GDCSeg-Net: general optic disc and cup segmentation network for multi-device fundus images

QIANLONG ZHU,[1] XINJIAN CHEN,[1,2] QINGQUAN MENG,[1] JIAHUAN SONG,[1] GAOHUI LUO,[1] MENG WANG,[1] FEI SHI,[1] ZHONGYUE CHEN,[1] DEHUI XIANG,[1] LINGJIAO PAN,[3] ZUOYONG LI,[4] AND WEIFANG ZHU[1,*]

[1] *MIPAV Lab, School of Electronics and Information Engineering, Soochow University, Jiangsu 215006, China*
[2] *State Key Laboratory of Radiation Medicine and Protection, Soochow University, Jiangsu 215006, China*
[3] *School of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou, Jiangsu Province, 213000, China*
[4] *College of Computer and Control Engineering, Minjiang University, Fujian 350108, China*
[*] *wfzhu@suda.edu.cn*

Abstract: Accurate segmentation of optic disc (OD) and optic cup (OC) in fundus images is crucial for the analysis of many retinal diseases, such as the screening and diagnosis of glaucoma and atrophy segmentation. Due to domain shift between different datasets caused by different acquisition devices and modes and inadequate training caused by small sample dataset, the existing deep-learning-based OD and OC segmentation networks have poor generalization ability for different fundus image datasets. In this paper, adopting the mixed training strategy based on different datasets for the first time, we propose an encoder-decoder based general OD and OC segmentation network (named as GDCSeg-Net) with the newly designed multi-scale weight-shared attention (MSA) module and densely connected depthwise separable convolution (DSC) module, to effectively overcome these two problems. Experimental results show that our proposed GDCSeg-Net is competitive with other state-of-the-art methods on five different public fundus image datasets, including REFUGE, MESSIDOR, RIM-ONE-R3, Drishti-GS and IDRiD.

1. Introduction

Optic disc (OD) and optic cup (OC), retinal vessel and macula are three most salient features in retinal fundus. The accurate segmentation of OD and OC in fundus images (as shown in Fig. 1) is crucial for the analysis of many retinal diseases, e.g., OD and OC segmentation based optic cup-to-disc ratio is one of the main criteria for clinical screening and diagnosis of glaucoma [1]. Due to the similar color and adjacent location, OD segmentation will greatly influence the atrophy segmentation, especially the peripapillary atrophy segmentation [2].
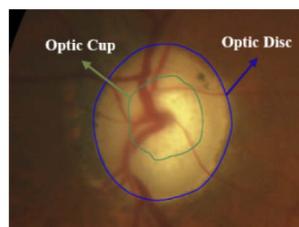
Fig. 1. Optic disc and optic cup in fundus image.

Figure 2 shows typical fundus images from 5 public datasets, including MESSIDOR [3], Drishti-GS [4], IDRiD [5], RIM-ONE-R3 [6] and REFUGE [7]. As can be seen from Fig. 2, the appearance differences between the images from different dataset, called as domain shift [8], are obvious because of the different acquisition devices and modes, which is one of the major reasons for the poor generalization ability of deep-learning-based OD and OC segmentation methods. In addition, the sample numbers of different public datasets are various. For example, REFUGE and MESSIDOR datasets both contain 1200 fundus images, while Drishti-GS, RIM-ONE-R3 and IDRID only contain 101, 159 and 81 ones respectively. Due to the insufficient training, the small sample training based segmentation network usually has bad segmentation performance and generality.

(a) MESSIDOR   (b) Drishti-GS   (c) IDRiD

(d) RIM-ONE-R3   (e) REFUGE-Train   (f) REFUGE-Test/Validation
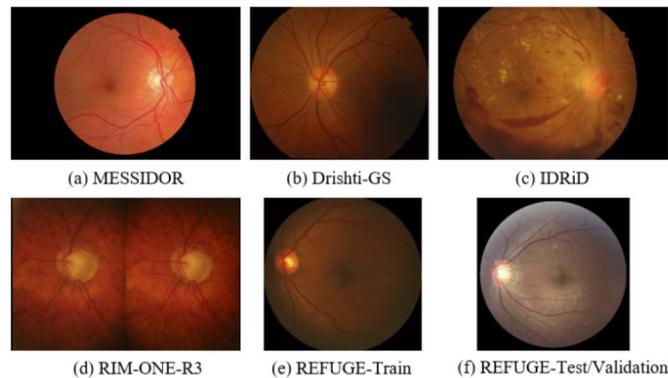
Fig. 2. Examples of different datasets. (a) MESSIDOR, (b) Drishti-GS, (c) IDRiD, (d) RIM-ONE-R3, (e) REFUGE-Train, (f) REFUGE-Test/Validation.

The segmentation of OD and OC has been one of the popular research topics in fundus images analysis for years. The methods can be classified as traditional algorithms and deep-learning-based algorithms. In traditional algorithms, Mittapalli et al. presented a glaucoma expert system based on the segmentation of OD and OC, in which OD was segmented with an implicit region based active contour model and OC was segmented based on its structural and gray level properties [9]. Morales et al. proposed a mathematical morphology and principal component analysis based method for the extraction of OD contour [10]. Aquino et al. used morphology and edge detection techniques and Hough circle transform to approximate a circular OD boundary [11]. Joshi et al. presented an automatic OD parameterization technique based on the segmentation of OD and OC in monocular retinal images [12]. In their following work, they proposed a depth discontinuity based approach to estimate OC boundary [13]. Cheng et al. proposed a super-pixel classification based OD and OC segmentation method for glaucoma screening [14].

In deep-learning-based algorithms, many convolutional neural network (CNN) based methods such as fully convolutional network (FCN) [15], U-Net [16] and its variants and adversarial learning based networks have been proposed for OD and OC segmentation in fundus images. Mohan et al. presented a CNN-based network names as Fine-Net for OD segmentation, in which full-resolution residual networks (FRRN) and atrous convolution were adopted for the efficient feature extraction [17]. In their following work, they introduced a prior CNN called as P-Net, which was cascaded with the Fine-Net, to generate a more accurate OD segmentation map [18]. Jiang et al. presented an end-to-end region-based convolutional neural network for the joint segmentation of OD and OC [19]. Since Ronneberger et al. proposed U-Net for medical image segmentation, there have been many CNN based algorithms for OD and OC segmentation that used U-Net as the baseline network and achieved good performances [20–23]. To learn discriminative representations and produce segmentation probability map, Fu et al. proposed a

multi-scale U-shape convolutional network with the side-output layer named as M-Net for OD and OC segmentation [20]. Gu at al. proposed a context encoder network (CE-Net) for 2D medical image segmentation [21]. Compared with M-Net, CE-Net performed better in OD segmentation. Shah et al. proposed a parameter-shared branched network (PSBN) to learn the optic disc and cup masks and weak region of interest model-based (WRoIM) segmentation network to jointly segment OD and OC [22]. Shankaranarayana et al. proposed a novel depth estimation guided OD and OC segmentation network [23]. In addition, to overcome the problem of domain shift between different fundus image datasets, adversarial learning based methods are introduced. Wang et al. presented a patch-based output space adversarial learning framework (pOSAL) to jointly segment OD and OC, which uses the DeepLabv3+ architecture [8]. In their following work, they presented an unsupervised boundary and entropy-driven adversarial learning (BEAL) framework to improve OD and OC segmentation performance [24]. Recently, graph convolution has also been applied in image segmentation. Tian et al. proposed a segmentation network based on graph convolution for OD and OC segmentation, which achieved good performances on REFUGE and Drishti-GS datasets [25].

