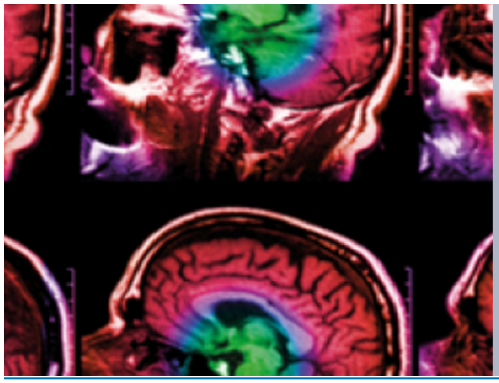
PAPER

Group-wise context selection network for choroid segmentation in optical coherence tomography



IOP Publishing Phys. Med. Biol. 66 (2021) 245010 https://doi.org/10.1088/1361-6560/ac3a23

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
4 June 2021

REVISED
3 November 2021

ACCEPTED FOR PUBLICATION
16 November 2021

PUBLISHED
9 December 2021

PAPER

Group-wise context selection network for choroid segmentation in optical coherence tomography

Fei Shi[1], **Xuena Cheng**[1], **Shuanglang Feng**[1], **Changqing Yang**[1], **Shengyong Diao**[1], **Weifang Zhu**[1], **Dehui Xiang**[1], **Qiuying Chen**[2], **Xun Xu**[2], **Xinjian Chen**[1,3,*] **and Ying Fan**[2,*]

[1] MIPAV Lab, the School of Electronics and Information Engineering, Soochow University, Suzhou 215006, People's Republic of China
[2] The First People's Hospital Affiliated to Shanghai Jiao Tong University, Shanghai 200080, People's Republic of China
[3] The State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou 215123, People's Republic of China
[*] Authors to whom any correspondence should be addressed.

E-mail: shifei@suda.edu.cn, 834614265@qq.com, 2470697802@qq.com, 649448381@qq.com, 20204228031@stu.suda.edu.cn, wfzhu@suda.edu.cn, xiangdehui@suda.edu.cn, Izzie_Qiu@yeah.net, drxuxun@sjtu.edu.cn, xjchen@suda.edu.cn and mdfanying@sjtu.edu.cn

Keywords: choroid segmentation, optical coherence tomography, channel attention, spatial attention, deep learning

Abstract

Choroid thickness measured from optical coherence tomography (OCT) images has emerged as a vital metric in the management of retinal diseases such as high myopia. In this paper, we propose a novel group-wise context selection network (referred to as GCS-Net) to segment the choroid of either normal or high myopia eyes. To deal with the diverse choroid thickness and the variable shape of the pathological retina, GCS-Net adopts the group-wise channel dilation (GCD) module and the group-wise spatial dilation module, which can automatically select group-wise multi-scale information under the guidance of channel attention or spatial attention, and enhance the consistency between the receptive field and the target area. Furthermore, a boundary optimization network with a new edge loss is incorporated to improve the resulting choroid boundary by deep supervision. Experimental results evaluated on a dataset composed of 1650 clinically obtained OCT B-scans show that the proposed GCS-Net can achieve a Dice similarity coefficient of 95.97 $\pm$ 0.54%, which outperforms some state-of-the-art segmentation networks.

1. Introduction

The choroid layer, sandwiching between the retinal pigment epithelium (RPE) and the sclera, has complex vasculature and performs critical physiological functions (Bill *et al* 1983, Norren and Tiemeijer 1986, Parver 1991, Alm and Nilsson 2009, Nickla and Wallman 2010). Changes of the choroid thickness and volume are the manifestation of various eye diseases, such as pathological myopia (PM), age-related macular degeneration (AMD), central serous choroiretinopathy (CSC), Vogt-Koyanagi-Harada syndrome, and choroiditis (Imamura *et al* 2009, Manjunath *et al* 2011, Esmaeelpour *et al* 2011, Sim *et al* 2013, Dhoot *et al* 2013). In recent decades, the emergence and development of optical coherence tomography (OCT) imaging has largely facilitated the diagnosis of retinal/choroidal diseases (Huang *et al* 1991, Fujimoto and Swanson 2016). With the advance of OCT techniques, the deeper structures of the eye become visible, and the whole choroid region can be visualized (Yasuno *et al* 2007). Accurate quantification of the choroid thickness from OCT scans is of great significance for the pathophysiology studies of diseases associated with choroid, and can help the prediction, diagnosis, and management of these diseases. In this paper, we propose an automatic method to segment the choroid region for either normal or high myopia retina, from wide-view OCT scans including both macula and optic nerve head (ONH) area, as shown in figure 1.

Many existing algorithms (Li *et al* 2012, Alonso-Caneiro *et al* 2013, Danesh *et al* 2014, Chen *et al* 2015, 2016, Shi *et al* 2016) divide choroid segmentation into two tasks, first detecting the Bruchs membrane (BM) as its upper boundary and then detecting the choroidal-scleral interface (CSI) as the lower boundary. The BM with high contrast can be segmented using gradient-based methods, while the CSI is more subtle and therefore more regional or textural constraints need to be added. These traditional algorithms have certain limitations. Some are

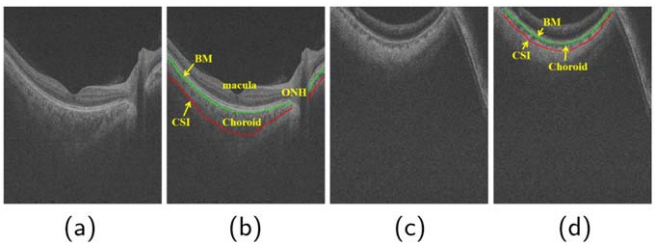


Figure 1. B-scans from wide-view volumetric OCT scans. (a) A B-scan of a normal eye. (b) The same B-scan in (a) with several structure manually labeled. (c) A B-scan from an eye with high myopia. (d) The same B-scan in (c) with several structure manually labeled.

only applicable to high-SNR B-scans obtained by compounding, and some are only applicable to normal retinas, or to macular centered OCT scans. For high myopia (HM) retina, low signal intensity, folding of the retina caused by the long axial length, and abnormal regions such as retinoschisis and choroidal neovascularization make detection of the retinal and choroidal structure more complicated and difficult. Moreover, for OCT scans containing the optical nerve head (ONH), gradient-based methods will give inaccurate results if no proper pre-processing is applied (Yu *et al* 2018). In this paper, we seek to use the deep learning method to directly segment the choroid region, avoiding any pre-processing steps.

