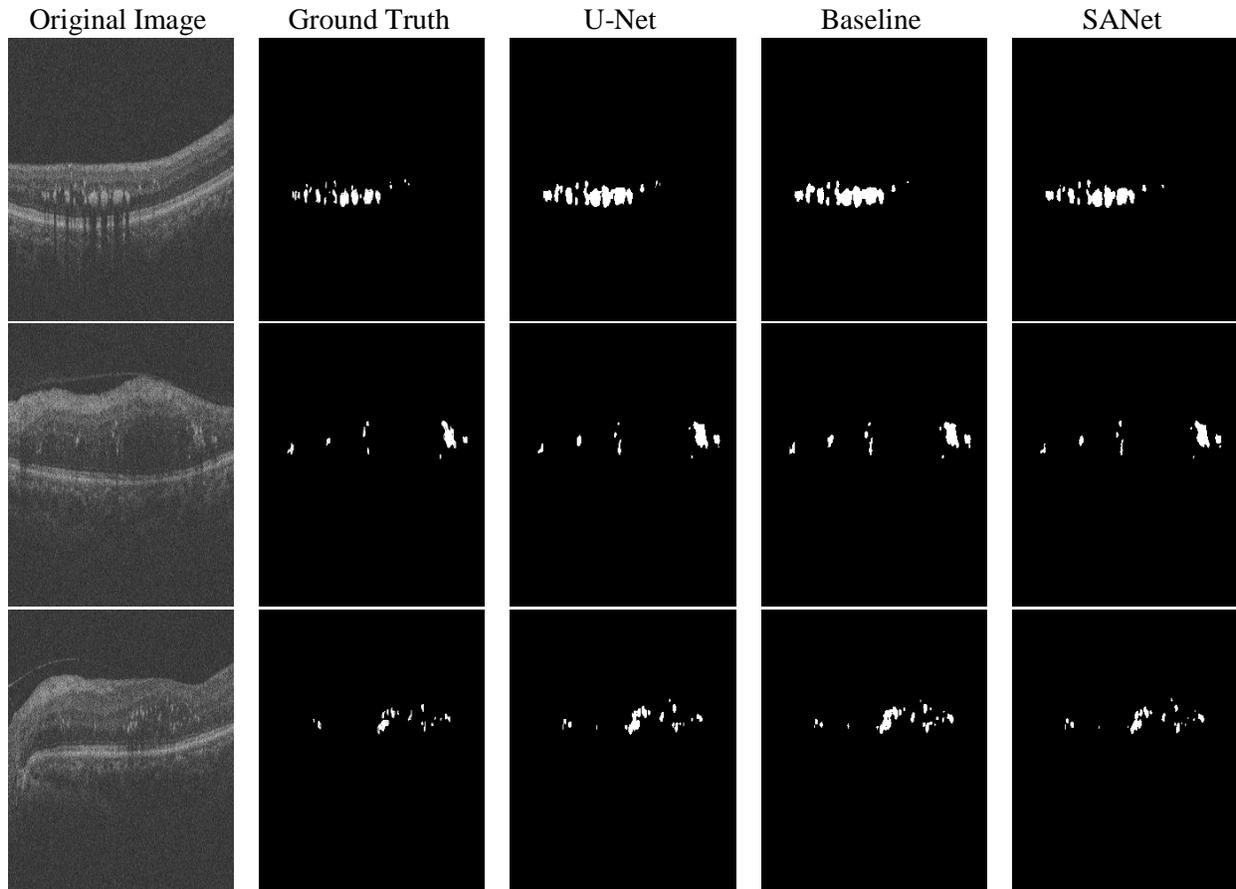


Fig. 5. Qualitative comparison between U-Net, Baseline, and SANet

In order to quantitatively evaluate the performance of the model, Dice similarity coefficient, Jaccard, sensitivity and accuracy are used as evaluation indicators. As shown in Table 1, FCN and U-Net, which are used as comparison experiments, are slightly less effective in this segmentation task. In contrast, Baseline has improved in multiple indicators. In ablation experiments, baseline with SAM but without deformable convolution is better than baseline in Dice similarity coefficient, Jaccard and accuracy, respectively. However, the effect of wide baseline is not very high, indicating that simply increasing the parameters can not improve the performance. SANet has achieved the best results on the four indicators, which verifies the practicality of the network.