As can be seen from Fig. 2, although the appearance differences such as image size, image resolution, color gamut and field of view (FOV) between different fundus images from different public datasets are obvious, the highlighted and circle-like characteristics of OD and OC are common. In this paper, based on the mixed training strategy of different datasets, an encoder-decoder structure based general OD and OC segmentation network is proposed, which can effectively overcome the problems of the appearance differences caused by different acquisition devices and inadequate training caused by small sample dataset. The major contributions of this paper can be included as follows:

-Mixed training strategy is adopted for the first time to overcome the problems of domain shift caused by different acquisition devices and modes and inadequate training caused by small sample dataset.

-An encoder-decoder structure based network with multi-scale information fusion and attention mechanism for general OD and OC segmentation in multi-device fundus images is proposed, named as GDCSeg-Net.

-A novel multi-scale weight-shared attention (MSA) module is proposed and embedded into the top layer of the encoder to integrate the multi-scale OD and OC feature information with channel and spatial attention mechanisms.

-A novel densely connected depthwise separable convolution (DSC) module is proposed and embedded as the output layer of the GDCSeg-Net, which fully fuses the multi-scale features extracted by depthwise separable convolution layer-by-layer via dense connections and leads the network to efficiently focus on the targets.

The remainder of the paper is organized as follows. In Section 2, the proposed method is described in detail. In Section 3, experimental results are shown and analyzed, and followed by the conclusions and discussions in Section 4.

2. Methods

2.1. Overview

Figure 3 shows the overall framework of the proposed OD and OC segmentation method, which mainly includes two parts: region of interest (ROI) extraction and the proposed GDCSeg-Net for OD and OC segmentation.

2.2. ROI extraction network

Motivated by Ref. [26], we use the pre-trained U-Net to segment OD roughly and extract the ROI. After the OD is roughly segmented based on the pre-trained U-Net, the centroid of OD is
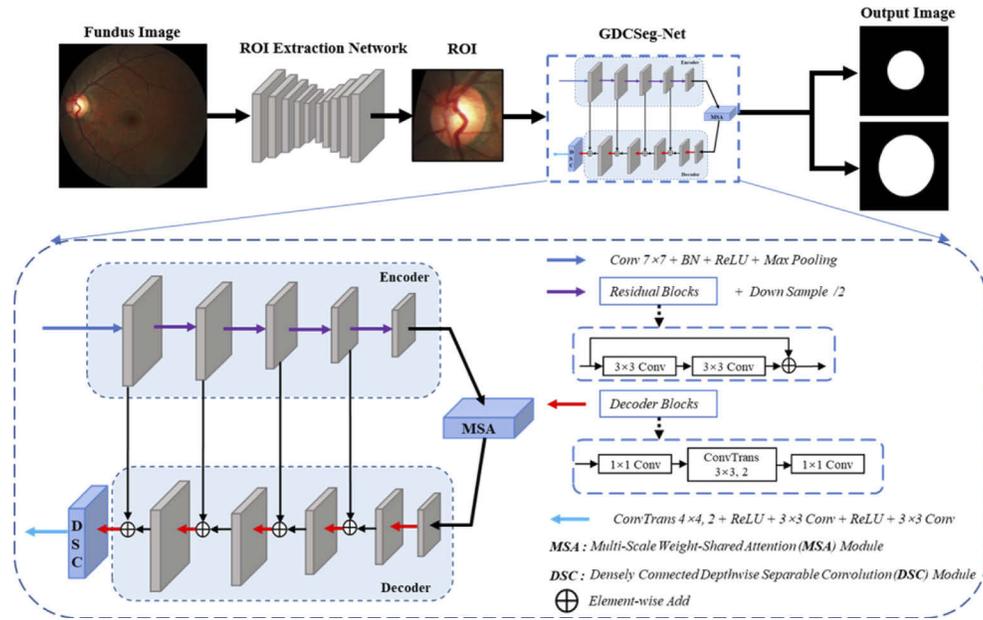
Fig. 3. Illustration of our segmentation framework, which mainly includes region of interest extraction and segmentation network.

located and the ROI with size of 512×512 is cropped around the centroid of OD, which is taken as the input of GDCSeg-Net.

2.3. GDCSeg-Net architecture

As shown in Fig. 3, with the U-shape structure, the proposed GDCSeg-Net mainly includes feature encoder, multi-scale weight-shared attention (MSA) module, densely connected depthwise separable convolution (DSC) module and feature decoder. The basic U-shape encoder-decoder model with pre-trained ResNet34 [27] backbone as feature extractor is taken as our Baseline network.

2.3.1. Multi-scale weight-shared attention (MSA) module

As can be seen from CE-Net, CPFNet [28] and DenseASPP [29], multi-scale feature information can improve the performance of semantic segmentation. However, how to further effectively utilize the multi-scale feature information is still worth studying. As shown in Fig. 4, motivated by the recent multi-scale feature and attention mechanism based approaches [31–34], we propose a novel multi-scale weight-shared attention (MSA) module, which includes the depthwise separable convolution based multi-scale feature extractor and channel and spatial attention modules, to obtain OD and OC feature information effectively.

