Attention-guided channel to pixel convolution network for retinal layer segmentation with choroidal neovascularization

Yang, Xiaoling, Chen, Xinjian, Xiang, Dehui

Xiaoling Yang, Xinjian Chen, Dehui Xiang, "Attention-guided channel to pixel convolution network for retinal layer segmentation with choroidal neovascularization," Proc. SPIE 11313, Medical Imaging 2020: Image Processing, 1131334 (10 March 2020); doi: 10.1117/12.2548940

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

Attention-guided Channel to Pixel Convolution Network for Retinal Layer Segmentation with Choroidal Neovascularization

Xiaoling Yang[a], Xinjian Chen[a], Dehui Xiang*[a]
[a]School of Electronics and Information Engineering, Soochow University, Suzhou, Jiangsu Province, 215006, China

ABSTRACT

This paper introduces an attention-guided channel to pixel convolution network for a fully automatic segmentation of retinal layers with choroidal neovascularization from optical coherence tomography (OCT) images. The proposed framework, consists of two new strategies for retinal layers segmentation: Channel to Pixel Block and Attention block. To deal with the contrast reduction of adjacent retinal layers caused by choroidal neovascularization, we firstly design a Channel to Pixel Block to convert particular channels into pixels in one bigger feature map, followed by a convolution layer optimized by a novel edge loss. Faced with large morphological changes of retinal layers, the attention mechanism is then introduced to extract more context information. The proposed method was trained on augmented 1280 OCT images and tested on 384 OCT images with choroidal neovascularization. The experimental results showed that the proposed method outperformed the state-of-art methods for retinal OCT image segmentation.

Keywords: Choroidal neovascularization, optical coherence tomography, attention mechanism, image segmentation

1. INTRODUCTION

Age-related macular degeneration (AMD) is the leading cause of blindness among older over 50 worldwide [1]. Choroidal neovascularization (CNV), characterized by the growth of neovascularization, usually between retinal pigments epithelial and choroid, is a typical characteristic of AMD in advanced stage [2]. Since CNV will lead to serious visual damage [1], it is necessary to accurately segment the retinal layers with CNV to assist the diagnosis.

Optical coherence tomography (OCT) has been widely used in CNV diagnosis recently due to its non-invasive and high resolution [3]. Computer-aided image segmentation becomes popular as manual segmentation is inefficient. As shown in Fig.1, (a) normal retinal layers are displayed in a B-scan slice OCT image, (b)-(d) large morphological changes and blurry

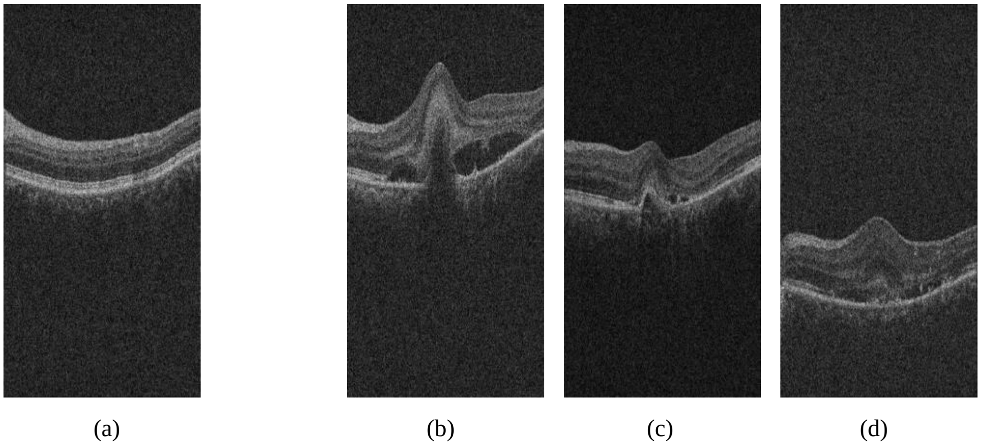
(a) (b) (c) (d)

Figure 1. Retinal layers in B-scan OCT images. (a) Normal retinal layer. (b) - (d) abnormal retinal layer: large morphological changes and blurry boundaries caused by CNV.

* Corresponding author: Dehui Xiang, E-mail: xiangdehui@suda.edu.cn

Medical Imaging 2020: Image Processing, edited by Ivana Išgum, Bennett A. Landman,
Proc. of SPIE Vol. 11313, 1131334 · © 2020 SPIE · CCC
code: 1605-7422/20/$21 · doi: 10.1117/12.2548940

boundaries appear due to CNV. Therefore, retinal layer segmentation in OCT images of CNV patients is a challenging task.

