# The Channel Attention Based Context Encoder Network for Inner Limiting Membrane Detection

Hao Qiu[1,2], Zaiwang Gu[3], Lei Mou[2], Xiaoqian Mao[2], Liyang Fang[2],
Yitian Zhao[2], Jiang Liu[3], and Jun Cheng[2](✉)

[1] School of Mechatronic Engineering and Automation,
Shanghai University, Shanghai, China
[2] Cixi Institute of Biomedical Engineering,
Ningbo Institute of Industrial Technology,
Chinese Academy of Sciences, Ningbo, China
Chengjun@nimte.ac.cn
[3] Department of Computer Science and Engineering,
Southern University of Science and Technology, Shenzhen, China

**Abstract.** The optic disc segmentation is an important step for retinal image based disease diagnosis such as glaucoma. The inner limiting membrane (ILM) is the first boundary in the OCT, which can help to extract the retinal pigment epithelium (RPE) through gradient edge information to locate the boundary of the optic disc. Thus, the ILM layer segmentation is of great importance for optic disc localization. In this paper, we build a new optic disc centered dataset from 20 volunteers and manually annotated the ILM boundary in each OCT scan as ground-truth. We also propose a channel attention based context encoder network modified from the CE-Net [1] to segment the optic disc. It mainly contains three phases: the encoder module, the channel attention based context encoder module, and the decoder module. Finally, we demonstrate that our proposed method achieves state-of-the-art disc segmentation performance on our dataset mentioned above.

**Keywords:** Disc segmentation · ILM layer detection · Channel attention based context encoder

## 1 Introduction

Glaucoma is the second leading cause of blindness globally, which may result in vision loss and irreversible blindness. The number of people suffering from glaucoma is estimated to increase to 80 million in 2020 [2]. As the disease progresses asymptomatic in the early stages, the majority of the patients are unaware until an irreversible visual loss occurs. Thus, early diagnosis and treatment for glaucoma is utmost essential for preventing the deterioration of vision. While there

H. Fu et al. (Eds.): OMIA 2019, LNCS 11855, pp. 104–111, 2019.
https://doi.org/10.1007/978-3-030-32956-3_13

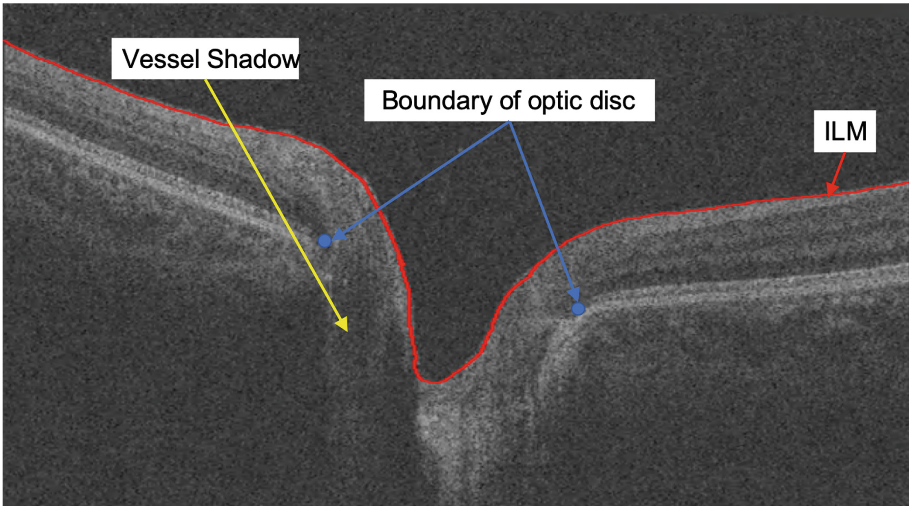

**Fig. 1.** Optic nerve head structure in a cropped OCT slice. The red curve denotes the ILM boundary. The blue points refer to the boundary points of the optic disc. ILM: Inner limiting membrane. (Color figure online)

are various approaches to diagnose glaucoma such as vessel distribution, FFT/B-spline coefficients, most of the known literature has endeavoured to assess the cup-to-disc ratio (CDR).