In the multi-scale feature extractor, we use four parallel depthwise separable convolutions with different dilation rates of 1, 3, 5 and 7 to capture multi-scale information. To reduce the model parameters and the risk of overfitting, these four depthwise separable convolutions have shared the weights. The output of multi-scale feature $F_D \in \mathbb{R}^{C \times H \times W}$ can be computed as:

$$F_D = Concat_{i=0}^{i=3}(D^{2*i+1}(F)) \tag{1}$$

where $F \in \mathbb{R}^{C \times H \times W}$ denotes the input feature map, C is the channel of feature map, H is the height of feature map and W is the width of feature map. *Concat* represents the concatenation operation, D^{2*i+1} represents the depthwise separable convolution with dilation rate of $2*i+1$.
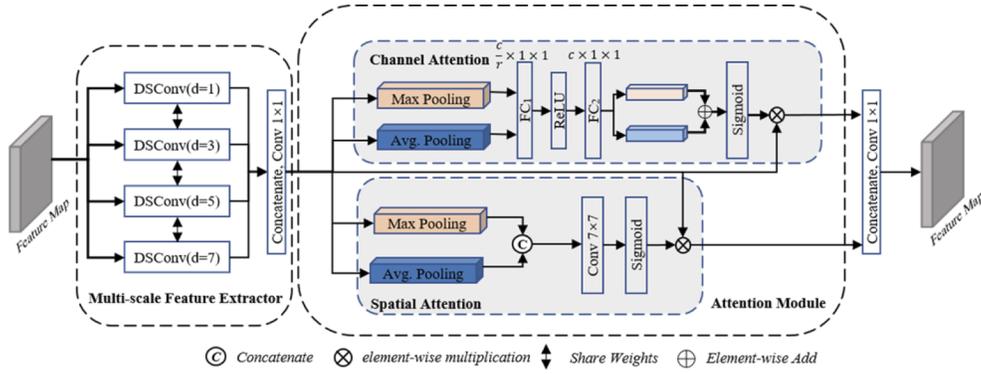
Fig. 4. Illustration of multi-scale weight-shared attention (MSA) module.

In the attention module, channel and spatial attention mechanisms are applied for the feature refinement. In channel attention module, the max-pooled and average-pooled features go through two fully connected layers, ReLU and sigmoid activation function to produce the channel attention map ($F_C' \in \mathbb{R}^{C \times 1 \times 1}$). The output of channel attention $F_C \in \mathbb{R}^{C \times H \times W}$ can be computed as:

$$F_C = F_C' \otimes F_D \tag{2}$$

$$F_C' = Sig(f_2(ReLU(f_1(Avg(F_D)))) + f_2(ReLU(f_1(Max(F_D))))) \tag{3}$$

where Sig denotes the sigmoid function, and $\otimes$ denotes element-wise multiplication. f_1 represents the first fully connected layer (FC$_1$) that compresses C channels into C/r ones, where r is the reduction ratio and is set to 16 in this paper. f_2 represents the second fully connected layer (FC$_2$) that restores channel to C channels.

In spatial attention module, a spatial attention map is produced according to the spatial relationship between features. Similar to the channel attention, the max-pooling and average-pooling operations are applied to generate the max-pooled and average-pooled features respectively. These two feature maps are concatenated and sent to a standard 7×7 convolution to generate a spatial attention map $F_S' \in \mathbb{R}^{1 \times H \times W}$. The output of spatial attention module $F_S \in \mathbb{R}^{C \times H \times W}$ can be computed as:

$$F_S = F_S' \otimes F_D \tag{4}$$

$$F_S' = Sig(f^{7 \times 7}(Concat(Avg(F_D); Max(F_D))) \tag{5}$$

where $Concat$ represents the concatenation operation, Sig denotes the sigmoid function, and $f^{7 \times 7}$ represents a 7×7 convolution operation.

The overall MSA module can be summarized as:

$$MSA(F) = Concat(F_C; F_S) \tag{6}$$

2.3.2. Densely connected depthwise separable convolution (DSC) module

Generally, in the output layer of most U-shape networks, simple bilinear interpolation based upsampling is adopted to output the final segmentation results [16,21,28]. As the feature maps shown in Fig. 5 (c), this simple upsampling method pays less attention to the target. In order to result in better response to the segmentation target, a densely connected depthwise separable convolution (DSC) module is presented and embedded as the output layer of the GDCSeg-Net, which is shown in Fig. 6. In the DSC module, considering the size of the input feature map, four depthwise separable convolutions with different dilation rates (1, 6, 12 and 18) are adopted to

capture different scale information. Through the dense connections, the multi-scale features can be fully fused layer-by-layer. As can be seen from Fig. 5 (d), DSC module focuses on the target features precisely, which will improve the segmentation performance.
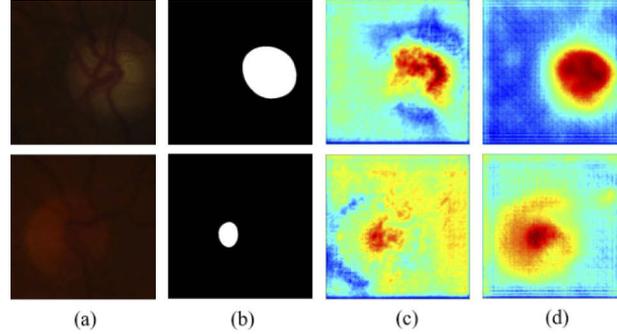
Fig. 5. Comparison of feature maps transferred by the output layer before and after insertion of DSC module. (a) Original image, (b) ground truth, (c) the feature map before inserting DSC module and (d) the feature map after inserting DSC module.
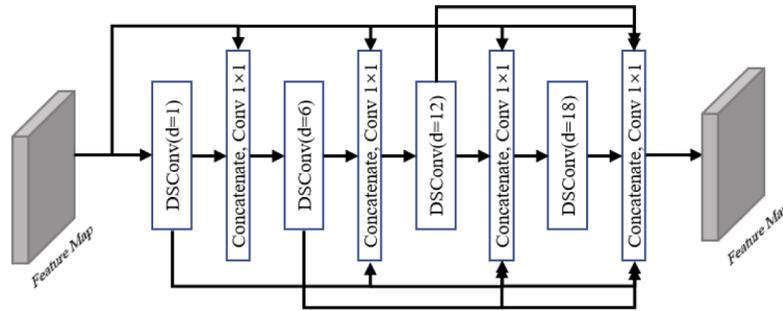
Fig. 6. Illustration of densely connected depthwise separable convolution (DSC) module.

2.4. Loss function

To effectively solve the data imbalance problem in the training process, the combination of Dice loss and binary cross-entropy (BCE) loss is adopted as the total loss function, which can be defined as follows:

$$L_{Total} = L_{Dice} + L_{BCE} \tag{7}$$

$$L_{Dice} = 1 - \frac{2 \sum_{i}^{N} \bar{y}_i y_i + \varepsilon}{\sum_{i}^{N} \bar{y}_i + \sum_{i}^{N} y_i + \varepsilon} \tag{8}$$

$$L_{BCE} = -\frac{1}{N} \sum_{i}^{N} \left(\begin{array}{c} y_i \log \bar{y}_i \\ +(1 - y_i) \log(1 - \bar{y}_i) \end{array} \right) \tag{9}$$

where N indicates the batch size, $\bar{y}_i \in [0, 1]$ and $y_i \in [0, 1]$ denote the predicted probability and ground truth label respectively. ε is a small smoothing factor.