Convolutional neural networks (CNNs) have been widely applied for medical image segmentation, with some applications on retinal OCT segmentation. In Fang *et al* (2017), a framework was proposed combining CNN and graph search method for automatic segmentation of nine retinal layer boundaries from OCT images. The U-Net (Ronneberger *et al* 2015) added skip-connection on the basis of fully convolutional network to integrate saliency prediction at multiple resolutions, and achieves high performance in medical image segmentation. ReLayNet (Roy *et al* 2017) used such encoder-decoder configuration and incorporate unpooling stages with skip connections for improved spatial consistency, and used a composite loss function to achieve end-to-end segmentation of retinal layers and fluid masses in eye OCT scans. CE-net (Gu *et al* 2019) improved U-Net using dense atrous convolution (DAC) block and residual multi-kernel pooling (RMP) block to capture more high-level features and preserve more spatial information, and achieved good results on several tasks including retinal layer segmentation. For choroid segmentation, in Sui *et al* (2017), a multi-scale CNN was proposed to learn the graph-edge weights needed for graph search methods. In Masood *et al* (2019), BM was segmented using a series of morphological operations, whereas the choroid layer was segmented using CNN. In Kugelman *et al* (2019), several patch-based or semantic segmentation CNN methods were tested and compared, and it was concluded that the semantic segmentation CNNs performed significantly better in detecting the CSI. In Tsuji *et al* (2020), the SegNet (Badrinarayanan and Cipolla 2017) was applied for choroid segmentation.

Despite that U-shaped networks have achieved high performance on different benchmarks, a significant disadvantage of such networks is that it perform a unified processing of all channels of feature maps within the same layer, which will result in the same receptive field and single-scale information in one layer. With the consecutive pooling and strided convolutional operations, this drawback becomes more prominent. Since each layer of the network only acquires single-scale local information, spatial information is gradually weakened as the feature map is downsampled layer by layer in the encoder side, and the global context information finally obtained by the encoder is gradually weakened as the feature map is upsampled layer by layer in the decoder side.

Aiming at such shortcomings, many networks were proposed to optimize feature extraction. For example, SegNet (Badrinarayanan and Cipolla 2017) utilized the saved pool indices to recover the reduced spatial information. The global convolutional network (Chao *et al* 2017) adopted large kernel size to expand the receptive field. Pyramid pooling module was proposed in PSPNet (Parver 1991) to fuse multi-scale features to tackle objects of different size. Recently, the attention module becomes an increasingly powerful tool for deep learning, which allows feature recalibration (Wang *et al* 2018, Fu *et al* 2019, Hu *et al* 2020). However, these networks achieve better performance at the expense of computational time and memory cost. The dilated/atrous convolution (Yu and Koltun 2016, Chen *et al* 2017, 2018) is another way to expand the receptive field without increasing the amount of parameters.

In this paper, motivated by the idea of the group convolution (Krizhevsky *et al* 2012, Chollet 2017, Gao *et al* 2021), we design a lightweight group-wise context selection network (GCS-Net) for choroid segmentation in OCT images. The core idea of GCS-Net is to group a number of feature maps, and to automatically select group-wise multi-scale information under the guidance of channel attention or spatial attention. This will enhance the consistency between the receptive field and the target area, thereby enhancing the learning of useful information.

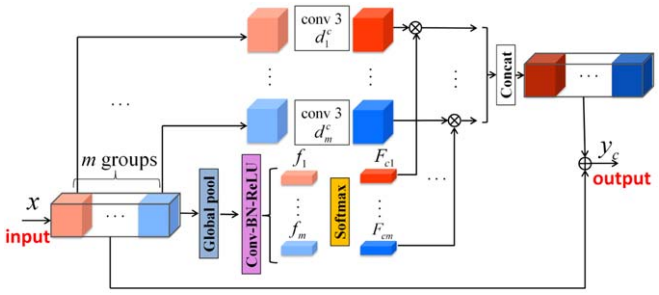


Figure 2. Components of the group-wise channel dilation (GCD) module.

The contributions of the paper are listed as follows:

- We propose a novel framework named GCS-Net, which can autonomously select group-wise context information that corresponds to different sizes of receptive field, so that targets with different scales can be segmented accurately. At the same time, GCS-Net is extremely lightweight with much less parameters than the state-of-the-art deep networks for segmentation.
- We design a group-wise channel dilation (GCD) module to be inserted in the skip-connection of the GCS-Net. The purpose of this module is to autonomously select the group-wise multi-scale information under the guidance of channel information.
- We design a group-wise spatial dilation (GSD) module to be placed in the decoding path of the GCS-Net. The purpose of this module is to autonomously select the group-wise multi-scale information under the guidance of spatial information.
- We design a new edge loss as part of the joint loss, which deals with the difficulty caused by blurred choroid boundary by the means of deep supervision.

The rest of the paper is organized as follows: section 2 details the structure of GCS-Net and the joint loss used by the network. In section 3, the dataset, implementation details and evaluation metrics are described and explained. Subsequently, results of ablation experiments and comparison experiments are given and discussed in section 4. Section 5 concludes the entire work and gives some discussions.

2. Methods

In this section, the self-attention mechanism is first introduced which motivates the two proposed modules for automatic selection of group-wise multi-scale information. Then the GCD module and GSD module are introduced in detail. Subsequently, the overall network structure is given with explanations on how the two modules work for our segmentation task. Finally, the joint loss is described.

2.1. Self-attention mechanism

Recently, self-attention modules are proposed to enhance the discriminating ability of feature representations by allowing feature recalibration with adaptive weights calculated from the feature maps. These modules can be categorized into channel attention and spatial attention modules. Used separately (Zhao *et al* 2017, Hu *et al* 2020) or combined (Woo *et al* 2018, Cao *et al* 2019), they achieved good results in computer vision tasks. The channel attention module spatially compresses a set of feature maps by global average pooling to obtain channel-wise importance, and thereby assigning different weights to each channel. The spatial attention module performs a series of compression operations on a group of feature maps along the channel axis to obtain a spatial feature description map, which is used to guide the optimization of the feature map. In this paper, we integrate multi-scale dilated convolution into these two types of modules, to further allow feature recalibration over different scales.

2.2. Group-wise channel dilation module

Figure 2 shows the proposed group-wise channel dilation module. The input feature maps go through global average pooling, convolution, activation to return m groups of feature maps.Then group-wise softmax regression is performed to return m groups of channel weights. Meanwhile, the input feature maps are evenly

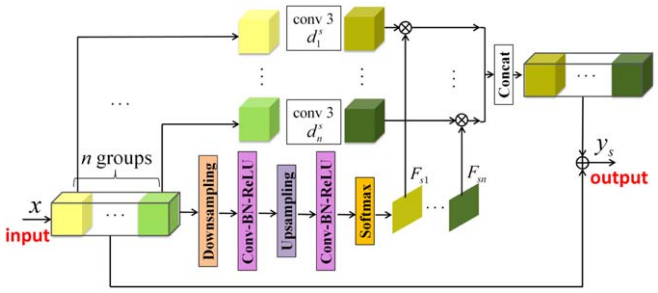


Figure 3. Components of the group-wise spatial dilation (GSD) module.

divided into m groups and dilated convolution with dilation ratios of $d_1^c,\cdots,d_m^c$ are applied to each group, respectively. Consequently, m groups of feature maps with different receptive field are obtained. Then the m groups of feature maps are multiplied with the corresponding channel weights to emphasize automatic selection of features of different scales. Finally, concatenation and residual operations are used to tackle performance degradation associated with deepening of the network. Therefore, the module automatically selects multi-scale information obtained from a group of feature maps under the guidance of the channel information.