There have been a number of attempts at automatically detecting the optic disc in ocular images. Many proposed optic disc detection approaches concentrate on segmenting the optic region in color fundus images. For example, Liu *et al.* [3] proposed Variational level set approach for segmentation of optic disc without reinitialization. Xu *et al.* [4] employed the deformable model technique through minimization of an energy function to detect the disc. Cheng *et al.* [5] used the state-of-the-art self-assessed disc segmentation method combined three methods to segment the disc. However, these proposed approaches face challenges when the optic disc does not have a distinct color in the fundus image.

Optical coherence tomography (OCT), an important retinal imaging method with non-invasive, high-resolution characteristics, provides the fine structure within the human retina [6]. A single image of OCT slice is shown in Fig. 1. Some optic disc segmentation methods are applied to 3-D OCT volumes. For example, Lee *et al.* [7] applied a K-NN classifier to segment the optic disc cup and neuroretinal. Fu *et al.* [8] provided a Low-rank reconstruction to automatically detect optic disc in OCT slices.

With the development of convolutional neural network (CNN) in image and video processing [9], automatic feature learning algorithms using deep learning have emerged as feasible approaches and are applied to handle the image analysis. Recently, some deep learning based segmentation algorithms have been proposed to segment medical images [10], [1]. Based on the U-Net, a recent

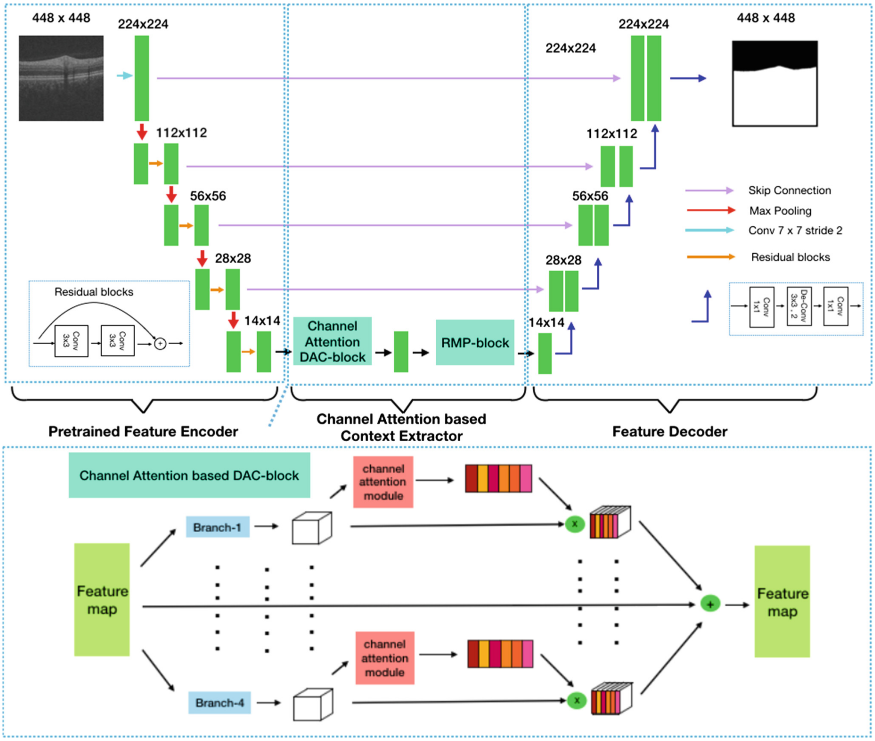

**Fig. 2.** Illustration of the proposed CACE-Net. Firstly, the images are fed into a feature encoder module, where the residual network (ResNet) block was employed as the backbone for each block, and then followed by a max-pooling layer to increase the receptive field for better extraction of global features. Then the features from the encoder module are fed into the proposed channel attention based context encoder module. Finally, the decoder module was used to enlarge the feature size and output a mask, the same size as the original input.

popular medical image segmentation architecture, CE-Net employs multi-scale atrous convolution and pooling operations to improve the segmentation performance. And it achieves some state-of-the-art performance in some medical image segmentation tasks, such as optic disc segmentation and OCT layers segmentation. The original context extractor module in CE-Net was consist of a dense atrous convolution (DAC) module and a residual multi-kernel pooling (RMP) module. However, the original DAC and RMP accounted for abundant channels to enrich the semantic features representations. Each channel of the features at the classification layer can be regarded as a specific-class response since we add the supervision signal on this layer. These abundant channels could be further embedded to produce the global distribution of channel-wise feature responses. In this paper, in order to extract more high-level semantic features, we introduce the channel attention mechanism to enhance the context extractor module of the

CE-Net, and propose a channel attention based context encoder network (called CACE-Net) for inner limiting membrane detection.