3.   Experiments and results

3.1.   Dataset

We carry out extensive validations for the proposed GDCSeg-Net on publicly available datasets including REFUGE, MESSIDOR, RIM-ONE-R3, Drishti-GS and IDRiD. A summary for each dataset including number of images, image resolution and availability of the OD and OC ground truth and the data division strategy for OD and OC segmentation experiments are shown in Table 1. Due to the high resolution of fundus image and the small target property of OD and OC, a 512×512 region of interest (ROI) is cropped by the ROI extraction network and taken as the input of the proposed GDCSeg-Net.

Table 1. An overview of the datasets and data division strategy.[a]

Dataset	Number of Images	Resolution	OD				OC			
			GT	Tra	Val	Tes	GT	Tra	Val	Tes
REFUGE	1200	2124×2056, 1634×1634	√	600	200	400	√	500	190	510
MESSIDOR	1200	2240×1488	√	600	200	400	×	-	-	-
RIM-ONE-R3	159	2144×1424	√	60	40	59	√	60	40	59
Drishti-GS	101	2047×2056	√	40	10	51	√	40	10	51
IDRiD	81	4288×2848	√	40	10	31	×	-	-	-
Total	2741	-	-	1340	460	941	-	600	240	620

[a](GT, Tra, Val and Tes represent ground truth, Training data, Validation data and test data, respectively)

3.2.   Implementation details

3.2.1.   Experiment setting

The implementation of the proposed GDCSeg-Net is based on the public platform PyTorch and a NVIDIA GTX3090 GPU with 24GB memory. In the training process, we use the 'poly' learning rate policy, where $lr = baselr \times \left(1 - \frac{iter}{total\_iter}\right)^{power}$, the basic learning rate $baselr$ is set to 0.01, $iter$ denotes the number of iterations, $total\_iter$ denotes the total number of iterations, and $power$ is set to 0.9. The stochastic gradient descent (SGD) algorithm with an initial learning rate of 0.01, momentum of 0.9 and weight decay of 0.0001 is used to optimize the network. Besides, the batch size is set to 4 and the number of epochs is 80. We have released our codes on Github [30].

3.2.2.   Data augmentation strategies

To increase the generalization of the model and reduce the risk of overfitting, we adopt online data augmentation strategies including left and right flipping, up and down flipping, random rotation (range from -30° to 30°) and additive Gaussian noise addition. For each round of training, 2-5 of these augmentation methods are used.

3.2.3.   Evaluation metrics

To quantitatively evaluate the segmentation performance, two common segmentation evaluation metrics including Dice coefficient (Dice) and intersection over union (IoU) are used, which are defined as follows:

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{10}$$

$$IoU = \frac{Area(Seg \cap GT)}{Area(Seg \cup GT)} \tag{11}$$

where *TP* denotes true positive, *FP* denotes false positive and *FN* denotes false negative. *Seg* and *GT* denote segmented mask and ground truth, respectively.

T-test with α=0.05 is adopted to evaluate the statistical differences between different methods.

3.3.   Optic disc segmentation

For OD segmentation, five datasets including REFUGE, MESSIDOR, RIM-ONE-R3, Drishti-GS and IDRiD are used. According to the mixed training strategy, a total of 2541 fundus images from five datasets are randomly splitted into training set (1340), validation set (460) and test set (741). The details about the data division for OD segmentation are listed in Table 1. Comparison experiments and ablation experiments are performed to verify the superiority of the proposed GDCSeg-Net compared with other state-of-the-art methods and the effectiveness of the proposed MSA and DSC modules, respectively.

(1) *Comparison experiments*

With the same data split strategy, we compare our method with other excellent CNN based methods, including FCN [15], U-Net [16], CE-Net [21], CPFNet [28], Attention U-Net [35], U-Net++ [36], Deep ResU-Net [37], ResU-Net++ [38], CS^{2}Net [39] and SegNet [40]. Table 2 presents the comparison results on REFUGE, MESSIDOR, Drishti-GS, RIM-ONE-R3 and IDRiD. As we can see, the proposed GDCSeg-Net outperforms the mentioned state-of-the-art CNN based methods. As can be seen from Table 2, all of the 11 networks perform well on REFUGE dataset. The major reason is that the contrast between OD and the background is obvious, which can be seen from Fig. 2 (e) and (f) and the first row of Fig. 7, making the OD segmentation relatively easy. On MESSIDOR dataset, our method achieves 0.9435 and 0.9700 in IoU and Dice respectively, better than other methods with statistical significance except the Dice index compared with CPFNet (p=0.064). On Drishti-GS dataset, although the proposed GDCSeg-Net performs slightly worse than CE-Net without statistical significance (the p-values for Dice and IoU are 0.387 and 0.396 respectively), the overall results still show that our GDCSeg-Net outperforms CPFNet and other methods. As can be seen from Fig. 2 (d) and the second row of Fig. 7, the contrast between OD and the background is low in the images of RIM-ONE-R3 dataset, which increases the difficulty of OD segmentation. So the performances of Deep ResU-Net, U-Net++, U-Net, Attention U-Net, CS^{2}Net, SegNet, FCN and ResU-Net++ degrade significantly on RIM-ONE-R3 dataset. The proposed GDCSeg-Net significantly outperforms the other methods except CPFNet (p=0.111 and p=0.103 for IoU and Dice respectively). IDRiD dataset consists of 81 images with four types of retinal lesions, includes hemorrhage (HE), microaneurysm (MA), hard exudate (EX) and soft exudate (SE). These lesions affect the OD segmentation to some extent. Besides, the dataset also suffers from low contrast between OD and the background. Therefore, although the proposed GDCSeg-Net outperforms the other methods, the improvement of IoU and Dice indexes are not statistically significant compared with FCN, U-Net, Attention U-Net and CE-Net.

(2) *Ablation experiments*

In order to verify the validity of the proposed MSA module and DSC module, four ablation experiments are conducted. The results of each test are shown in Table 3, in which "Baseline" represents the U-shape encoder-decoder model with pre-trained ResNet34 backbone.

As shown in Table 3, the embedding of the proposed DSC module (Baseline + DSC) achieves substantial improvement over the Baseline in Dice and IoU metrics, especially on the Drishti-GS dataset. Meanwhile, the embedding of MSA module (Baseline + MSA) also helps to improve the performance. For example, compared with Baseline, the Dice index increases 0.94% and reaches 0.9532 on the RIM-ONE-R3 dataset. As shown in Fig. 7, the proposed method obtains more accurate segmentation results than Baseline, especially on the IDRiD, Drishti-GS and
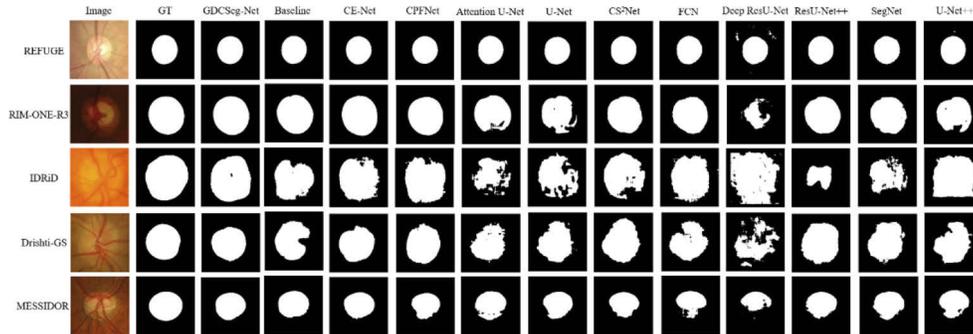
Fig. 7. Examples of optic disc segmentation. From left to right: original image, ground truth (GT), the proposed GDCSeg-Net, Baseline, CE-Net, CPFNet, Attention U-Net, U-Net, CS^{2}Net, FCN, Deep ResU-Net, ResU-Net++, SegNet and U-Net++.