2.3. Group-wise spatial dilation module

Figure 3 shows the proposed group-wise spatial attention module. The input feature maps go through a series of operations to obtain n feature maps with spatial weights. Among these operations, the downsampling serves to get more global information, and the upsampling restores the size of the feature map. Meanwhile, the input feature maps are divided into n groups. Note that in this module it is not necessary to have exactly the same number of channels in each group. Dilated convolution with dilation ratios of $d_1^s,\cdots,d_n^s$ are applied to each group, respectively. Consequently, n groups of feature maps with different scales are obtained. Then these feature maps are multiplied with the corresponding spatial weights. Finally, concatenation and residual operations are used to obtain the output feature maps. Therefore, the module automatically selects multi-scale information obtained from a group of feature maps under the guidance of spatial information.

2.4. Loss function

One major difficulty of choroid segmentation is that the lower boundary (CSI), defined between the choroid vasculature and the sclera, is often blurred. Meanwhile, the high level of noise and low contrast of pathological OCT scans can cause deflected results for both upper and lower boundaries. Therefore, a new edge loss L_{EDice}, which is the Dice loss calculated by edge probabilities, is used to make the network more sensitive to choroid boundary information, defined as in (1)

$$L_{EDice} = 1 - \frac{2\sum_i^N p_i' Sobel(g_i) + \epsilon}{\sum_i^N p_i' + \sum_i^N Sobel(g_i) + \epsilon}, \tag{1}$$

where N is the total size of the prediction map, $p_i' \in [0,\ 1]$ denotes the predicted probability for choroid boundaries, $g_i \in \{0, 1\}$ denote the ground truth label for choroid region, and the Sobel edge detection operator is used to extract the edge ground truth, and ϵ is a small smoothing factor.

In this paper, the proposed edge loss as well as the binary cross-entropy (BCE) and Dice loss commonly used in the segmentation tasks, comprise the total loss to jointly optimize the network. Among them, BCE can solve the problem of data imbalance in medical images, while Dice loss is more focused on small target segmentation

$$L_{BCE} = -\frac{1}{N}\sum_i^N [g_i \log(p_i) + (1 - g_i)\log(1 - p_i)], \tag{2}$$

$$L_{Dice} = 1 - \frac{2\sum_i^N p_i g_i + \epsilon}{\sum_i^N p_i + \sum_i^N g_i + \epsilon}, \tag{3}$$

where $p_i \in [0, 1]$ denotes the predicted probability for the choroid region, and $g_i \in \{0, 1\}$ denote the ground truth label for choroid region.

2.5. Network architecture

With GCD and GSD, we propose our group-wise context selection network for choroid segmentation as illustrated in figure 4. Since GCD and GSD are powerful in acquiring multi-scale contexts, we use a lightweight U-Net with only four layers of encoder-decoder architecture as our baseline. Two 3 × 3 convolutions and max

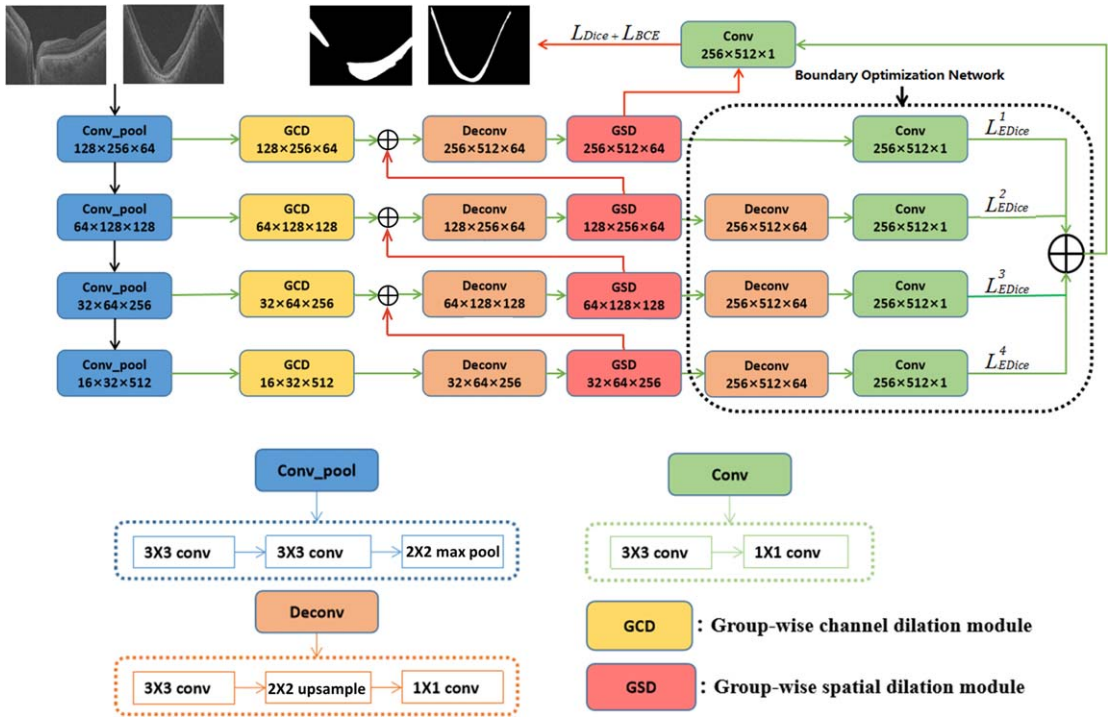


Figure 4. An overview of the group-wise context selection network (GCS-Net)

pooling are used in each layer of the encoding path to quickly acquire feature maps of different resolutions. Multiple simple decoder blocks are used in decoder path to restore the high resolution feature maps efficiently. Except for the last two convolutional layers for final prediction output, convolutional layers in the network are followed by batch normlization (BN) and rectified linear unit (ReLU), which are not shown in figure 4 for compact illustration. The GCD modules are applied in the skip-connection. The purpose is to obtain useful multi-scale information and transfer it to the decoder side. The GSD modules are applied in the decoder side, inserted between the deconvolution layers. The purpose is both to acquire multi-scale information and to compensate for the loss of global information caused by upsampling in the decoding process.

The boundary optimization network (BON) is a subnet which emphasizes boundary information in the final prediction map, and achieves deep supervision by the proposed edge loss. In BON, four edge prediction maps are generated, for which the edge loss is calculated respectively. Note that after the last three GSD modules, deconvolutions are added to unify the size of the feature map. The obtained semantic boundary information of different scales is added and merged into the last convolution of the network, so that the edges in the predicted map can be refined.