The major contributions of this work are summarized as follows:

(1) We annotate 20 3D-OCT scans (both of them are right eye scans) centered at optic disc.
(2) we leverage the ability of CACE-Net to accurately segment the inner limiting membrane (ILM) in our dataset, which is defined as the boundary between the retina and the vitreous body. This is necessary for our further work to detect the optic disc boundary points. The segmentations on database of OCT images are demonstrated to be superior to those from some known state-of-the-art methods. And we will release our code and dataset on Github later.

## 2 Proposed Method

The CE-Net [1] achieves the state-of-the-art performances in some 2D medical image segmentation tasks, such as optic disc segmentation, retinal vessel detection, lung segmentation and cell contour extraction. The proposed CACE-Net is modified from the CE-Net, which mainly contains three phases: the encoder module, the channel attention based context encoder module, and the decoder module, as shown in Fig. 2. The feature encoder module includes four encoder blocks, and the residual network (ResNet) block was employed as the backbone for each block, and then followed by a max-pooling layer to increase the receptive field for better extraction of global features. Then the features from the encoder module are fed into the proposed channel attention based context encoder module. Finally, the decoder module was used to enlarge the feature size and output a mask, the same size as the original input.

### 2.1 Channel Attention Based Context Extractor Module

The original context extractor module in CE-Net [1] employed four cascade branches with multi-scale atrous convolution to capture multi-scale semantic features, followed by various size pooling operations to further encode the multi-scale context features. This module accounts for abundant channels to enrich the semantic features representations, which could be further embedded to generate the global distribution of channel-wise feature responses. Therefore, motivated by the SE-Net [11], we propose a channel attention based context extractor module, introducing the relationship between channels.

In this section, we mainly introduce how to exploit the interdependencies of channel maps, as illustrated in Fig. 2. The proposed channel attention based context extractor module employs channel attention mechanism to allow the network to perform feature recalibration of aggregated context features, with the basis of original DAC block. Specially, the CACE module utilizes four cascade branches with multi-scale atrous convolution and channel attention module, to gain high-level features.

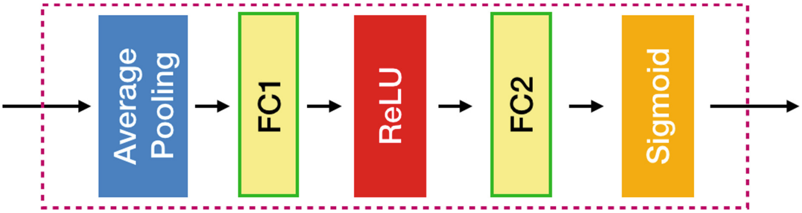

**Fig. 3.** Illustration of the channel attention module.

As illustrated in Fig. 3, the extracted feature map $F \in \mathbb{R}^{C\times H\times W}$ in channel attention module is first calculated directly by the global average pooling to generate channel-wise statistics $z \in \mathbb{R}^{C}$:

$$z_c = \frac{1}{H \times W} \Sigma_{i=1}^{H} \Sigma_{j=1}^{W} f_c(i,j) \tag{1}$$

where $H\times W$ represents the spatial dimensions of features and $C$ is the number of channels. Then, the two linear transformations $W_1, W_2$ and a sigmoid activation function $\sigma$ are employed to obtain the squeeze and excitation statistics $s \in \mathbb{R}^{C}$:

$$s_c = \sigma(W_2 \delta(W_1 z_c)) \tag{2}$$

where $\delta$ refers to the ReLU function, $W_1 \in \mathbb{R}^{\frac{C}{r}\times C}$ and $W_2 \in \mathbb{R}^{C\times \frac{C}{r}}$. Finally, a matrix multiplication between the statistics $s \in \mathbb{R}^{C}$ and the feature $F \in \mathbb{R}^{C\times H\times W}$ is added to obtain the final output in each branch of the proposed channel attention DAC module, followed by the RMP block for further context information with multi-scale pooling operations.