Table 2. The results of comparison experiments for OD segmentation on the test set (p represents p-value).

Methods	REFUGE		MESSIDOR		Drishti-GS		RIM-ONE-R3		IDRiD	
	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
	p	p	p	p	p	p	p	p	p	p
Deep ResU-Net	0.8913	0.9408	0.8331	0.8935	0.7921	0.8634	0.6076	0.7201	0.8657	0.9263
	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SegNet	0.9179	0.9568	0.9159	0.9533	0.9203	0.9577	0.8238	0.8959	0.9028	0.9469
	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.009	0.014
CS^{2}Net	0.9216	0.9588	0.9182	0.9533	0.9346	0.9657	0.8189	0.8917	0.9192	0.9573
	<0.001	<0.001	<0.001	<0.001	0.001	0.001	<0.001	<0.001	0.013	0.013
FCN	0.9219	0.9589	0.9291	0.9620	0.9336	0.9653	0.8405	0.9051	0.9288	0.9628
	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.276	0.272
U-Net	0.9223	0.9591	0.9145	0.9504	0.9190	0.9559	0.7486	0.8222	0.9192	0.9570
	<0.001	<0.001	<0.001	<0.001	0.001	0.002	<0.001	<0.001	0.05	0.055
U-Net++	0.9228	0.9594	0.9114	0.9485	0.9326	0.9645	0.7414	0.8187	0.9209	0.9585
	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	<0.001	<0.001	0.019	0.02
ResU-Net++	0.9250	0.9605	0.9250	0.9596	0.9300	0.9647	0.8617	0.9223	0.8848	0.9298
	<0.001	<0.001	<0.001	<0.001	0.002	0.002	<0.001	<0.001	0.03	0.04
Attention U-Net	0.9258	0.9611	0.9201	0.9542	0.9283	0.9610	0.7709	0.8416	0.9115	0.9509
	<0.001	<0.001	<0.001	<0.001	0.012	0.021	<0.001	<0.001	0.071	0.078
CPFNet	0.9305	0.9637	0.9419	0.9692	0.9446	0.9714	0.9110	0.9523	0.9256	0.9611
	0.163	0.191	0.041	0.064	0.005	0.005	0.111	0.103	0.026	0.028
CE-Net	0.9310	0.9640	0.9387	0.9673	**0.9506**	**0.9746**	0.9046	0.9488	0.9271	0.9620
	0.301	0.318	<0.001	<0.001	**0.387**	**0.396**	0.008	0.008	0.085	0.089
GDCSeg-Net	**0.9314**	**0.9642**	**0.9435**	**0.9700**	0.9501	0.9743	**0.9167**	**0.9560**	**0.9313**	**0.9642**

Table 3. The results of ablation experiments for OD segmentation on the test set.

Methods	REFUGE		MESSIDOR		Drishti-GS		RIM-ONE-R3		IDRiD	
	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
Baseline	0.9270	0.9618	0.9393	0.9675	0.9424	0.9701	0.8976	0.9443	0.93040	0.9635
Baseline + DSC	0.9276	0.9620	0.9394	0.9677	0.9472	0.9727	0.9018	0.9470	0.9310	0.9640
Baseline + MSA	**0.9317**	**0.9644**	0.9398	0.9679	0.9477	0.9730	0.9124	0.9532	0.9299	0.9634
Proposed	0.9314	0.9642	**0.9435**	**0.9700**	**0.9501**	**0.9743**	**0.9167**	**0.9560**	**0.9313**	**0.9642**

MESSIDOR datasets. With the addition of both DSC and MSA modules, the Dice of the proposed GDCSeg-Net reaches 0.9642, 0.9700, 0.9743, 0.9560 and 0.9642 on REFUGE, MESSIDOR, Drishti-GS, RIM-ONE-R3 and IDRiD respectively, which are significantly better than the ones of the Baseline. The results of ablation experiments show that the proposed DSC and MSA modules are beneficial for OD segmentation.

3.4. Optic cup segmentation

For OC segmentation, we use three datasets including REFUGE, RIM-ONE-R3 and Drishti-GS. According to the mixed training strategy, a total of 1460 fundus images from three datasets are randomly divided into training set (600), validation set (240) and test set (620).The details about data division are listed in Table 1. Similar to OD segmentation, comparison experiments and ablation experiments are also performed and analyzed.

(1) *Comparison experiments*

With the same data split strategy, we compare our method with other excellent CNN based methods, including FCN, U-Net, CE-Net, CPFNet, Attention U-Net, U-Net++, Deep ResU-Net, ResU-Net++, CS^{2}Net and SegNet. Table 4 presents the segmentation results on the REFUGE, RIM-ONE-R3 and Drishti-GS.

Compared with OD segmentation, OC segmentation is more difficult due to the blurred boundary between OC and OD as well as the smaller OC region. As can be seen from Table 4, on REFUGE dataset, the proposed GDCSeg-Net significantly outperforms the other methods except CPFNet (p=0.206 and p=0.22 for IoU and Dice respectively). On RIM-ONE-R3 dataset, CPFNet achieves the best performance in Dice, while the proposed GDCSeg-Net achieves the best performance in IoU. Through T-Test analysis, there is no significant differences between the proposed GDCSeg-Net, CPFNet and CE-Net on the RIM-ONE-R3 dataset. On Drishti-GS dataset, the proposed GDCSeg-Net significantly outperforms the other methods.

Figure 8 shows five OC segmentation results of different methods, which reveal that the proposed method obtains more accurate segmentation results, especially on the RIM-ONE-R3 and Drishti-GS datasets.

(2) *Ablation experiments*

In order to verify the validity of the proposed MSA and DSC module, we also conduct four ablation experiments on OC segmentation. As shown in Table 5, compared with the Baseline, the IoU and Dice of OC segmentation increase obviously with the addition of DSC and MSA modules. Especially, the IoU and Dice increase from 0.6919 and 0.7975 to 0.7237 and 0.8237 on images from RIM-ONE-R3 dataset, in which the boundaries between OC and OD are very blurred. The results show that the proposed DSC and MSA modules are beneficial for OC segmentation as well.

