Segmentation of choroid neovascularization in OCT images based on convolutional neural network with differential amplification blocks

Su, Jinzhu, Chen, Xinjian, Ma, Yuhui, Zhu, Weifang, Shi, Fei, et al.

Jinzhu Su, Xinjian Chen, Yuhui Ma, Weifang Zhu, Fei Shi, Ivica Kopriva, "Segmentation of choroid neovascularization in OCT images based on convolutional neural network with differential amplification blocks," Proc. SPIE 11313, Medical Imaging 2020: Image Processing, 1131320 (10 March 2020); doi: 10.1117/12.2548273

SPIE.

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

Segmentation of choroid neovascularization in OCT images based on convolutional neural network with differential amplification blocks

Jinzhu Su [1,#] , Xinjian Chen [1,2,#] , Yuhui Ma [1] , Weifang Zhu [1] , Fei Shi [1,*]

[1]School of Electronics and Information Engineering, Soochow University, Suzhou, 215006, China

[2]State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, 215123, China

ABSTRACT

Choroidal neovascularization(CNV) refers to abnormal choroidal vessels that grow through the Bruch's membrane to the bottom of retinal pigment epithelium (RPE) or retinal neurepithelium (RNE) layer, which is the pathological characterization of age-related macular degeneration (AMD) and pathological myopia (PM). Nowadays, optical coherence tomography (OCT) is an important imaging modality for observing CNV. This paper creatively proposes a convolutional neural network with differential amplification blocks (DACNN) to segment CNV in OCT images. There are two main contributions. (1) A differential amplification block (DAB) is proposed to extract the contrast information of foreground and background. (2) The DAB is integrated into the U-shaped convolutional neural network for CNV segmentation. The method proposed in this paper was verified on a dataset composed of 886 OCT B-scans. Compared with manual segmentation, the mean Dice similarity coefficient can reach 86.40%, outperforming some existing deep networks for segmentation.

KEYWORDS: Optical coherence tomography, choroidal neovascularization segmentation, differential amplification block, U-Net

1. INTRODUCTION

Choroidal neovascularization (CNV) is a pathological manifestation of retinal-choroidal diseases such as age-related macular degeneration and pathological myopia, which can damage central vision and lead to blindness in severe cases. Fundus fluorescein angiography (FA) and indocyanine green angiography (ICGA) can only give the position and shape of CNV in the 2D image, while optical coherence tomography (OCT) is non-invasive and can show the lesion in 3D. As the quantitative measurement of CNV is of great value to clinical diagnosis and treatment planning, automatic segmentation of CNV from OCT images is needed.

In clinical practice, CNV can be divided into two types according to the location related to retinal pigment epithelium (RPE) layer[1]. The first type of CNV refers to abnormal choroid vessels growing under the RPE, and the second type refers to those growing through the RPE layer into the retinal neurepithelium layer. In structural OCT images, the abnormal vessels are often obscured by hemorrhage and fluid exudates, or even hidden in the choroid. What can be observed and quantified are the pathologic regions manifested by the uplifted or partly missing RPE, with hyperreflective or partly hyperreflective regions underneath. Therefore, the "CNV" regions we aim to segment in this paper refer to these abnormal regions, as shown in Figure 1. Different size, shape, intensity, and the blurry or missing boundaries make the segmentation challenging.

*Corresponding author: E-mail: shifei@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, 1131320 · © 2020 SPIE · CCC
code: 1605-7422/20/$21 · doi: 10.1117/12.2548273