## 2.2 Feature Decoder

Instead of directly upsampling the features to the original image dimensions, we follow the CE-Net [1] to introduce a feature decoder module that restores the dimensions of the high level semantic features layer by layer. In each layer, we use ResNet block as the backbone of the decoder block which is followed by a 1 × 1 convolution, a 3 × 3 transposed convolution, a 1 × 1 convolution. Similar to U-Net [12], we add a skip connection between each layer of the encoder and decoder. Finally, the feature decoder module could generate the prediction of the same size as the original input.

## 2.3 Boundary Extractor

The main goal of this method is to detect internal limiting membrane. Therefore, we need to turn the segmentation prediction to a boundary line, which corresponds to the internal limiting membrane. We remove the small connected components to denoise the segmentation prediction, adopting the morphology method. After this post processing operation, we achieve the final boundary corresponding to the internal limiting membrane between the retina and the vitreous body.

### 2.4 Loss Function

In this method, we choose binary cross-entropy loss as our loss function $\mathcal{L}_B$, since the method just needs to predict the binary outputs. The binary cross-entropy loss is as follows:

$$\mathcal{L}_B = -\mathbb{E}_{\boldsymbol{x} \sim p_{data}}[\boldsymbol{y} \cdot \log(D(\boldsymbol{x})) + (1 - \boldsymbol{y}) \cdot \log(1 - D(\boldsymbol{x}))], \tag{3}$$

where $\boldsymbol{y}$ represents the ground truth, and $D(\boldsymbol{x})$ is the prediction.

## 3 Experiment Results

### 3.1 Dataset and Metric

20 3D-OCT scans (both of them are right eye scans) centered at optic disc were collected from 20 volunteers. Each OCT scan consisted of 885 × 512 image resolution. While there exist methods for extracting multiple retinal layers from OCT slices, only ILM layer boundaries is needed in our paper. The ILM is defined as the boundary between the retina and the vitreous body, which is the first boundary of retinal OCT. The ground-truth optic disc boundary of a 3D-OCT volume is obtained by first manually labeling the optic disc points in each optic disc centered slice (with a trained labeler and two experts for quality control). These labeled points were then to generate the ground-truth optic disc boundary. In our paper, we also randomly take 10 people's images for training, and others for testing. In this paper, we follow the same partition of the data set to train and test our models.

Following the previous approaches [1], we compute the mean absolute error (mae) between prediction and ground truth as the metric to evaluate the accuracy of segmentation algorithms.

$$error = \frac{1}{n}\sum_{i=1}^{n}|y_i - Y_i| \tag{4}$$

where $y_i$ represents the $i_{th}$ pixel predicted value of one surface, and $Y_i$ represents that of ground truth.

### 3.2 Implementation Details

The proposed CACE-Net was implemented on PyTorch library with the NVIDIA GPU. We choose stochastic gradient descent (SGD) optimization, other than adaptive moment estimation (Adam) optimization. We use SGD optimization since recent studies [13] show that SGD often achieves a better performance, though the Adam optimization convergences faster. The initial learning rate is set to 0.001 and a weight decay of 0.0001. We use poly learning rate policy where the learning rate is multiplied by $(1 - \frac{iter}{max\_iter})^{power}$ with power 0.9. All training images are rescaled to 448 × 448.

In order to demonstrate conclusively the superiority of the proposed method over the other methods, we compare our method with two algorithms for the ILM segmentation:

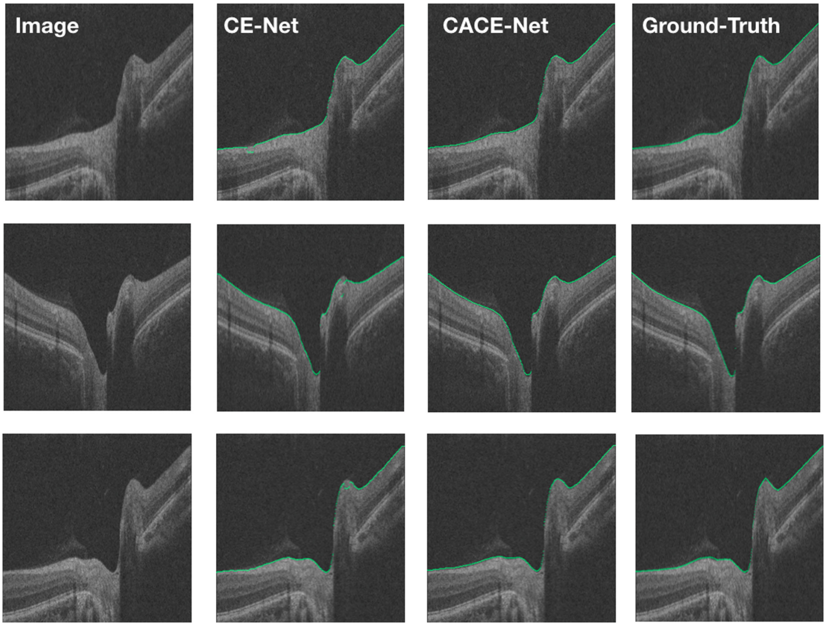

**Fig. 4.** Sample results of the ILM segmentation. From left to right: original images, CE-Net, CACE-Net and Ground-Truth

(1) U-net, a popular neural network architecture for biomedical image segmentation tasks.
(2) CE-Net [1], which achieves the state-of-the-art performances in some 2D medical image segmentation tasks, such as optic disc segmentation, retinal vessel detection, lung segmentation and cell contour extraction.

### 3.3 Results and Discussion

As can be seen in Table 1, we show the performances of three optic disc segmentation algorithms. Compared with other state-of-the-art optic disc segmentation methods, our CACE-Net outperforms the other algorithms based on deep learning image processing method. From the comparison shown in Table 1, the CACE-Net achieves 2.199 in the mean absolute error, better than the U-Net. From the comparison between CE-Net [1] and our CACE-Net, we also observe that there is a drop of the mean absolute error by 10.8% from 2.467 to 2.199.

**Table 1.** Performance comparison of the ILM detection (mean ± standard deviation)

| Method | U-Net | CE-Net | CACE-Net |
|---|---|---|---|
| *error* | 6.404 ± 16.407 | 2.467 ± 1.989 | 2.199 ± 1.471 |

We also show three sample results in Fig. 4 to visually compare our method with the most competitive methods, CE-Net. The comparison images show that our method obtain more accurate segmentation results.

## 4 Conclusion

In this paper, we have built a manually labeled OCT dataset and proposed an effective architecture for segmenting the ILM layer in our OCT dataset. The proposed CACE-Net achieves the mean absolute error of 2.199 in our dataset, better than other methods.

## References

1. Gu, Z., et al.: CE-NET: context encoder network for 2D medical image segmentation. IEEE Trans. Med. Imaging (2019)
2. Tham, Y.C., Li, X., Wong, T.Y., Quigley, H.A., Aung, T., Cheng, C.Y.: Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology **121**, 2081–2090 (2014)
3. Liu, J., et al.: Automatic glaucoma diagnosis through medical imaging informatics. J. Am. Med. Inform. Assoc. **20**, 1021–1027 (2013)
4. Xu, J., Chutatape, O., Sung, E., Zheng, C., Kuan, P.C.T.: Optic disk feature extraction via modified deformable model technique for glaucoma analysis. Pattern Recogn. **40**, 2063–2076 (2007)
5. Cheng, J., Yin, F., Wong, D.W.K., Tao, D., Liu, J.: Sparse dissimilarity-constrained coding for glaucoma screening. IEEE Trans. Biomed. Eng. **62**, 1395–1403 (2015)
6. Schmitt, J.M.: Optical coherence tomography (OCT): a review. IEEE J. Sel. Top. Quantum Electron. **5**, 1205–1215 (1999)
7. Lee, C.S., Tyring, A.J., Deruyter, N.P., Wu, Y., Rokem, A., Lee, A.Y.: Deep-learning based, automated segmentation of macular edema in optical coherence tomography. Biomed. Opt. Express **8**, 3440–3448 (2017)
8. Fu, H., Xu, D., Lin, S., Wong, D.W.K., Liu, J.: Automatic optic disc detection in OCT slices via low-rank reconstruction. IEEE Trans. Biomed. Eng. **62**, 1151–1158 (2014)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
10. Gu, Z., et al.: DeepDisc: optic disc segmentation based on atrous convolution and spatial pyramid pooling. In: Stoyanov, D., et al. (eds.) OMIA/COMPAY -2018. LNCS, vol. 11039, pp. 253–260. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00949-6_30
11. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
13. Keskar, N.S., Socher, R.: Improving generalization performance by switching from Adam to SGD. arXiv preprint arXiv:1712.07628 (2017)