Choroidal atrophy segmentation based on deep network with deep-supervision and EDT-auxiliary-loss

Lu, Ruyi, Zhu, Weifang, Cheng, Xuena, Chen, Xinjian, Kopriva, Ivica

Ruyi Lu, Weifang Zhu, Xuena Cheng, Xinjian Chen, Ivica Kopriva, "Choroidal atrophy segmentation based on deep network with deep-supervision and EDT-auxiliary-loss," Proc. SPIE 11313, Medical Imaging 2020: Image Processing, 113131X (10 March 2020); doi: 10.1117/12.2547763

SPIE.

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

Choroidal atrophy segmentation based on deep network with deep-supervision and EDT-auxiliary-loss

Ruyi Lu[1], Weifang Zhu1[1,2], Xuena Cheng[1], Xinjian Chen[1,3*]

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

[2]Collaborative Innovation Center of IoT Industrialization and Intelligent Production, Minjiang University, Fuzhou 350108, China

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

ABSTRACT

The prevalence of myopia is rapidly increasing worldwide. Along with the deepening of myopia, there will be various pathological changes of retina, such as choroidal atrophy, choroidal neovascularization, etc. In this paper, a U-Net based deep network is proposed to automatically segment choroidal atrophy in fundus images. We use U-Net as the main structure, which can learn rich hierarchical feature representations. In the decoder path, Squeeze-and-Excitation (SE) block is employed before each deconvolution to adaptively recalibrate channel feature response. We introduce deep-supervision mechanism and merge all the early prediction maps to obtain final prediction map. To ensure the smoothness of segmentation results, we propose a new loss function, which is termed EDT-auxiliary-loss (Euclidean distance transformation auxiliary loss). EDT-auxiliary-loss consists of Dice loss for ground truth and mean square error (MSE) loss for distance map. Another strategy for performance improvement is utilizing the information of optic disc (OD), which is usually adjacent to atrophy. The proposed method was evaluated on ISBI 2019 Pathologic Myopia Challenge dataset, which consists of 400 fundus images from161 normal eyes, 26 high myopia eyes and 213 pathologic myopia eyes. The proposed network was validated with four-fold cross validation. The experiment results show that the proposed method can successfully segment choroidal atrophy and achieve better performance than traditional U-Net.

KEYWORDS: Choroidal atrophy segmentation, deep network, Squeeze-and-Excitation, deep-supervision, EDT-auxiliary-loss

1. INTRODUCTION

Pathological myopia is a major cause of blindness around the world[1,2]. The pathological fundus changes of myopia patients are closely associated with their visual acuity. Choroidal atrophy[3] and parapapillary atrophy (PPA) [4] are the earliest pathological changes of pathological myopia. Lu made use of dual-channel color morphology and snake model to detect the boundaries of OD and PPA, which is based on visual appearance and anatomical features[5]. This is the first method proposed for choroidal atrophy segmentation, but the segmentation performance will be poor if the anatomical features are not obvious. H Li *et al* proposed a method to segment PPA by using evenly-oriented radial line segments and ellipse fitting[4]. However, this method is not effective when there are parapapillary hyperpigmentation or indistinct PPA boundaries. Septiarini *et al* proposed a subsection-based PPA segmentation method, which utilized statistical characteristics and back propagation neural network (BPNN) [6]. To our best knowledge, there are no literatures about choroidal atrophy segmentation in myopia fundus images.

The segmentation of myopia choroidal atrophy is challenging. High myopia and pathological myopia fundus are usually tessellated, caused by the tessellated choroid vessels, as shown in Fig.1(b). This may lead to the false positives. If the atrophy, which is adjacent to the optics disk (OD), is too severe that the sclera is exposed,

*Corresponding author: E-mail: xjchen@suda.edu.cn

Medical Imaging 2020: Image Processing, edited by Ivana Išgum, Bennett A. Landman,
Proc. of SPIE Vol. 11313, 113131X · © 2020 SPIE · CCC
code: 1605-7422/20/$21 · doi: 10.1117/12.2547763

the boundaries between the atrophy and OD will be blurry and hard to identify. In severe cases, atrophy areas are big enough to merge and change the morphology of fundus, as shown in Fig.1(c). Another factor that affects segmentation performance is the various sizes and shapes of atrophy. In early stage, choroidal atrophy areas are very small and even invisible, as shown in Fig.1(d).