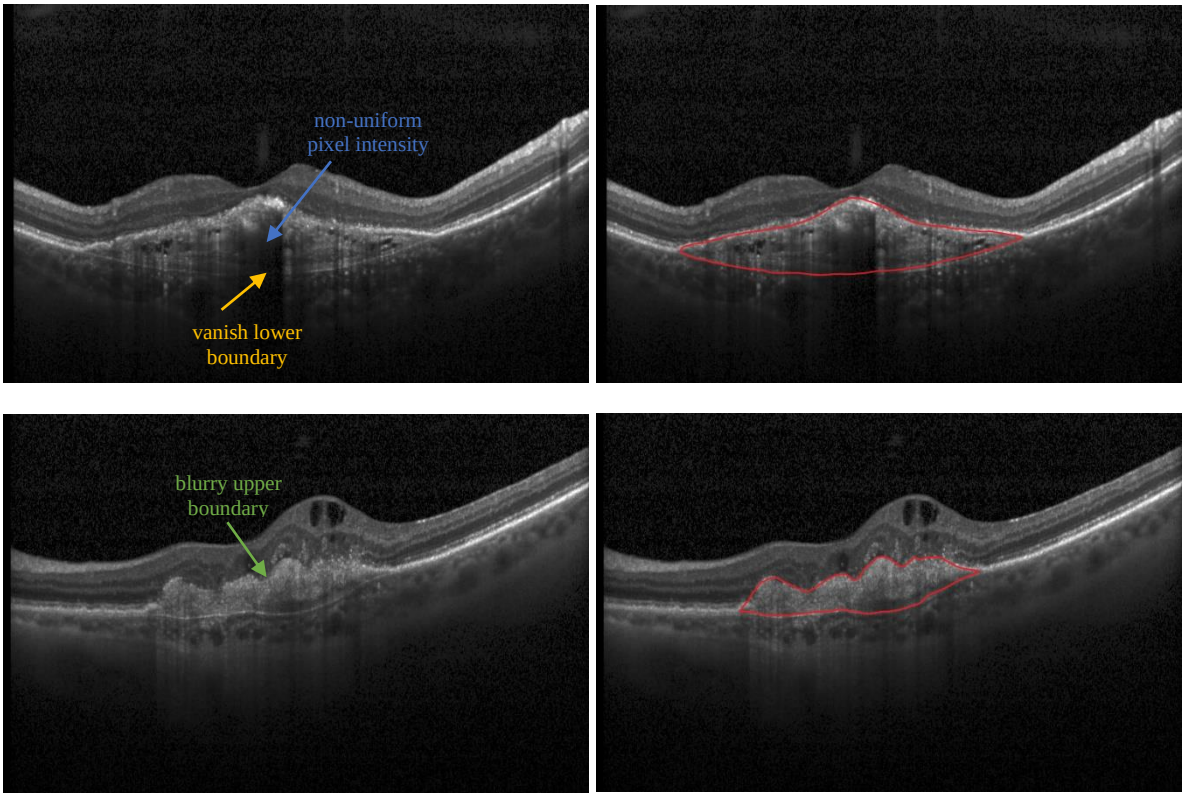


Fig. 1. Pathological features of CNV. The left column is the original image, and the red boundaries in the right column indicate the CNV lesion area.

In recent years, some effective methods have been used to detect and segment CNV. Li *et al.* [2] constructed a gradient-oriented 3D histogram (3d-HOG) feature to generate a random forest model to detect the CNV region. Xi *et al.* [3] proposed multi-scale convolutional neural networks with structure prior to segment CNV from OCT data. Zhu *et al.* [4] predicted the growth trend of CNV by using reaction diffusion model. With the development of deep learning, it is possible to achieve better generalization ability without too much complex image preprocessing. Zhang *et al.* [5] designed a multi-scale parallel branch convolutional neural network (CNN) for choroidal neovascularization segmentation in SD-OCT images. However, as the edges of CNV regions are often blurry or missing, many existing methods do not perform well near the region boundaries.

To deal with this problem, we propose a convolutional neural network with differential amplification blocks (DACNN) which can achieve automatic end-to-end segmentation of CNV lesion regions from OCT images. A differential amplification block (DAB) is proposed and integrated into a U-shaped CNN. Experiments shows the DAB helps to extract features that emphasize the edges, and the proposed network outperforms the baseline and some existing deep networks.

2. METHODS

We introduce the proposed method in the following three parts: the structure of the proposed deep network, the differential amplifier block (DAB) and the loss function.