In this paper, we introduce an attention-guided channel to pixel convolution network to segment the retinal layer into background and 7 layers. The proposed network provides several advantages: 1) the edge of the retinal layer can be segmented accurately; 2) layers with large morphological changes can be completely divided by studying abundant context information.

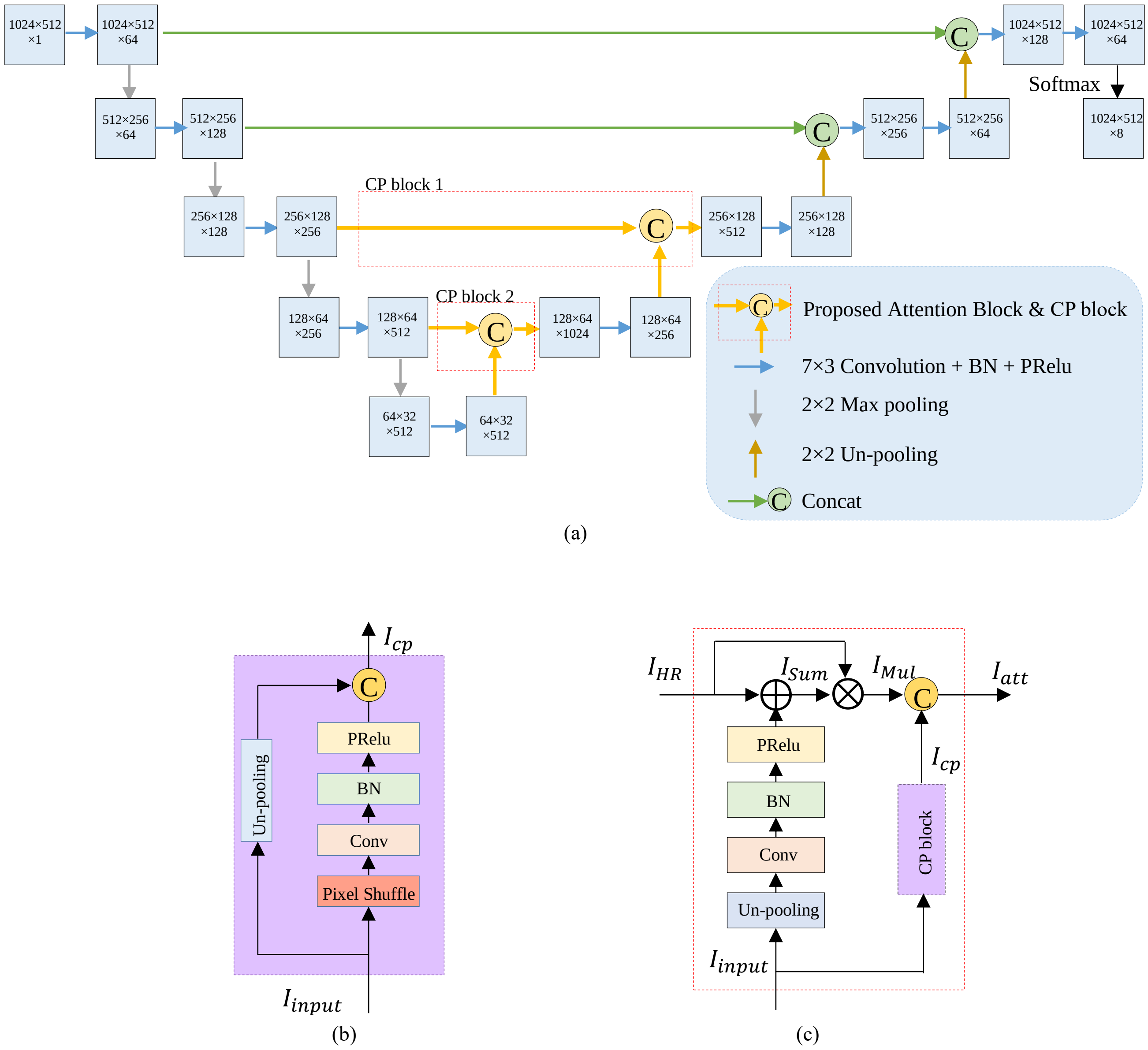

Figure 2. Overview of the proposed network. (a) The architecture of the proposed Attention-guided Channel to Pixel Convolution Network architecture. (b) CP block: pixel shuffle is used to help up-scaling. (c) Attention Block, ⨁ denotes element-wise sum and ⨂ denotes spatial element-wise multiplication.