Table 4. **The results of comparison experiments for OC segmentation on the test set (p represents p-value).**

Methods	REFUGE		RIM-ONE-R3		Drishti-GS	
	IoU	Dice	IoU	Dice	IoU	Dice
	p	p	p	p	p	p
Deep ResU-Net	0.7396	0.8433	0.4466	0.5655	0.6639	0.7791
	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SegNet	0.7833	0.8748	0.6183	0.7289	0.7436	0.8424
	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
CS^{2}Net	0.7844	0.8748	0.5961	0.7215	0.7249	0.8250
	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
FCN	0.7784	0.8706	0.6516	0.7648	0.7447	0.8365
	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
U-Net	0.7869	0.8769	0.5609	0.6759	0.7416	0.8378
	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
U-Net++	0.7858	0.8754	0.5363	0.6538	0.7379	0.8347
	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
ResU-Net++	0.7882	0.8780	0.6201	0.7426	0.7477	0.8456
	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Attention U-Net	0.7941	0.8814	0.5943	0.7024	0.7577	0.8493
	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
CPFNet	0.8110	0.8929	0.7217	**0.8262**	0.8168	0.8890
	0.206	0.22	0.408	**0.383**	0.004	0.009
CE-Net	0.8035	0.8881	0.7134	0.8195	0.8213	0.8922
	<0.001	<0.001	0.23	0.381	0.008	0.014
GDCSeg-Net	**0.8127**	**0.8938**	**0.7237**	0.8237	**0.8344**	**0.9003**
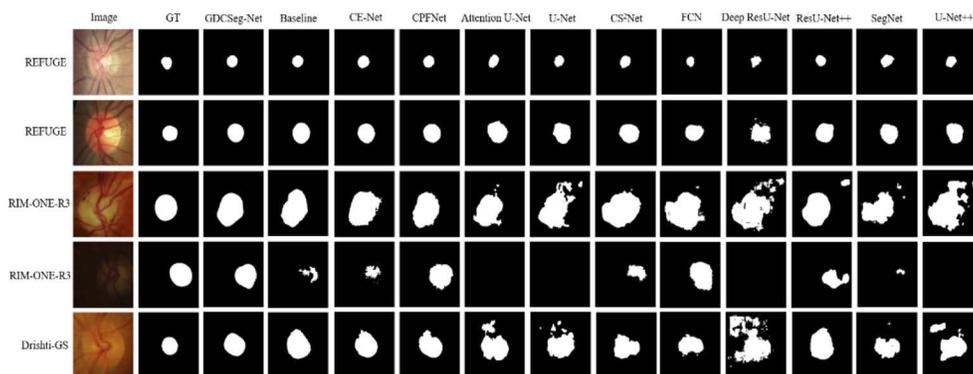
Fig. 8. Examples of optic cup segmentation. From left to right: original image, ground truth (GT), the proposed GDCSeg-Net, Baseline, CE-Net, CPFNet, Attention U-Net, U-Net, CS^{2}Net, FCN, Deep ResU-Net, ResU-Net++, SegNet and U-Net++.

Table 5. The results of ablation experiments for OC segmentation on the test set.

Methods	REFUGE		Drishti-GS		RIM-ONE-R3	
	IoU	Dice	IoU	Dice	IoU	Dice
Baseline	0.8044	0.8886	0.8206	0.8912	0.6919	0.7975
Baseline + DSC	0.8080	0.8908	0.8187	0.8905	0.7101	0.8111
Baseline + MSA	0.8047	0.8889	0.8304	0.8979	0.7202	0.8237
Proposed	**0.8127**	**0.8938**	**0.8344**	**0.9003**	**0.7237**	**0.8237**

3.5. Generalization experiments

In order to verify the effectiveness of mixed training strategy for small sample datasets such as Drishti-GS, RIM-ONE-R3, IDRiD and an in-house dataset (144 fundus images with myopic from the First People's Hospital Affiliated to Shanghai Jiao Tong University, in which the OD and OC ground truth were annotated under the supervision of an experienced ophthalmologist.), we compare the results of single dataset training based GDCSeg-Net (named as "Single training" in Table 6 and Table 7) with the results of mixed training based GDCSeg-Net. For the single dataset training strategy on OD and OC segmentation, the divisions of training, validation and test set are as follows: (1) In Drishti-GS, 101 images are divided into 40 for training, 10 for validation and 51 for test. (2) In RIM-ONE-R3, 159 images are divided into 60 for training, 40 for validation and 59 for test. (3) In IDRiD, 81 images are divided into 40 for training, 10 for validation and 31 for test. (4) In in-house dataset, 144 images are divided into 80 for training, 20 for validation and 44 images for test. As shown in Table 6 and Table 7, the results show that the mixed training strategy has a significant improvement in the OD and OC segmentation of small sample datasets. In particular, the IoU index improves from 0.9302 to 0.9501 for OD segmentation and from 0.7727 to 0.8344 for OC segmentation on Drishti-GS dataset.

Table 6. The results of generalization experiments for OD segmentation on the test set.

Methods	Drishti-GS		RIM-ONE-R3		IDRiD		In-house dataset	
	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
Single Training	0.9302	0.9627	0.9155	0.9548	0.9147	0.9550	0.9131	0.9542
Proposed	**0.9501**	**0.9743**	**0.9167**	**0.9560**	**0.9313**	**0.9642**	**0.9244**	**0.9605**

Table 7. The results of generalization experiments for OC segmentation on the test set.

Methods	Drishti-GS		RIM-ONE-R3		In-house dataset	
	IoU	Dice	IoU	Dice	IoU	Dice
Single Training	0.7727	0.8599	0.7217	**0.8277**	0.6864	0.8055
Proposed	**0.8344**	**0.9003**	**0.7237**	0.8237	**0.7075**	**0.8212**

3.6. Comparison of the state-of-the-art OD and OC segmentation methods

To further prove the effectiveness of the proposed method, we compare the performance of the proposed method with the state-of-the-art OD and OC segmentation methods. As shown in Table 8, the results indicate that our proposed GDCSeg-Net is competitive with other state-of-the-art methods on five different public fundus image datasets. Among them, Tian et al. achieved the best performances on REFUGE and Drishti-GS datasets, especially in OC segmentation. The possible reasons are as follows: first, the cropped images input for OC segmentation is only 70% of those for OD segmentation in the training, which greatly reduces the interference of OD and

background. Second, graph convolution can be used to predict the object contour with obvious boundary. As can be seen from Fig. 2 (b), (e) and (f), the boundaries of OC and OD in images from REFUGE and Drishti-GS datasets are relatively obvious. Shankaranarayana et al. achieved the best OD and OC segmentation performances on RIM-ONE-R3. The possible reason is that they used the initial weights obtained from ORIGA dataset (650 images, not available now) [41] to continuously train their network for RIM-ONE-R3 dataset. The proposed GDCSeg-Net achieves best OD segmentation performances on MESSIDOR, Drishti-GS and IDRiD datasets.