Then the final joint loss function is defined as:

$$L_{Total} = L_{BCE} + \alpha L_{Dice} + \beta \sum_{i=1}^{4} L^{i}_{EDice}, \tag{4}$$

where α and β are predefined weights.

3. Experimental settings

3.1. Dataset

The dataset used in this paper contained B-scans from volumetric OCT scans acquired by a Topcon Atlantis DRI-1 swept source OCT scanner (Topcon Corp., Tokyo, Japan) at the First People's Hospital Affiliated to Shanghai Jiao Tong University, China. The scan range included the macular center and the ONH region. The volumetric image size was $512 \times 992 \times 256$ (width $\times$ height $\times$ B-scans), which corresponded to a $12 \times 2.6 \times 9\,\text{mm}^3$ volume. The collection and analysis of image data were approved by the Institutional Review Board of the First People's Hospital Affiliated to Shanghai Jiao Tong University and adhered to the tenets of the Declaration of Helsinki. An informed consent was obtained from each subject. The dataset included a total of 1650 OCT B-scans with ground truth. Among them, 1150 B-scans came from 115 normal eyes, 500 B-scans came from 50 eyes with high myopia, where ten evenly spaced B-scans were extracted from each OCT volume.

Table 1. Results of ablation experiments for model complexity.

Methods	IoU(%)	DSC(%)	Sen(%)	Spe(%)	#params
Baseline_4layer	90.38 ± 0.80	94.67 ± 0.59	95.36 ± 0.95	94.31 ± 1.27	4.88M
Baseline_5layer	90.56 ± 0.38	94.78 ± 0.22	96.23 ± 1.32	93.68 ± 1.23	9.87M
GCS-Net_4layer(w/o BON)	**91.68 ± 1.01**	**95.42 ± 0.68**	**96.29 ± 0.67**	94.95 ± 0.47	7.50M
GCS-Net_5layer(w/o BON)	91.58 ± 0.88	95.35 ± 0.57	96.18 ± 0.67	**94.97 ± 0.67**	17.65M

The upper and lower boundaries of choroid were delineated by a clinical professional using the software ITK-SNAP (version 3.4.0) (Yushkevich *et al* 2006). Adjacent B-scans might be used for reference to assist delineation of subtle boundaries. The B-scans were evenly divided into five folds, and it was made sure all B-scans from one subject only occur in one of the five folds, so that training and testing data were from different subjects. Five-fold cross validation was performed to get all the reported results.

3.2. Implementation details

The experiments were performed on the public platform PyTorch with NVIDIA Tesla K40 GPU with 12 GB memory. In the training process, the batch size was 8, and the iteration number was 60. The stochastic gradient descent algorithm with an initial learning rate of 0.01, the momentum of 0.9 and weight decay rate of 0.0001 was used to optimize the network.

In the training process, we used random left-right flips for data augmentation, considering the symmetry of left and right eyes. To improve computational efficiency, we resized the input image size from 992 × 512 to 512 × 256, which almost kept the aspect ratio of the original B-scan.

3.3. Evaluation metrics

To quantitatively evaluate the performance of the proposed model, four segmentation evaluation metrics: intersection-over-union (IoU), Dice similarity coefficient (DSC), sensitivity (Sen) and specificity (Spe), are used in ablation study. In comparison with other methods, we add three additional indicators for the evaluation of choroid segmentation: the mean absolute boundary difference (ABD) of BM and CSI, and the mean thickness difference (TD) of the choroid (Shi *et al* 2016).

3.4. Parameter selection

We empirically selected the number of groups and the dilation ratios in the GCD and GSD modules. Various combinations of group numbers $m = 1, 2, 4$ and $n = 1 \sim 4$ were tested. (Note that $m = 3$ is not applicable to ensure the same group size in the GCD module.) For dilation ratios, $d_1^c = 1$ and $d_1^s = 1$ were fixed, representing non-dilated convolution for the first group, and various combinations of dilation ratios of 2, 4, 6 were tested for the other groups. The combination of $m = 2$ with $d_1^c = 1$, $d_2^c = 2$, and $n = 3$ with $d_1^s = 1$, $d_2^s = 2$, $d_3^s = 4$ resulted in the highest IoU index, and were used to obtain all reported results in the next section.

We also empirically selected the weights in the loss function as $\alpha = 1$ and $\beta = 1$. Larger β makes it more difficult for the model to converge, because small perturbation in the edge location can result in large variation of the edge loss.

4. Results

4.1. Ablation study

In this subsection, we present several plausible variations of the proposed GCS-Net, in order to highlight the importance of each of the proposed contributions.

4.1.1. Ablation study for model complexity

In order to verify that the performance improvement of the proposed model is not caused by the increase in network parameters, we deepened the baseline and GCS-Net (w/o BON) from four layers to five layers. The results and amounts of model parameters are listed in table 1. It can be seen that the proposed model has less parameters but better performance than the 5-layer baseline. Interestingly, for GCS-Net (w/o BON), theres no improvement of performance when the number of layers increases. There are probably two reasons for this. First, as the network is applied for segmentation of a single type of region, the number of parameters with four layers are already enough to capture the features of choroid regions. Too many parameters may cause some useful information to be overwhelmed during the encoding process, which will cause the network to fail to

Table 2. Results of ablation experiments for GCD and GSD modules.

Methods	IoU(%)	DSC(%)	Sen(%)	Spe(%)
Baseline	90.38 ± 0.80	94.67 ± 0.59	95.36 ± 0.95	94.31 ± 1.27
Baseline+GCD(non-dilated)	91.36 ± 0.99	95.22 ± 0.67	95.83 ± 0.72	95.03 ± 0.61
Baseline+GCD(no-channel)	91.23 ± 1.05	95.13 ± 0.73	95.93 ± 0.78	94.86 ± 0.54
Baseline+GCD	91.43 ± 1.01	95.24 ± 0.71	95.89 ± 0.78	95.11 ± 0.41
Baseline+GSD(non-dilated)	91.33 ± 1.11	95.21 ± 0.76	95.95 ± 0.88	94.90 ± 0.73
Baseline+GSD(non-spatial)	91.26 ± 1.11	95.13 ± 0.78	95.99 ± 0.74	94.92 ± 0.56
Baseline+GSD	91.40 ± 1.08	95.25 ± 0.74	96.07 ± 0.63	94.83 ± 0.66
GCS-Net(w/o BON)	91.68 ± 1.01	95.42 ± 0.68	96.29 ± 0.67	94.95 ± 0.47
GCS-Net	**92.59 ± 0.83**	**95.97 ± 0.54**	**96.66 ± 0.61**	**95.56 ± 0.36**

recover more semantic information during the decoding process. Moreover, increasing the depth of the network is nothing more than increasing the ability of acquiring multi-scale features. However, with multiple dilated convolutions, the proposed GCS-Net (w/o BON) is already powerful in multi-scale representations, making deeper structure unnecessary. Therefore, we adopt the four-layer GCS-Net (w/o BON), which is lightweight.