2. METHODS

The proposed deep learning architecture is inspired by U-Net [4] and ReLayNet [5] architectures, as shown in Fig.2 (a). To further improve the segmentation performance especially in images with CNV, we propose firstly a Channel to Pixel block (CP block) with an edge loss function, helping segmenting the blurry boundary between two retinal layers. Secondly, the attention mechanism is integrated into the network to deal with large morphological changes of retinal layers. The specific description is as follows:

2.1 Channel to Pixel Block (CP block)

Un-pooling is the most commonly used method for up-sampling, while it does not fully utilize the channel information. We design a CP block, as shown in Fig.2 (b) to extract more information from different channels. The proposed CP block contains two branch networks. One branch network consists of one-pixel shuffle layer, one convolutional layer, one batch normalization layer (BN), and one PReLU activation function layer. Another branch network consists of one un-pooling layer. The whole process of CP block operation can be described as,

$$I_{cp} = concat\left[\sigma\left(W_{cp} \cdot PS\left(I_{input}\right) + b_{cp}\right),\ f_{unpool}\left(I_{input}\right)\right] \tag{1}$$

where $I_{input} \in \mathbb{R}^{H\times W\times r^2C}$ indicates the input; $I_{cp} \in \mathbb{R}^{H\times W\times \frac{5}{4}r^2C}$ indicates the output; σ represents batch normalization and PReLU activation function; $PS(\cdot)$ is the pixel shuffle operation; $W_{cp} \in \mathbb{R}^{H\times W\times \frac{1}{4}r^2C}$ and $b_{cp} \in \mathbb{R}^{1\times 1\times \frac{1}{4}r^2C}$ are the convolution's (kernel $r\times r$, stride $\frac{r}{2}$) weight and bias; $f_{unpool}(\cdot)$ is the un-pooling operation, and $concat[\cdot]$ is concatenation computation of feature maps in the channel domain. The shape change of the feature map is shown in Table1.

Table 1. Shape change of feature maps in CP block.

Input feature map shape	Layer Name	Output feature map shape
$H\times W\times r^2C$	Pixel Shuffle layer	$rH\times rW\times C$
$rH\times rW\times C$	Convolution layer	$2H\times 2W\times \frac{r^2}{4}\ C$
$2H\times 2W\times \frac{r^2}{4}\ C$	BN&PReLU	$2H\times 2W\times \frac{r^2}{4}\ C$
$H\times W\times r^2C$	Un-pooling layer	$2H\times 2W\times r^2C$
$2H\times 2W\times r^2C\&2H\times 2W\times \frac{r^2}{4}\ C$	Concatenation computation	$2H\times 2W\times \frac{5r^2}{4}C$

In details, the pixel shuffle layer, as shown in Fig.3, shuffles the pixels periodically from different channels into a bigger feature map by an upscale factor r^2.

Notice that Fig.3 is only an example to explain how Pixel Shuffle Layer works, the actual input feature map size and upscale factor depends on specific layers. Assumed that the shape of the input feature map for CP block is $H\times W\times r^2C$, H refers to the height of the input feature map; W refers to the width. r^2C refers to the number of channels. The specific position transformation relationship of pixel shuffling operation can be described as,

$$I_{ps}[i,j,c] = I_{input}[floor(i/r),\ floor(j/r), r^2c + \left(mod(i,r)\right)r + mod(j,r)\Big] \tag{2}$$

$$I_{ps} = PS\left(I_{input}\right) \tag{3}$$

I_{ps} is the output of pixel shuffle; $mod(x,y) = x\%y$; i ∈ $[0,2H]$, j ∈ $[0,2W]$, c ∈ $[0,C]$. The Pixel Shuffle Layer shuffle pixels from particular channels into pixels in one particularly bigger feature map without loss of channel information.

A convolution operation is then used to extract features from the enlarged feature map. We force the convolution output has the same shape with un-pooling layer in parallel by setting the convolution stride to $\frac{r}{2}$.