In order to solve the mentioned difficulties, we propose a U-Net[7] based deep neural network with improvements by Squeeze-and-Excitation block (SE-block) [8] and deep-supervision[9]. A new loss function, combining Dice loss with mean square error (MSE) loss, is used to improve the smoothness of segmentation results. To restrain the interference of OD, OD and choroidal atrophy are segmented simultaneously.

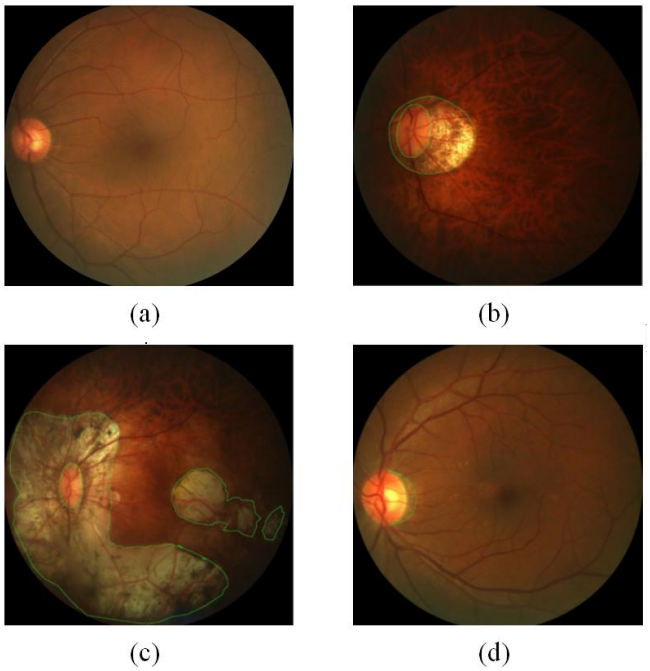

Fig.1 Examples of different kinds of fundus images. The region with green boundary represents choroidal atrophy. (a) Normal fundus. (b)Tessellated fundus. (c) PM fundus with severe choroidal atrophy. (d) Fundus with choroidal atrophy in early stage.

2. METHODS

2.1 Overall structure of the network

In recent years, deep networks based on U-Net have been widely used in medical image segmentation. U-Net consists of encoder and decoder path. We use U-Net as the main structure of our network and improve its decoder path. Fig.2 illustrates the overall structure of the proposed network. The encoder path contains five double convolutions and four max-poolings. Successive encoding helps the network to abstract rich hierarchical feature maps[10]. In the decoder path, four deconvolution layers and three SE-blocks are employed in a cascaded way to restore the resolution of the feature maps. One SE-block, which can make the feature maps more effective, connects two deconvolution layers. Fig.3 depicts the structure of SE-block. Skip connections copy the feature maps of encoder path and merge them with corresponding decoder feature maps.

Along with each decoder layer, there is an early prediction map. The proposed EDT-auxiliary-loss (Euclidean distance transformation auxiliary loss) function is applied to each prediction map. As shown in Fig.4, to calculate the EDT-auxiliary-loss, the dimension of prediction map is reduced by two parallel 1×1

convolutions and two independent feature channels are generated. Then, softmax activation and sigmoid activation are employed, respectively. After these operations, we not only obtain the multi-task segmentation results that are pixel-wise predictions of OD and atrophy, but also obtain their respective prediction distance maps. Final prediction map is the fusion result of the early prediction maps. The final choroidal atrophy segmentation result is extracted from the multi-task segmentation result of the final prediction map.