4.1.2. Ablation study for GCD module

We first studied a model variation with only the GCD modules integrated in the baseline, called Baseline+GCD. We also designed another two model variations for comparison, called Baseline+GCD (non-dilated) and Baseline+GCD (no-channel). In Baseline+GCD (non-dilated), all dilation ratios of the dilated convolution were set as 1, so that the module did not extract multi-scale information. In Baseline+GCD (no-channel), the channel attention mechanism was removed. As shown in table 2, all these variations have higher indices over the baseline, and GCD with both dilated convolution and channel attention achieves highest IoU, DSC and Spe. This shows both of these two components contribute to the final performance.

4.1.3. Ablation study for GSD module

Similarly, we designed three model variations called Baseline+GSD, Baseline+GSD (non-dilated) and Baseline +GSD (non-spatial). As shown in table 2, all these variations have higher IoU, DSC, Sen and Spe over the baseline, and GSD with both dilated convolution and channel attention achieves highest IoU, DSC, and Sen. This shows multi-scale dilated convolution and spatial attention contribute to the final performance.

4.1.4. Ablation study for the whole network

Firstly we compare Baseline+GCD, Baseline+GSD, and GCS-Net (w/o BON) with both modules integrated. As shown in table 2, compared with the baseline, the GCD module improves the IoU by 1.05% and the GSD module improves the IoU by 1.02%. Adding both GCD and GSD to the baseline improves the IoU by 1.30%. The DSC, Sen, and Spe are improved by 0.75%, 0.93%, and 0.64%, respectively. This shows both the GCD and GSD contribute to the final performance. Finally, we compare GCS-Net with GCS-Net (w/o BON), As show from the last two rows of table 2, the IoU, DSC, Sen, and Spe are improved by 0.91%, 0.55%, 0.37%, and 0.61%, respectively. This shows that the Boundary Optimization Network and the edge loss are effective in optimizing the boundary information of choroid by deep supervision.

Figure 5 shows the average feature maps output by the decoder from low to high resolutions for two B-scans by the baseline network and the GCS-Net. It shows that with the proposed modules, the GCS-Net can extract features more focused on the choroid region, especially for higher resolutions. Moreover, the model adapts well to both thick and thin choroids.

4.2. Comparisons to other methods

We compared the proposed GCS-Net with some state-of-the-art segmentation algorithms: scSE-Unet (Kugelman *et al* 2019), Res-Unet (Kugelman *et al* 2019), SegNet (Badrinarayanan and Cipolla 2017, Tsuji *et al* 2020), Deeplab (Chen *et al* 2018), PSPNet (Zhao *et al* 2017), and CE-Net (Gu *et al* 2019). Among these, scSE-Unet, Res-Unet, and SegNet were applied for choroid segmentation tasks (Kugelman *et al* 2019, Tsuji *et al* 2020), and CE-Net were applied for retinal layer segmentation tasks (Gu *et al* 2019), and were reported to achieve good performance. Especially, the scSE-Unet adopts spatial and channel squeeze-and-excitation blocks (Zhao *et al* 2017), as a form of self-attention. The performance comparisons are summarized in table 3, where the average indices calculated on the normal eyes, high myopia (HM) eyes, and the total dataset, are listed respectively. It can be seen that our method achieves the highest indices in most cases, and reaches 92.59%, 95.97%, 96.66%, and 95.56% in average IoU, DSC, Sen, and Spe, respectively. Regarding the choroid segmentation indicators, the

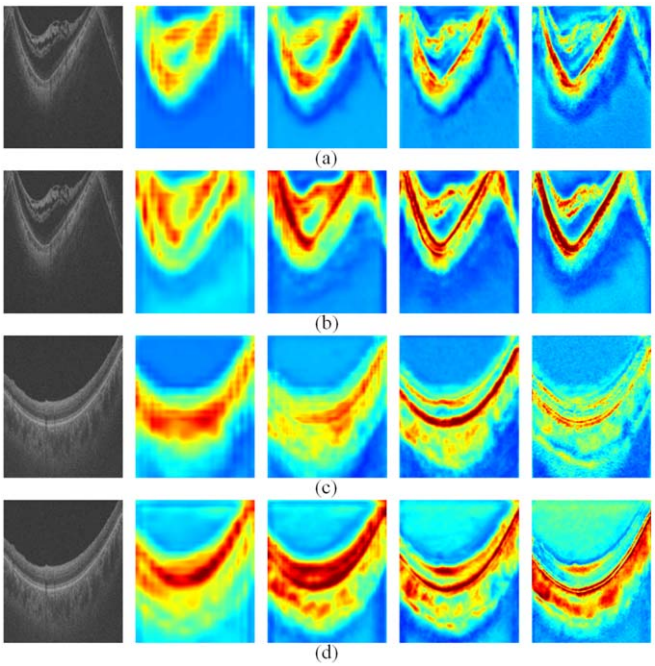


Figure 5. Comparison of feature maps. (a) Feature maps extracted by the baseline network for a high myopia subject. (b) Feature maps extracted by the GCS-Net for a high myopia subject. (c) Feature maps extracted by the baseline network for a normal subject. (d) Feature maps extracted by the GCS-Net for a normal subject. First column: input B-scans, second to fifth column: average feature maps output of the decoder from low to high resolutions.

mean ABD _BM, ABD_CSI, and TD are the lowest. At the same time, the proposed GCS-Net is lightweight with much less parameters than methods with competing performance. The total inference time for the test set is also listed in table 3. The proposed GCS-Net costs slightly longer than the Deeplab and the CE-Net, while being faster than scSE-Unet, SegNet and PSPNet. Figure 6 shows some B-scans with the results overlaid. Compared with other methods, the proposed method delineates choroidal boundaries more accurately both for normal retinas and retinas with high myopia, even in low-contrast regions or when the choroid is extremely thin. Moreover, the proposed method is not affected by pathological regions with similar intensities to the choroid (the fourth row in figure 6). Therefore, both quantitative and qualitative analysis of the results indicate that the ability of our proposed GCS-Net to automatically selects group-wise multi-scale information is very powerful and robust.

4.3. Quantitative analysis of choroid

With the proposed method, automatic quantification of the choroid thickness can be carried out. We calculated the average choroid thickness of a group of 44 eyes with pathological myopia (mean age 62.57 $\pm$ 3.99) and a control group of 96 normal eyes (mean age 62.30 $\pm$ 3.43). The average cho-roidal thickness was 97.95 μm for PM eyes versus 172.27 μm for normal eyes, showing significant choroid atrophy associated with PM. Figure 7 shows the choroid thickness maps of one normal eye and one eye with PM, respectively. Substantial difference in choroid thickness can be observed. From the thickness map, the thickness of different regions around the macula or the ONH, and their variations, can be further studied. This will support the research of related pathologies, and also be helpful for pathology prediction and treatment planning in clinical practice.