Table 8. Performance comparison of the proposed method for OD and OC segmentation with the state-of-the-art methods.

Methods	Datasets	OD		OC	
		IoU	Dice	IoU	Dice
Wang et al.(2019a) [8]	RIM-ONE-R3	-	0.8650	-	0.7870
	REFUGE	-	0.9602	-	0.8826
	Drishti-GS	-	0.9650	-	0.8580
Wang et al.(2019b) [24]	RIM-ONE-R3	-	0.898	-	0.810
	Drishti-GS	-	0.961	-	0.862
Mohan et al. (2018) [17]	MESSIDOR	0.920	0.957	-	-
	Drishti-GS	0.931	0.964	-	-
Mohan et al. (2019) [18]	MESSIDOR	0.939	0.968	-	-
	Drishti-GS	0.947	0.9713	-	-
Shah et al. (2019) (PSBN) [22]	RIM-ONE-R3	0.84	0.91	0.60	0.75
	Drishti-GS	0.91	0.95	0.80	0.88
Shah et al. (2019) (WRoIM) [22]	RIM-ONE-R3	0.90	0.94	0.71	0.82
	Drishti-GS	0.93	0.96	0.80	0.89
Shankaranarayana et al. (2019) [23]	RIM-ONE-R3	**0.942**	**0.970**	0.716	**0.872**
	Drishti-GS	0.929	0.963	0.760	0.848
Tian et al. (2020) [25]	REFUGE	**0.9564**	**0.9776**	**0.9160**	**0.9558**
	Drishti-GS	-	**0.978**	-	**0.957**
Dey et al. (2019) [42]	RIM-ONE-R3	0.880	0.933	-	-
	Drishti-GS	0.921	0.958	-	-
	IDRiD	0.896	0.943	-	-
	MESSIDOR	0.857	0.912	-	-
Yu et al. (2019) [43]	RIM-ONE-R3	0.9256	0.9610	**0.7429**	0.8445
	Drishti-GS	0.9492	0.9738	0.8042	0.8877
Proposed	RIM-ONE-R3	0.9167	0.9560	0.7237	0.8237
	REFUGE	0.9314	0.9642	0.8127	0.8938
	Drishti-GS	**0.9501**	0.9743	**0.8344**	0.9003
	MESSIDOR	**0.9435**	**0.9700**	**-**	**-**
	IDRiD	**0.9313**	**0.9642**	-	-

4.   Conclusion and discussions

The OD and OC segmentation in fundus image is an important basis for the analysis of glaucoma. In this paper, we adopt the mixed training strategy of different datasets for the first time. Based on the U-shape encoder-decoder structure, a general domain-adaptive OD and OC segmentation network (GDCSeg-Net) is proposed, which effectively overcomes the problems of domain shift

caused by different acquisition devices and modes and inadequate training caused by small sample dataset. The proposed MSA module is embedded into the top layer of the encoder to integrate the multi-scale OD and OC feature information with channel and spatial attention mechanisms. The DSC module is proposed and embedded as the output layer of the GDCSeg-Net, which fully fuses the multi-scale features extracted by depthwise separable convolution layer-by-layer via dense connections and leads the network to efficiently focus on the targets. The proposed MSA and DSC modules are effective and universal, which can be easily introduced into other encoder-decoder networks.

The comparison experimental results show that the proposed GDCSeg-Net achieves the best OD and OC segmentation performance on five fundus image datasets, including REFUGE, MESSIDOR, RIM-ONE-R3, Drishti-GS and IDRiD datasets. Although CE-Net has achieved comparable performance with the proposed GDCSeg-Net in OD segmentation on REFUGE, IDRiD and Drishti-GS datasets, it is unable to perform well in OC segmentation on REFUGE, RIM-ONE-R3 and Drishti-GS datasets. Similarly, CPFNet performs comparably to the proposed GDCSeg-Net in OD segmentation on REFUGE and MESSIDOR, it does not perform well in OD and OC segmentation on Drishti-GS dataset. These results suggest that the proposed GDCSeg-Net is more general and effective than the state-of-the-art segmentation networks in OD and OC segmentation task.

In addition, compared with state-of-the-art OD and OC segmentation methods, our method has achieved competitive performance in OD and OC segmentation on five fundus image datasets. As one of our future focuses, we will try to improve the performance of OC segmentation by integrating the newly proposed self-attention-based transformer structure [44,45] into our proposed GDCSeg-Net, which may focus on the blurred boundary between OD and OC. To validate the generality of the proposed GDCSeg-Net, other segmentation tasks such as diabetic retinopathy related fundus lesions and retinal vessel segmentation in fundus images with the proposed GDCSeg-Net will be explored in the future.

Disclosures. The authors declare that there are no conflicts of interest related to this article.

Data availability. REFUGE, MESSIDOR, RIM-ONE-R3, Drishti-GS and IDRiD datasets underlying the results presented in this paper are available at [3–7]. The in-house dataset underlying the results presented in this paper is not publicly available at this time due to privacy reasons.

References

1. J. B. Jonas, A. Bergua, P. S. Valckenberg, K. I. Papastathopoulos, and W. M. Budde, "Ranking of optic disc variables for detection of glaucomatous optic nerve damage," Invest. Ophthalmol. Vis. Sci. 41(7), 1764–1773 (2000).
2. R. Lu, W. Zhu, X. Cheng, X. Chen, and I. Kopriva, "Choroidal atrophy segmentation based on deep network with deep-supervision and EDT-auxiliary-loss," Proc. SPIE **11313**, 113131X (2020).
3. E. Decencire, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, and J. Klein, "Feedback on a publicly distributed database: the messidor database," Image Anal. Stereol. **33**(3), 231–234 (2014).
4. J. Sivaswamy, S. R. Krishnadas, G. Datt Joshi, M. Jain, and A. U. Syed Tabish, "Drishti-GS: Retinal image dataset for optic nerve head (ONH) segmentation," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)* (IEEE, 2014), pp.53–56.
5. P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research," Data **3**(3), 25 (2018).
6. F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "Rim-one: an open retinal image database for optic nerve evaluation," in *2011 24th International Symposium on Computer-Based Medical Systems (CBMS)* (IEEE, 2011), pp.1–6.
7. J. I. Orlando, H. Fu, J. B. Breda, K. van Keer, D. R. Bathula, A. Diaz-Pinto, R. Fang, P.-A. Heng, J. Kim, and J. Lee, "REFUGE Challenge: a unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," Med. Image Anal. **59**, 101570 (2020).