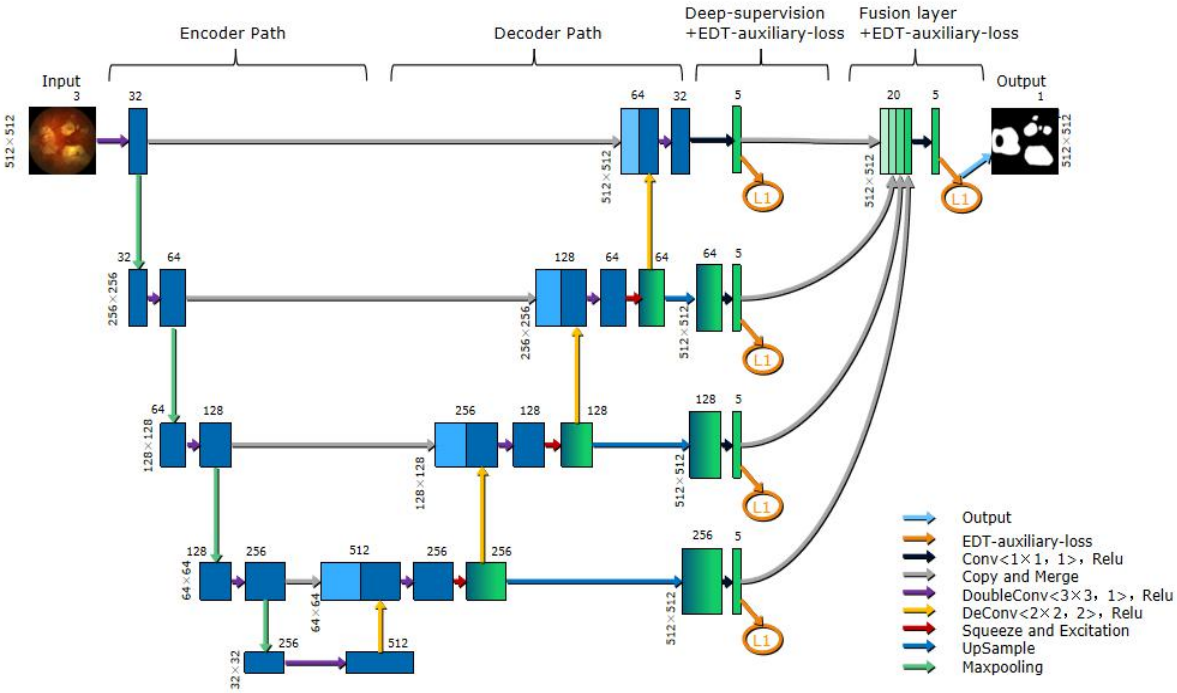

Fig.2 Overall structure of the network.

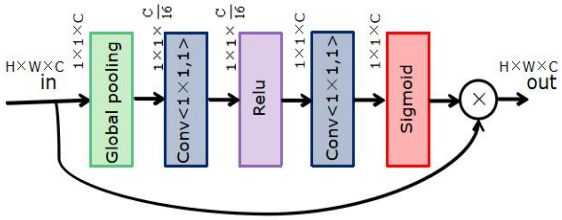

Fig.3 Squeeze-and-Excitation block.

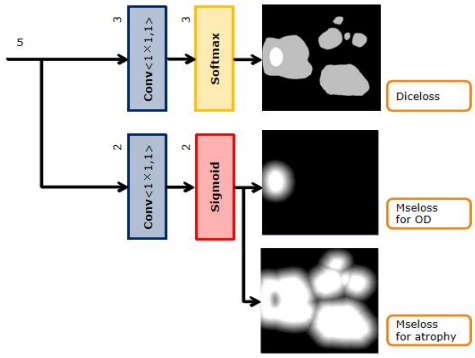

Fig.4 Preparing for EDT-auxiliary-loss calculation.

2.2 EDT-auxiliary-loss function

Most of the deep networks derived from U-Net produce coarse and unsmooth segmentation results, which may due to the drawbacks of architecture and the improper loss function[11,12]. On one hand, the continuous

down-samplings in the encoder path cause the loss of spatial information. On the other hand, Dice loss[13], the common used loss function in medical image segmentation, doesn't consider the spatial information. To overcome this issue, EDT-auxiliary-loss function is proposed, which is a weighted sum of Dice loss for ground truth and mean square error (MSE) loss for distance map. Distance map is a Euclidean distance transformation (EDT) of ground truth. The pixels in distance map represent the proximity of the pixel to the mask boundary. Specially, the pixels in mask corresponding region are set to 255. Hence, as shown in the low right corner of Fig.4, distance map has smooth boundaries, which provide numerical continuity from foreground to background and produce constrains for spatial information.