2.1 Network structure of DACNN

It is well known that the U-Net[6] achieves good results in medical image segmentation. This paper also uses the idea of encoding and decoding network. In this paper, VGG16 [7] is used to extract image features. Therefore, five layers of sub-sampling are adopted instead of the four layers in the original U-Net, and the semantic information extracted is richer

than the original U-Net. As shown in Figure 2, in the encoder path, the blue layers correspond to the convolution blocks of VGG16 respectively. The size of each convolution kernel is 3×3, and the number of convolution kernels is 64, 128, 256, 512 and 512 successively. A maxpooling layer is connected to each convolution layer. The skip connections are improved by integrating the differential amplification module. The features extracted by the encoder are further enhanced by the DABs before passing to the decoder. In the decoder path, to simplify the network, only deconvolutions are carried out before the final convolution layer for output.

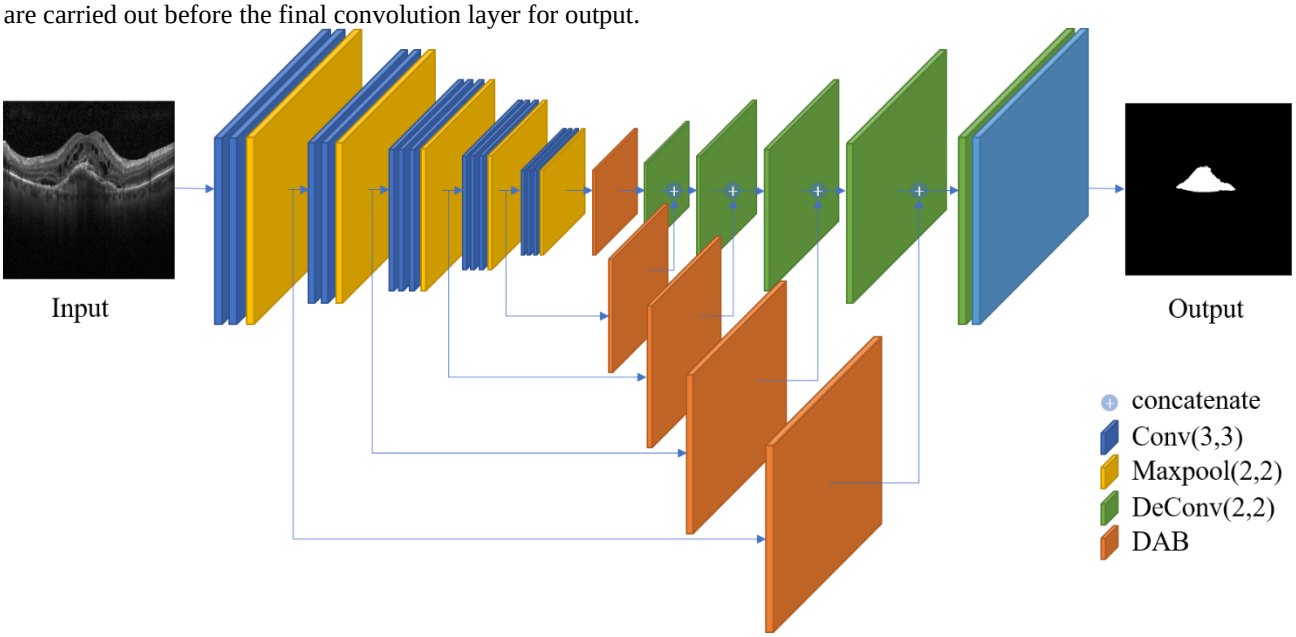


Fig. 2. An overview of the proposed DACNN

2.2 Differential amplification block(DAB)

The design of the block is motivated by the differential amplifier in electronic circuits, that assigns different gains to the difference and mean values of the two input terminals, respectively. Applying this idea to images, we propose a network module to extract the high-frequency and low-frequency information from the images and process them as different channels. The high frequency information of the image refers to the edge of the image, while the low frequency information refers to the homogeneous regions.