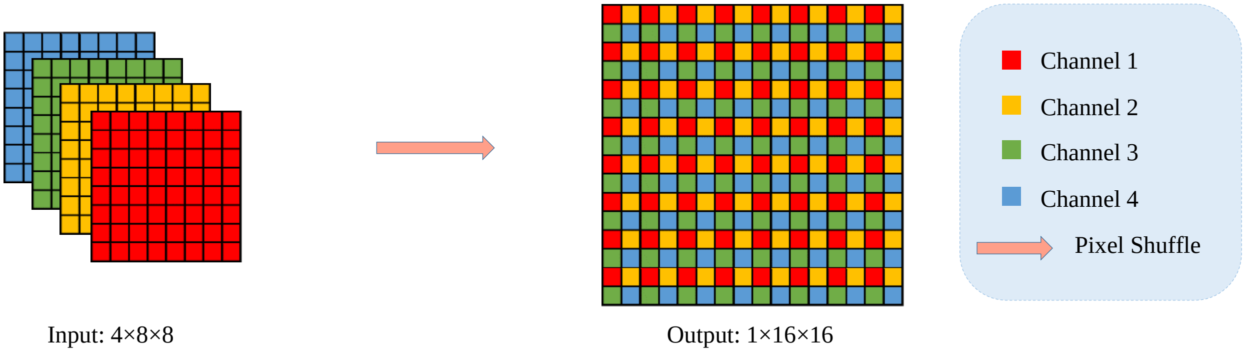

Figure 3. An example of the Pixel Shuffle, in which we convert pixels from 4 channels into a bigger feature map with 1 channel. As the feature map is magnified 2^2 after Pixel Shuffle, the upscale factor r^2 equals to 2^2 here.

The adjacent retinal layers' boundary in OCT images of objects suffering from CNV, is blurry and difficult to distinguish. In order to guide the CP block to extract effective edge features, we propose an auxiliary edge loss as

$$Loss_{edge} = -log \left| \frac{sobel(I_p)}{sobel(I_{cGAN})} \right| \tag{4}$$

$$I_p = \sigma(W_{cp} \cdot PS(I_{input}) + b_{cp}) \tag{5}$$

In which I_{cGAN} indicates the speckle noise reduction result of raw input [6]. The retinal layers boundary of I_{cGAN} get clearer than the raw image. The $Loss_{edge}$ measures the similarity of boundary gradient between I_p and the target I_{cGAN}. By optimizing the edge loss, the adjacent retinal layers' gradient becomes bigger so that the edge of adjacent layers of the retina becomes clearer.

2.2 Proposed Attention Module

The proposed attention module adds the feature map I_{input} at low resolution to the feature map I_{HR} at high resolution, and then multiply with high resolution input I_{HR}, as shown in Fig.2 (c). The output feature map contains both local semantic information and global structural information. The output of the attention module can be represented as follows,

$$I_{Sum} = I_{HR} \oplus \sigma\left(W_{att} \cdot f_{unpool}\left(I_{input}\right) + b_{att}\right) \tag{6}$$

$$I_{Mul} = I_{HR} \otimes I_{Sum} \tag{7}$$

$$I_{Att} = concat\left[I_{Mul},\ I_{cp}\right] \tag{8}$$

where $W_{att} \in \mathbb{R}^{H \times W \times r^2 C}$ and $b_{att} \in \mathbb{R}^{1 \times 1 \times r^2 C}$ are the convolution (kernel 3×3) weight and bias parameters. Here, $\oplus$ denotes element-wise sum and $\otimes$denotes spatial element-wise multiplication. $I_{Mul} \in \mathbb{R}^{2H \times 2W \times r^2 C}$, and $I_{Att} \in \mathbb{R}^{2H \times 2W \times \frac{9}{4} r^2 C}$.

As the retina with CNV has large morphological changes, it is very important to enhance the location information of the layer where the abnormal neovascularization appears. Low resolution feature map I_{input}, has rich retinal structure information. The structure information, especially the location of the layer with abnormal vessels, will be very useful for the complete segmentation of the retina.

The high-resolution feature map I_{HR} contains rich detail information, such as retinal boundary. In Eq. (6), we add the high-resolution feature map I_{HR} and the low-resolution feature map I_{input} , which shape has transformed as large as I_{HR} by an un-pooling layer and convolutional layer.

Therefore, I_{Sum} has both layer position information and layer boundary information. Then, I_{Sum} can be considered as the attention information, multiply with I_{HR}, effectively increasing the spatial structure information of I_{HR}. Finally, I_{Mul} and the result of the CP block I_{cp} are concated. I_{cp} has rich edge information, so the output of the attention module I_{Att} has abundant global and local information.