The EDT-auxiliary-loss can be expressed as,

$$\mathrm{L}_{\mathrm{joint}} = \alpha_1 L_{\mathrm{dice}} + \alpha_2 L_{\mathrm{mse}} \tag{1}$$

$$\mathrm{L}_{\mathrm{dice}} = 1 - \sum_{k}^{\mathrm{K}} \frac{2\lambda_k \sum_i^N p_{k,i} g_{k,i}}{\sum_i^N p_{k,i} + \sum_i^N g_{k,i}} \quad , \quad \mathrm{L}_{\mathrm{mse}} = \sum_{k}^{K} \frac{\omega_k \sum_i^N (Dp_{k,i} - Dg_{k,i})^2}{N} \tag{2}$$

Where N is the number of pixels. K is the number of class. L_{dice} and L_{mse} represent Dice loss and MSE loss, respectively. $p_{k,i}$ and $g_{k,i}$ denote the prediction probability and ground truth for class k, respectively. $Dp_{k,i}$ and $Dg_{k,i}$ denote prediction proximity and true proximity in distance maps for class k, respectively. λ_k and ω_k denote the class weights. α_1 and α_2 are the scale factors. In experiments, we set $K = 2$ for OD and atrophy segmentation. We set $\alpha_1 = 1/3$ and $\alpha_2 = 2/3$ empirically. To highlight the importance of atrophy segmentation, we set $\lambda_k = \omega_{\mathrm{k}} = 1/3$ for OD and $\lambda_k = \omega_k = 2/3$ for atrophy.

The mask prediction and distance map prediction share the parameters of network. The segmentation task pays more attention to smooth boundaries when extracting feature information. For false positives near the foreground, the EDT-auxiliary-loss is relatively small. The EDT-auxiliary-loss can be applied to other medical image segmentation task.

2.3 Improved Deep-supervision

Inspired by the side-output layer of M-Net[14], deep-supervision mechanism is introduced and improved in our network (shown in the right side of Fig.2). Deep supervision mechanism is first proposed in Deeply-Supervised Nets (DSN), which alleviates the problem of gradient vanishing and improves the discriminability of feature maps. Specifically, the output of each decoder layer is up-sampled to original size of input image, and the corresponding early prediction map is generated by 1×1 convolution. Different from M-Net, which directly takes the average of all the side-output maps as final output, we merge all early prediction maps and produce final prediction map by convolution. In other words, we introduce a fusion layer to achieve feature fusion for decoder path. The proposed EDT-auxiliary-loss is applied to all prediction maps. The average of the losses is back-propagated to the deep network.

3. RESULT

3.1 Dataset

The experimental fundus images are acquired from ISBI 2019 Pathologic Myopia Challenge, which contain 161 images from normal eyes, 26 images from high myopia eyes and 213 images from pathological myopia eyes. Online data augmentation strategies are used during training, including random cropping, random rotation and random horizontal flipping. The proposed method is evaluated with four-fold cross validation.

3.2 Experimental settings

Our deep network is implemented with PyTorch. In the training, we employ stochastic gradient descent (SGD) with momentum (momentum=0.9) as optimizer. Learning rate is initialized to 0.02 and updated with the poly mechanism,

$$lr = in\_lr(1 - step/epoch)^{0.9} \tag{3}$$

Where in_lr denotes initial learning rate. $step$ represents the current training round. We set $epoch = 100$, which is the total training rounds for one fold validation. In each $epoch$, the learning rate decreases with the increase of $step$.

We use a batch size of 4 and weight decay of 0.002. All input images are resized to 512×512.

3.3 Evaluation metrics

To quantitatively evaluate the segmentation performance, we use three common segmentation evaluation metrics: Dice similarity coefficient (DSC), Jaccard index (Jac) and Accuracy (Acc).

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{4}$$

$$Jac = \frac{TP}{TP + FP + FN} \tag{5}$$

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

Where TP, FP, TN and FN denote true positive, false positive, true negative and false negative, respectively.

3.4 Results

The comparison results between the proposed network and other networks are listed in Table 1. Paired Wilcoxon signed-rank tests were used to compare the results of different networks. As shown in Table 1, the performance of the proposed method is significantly improved than U-Net (with p-values <0.001 for DSC, Jac, and Acc) and Multi-task U-Net (with p-values of 0.048 for DSC, 0.033 for Jac, and 0.01 for Acc). The networks without EDT-auxiliary-loss function utilize Dice loss function.