5. Conclusions and discussions

In this paper, we design a novel lightweight group-wise context selection network for choroid segmentation in B-scans from volumetric swept source OCT data. Multiple dilated convolutions are used to obtain multi-scale feature maps. Meanwhile, based on the self-attention mechanism, adaptive weights are calculated and used for automatic recalibration of multi-scale information. Thus, the feature maps whose receptive fields best fit the scale of the segmentation target area will be most emphasized in decision. Two modules, the GCD and GSD blocks, are designed to achieve group-wise multi-scale information selection in two different ways. By integrating these modules into a U-shaped baseline network, multi-scale information is obtained in each layer of the network, making the network powerful in segmenting both thick and thin choroid areas. A boundary

Table 3. Results compared with other methods.

Metrics		scSE-Unet	Res-Unet	SegNet	Deeplab	PSPNet	CE-Net	GCS-Net
IoU(%)	Total	86.88 ± 1.54	88.21 ± 1.03	91.04 ± 0.79	88.15 ± 1.22	90.10 ± 0.92	91.26 ± 0.86	**92.59 ± 0.83**
	Normal	90.63 ± 1.09	91.28 ± 0.69	93.51 ± 0.73	92.70 ± 0.93	92.76 ± 1.01	93.75 ± 0.81	**94.47 ± 0.75**
	HM	78.26 ± 4.10	81.14 ± 3.67	85.36 ± 2.20	77.70 ± 3.44	83.99 ± 2.00	85.54 ± 1.92	**88.29 ± 1.79**
DSC(%)	Total	92.33 ± 1.08	93.32 ± 0.72	95.00 ± 0.55	93.13 ± 0.90	94.49 ± 0.59	95.21 ± 0.56	**95.97 ± 0.54**
	Normal	94.94 ± 0.64	95.29 ± 0.33	96.56 ± 0.44	96.12 ± 0.56	96.15 ± 0.58	96.71 ± 0.47	**97.08 ± 0.43**
	HM	86.31 ± 3.15	88.79 ± 2.58	91.41 ± 1.60	86.24 ± 2.67	90.67 ± 1.42	91.76 ± 1.29	**93.42 ± 1.16**
Sen(%)	Total	93.95 ± 1.20	94.46 ± 2.08	95.91 ± 0.73	94.18 ± 0.86	95.49 ± 1.35	95.58 ± 0.46	**96.66 ± 0.61**
	Normal	97.08 ± 0.92	96.95 ± 1.71	96.97 ± 0.65	96.52 ± 0.65	96.80 ± 1.09	96.96 ± 0.58	**97.26 ± 0.52**
	HM	86.76 ± 3.22	88.75 ± 3.89	93.46 ± 1.54	88.80 ± 2.29	92.47 ± 2.42	92.40 ± 1.33	**95.27 ± 1.11**
Spe(%)	Total	91.84 ± 1.82	92.77 ± 2.10	94.71 ± 0.68	92.72 ± 1.05	93.98 ± 1.59	95.12 ± 0.65	**95.56 ± 0.36**
	Normal	93.26 ± 1.11	94.13 ± 2.01	96.24 ± 0.72	95.80 ± 0.68	95.63 ± 1.62	96.57 ± 0.53	**97.02 ± 0.36**
	HM	88.58 ± 4.30	89.65 ± 2.90	91.20 ± 1.34	85.64 ± 2.92	90.20 ± 1.85	91.78 ± 1.41	**92.20 ± 1.30**
ABD_BM(μm)	Total	4.09 ± 0.74	3.30 ± 0.63	2.69 ± 0.08	4.23 ± 0.28	3.47 ± 0.32	2.83 ± 0.14	**2.48 ± 0.20**
	Normal	3.27 ± 0.61	2.78 ± 0.50	2.27 ± 0.09	3.01 ± 0.14	2.87 ± 0.36	2.22 ± 0.13	**2.18 ± 0.14**
	HM	5.97 ± 1.66	4.50 ± 1.00	3.67 ± 0.26	7.04 ± 0.76	4.85 ± 0.49	4.22 ± 0.46	**3.16 ± 0.39**
ABD_CSI(μm)	Total	13.59 ± 1.86	12.77 ± 1.27	8.36 ± 1.06	9.39 ± 1.34	9.00 ± 1.19	8.15 ± 1.18	**6.95 ± 1.18**
	Normal	13.08 ± 2.12	12.57 ± 1.08	9.02 ± 1.51	9.44 ± 1.78	9.70 ± 1.62	8.68 ± 1.56	**7.50 ± 1.58**
	HM	14.76 ± 2.48	13.23 ± 3.06	6.83 ± 0.25	9.28 ± 0.79	7.38 ± 0.40	6.94 ± 0.42	**5.70 ± 0.40**
TD(μm)	Total	13.89 ± 1.84	12.89 ± 1.12	9.01 ± 1.13	10.90 ± 1.44	10.12 ± 1.37	8.96 ± 1.19	**7.84 ± 1.18**
	Normal	13.98 ± 2.25	13.34 ± 0.81	9.70 ± 1.58	10.35 ± 1.83	10.75 ± 1.83	9.35 ± 1.62	**8.30 ± 1.58**
	HM	13.67 ± 1.10	11.84 ± 2.62	7.42 ± 0.36	12.17 ± 1.15	8.66 ± 0.66	8.06 ± 0.61	**6.78 ± 0.60**
#params		5.0M	7.4M	29M	42M	45M	50M	8.4M
Inference time(1650 B-scans)(s)		248	165	294	151	265	140	162