As shown in Figure 3, we regard the Avgpool operation of 2×2 as a low-pass filter (averaging the pixels in the region of 2×2 is equivalent to a smoothing operation), and the difference operation as a high-pass filter, which can extract the edge information in the feature maps. Then the outputs of the two parts are further processed by convolution. After that, the output features of the two parts are concatenated, and then go through a convolution layer followed by a 2×2 up-sampling operation so that the number of feature maps and their size become the same as the input feature channels. Finally, these features are further concatenated with the original input features. With this structure, we aim to enrich the input features by separating the edge or background information, optimizing them respectively to produce additional features.

To design a trainable difference block to extract high-frequency information from feature maps, a new idea isproposed as shown in Figure 4. Maxpool operation is used to obtain the maximum pixel value in the 2×2 area of the image. The input feature is also processed by a convolution operation with a convolution kernel of 2×2 and a step size of 2 to ensure the size of the two feature blocks is the same. Then, a subtraction operation is performed to extract the high-frequency information.

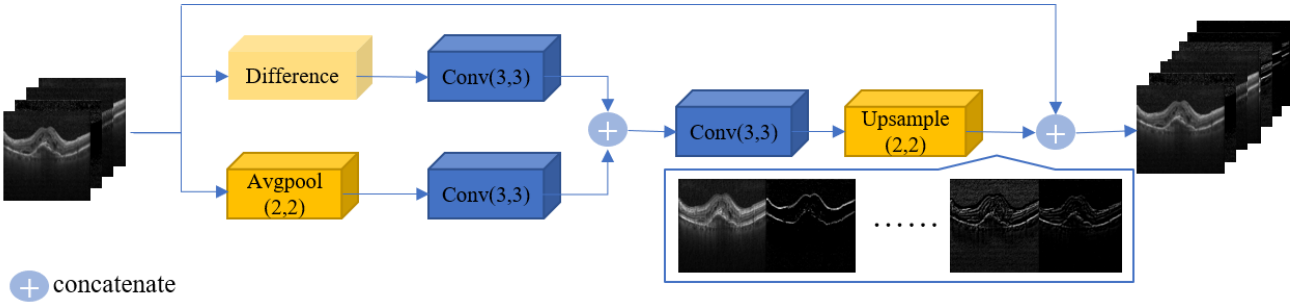


Fig. 3. The structure of differential amplifier block (DAB).

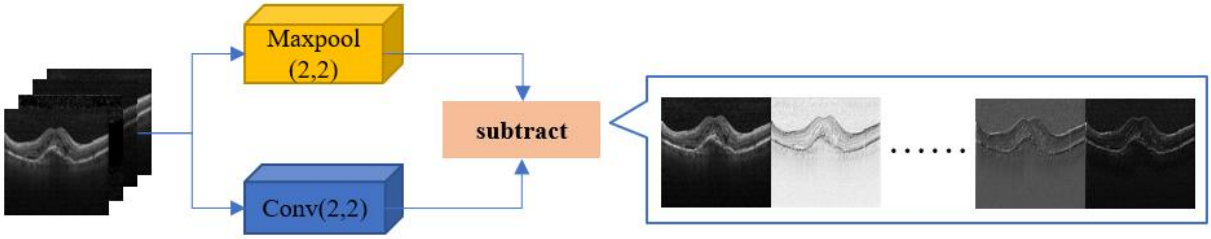


Fig. 4. Illustration of the difference block.

By using DAB, features with richer spatial information are fused with the original features. It will enable the network to pay different attention to low and high frequency information during the training process, which helps to improve segmentation for CNV. Figure 5 shows the original feature maps and the feature maps after the processing of the differential amplification module (before final concatenation). It can be seen from the figures that this module does have a strong response to the edge information of the image. But some of the low-frequency part of the image information is lost. Therefore, in order to provide more completed spatial information for the decoder operator, it is necessary to concatenate the original feature maps with the output of the differential amplification module.