To test the usefulness of deep-supervision and fusion layer, we chose Multi-task U-Net with deep-supervision (denoted as DS for short) and Multi-task U-Net with DS and fusion layer for comparison. The final result of Multi-task U-Net with DS is achieved by the average of local outputs. To test the usefulness of our proposed EDT-auxiliary-loss, we chose Multi-task U-Net with EDT-auxiliary-loss for comparison. The experiment results shown in Table 1 show the effectiveness of deep-supervision, fusion layer and EDT-auxiliary-loss for Multi-task U-Net.

To further test the effectiveness of each component employed in our method, we built models by removing every component or learning strategy from our method. As shown in Table 1, the proposed method has an enhanced performance compared with any other models. It is interesting to note that proposed method degrades without EDT-auxiliary-loss with a statistical significant margin (p-value<0.05), although the segmentation performance of Multi-task U-Net with EDT-auxiliary-loss is not enhanced significantly. These comparison results demonstrate the benefit of each component or learning strategy. Moreover, our method improves segmentation performance with just a little increase in the size of network parameters.

Fig.5 shows some results of choroidal atrophy segmentation with different methods. The first row is normal eye, the second row is high myopia eye, the third row is pathological myopia eye just with PPA and the fourth

row is more serious case of pathological myopia. For U-Net, the segmented choroidal atrophy is smaller than ground truth and the internal area of OD may be segmented as choroidal atrophy mistakenly. For multi-task U-Net, the segmentation is rough and there are many spatially isolated segmentation errors. In comparison, the results of our method are more consistent with ground truth.

Table 1 Comparison experiments and ablation experiments (w/o means without the following component).

	DSC(%)	Jac(%)	Acc(%)	Params size(M)
U-Net	71.32±1.11**$^{++}$	63.96±1.49**$^{++}$	97.76±0.26*$^{+}$	29.63
Multi-task U-Net	74.65±0.78*	67.25±0.88*	97.85±0.29*	29.63
Multi-task U-Net with DS	74.96±0.96	67.69±1.44^{+}	97.93±0.20^{+}	29.63
Multi-task U-Net with DS and fusion layer	75.20±0.49	68.07±0.38	97.99±0.27^{+}	29.63
Multi-task U-Net with EDT-auxiliary-loss	75.43±0.71	68.26±0.80	97.91±0.23	29.63
Our method (w/o EDT-auxiliary-loss)	75.64±0.89*	68.59±1.25*	98.12±0.25*	29.68
Our method (w/o SE block)	75.63±0.71	68.29±0.75*	97.86±0.30*	29.64
Our method (w/o fusion layer)	76.57±0.61	69.49±0.87	98.18±0.28	29.68
Our method (w/o DS and fusion layer)	75.76±0.35	68.22±0.76*	98.10±0.29	29.67
Our method	**76.62±0.89**	**69.58±0.55**	**98.19±0.30**	29.68

*for p-value<0.05 (compared with our method) **for p-value<0.001(compared with our method)

$^{+}$for p-value<0.05 (compared with Multi-task U-Net) $^{++}$for p-value<0.001 (compared with Multi-task U-Net)