8. S. Wang, L. Yu, X. Yang, C. -W. Fu, and P. -A. Heng, "Patch-based output space adversarial learning for joint optic disc and cup segmentation," IEEE Trans. Med. Imaging **38**(11), 2485–2495 (2019).
9. P. S. Mittapalli and G. B. Kande, "Segmentation of optic disk and optic cup from digital fundus images for the assessment of glaucoma," Biomed. Signal Process. Control **24**(47), 34–46 (2016).
10. S. Morales, V. Naranjo, J. Angulo, and M. Alcañiz, "Automatic detection of optic disc based on PCA and mathematical morphology," IEEE Trans. Med. Imaging **32**(4), 786–796 (2013).
11. A. Aquino, M. E. Gegúndez-Arias, and D. Marín, "Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques," IEEE Trans. Med. Imaging **29**(11), 1860–1869 (2010).
12. G. D. Joshi, J. Sivaswamy, and S. R. Krishnadas, "Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment," IEEE Trans. Med. Imaging **30**(6), 1192–1205 (2011).
13. G. D. Joshi, J. Sivaswamy, and S. R. Krishnadas, "Depth discontinuity-based cup segmentation from multiview color retinal images," IEEE Trans. Biomed. Eng. **59**(6), 1523–1531 (2012).
14. J. Cheng, J. Liu, Y. Xu, F. Yin, D. W. K. Wong, N. Tan, D. Tao, C. Cheng, T. Aung, and T. Wong, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," IEEE Trans. Med. Imaging **32**(6), 1019–1032 (2013).
15. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2015), pp. 3431–3440.
16. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015), pp. 234–241.
17. D. Mohan, J. R. Harish Kumar, and C. Sekhar Seelamantula, "High-performance optic disc segmentation using convolutional neural networks," in *25th IEEE International Conference on Image Processing (ICIP)* (IEEE, 2018), pp. 4038–4042.
18. D. Mohan, J. R. Harish Kumar, and C. Sekhar Seelamantula, "Optic disc segmentation using cascaded multiresolution convolutional neural networks," in *2019 IEEE International Conference on Image Processing (ICIP)* (IEEE, 2019), pp. 834–838.
19. Y. Jiang, L. Duan, J. Cheng, Z. Gu, H. Xia, H. Fu, C. Li, and J. Liu, "JointRCNN: a region-based convolutional neural network for optic disc and cup segmentation," IEEE Trans. Biomed. Eng. **67**(2), 335–343 (2020).
20. H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," IEEE Trans. Med. Imaging **37**(7), 1597–1605 (2018).
21. Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: context encoder network for 2D medical image segmentation," IEEE Trans. Med. Imaging **38**(10), 2281–2292 (2019).
22. S. Shah, N. Kasukurthi, and H. Pande, "Dynamic region proposal networks for semantic segmentation in automated glaucoma screening," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI)* (IEEE, 2019), pp. 578–582.
23. S. M. Shankaranarayana, K. Ram, K. Mitra, and M. Sivaprakasam, "Fully convolutional networks for monocular retinal depth estimation and optic disc-cup segmentation," IEEE J. Biomed. Health Inform. **23**(4), 1417–1426 (2019).
24. S. Wang, L. Yu, K. Li, X. Yang, C. Fu, and P. Heng, "Boundary and entropy-driven adversarial learning for fundus image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2019), pp. 102–110.
25. Z. Tian, Y. Zheng, X. Li, S. Du, and X. Xu, "Graph convolutional network based optic disc and cup segmentation on fundus images," Biomed. Opt. Express **11**(6), 3043–3057 (2020).
26. H. Fu, J. Cheng, Y. Wu, C. Zhang, D. W. K. Wong, J. Liu, and X. Cao, "Disc-aware ensemble network for glaucoma screening from fundus image," IEEE Trans. Med. Imaging **37**(11), 2493–2501 (2018).
27. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 770–778.
28. S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, and X. Chen, "CPFNet: context pyramid fusion network for medical image segmentation," IEEE Trans. Med. Imaging **39**(10), 3008–3018 (2020).
29. M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "DenseASPP for semantic segmentation in street scenes," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), pp. 3684–3692.
30. https://github.com/hahabingo/GDCSeg-Net
31. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), pp. 4510–4520.
32. S. Woo, J. Park, J. -Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 3–19.
33. X. Xi, X. Meng, Z. Qin, X. Nie, Y. Yin, and X. Chen, "IA-net: informative attention convolutional neural network for choroidal neovascularization segmentation in OCT images," Biomed. Opt. Express **11**(11), 6122–6136 (2020).
34. X. Xi, X. Meng, Y. Lu, X. Nie, G. Yang, H. Chen, X. Fan, Y. Yin, and X. Chen, "Automated segmentation of choroidal neovascularization in optical coherence tomography images using multi-scale convolutional neural networks with structure prior," Multimedia Syst. **25**(2), 95–102 (2019).

35. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: learning where to look for the pancreas," arXiv preprint, arXiv:1804.03999 (2018).

36. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested u-net architecture for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2018), pp. 3–11.

37. Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," IEEE Geosci. Remote Sens. Lett. **15**(5), 749–753 (2018).

38. D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, and H. D. Johansen, "ResUNet++: An advanced architecture for medical image segmentation," in *2019 IEEE International Symposium on Multimedia (ISM)*, (2019), pp. 225–2255.

39. L. Mou, Y. Zhao, H. Fu, Y. Liu, J. Cheng, Y. Zheng, P. Su, J. Yang, L. Chen, A. F. Frangi, M. Akiba, and J. Liu, "CS^2-Net: deep learning segmentation of curvilinear structures in medical imaging," Med. Image Anal. **67**, 101874 (2021).

40. V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017).

41. Z. Zhang, F. S. Yin, J. Liu, W. K. Wong, N. M. Tan, B. H. Lee, J. Cheng, and T. Y. Wong, "Origa-light: an online retinal fundus image database for glaucoma analysis and research," in *Engineering in Medicine and Biology Society (EMBC)* (IEEE, 2010), pp. 3065–3068.

42. S. Dey, K. Tahiliani, J. R. Harish Kumar, A. K. Pediredla, and C. S. Seelamantula, "Automatic segmentation of optic disc using affine snakes in gradient vector field," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2019), pp. 1204–1208.

43. S. Yu, D. Xiao, S. Frost, and Y. Kanagasingam, "Robust optic disc and cup segmentation with deep learning for glaucoma detection," Comp. Med. Imag. and Graphics **74**, 61–71 (2019).

44. J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "TransUNet: transformers make strong encoders for medical image segmentation," arXiv preprint, arXiv: 2102.04306 (2021).

45. Y. Gao, M. Zhou, and D. Metaxas, "UTNet: a hybrid transformer architecture for medical image segmentation," arXiv preprint, arXiv: 2107.00781 (2021).