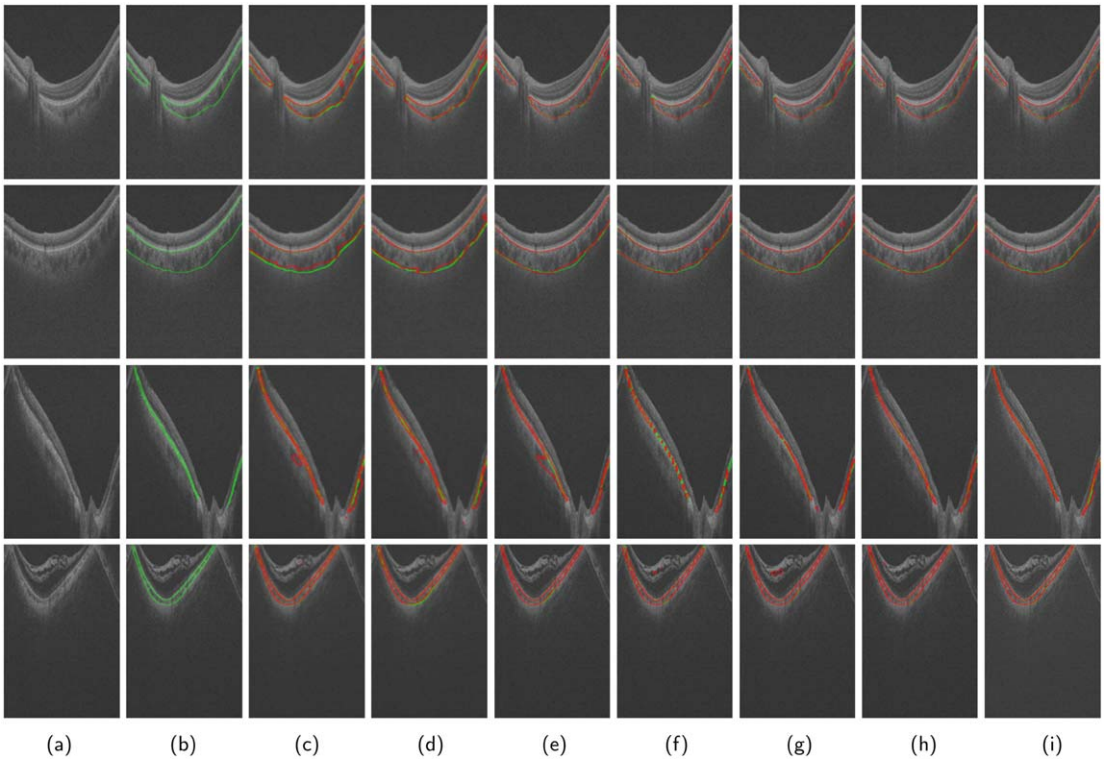


Figure 6. Example B-scans with choroid segmentation results overlaid. The green and red outline represents the ground truth and the prediction, respectively. The first and second rows are B-scans from normal subjects, while the third and fourth rows are B-scans from high myopia subjects. (a) Original images. (b) Ground truth. (c) scSE-UNet (Kugelman *et al* 2019). (d) Res-UNet (Kugelman *et al* 2019). (e) SegNet (Badrinarayanan and Cipolla 2017). (f) Deeplab (Chen *et al* 2018). (g) PSPNet (Hu *et al* 2020). (h) CE-Net (Gu *et al* 2019). (i) The proposed GCS-Net.

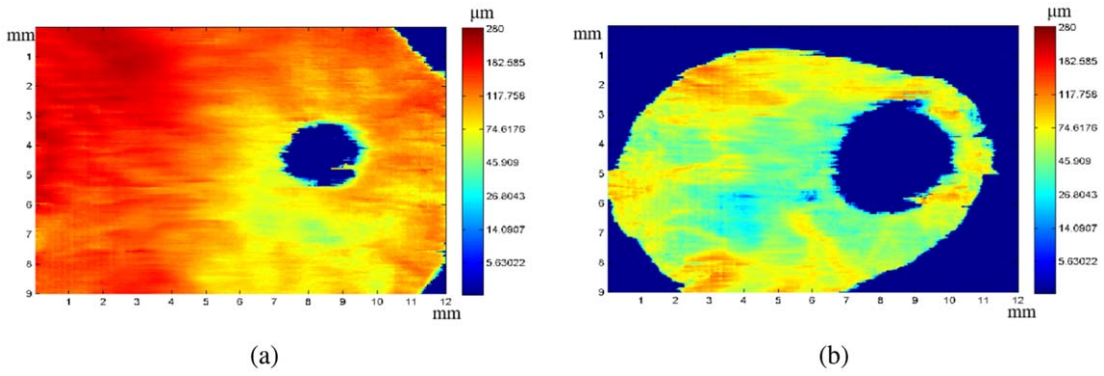


Figure 7. Example choroid thickness maps (a) normal eye (b) PM.

optimization network and a new edge loss are proposed to further tackle the difficulty of detecting weak choroid boundaries by deep supervision.

By comparing model variations, the experimental results show that both the multi-scale dilated convolution and the attention mechanism, both GCD and GSD blocks, and the boundary optimization network with edge loss contribute to the improvement of segmentation performance. The experimental results also show that the proposed method can accurately segment choroid of either normal or high myopia retinas and outperforms other state-of-the-art deep networks for segmentation.

Graph search based methods are another category of state-of-the-art choroid segmentation methods (Danesh *et al* 2014, Chen *et al* 2015, 2016, Shi *et al* 2016, Sui *et al* 2017), but they will generally fail when directly applied to our test images, where the existence of ONH compromises layer continuity and smoothness, and the folding of the retina and the pathological regions disrupt the spacial constraints. Nevertheless, the proposed

method can deal with these cases well. It automatically avoids the ONH region and the invalid region caused by image folding, and is not affected by the pathological regions. The method is efficient and effective with no pre-processing needed. It can meet the needs of quantitative analysis of choroid regions in both pathology study and clinical diagnosis and treatment.

In the proposed network, the network parameters such as the number of groups, the size of dilated convolutions, and the depth of the network were empirically selected, and suited the specific segmentation task. For more complicated segmentation tasks, more dilated convolutions with different sizes, and a deeper network structure may be adopted. In the future, we will optimize the network to further improve the choroid segmentation performance on pathological data. We will further investigate on how GCS-Net or its variations work for other segmentation tasks. We will also explore the usage of GCD and GSD blocks in other medical image analysis tasks, such as classification.

Acknowledgments

This work was 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 61 971 298 and 61 771 326.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alm A and Nilsson S F 2009 Uveoscleral outflow-a review *Exp. Eye Res.* **88** 760–8

Alonso-Caneiro D, Read S A and Collins M J 2013 Automatic segmentation of choroidal thickness in optical coherence tomography *Biomed. Opt. Express* **4** 2795–812

Bill A, Sperber G and Ujiie K 1983 Physiology of the choroidal vascular bed *Int. Ophthalmol.* **6** 101–7

Cao Y, Xu J, Lin S, Wei F and Hu H 2019 GCnet: non-local networks meet squeeze-excitation networks and beyond *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)* (https://doi.org/10.1109/ICCVW.2019.00246)

Chao P, Zhang X, Gang Y, Luo G and Jian S 2017 Large kernel matters—improve semantic segmentation by global convolutional network *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 1743–51

Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L 2018 Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs *IEEE Trans. Pattern Anal. Mach. Intell.* **40** 834–48

Chen L C, Papandreou G, Schroff F and Adam H 2017 Rethinking atrous convolution for semantic image segmentation arXiv:1706.05587

Chen Q, Fan W and Niu S 2015 Automated choroid segmentation based on gradual intensity distance in HD-OCT images *Opt. Express* **23** 8974–94

Chen Q, Niu S, Yuan S, Fan W and Liu Q 2016 Choroidal vasculature characteristics based choroid segmentation for enhanced depth imaging optical coherence tomography images *Med. Phys.* **43** 1649–61