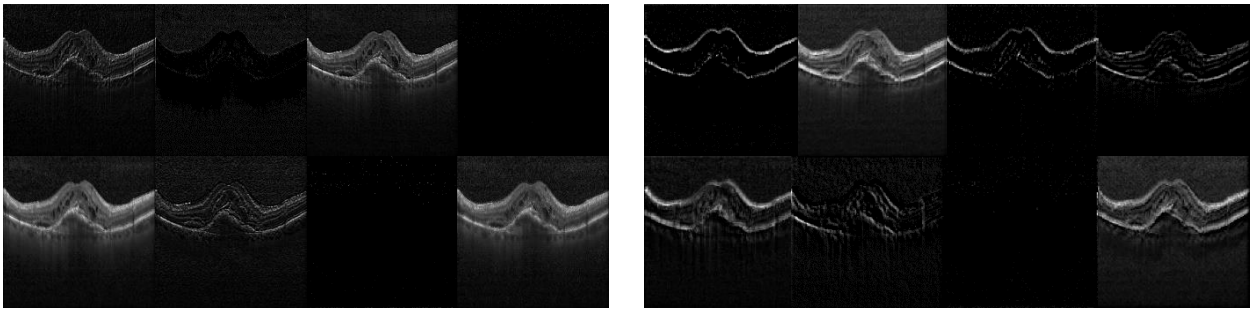


Fig. 5. Input(left) and the results(right) of the differential amplifier block

2.3 Loss function

The binary cross entropy (BCE) can be used in neural networks as a loss function for single class segmentation tasks. It can measure the similarity between the ground truth and the predicted output. Since the background pixels of the image processed are much more than the foreground pixels, the BCE loss is no longer sufficient. The Dice coefficient loss can make the network pay more attention to small targets[8]. Therefore, we choose the sum of BCE-loss and Dice-loss as the loss function, which can effectively solve the problem of unbalanced data in training process. Let x represent the pixel value in the ground truth, y represents the value predicted by the network, n represents the total number of pixels in the image, and i represents the *i-th* pixel. The loss function can be defined as follows:

$$loss_{bce} = -\frac{1}{n}\sum_{i}^{n}[x_i \ln y_i + (1-x_i)\ln(1-y_i)] \tag{1}$$

$$loss_{dice} = 1 - \frac{1}{n}\sum_{i=1}^{n}\frac{2x_i y_i}{x_i^{\,2} + y_i^{\,2}} \tag{2}$$

$$loss = loss_{bce} + loss_{dice} \tag{3}$$

3. RESULTS

3.1 Datasets

The dataset used in the experiments was a subset from the dataset used for disease identification [9]. The original dataset contains a total of 84484 OCT B-scans with class labels (37206 CNV, 11349 DME, and 8617 drusen and 26565 normal), obtained by Spectralis OCT (Heidelberg Engineering, Germany) [10]. Considering the huge efforts required for manual labelling, a small subset of 886 OCT B-scans labeled as CNV from the training dataset were used as our training dataset. Then 30 B-scans labeled as CNV from the test dataset were used as our test dataset. The image size was 768×496.

3.2 Parameter settings

Our proposed network was trained in an end-to-end way. In order to facilitate training, we resized the image and ground truth to 512×512 before input, and resized the predicted result to 768×496 after the prediction is completed. We used Adam algorithm with an initial learning rate of 0.00001 to optimize the weight of the network in the training process. In our experiment, the batchsize was set to 2 and our model was trained for 50 epochs. The models are implemented based on the Keras framework.

3.3 Evaluation metrics

We choose the following five evaluation metrics to evaluate our network: pixel accuracy (PA), true positive rate (TPR), false positive rate (FPR), Dice similarity coefficient (DSC) and intersection over union (IoU). PA is the proportion of the properly labeled pixels in the total number of pixels. TPR represents the proportion of pixels in the original foreground that are correctly marked as foreground. FPR refers to the proportion of pixels in the original background that are misclassified as foreground pixels. The IoU is the ratio of the intersection and union of the foreground in the ground truth and the foreground in the predict results.

The calculation of PA, TPR, FPR is shown as follows:

$$PA = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$TPR = \frac{TP}{TP + FN} \tag{5}$$

$$FPR = \frac{FP}{TN + FP} \tag{6}$$

where TP, TN, FP, FN represents the number of true positive, true negative, false positive, and false negative predictions, respectively.

