Joint Optic Disc and Dup Segmentation Based on Densely Connected Depthwise Separable Convolution Deep Network

Bingyan Liu

South China Normal University

Daru Pan (✉ pandr@scnu.edu.cn)

South China Normal University

Hui Song

South China Normal University

Research

Keywords: Deep learning, optic disc segmentation, optic cup segmentation, depthwise separable convolution, densely connected

Posted Date: August 12th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-54643/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

RESEARCH

Joint Optic Disc and Cup Segmentation Based on Densely Connected Depthwise Separable Convolution Deep Network.

Bingyan Liu, Daru Pan^{*} and Hui Song

^{*}Correspondence:

pandr@scnu.edu.cn

South China Normal University,
510006 Guangzhou, China

Full list of author information is
available at the end of the article
†

Abstract

Background: Glaucoma is an eye disease that causes vision loss and even blindness. The cup to disc ratio is the main indicator used to screen and diagnose glaucoma, optic disc and optic cup segmentation can assist computer in diagnosing glaucoma. Therefore, accurate segmentation of optic disc and optic cup is beneficial to the screening and diagnosis of glaucoma and helps patients diagnose and treat early.

Method: In this paper, we consider the segmentation of the optic disc and the optic cup as a multi-category semantic segmentation task, and proposed a deep learning-based model named DDSC-Net (densely connected depthwise separable convolution network) to extract the optic disc and the optic cup. The backbone network is composed of densely connected deep separable convolution blocks to form a deeper network and the image pyramid input is introduced into the input layer to widen the network. In the post-process, we apply the morphological method to refine the output segmentation results.

Result: The proposed method is evaluated on two publicly available datasets, DRISHTI-GS dataset and REFUGE dataset. And the experiment results show that our DDSC-Net outperforms state-of-the-art optic disc and cup segmentation methods in terms of disc coefficients and Jaccard score. Furthermore, our method achieves the best result on the more challenging optic cup segmentation task.

Conclusion: The promising segmentation performances reveal that our method has potential in assisting the screening and diagnosis of the glaucoma.

Keywords: Deep learning; optic disc segmentation; optic cup segmentation; depthwise separable convolution; densely connected

Background

Glaucoma is an eye disease that damages the optic nerves and causes irreversible vision loss [1]. It has been estimated that 60.5 million people globally were affected by glaucoma in 2010 and predicted to affect almost 80 million people worldwide by 2020 [2]. Since vision loss is irreversible, early detection and diagnosis are very important to prevent vision loss and has been shown to decrease the rate of blindness by around 50% [3]. Hence it is essential to have a glaucoma screening technique to identify the glaucomatous and healthy eyes. Intraocular pressure assessment (IOP), visual field test and optic nerve head (ONH) assessment are three main techniques to detect glaucoma, in which ONH evaluation is the most clinically significant screening technique for glaucoma. For ONH evaluation, cup to disc ratio (CDR), means optic

nerve rim to disc ratio in diameters, which healthy people are within 0.3, is one of the most important indexes. In color fundus images, the optic disc which appears as a bright yellow oval region, can be divided into two regions: a central bright region named the optic cup and a peripheral zone named the neuroretinal rim. Accurate segmentation of optic disc (OD) and optic cup (OC) is very essential for the calculation of CDR. However, in the past this part of the evaluation work was done by clinicians, which was time-consuming and costly in the case of a large number of samples. Therefore, computer aided diagnosis (CAD) methods for large-scale fundus image screening are needed. Segmenting the optic disc and optic cup is the preliminary step in CDR measurement and evaluating glaucoma. A number of works have been proposed to segment the optic disc and cup from the fundus image to assist clinicians diagnose glaucoma more effectively.

The main segmentation techniques include template-based methods [4], [5], boundary detection [6], [7], hand-crafted visual feature approach, and deep learning segmentation methods [8], [9], [10], [11]. In these methods, the template-based method models the OD as a circular or elliptical object and use a circular Hough transforms [4], [5] or sliding band filter [6] to obtain an approximate boundary of optic disc. The method based on boundary detection needs to mark multiple landmark points [6], or requires each pixel to have a direct edge in the 15-pixel neighborhood, and does not consider depth information [7]. The hand-crafted visual feature approaches convert boundary problems into pixel classification problems, and obtain satisfactory results. However, due to the colour change of the fundus images, intensity inhomogeneity, the shape of the intervertebral disc, and the presence of internal and surrounding blood vessels, precise boundary of optic disc and optic cup cannot be robustly obtained through these approaches. In addition, most of the aforementioned methods use the hand-crafted features of optic disc and optic cup to obtain segmentation results. However, good manual features are difficult to design and usually cannot achieve good classification performance.

It has been shown that deep learning-based techniques, represented by convolution neural network (CNN) [12], achieve promising results for optic disc and optic cup segmentation [8], [9], [10], [11]. Compared with the aforementioned hand-crafted design feature extraction methods, deep learning networks based on convolutional neural networks can automatically extract complex features from the input data. In [8], an optic disc and optic cup segmentation method based on CNN which used an entropy-based sampling technique to reduce computational complexity is proposed. Although the segmentation performance of this approach is better than that of the method of using hand-crafted features, it leaves much to desire in terms of time consuming. In [9], a fully-convolutional neural network [13] based on VGG-16 net [14] is proposed to segment the optic disc. This method can segment retinal vessels and optic disc simultaneously, without segmenting the more challenging OC. In 2015, Ronneberger [15] et al. propose an architecture called U-Net, which has been widely used in biomedical image segmentation and has achieved good results. And many biological image segmentation networks are modified based on the U-Net convolutional network. In [10], a modified version of the U-Net convolutional network is presented for automatic optic disc and cup segmentation. Although this method has the advantages of fast processing speed and less parameters, it fails

to make full of the context information of the network, which results in the poor segmentation of OC. In [11], the authors utilize fully convolutional networks with adversarial training for jointly segment the OD and OC, but assumes that the region of interest (ROI) can be accurately extracted by preprocessing. Furthermore, Fu et al [16], proposes a deep learning architecture named M-Net which used a polar transformation with the multi-label deep learning concept. Although M-Net uses the same network architecture to segment the optic disc and the cup, it still considered these two segmentation tasks as two independent problems, and also without making full use of the prior information that OC is located within the OD.