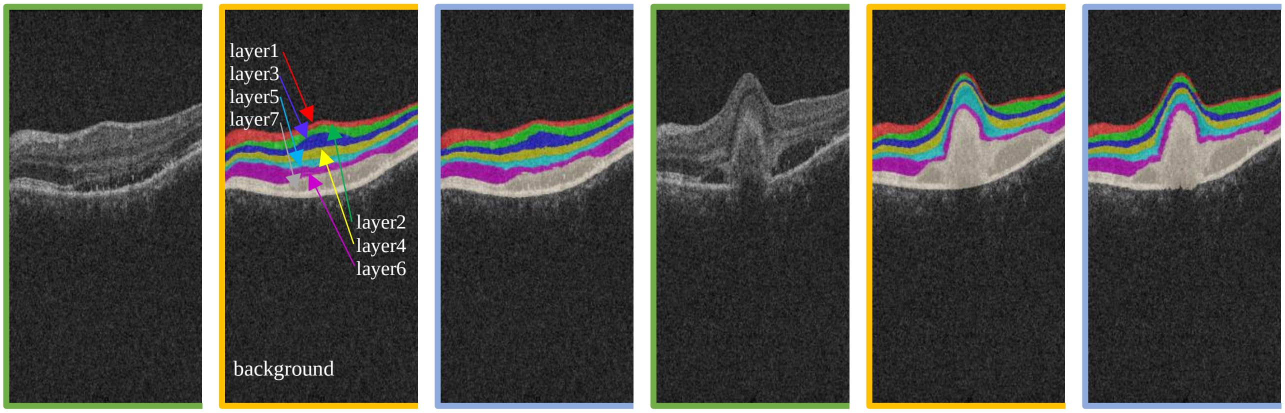

Figure 4. Segmentation result of retinal layers with CNV. Green boxes: original images; Orange boxes: ground truth; Blue boxes: segmentation results.

2.3 Loss Function

The loss function is inspired by ReLayNet. Total loss in this paper includes weighted multi-class logistic loss [5], Dice loss [5] and proposed Edge loss.

The total loss function is shown as follows,

$$Loss_{overall} = \lambda_1 Loss_{logloss} + \lambda_2 Loss_{dice} + \lambda_3 Loss_{edge\_CP1} + \lambda_4 Loss_{edge\_CP2} \tag{9}$$

where $Loss_{logloss}$, $Loss_{dice}$, $Loss_{edge\_CP1}$ and $Loss_{edge\_CP2}$ represent the weighted multi-class logistic loss, the Dice loss, Edge loss in CP block 1 and Edge loss in CP block 2, respectively. The weight terms in loss function in Eq. (9) are set as $\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 1e^{-5}, \lambda_4 = 1e^{-6}$.

In this paper, region weights of weighted multi-class logistic loss are defined as shown in Eq. (10),

$$\omega_{logloss}(x) = \begin{cases} 10, & x \in retinal\ layer's\ boundary \\ 5, & x \in\ retina \\ 1, & others \end{cases} \tag{10}$$

Here, x is the pixel of input image; $\omega_{logloss}$ is the region weights. Pixels belong to $retina$ region bigger weights than background region. Pixels on retinal layer’s boundary the biggest weight.

3. RESULTS

The dataset contained 1664 OCT B-scan images acquired by a Topcon DRI-OCT. The training set includes 1280 images, augmented per training epoch. The testing set has 384 images. Each OCT images contain 512 × 128 pixels, with pixel size of 11.74 × 47.24 × 1.96 μm^3. The kernel size of convolution in the proposed network is 3×5, and the stride is 2. The upscale factor r^2 used sets to 16^2 in our network, both in CP block 1 and CP block 2.

3.1 Qualitative comparison of proposed network with comparative methods

The representative images obtained by the proposed method is shown in Fig.4. Images in green boxes are the original images, images in orange boxes are ground truth, and images in blue boxes are segmentation results.

To quantitatively evaluate the performance of our proposed method, we compared the proposed method segmentation results with ReLayNet and U-Net at Dice coefficient, Intersection over Union (IoU) and Accuracy of each class. As shown in Table 2, the best performance is shown by **bold**. Overall, Attention-guided Channel to Pixel Convolution Network showed the excellent segmentation in the background and other seven retinal layers. Especially under the Dice coefficient, the proposed network showed the best performance in various categories.