Table 1. The performance of segmentation with four evaluation indicators.

Methods	Dice(%)	Jaccard(%)	Sensitivity(%)	Precision(%)
FCN	49.66±1.33	34.23±1.21	57.96±1.71	50.11±2.43
U-Net	71.16±1.88	56.00±2.18	71.25±3.51	74.08±0.85
Baseline	71.38±1.28	56.26±1.65	74.46±2.45	71.49±0.71
Baseline_Wide	72.66±0.65	57.81±0.83	74.25±1.74	73.44±1.57
Baseline_SAM_w/o_DC	73.08±0.54	58.33±0.80	74.19±0.74	73.81±1.82
SANet	73.69±0.72	59.17±1.00	74.57±1.16	75.54±1.35

4. CONCLUSIONS

In this paper, we design a novel self-adaptive network (SANet) for HRF segmentation in retinal OCT images. Two modules including dual residual module (DRM) and self-adaptive module (SAM), are designed to improve the segmentation performance. DRM replaces the ordinary convolution block in the encoder structure to simplify the learning process and enhance the gradient propagation. SAM is embedded in the deep layer of the model, which enables the network to integrate local features and global dependencies adaptively and makes it adaptive to the various and irregular shapes of HRF. The experimental results show that the proposed method implements the adaptive HRF segmentation in retinal OCT images and outperforms some other state-of-the-art deep networks.

5. ACKNOWLEDGEMENTS

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0701700, in part by the National Nature Science Foundation of China under Grant 61622114, and in part by the National Basic Research Program of China under Grant 2014CB748600.

6. REFERECENS

- [1] M. Wojtkowski, T. Bajraszewski, P. Targowski, et al. "Real-time in vivo imaging by high-speed spectral coherence tomography," *OPTICS LETTERS* , vol. 28, no. 19, pp. 1745-1747 (2003).
- [2] Y. Pan, L. Pan. "Detection of Hard Exudates in Fundus Image Based on SVM," *Computer and Modernization*, vol. 4, pp. 33-37 (2014).
- [3] T. Walter, J. C. Klein, P. Massin, et al. "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* , vol. 21, no. 10, pp. 1236-1243 (2002).
- [4] J. Lammer, M. Bolz, B. Baumann, et al. "Detection and analysis of hard exudates by polarization-sensitive optical coherence tomography in patients with diabetic maculopathy," *INVESTIGATIVE OPHTHALMOLOGY & VISUAL SCIENCE* , vol. 55, pp. 1564-1571 (2014).
- [5] S. Xie, C. Yu, Q. Chen. "Segmentation of bright speckles in 3D SD-OCT diabetic retinal images based on self-adaption threshold," *Computer Engineering and Applications*, vol. 54, no. 22, pp. 211-215 (2018).
- [6] L. Varga., A. Kovács, T. Grósz, et al. "Automatic segmentation of hyperreflective foci in OCT images," *COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE*, vol. 178, pp. 91–103 (2019).
- [7] O. Ronneberger , P. Fischer, T. Brox. "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234-241 (2015).
- [8] J. Fu, J. Liu, H. Tian, et al. "Dual Attention Network for Scene Segmentation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3146-3154 (2019).
- [9] J. Hu, L. Shen, G. Sun, et al. "Squeeze-and-Excitation Networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7132–7141 (2018).
- [10] J. Dai, H. Qi, Y. Xiong, et al. "Deformable convolutional networks," *IEEE International Conference on Computer Vision (ICCV)*, pp. 764–773 (2017).
- [11] X. Zhu, H. Hu, J. Dai, et al. "Deformable ConvNets v2: More deformable, better results," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9308–9316 (2018).
- [12] F. Milletari, N. Navab, S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *IEEE International Conference on 3d Vision (3DV)*, pp. 565-571 (2016).