Chollet F 2017 Xception: deep learning with depthwise separable convolutions *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Honolulu, HI, 21–26 July 2017)* (Picastaway, NJ: IEEE) pp 1800–7

Danesh H, Kafieh R, Rabbani H and Hajizadeh F J 2014 Segmentation of choroidal boundary in enhanced depth imaging OCTs using a multiresolution texture based modeling in graph cuts *Comput. Math. Methods Med.* **2014** 479268

Dhoot D S, Huo S, Yuan A, Ehlers S S, Traboulsi J P and Kaiser, P K E 2013 Evaluation of choroidal thickness in retinitis pigmentosa using enhanced depth imaging optical coherence tomography *Br. J. Ophthalmol.* **97** 66–9

Esmaeelpour M, Zay B P, Hermann B, Hofer B, Kajic V, Hale S L, Hale R V, Drexler W and Sheen N J L 2011 Mapping choroidal and retinal thickness variation in type 2 diabetes using three-dimensional 1060-nm optical coherence tomography *Investigative Opthalmol. Vis. Sci.* **52** 5311–6

Fang L, Cunefare D, Wang C, Guymer R H, Li S and Farsiu S 2017 Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search *Biomed. Opt. Express* **8** 2732–44

Fu J, Liu J, Tian H, Fang Z and Lu H 2019 Dual attention network for scene segmentation *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Long Beach, CA, 15–20 June 2019)* (Picastaway, NJ: IEEE) pp 3146–54

Fujimoto J and Swanson E 2016 The development, commercialization, and impact of optical coherence tomography *Investigative Ophthalmol. Vis. Sci.* **57** 1–13

Gao S H, Cheng M M, Zhao K, Zhang X Y, Yang M H and Torr P 2021 Res2net: a new multi-scale backbone architecture *IEEE Trans. Pattern Anal. Mach. Intell.* **43** 652–62

Gu Z, Cheng J, Fu H, Zhou K and Liu J J 2019 CE-Net: context encoder network for 2D medical image segmentation *IEEE Trans. Med. Imaging* **38** 2281–92

Hu J, Shen L, Albanie S, Sun G and Wu E 2020 Squeeze-and-excitation networks *IEEE Trans. Pattern Anal. Mach. Intell.* **42** 2011–23

Huang D, Swanson E A, Lin C P, Schuman J S, Stinson W G and Chang W 1991 Optical coherence tomography *Science* **254** 1178–81

Imamura Y, Fujiwara T, Margolis R, Xu D and Spaider R F 2009 Enhanced depth imaging optical coherence tomography of the choroid in central serous chorioretinopathy *Retina* **29** 1469–73

Krizhevsky A, Sutskever I and Hinton G E 2012 Image Net classification with deep convolutional neural networks *Int. Conf. on Neural Information Processing Systems (NIPS)* (https://proceedings.neurips.cc/paper/2012)

Kugelman J, Alonso-Caneiro D, Read S A, Hamwood J, Vincent S J, Chen F K and Collins M J 2019 Automatic choroidal segmentation in OCT images using supervised deep learning methods *Sci. Rep.* **9** 13298

Li Z, Lee K, Niemeijer M, Mullins R F and Sonka M 2012 Automated segmentation of the choroid from clinical SD-OCT *Investigative Opthalmol. Vis. Sci.* **53** 7510–9

Manjunath V, Goren J, Fujimoto J G and Duker J S 2011 Analysis of choroidal thickness in age-related macular degeneration using spectral-domain optical coherence tomography *Am. J. Ophthalmol.* **152** 663–8

Masood S *et al* 2019 Automatic choroid layer segmentation from optical coherence tomography images using deep learning *Sci. Rep.* **9** 3058

Nickla D L and Wallman J 2010 The multifunctional choroid *Prog. Retinal Eye Res.* **29** 144–68

Norren D V and Tiemeijer L 1986 Spectral reflectance of the human eye *Vis. Res.* **26** 313–20

Parver L M 1991 Temperature modulating action of choroidal blood flow *Eye* **5** 181–5

Ronneberger O, Fischer P and Brox T 2015 U-Net: convolutional networks for biomedical image segmentation *Int. Conf. on Medical Image Computing & Computer-assisted Intervention* (https://doi.org/10.1007/978-3-319-24574-4_28)

Roy A G, Conjeti S, Karri S P, Sheet D, Katouzian A and Wachinger C 2017 ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks *Biomed. Opt. Express* **8** 3627–42

Shi F, Tian B, Zhu W, Xiang D, Zhou L, Xu H and Chen X 2016 Automated choroid segmentation in three-dimensional 1-μm wide-view oct images with gradient and regional costs *J. Biomed. Opt.* **21** 126017

Sim D A, Keane P A, Mehta H, Fung S, Zarranz J, Fruttiger M, Patel P J and Egan C 2013 Repeatability and reproducibility of choroidal vessel layer measurements in diabetic retinopathy using enhanced depth optical coherence tomography *Investigative Ophthalmol. Vis. Sci.* **54** 2893–901

Sui X, Zheng Y, Wei B, Bi H, Wue J, Pang X, Yinh Y and Zhang S 2017 Choroid segmentation from optical coherence tomography with graph-edge weights learned from deep convolutional neural networks *Neurocomputing* **237** 332–41

Tsuji S, Sekiryu T, Sugano Y, Ojima A, Kasai A, Okamoto M and Eifuku S 2020 Semantic segmentation of the choroid in swept source optical coherence tomography images for volumetrics *Sci. Rep.* **10** 1088

V Badrinarayanan A K and Cipolla R 2017 SegNet: a deep convolutional encoder-decoder architecture for image segmentation *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 2481–95

Wang X, Girshick R, Gupta A and He K 2018 Non-local neural networks *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 7794–803

Woo S, Park J, Lee J Y and Kweon I S 2018 CBAM: convolutional block attention module *The European Conf. on Computer Vision (ECCV)* pp 3–19

Yasuno Y, Hong Y, Makita S, Yamanari M, Akiba M and Miura M 2007 *In vivo* high-contrast imaging of deep posterior eye by 1 μm swept source optical coherence tomography and scattering optical coherence angiography *Opt. Express* **15** 6121–39

Yu F and Koltun V 2016 Multi-scale context aggregation by dilated convolutions arXiv:1511.07122

Yu K, Shi F, Gao E, Zhu W, Chen H and Chen X 2018 Shared-hole graph search with adaptive constraints for 3D optic nerve head optical coherence tomography image segmentation *Biomed. Opt. Express* **9** 962–83

Yushkevich P A, Piven J, Cody Hazlett H, Gimpel Smith R, Ho S, Gee J C and Gerig G 2006 User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability *Neuroimage* **31** 1116–28

Zhao H, Shi J, Qi X, Wang X and Jian J 2017 Pyramid scene parsing network *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Honolulu, HI, 21–26 July 2017)* (Picastaway, NJ: IEEE) pp 2881–90