In this challenging layer7, where the neovascularization appears, the proposed network surprisingly achieved than 94.56% under Dice coefficient, while the second best performance by ReLayNet only got 89.12%. Under IoU, our method got 70.37% in layer5, 5.01% above the second-best performance by U-Net.

Table 2. Comparisons with comparative methods. The best performance is shown by **bold**.

	method	background	layer1	layer2	layer3	layer4	layer5	layer6	layer7
Dice	Proposed	**0.9980**	**0.9374**	**0.8715**	**0.8490**	**0.8756**	**0.8258**	**0.8964**	**0.9456**
	ReLayNet	0.9949	0.9286	0.8526	0.8264	0.8460	0.7828	0.8866	0.8912
	U-Net	0.9941	0.9299	0.8522	0.8119	0.8295	0.7679	0.8646	0.8708
IoU	Proposed	0.9961	**0.8822**	**0.7723**	**0.7379**	**0.7790**	**0.7037**	**0.8123**	**0.8969**
	ReLayNet	**0.9989**	0.8666	0.7433	0.7049	0.7337	0.6441	0.7969	0.8052
	U-Net	0.9949	0.8703	0.7566	0.7299	0.7770	0.6536	0.7762	0.8582
Acc	Proposed	**0.9977**	**0.9447**	**0.8449**	0.8455	**0.9168**	0.7823	0.8840	**0.9624**
	ReLayNet	0.9917	0.9266	0.8240	0.8507	0.8947	**0.7899**	0.8786	0.9244
	U-Net	0.9960	0.9362	0.8383	**0.8614**	0.8977	0.8140	**0.9244**	0.9059

3.2 Ablation study

We do ablation study to reveal effect of CP block, the attention mechanism and the edge loss function.

Table 3. Comparison with baselines. The best performance is shown by **bold**.

method	CP block	Attention module	edge loss	Per class dice	mIoU	mPA
Proposed	√	√	√	**0.8999**	**0.8226**	**0.8973**
CPANet	√	√	×	0.8826	0.7971	0.8882
CPNet	√	×	×	0.8823	0.8015	0.8858
Baseline	×	×	×	0.8761	0.7855	0.8850

In Table 3, CPANet refers to network with CP block and Attention Module; CPNet refers to network with CP block; Baseline refers to network without CP block, Attention Module and edge loss. "√" means the method has this block, while "×" means the method does not has this block.

Per class dice, mean Intersection over Union (mIoU), mean pixel accuracy (mPA) were used to evaluate segmentation performance of different methods. It can be found from Table 3 that the three modules proposed in this paper are proved to be valid by comparing network in pairs. Trying more different ways of attention will be our follow-up.

4. CONCLUSIONS

In this paper, we proposed an attention-guided channel to pixel convolution network for retinal layers segmentation in OCT images with CNV. The method includes two new blocks: a channel to pixel block and an attention module. Due to data augmentation with flipping and translation in four directions, the proposed method needs only a few labeled images. Experiments showed the proposed method can accurately segment the retinal layers.

5. ACKNOWLEDGEMENTS

This work has been supported in part by the National Key R&D Program of China under Grant 2018YFA0701700, and in part by the National Natural Science Foundation of China (NSFC) under Grant 61971298.

REFERENCES

[1] Kawther Y., Yasmina C., Alexandra M., Donato C., Eric S., Eric P., " Automated quantification of choroidal neovascularization on Optical Coherence Tomography Angiography images," SID Int. Symp. Computers in Biology and Medicine 114 (2019) 103450.

[2] Grossniklaus, Hans E., and W. R. Green. "Choroidal neovascularization." American Journal of Ophthalmology 137.3(2004):0-503.

[3] Zhu S, Shi F, Xiang D, et al. Choroid Neovascularization Growth Prediction with Treatment Based on Reaction-Diffusion Model in 3D OCT Images. IEEE Journal of Biomedical and Health Informatics, 2017:1-1.

[4] Ronneberger O, Fischer P, Brox T. "U-Net: Convolutional Networks for Biomedical Image Segmentation". International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.

[5] Roy, Abhijit Guha, et al. "ReLayNet: Retinal Layer and Fluid Segmentation of Macular Optical Coherence Tomography using Fully Convolutional Network." Biomedical Optics Express (2017):3627.

[6] YH. Ma, XJ. Chen, et al. "Speckle noise reduction in optical coherence tomography images based on edge-sensitive cGAN." Biomedical Optics Express (2018):5129