While the above metrics are used to calculate the accuracy in pixel level, DSC and IoU are the measurements of similarity between two regions. Assuming that LA_{pred} and LA_{gt} represent the predicted results of this method and the lesion area in the ground truth respectively, and |.| means the size of the pixel set. DSC and IoU can be calculated as follows:

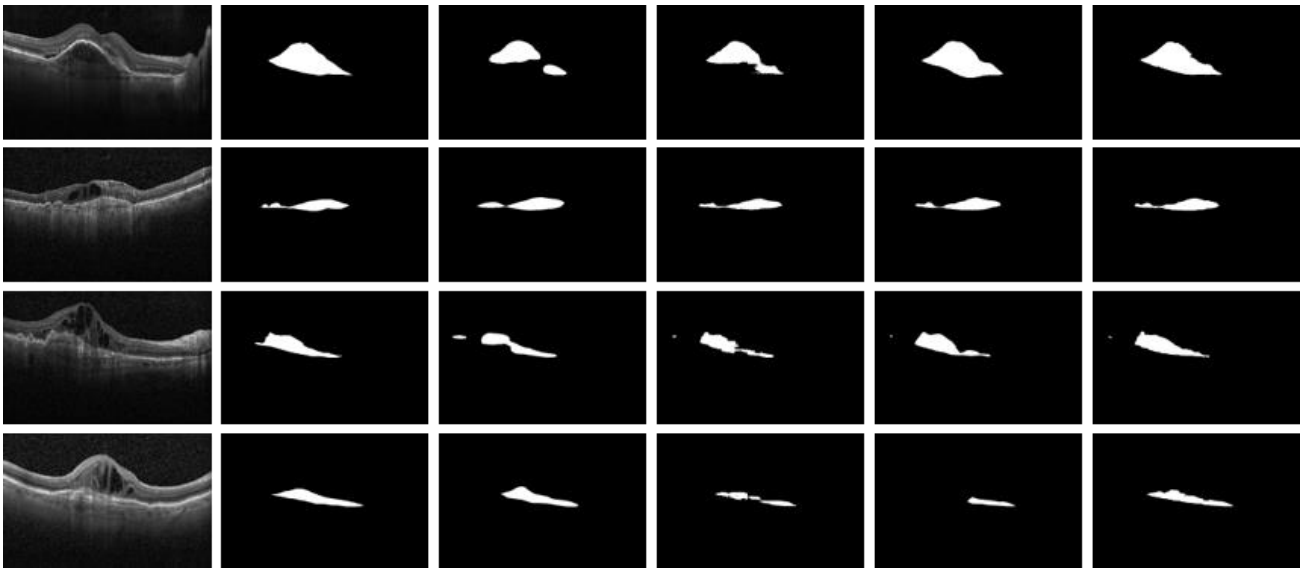
$$\mathrm{DSC} = 2\frac{|LA_{pred} \cap LA_{gt}|}{|LA_{pred}| + |LA_{gt}|} \tag{7}$$

$$\mathrm{IoU} = \frac{|LA_{pred} \cap LA_{gt}|}{|LA_{pred} \cup LA_{gt}|} \tag{8}$$

3.4 Experimental results

In order to demonstrate the effectiveness of our approach, we compare our network with some other existing networks. First, with BCE loss, we compare the proposed DACNN to the original U-Net which has four layers of downsampling, and U-NetV where the encoder of U-Net adopts VGG16 structure, which can extract more semantic information. After that, with the joint BCE and Dice loss, we compare our DACNN with Seg-Net[11] and PSP-Net[12].

As can be seen from Table 1, using VGG16 as feature encoder achieves better segmentation results than the original U-Net. And after the differential amplification block is added, the performance is further improved. For the proposed DACNN, by using the joint loss instead of BCE loss, all indices are improved, and the Dice coefficient reaches 0.8640. With the same joint loss, DACNN performs better than Seg-Net and PSP-Net regarding most metrics.