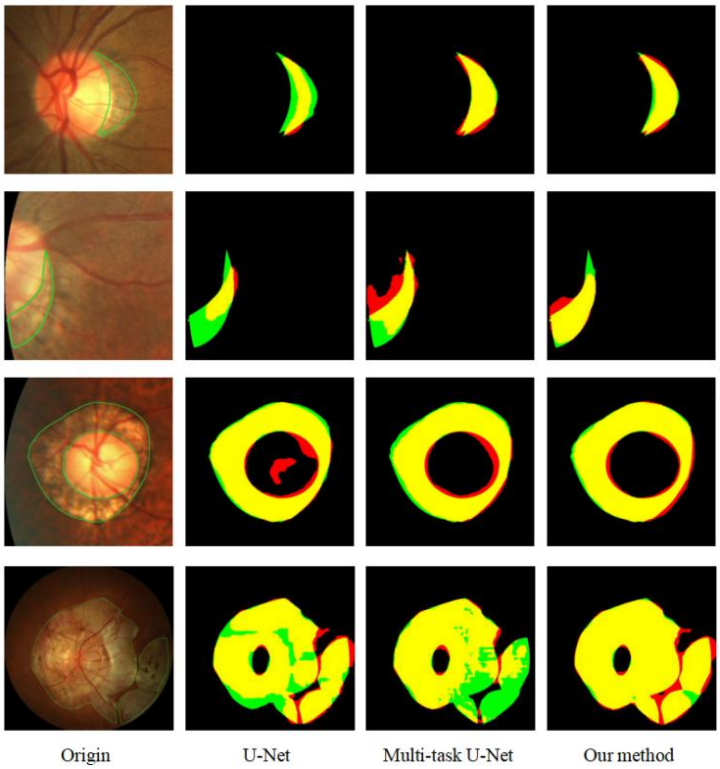

Fig.5 Examples of choroidal atrophy segmentation results. Green lines in the origin images indicate ground truth boundaries. In the segmentation results images of different methods, green, red and yellow indicate FN, FP and TP, respectively.

4. CONCLUSIONS

In this paper, we propose an improved deep network based on multi-task U-Net for choroidal atrophy segmentation in myopia fundus images. The improvements of our proposed deep network include three aspects. First, we improve U-Net to a multi-task segmentation network, which can segment OD and choroidal atrophy simultaneously. To a certain degree, mask information of OD helps choroidal atrophy segmentation. Second, we introduce deep-supervision mechanism and optimize it through fusion layer. Third, a new loss function termed as EDT-auxiliary-loss is proposed to improve the smoothness of segmentation. The experiment results demonstrate that our method is more effective than original U-Net.

5. ACKNOWLEDGEMENTS

This work was supported in part by the National Key R&D Program of China under Grant 2018YFA0701700, and in part by the National Basic Research Program of China (973 Program) under Grant 2014CB748600, the National Natural Science Foundation of China (NSFC) under Grant 61622114, 81401472, and in part by Collaborative Innovation Center of IoT Industrialization and Intelligent Production, Minjiang University under Grant IIC1702.

6. REFERENCE

[1] Ohno-Matsui K, Lai T Y Y, Lai C C, *et al*. "Updates of pathologic myopia," Progress in Retinal and Eye Research, 52: 156-187 (2016).
[2]Elnahry, Ayman & H Talbet, Joseph.. "Posterior Segment Manifestations of Pathological Myopia: A Review". (2019) .
[3] Ohno-Matsui K, Jonas J B, Spaide R F. "Macular Bruch membrane holes in highly myopic patchy choroidal atrophy," American journal of ophthalmology, 166: 22-28 (2016).
[4] Li H, Kang J, Feng Y, et al. "Automatic segmentation of PPA in retinal images,"IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, 1791-1796 (2018).
[5] A. Laude, C. K. Lu, T. B. Tang, A. F. Murray, R. D. Henderson, I. J. Deary, B. Dhillon, "Boundary detection of optic disc and parapapillary atrophy from color fundus images using dual-channel color morphology and snakes," INVEST OPHTH VIS SCI, vol. 51, pp. 1800-1800, (2010).
[6] Septiarini A, Harjoko A, Pulungan R, et al. "Automatic detection of peripapillary atrophy in retinal fundus images using statistical features," Biomedical Signal Processing and Control, 45: 151-159 (2018).
[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in MICCAI, pp. 234-241, Springer, (2015).
[8] Hu J, Shen L, Sun G. "Squeeze-and-excitation networks," Proceedings of the IEEE conference on computer vision and pattern recognition, 7132-7141 (2018).
[9]C. Y. Lee, S. Xie, P. W. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in Proc. Int. Conf. Artif. Intell. Statist. , pp. 562-570 (2015).
[10]J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in CVPR, pp. 3431–3440(2015).
[11] C. Tan, L. Zhao, Z. Yan, K. Li, D. Metaxas, and Y. Zhan, "Deep multi-task and task-specific feature learning network for robust shape preserved organ segmentation," in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 1221-1224 (2018).
[12] Murugesan B, Sarveswaran K, Shankaranarayana S M, et al. "Psi-Net: Shape and boundary aware joint multi-task deep network for medical image segmentation, " arXiv preprint arXiv:1902. 04099, (2019).
[13] F. Milletari, N Navab and S. -A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," arXiv: 1606. 04797 (2016).
[14] Fu H, Cheng J, Xu Y, et al. "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," IEEE Transactions on Medical Imaging, 37(7): 1597-1605. 32 (2018).