(a) (b) (c) (d) (e) (f)

Fig. 6. Segmentation results for some B-scans. (a)Raw image. (b)Ground truth. (c)Seg-Net. (d)U-NetV. (e)PSP-Net. (f)DACNN

Figure 6 shows some OCT B-scans and the segmentation results. The first row and the remaining three rows belong to two different types of CNV lesions. It can be seen that our method can segment the CNV lesion region well for the two different types. Although the CNV lesions in different images have different sizes, shapes and intensities, our method still can achieve good results. When compared with other methods, the boundaries are more accurate, indicating that the proposed model makes the network more focused on high frequency features.

Table 1 the mean PA and TPR, FPR and the mean Dice and IoU compared with different methods

	PA	TPR	FPR	Dice	IoU
U-Net(BCE)	0.9899	0.8342	0.0059	0.8201	0.7132
U-NetV(BCE)	0.9906	0.7846	0.0021	0.8348	0.7318
DACNN(BCE)	0.9914	0.8204	0.0024	0.8573	0.7631
Seg-Net (BCE-Dice)	0.9859	**0.8414**	0.0088	0.7711	0.6434
PSP-Net(BCE-Dice)	0.9913	0.7948	0.0028	0.8226	0.7261
DACNN(BCE-Dice)	**0.9920**	0.8263	**0.0020**	**0.8640**	**0.7738**

4. CONCLUSIONS

In this paper, we propose a differential amplification block that extracts the low and high frequency features respectively. Furthermore, we design a DACNN that integrates the DAB with a U-shaped structure for the segmentation of choroid neovascularization in OCT B-scans, which can effectively eliminate the uncertainty of boundary pixels and achieve more accurate segmentation results. The method is useful in providing quantitative information for analysis of CNV in clinical practice.

5. ACKNOWLEDGEMENTS

This work was supported in part by the National Key R&D Program of China under Grant 2018YFA0701700, the National Basic Research Program of China (973 Program) under Grant 2014CB748600, and in part by the National Natural Science Foundation of China (NSFC) under Grant 61622114, 61971298, 61771326.

6. REFERENCE

[1] C. L. Tsai, Y. L. Yang, S. J. Chen, *et al*, "Automatic characterization and segmentation of classic choroidal neovascularization using Adaboost for supervised learning." *IEEE Nuclear Science Symposium Conference Record.* 2010.

[2] Y. Li, S. Niu, Z. Ji, *et al*, "Automated choroidal neovascularization detection for time series SD-OCT images." *Medical Image Computing and Computer Assisted Intervention*, 2018.

[3] X. Xi, X. Meng, L. Yang, *et al*, "Automated segmentation of choroidal neovascularization in optical coherence tomography images using multi-scale convolutional neural networks with structure prior." *Multimedia Systems*, 2018.

[4] S. Zhu, F. Shi, D. Xiang, *et al*, "Choroid neovascularization growth prediction with treatment based on reaction-diffusion model in 3D OCT images. " *IEEE Journal of Biomedical and Health Informatics*, 2017.

[5]Y. Zhang, Z. Ji, Y. Wang, *et al*, "MPB-CNN: a multi-scale parallel branch CNN for choroidal neovascularization segmentation in SD-OCT images." *OSA Continuum*, 2019.

[6]O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer Assisted Intervention*, 2015.

[7] S. Karen, Z. Andrew, "Very deep convolutional networks for large-scale image recognition." *Computer Science*, 2014.

[8] F. Milletari, N. Navab and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation." arXiv:1606.04797, 2016.

[9] D. S. Kermany, M. Goldbaum, W. Cai, *et al*. "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, 172(5):1122-1131.e9. 2018.

[10]https://www.kaggle.com/paultimothymooney/kermany2018

[11] V. Badrinarayanan, A. Kendall, R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for scene segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[12] H. Zhao, J. Shi, X Qi, *et al*, "Pyramid scene parsing network," *IEEE Conference on Computer Vision and Pattern Recongnition* 2017.