At present, most segmentation methods consider the process of detection-then-segmentation mechanism. However, many location methods are susceptible to the influence of the image quality and pathological area [8], [10], resulting in wrong location. Furthermore, the accurate segmentation of the optic cup is a more challenging task in the task of disc and cup segmentation. Although most of the disc and cup segmentation networks based on deep learning can obtain satisfactory OD segmentation results, they cannot produce more accurate optic cup segmentation results [8], [9], [10], [11], [16]. For the above problems, we propose a deep learning-based method to make full use of the contextual information and prior knowledge of fundus images to improve the segmentation accuracy of OD and OC, especially for more challenging OC segmentation tasks. In our method, we design a CNN network and Hough transform method to locate the optic disc, then a novel end-to-end multi-label densely connected depthwise separable convolution network is utilized to segment optic disc and optic cup simultaneously. Our method obtains 100% accuracy of optic disc localization results on the DRISHTI-GS and REFUGE datasets, and achieves start-of-the-art performances in optic disc and optic cup segmentation, especially in the more difficult optic cup segmentation task.

The contributions of this paper are as follow:

- 1 we propose a novel computer aided diagnosis (CAD) method for joint optic disc and optic cup segmentation using a densely connected depthwise separable convolution network, named DDSC-Net. The presented DDSC-Net is an deep learning-based system which contains DDSC blocks to create a deeper network to learn more semantic information and produces OD and OC segmentation probability map.
- 2 We use a post-processing method based on morphological closed operation. This post processing methods improves the segmentation results of optic disc and cup, which improves the index of dice coefficient in REFUGE dataset from 0.9592 to 0.9601 in OD and 0.8894 to 0.8903 in OC.
- 3 We conduct experiments on two public retinal image datasets, DRISHTI-GS dataset and REFUGE dataset, to evaluate the effectiveness of the proposed DDSC-Net. The experimental results show that Our DDSC-Net achieves state-of-the-art segmentation performance, with the dice coefficients of 0.9780, 0.9123 for OD and OC on DRISHTI-GS dataset and 0.9601, 0.8903 on REFUGE dataset.

Experiments and Results

Dataset

In this paper, we conduct experiments on the DRISHTI-GS [17] and REFUGE [18] dataset. The number of dataset is shown in the Table 2.

DRISHTI-GS dataset

The DRISHTI-GS dataset consists of 101 retinal fundus images which were collected at Aravind eye hospital, Madurai. The resolution of these images is 2986*1944 and stored in uncompressed PNG format. And the ground truth of these images was marked by 4 ophthalmologists with different clinical experience and divided into 50 training and 51 testing images.

REFUGE dataset

The Retinal Fundus Glaucoma Challenge (REFUGE) dataset consists of 1200 images which include 120 glaucomatous and 1080 non-glaucoma images. The REFUGE dataset is divided into three parts: 400 training images, 400 validation images and 400 testing images, in which the validation and testing images are acquired with the same cameras.

Image processing and data augmentation

In order to segment the optic disc and optic cup based on the original resolution and save computing resources, we employed the ROI extractor to locate the position of the optical disc, and then cut the image into a small image size, in which DRISHTI-GS dataset is crop to 560×560 while REFUGE dataset 480×480.

Because the dataset used for training has fewer pictures, and too few pictures for network training may lead to overfitting, so we utilized data augmentation to expand training images to prevent this problem. The augmentation methods include translation, rotation, noise addition, and brightness adjustment. Among them, the images used for training in the DRISHTI-GS data set were expanded to 6,750, and the REFUGE dataset was expanded to 30,000.

Evaluation metrics

In our paper, we adopt the Dice coefficients (DC), Jaccard (JAC), Sensitivity (SEN) and Precision (PRE) to evaluate the segmentation performance of the presented method. The criteria are defined as:

$$Dice = \frac{2 * N_{tp}}{2 * N_{tp} + N_{fp} + N_{fn}}$$

$$Jaccard = \frac{N_{tp}}{N_{tp} + N_{fp} + N_{fn}}$$

$$Sensitivity = \frac{N_{tp}}{N_{tp} + N_{fn}}$$

$$Precision = \frac{N_{tp}}{N_{tp} + N_{fp}}$$

Where the N_{tp} , N_{fp} , N_{fn} represent the number of true positive, false positive and false negative pixels, respectively.

Evaluation

All experiments were implemented with python based on pytorch backend on the workstation of Intel i7-8700K, 16G RAM and a Nvidia 1080Ti GPU was used to train the network. After data augmentation we split the dataset into 90% for training and 10% for validation. The image is resized into the resolution of 240×240 , and then fed to the network for training. The final segmentation results are restored to the same resolution as the ROI image. Figure 5 and Figure 6 show the visual examples of the segmentation in DRISHTI-GS and REFUGE dataset, respectively.

In order to verify that our method is better than other methods, we compare our proposed method with the state-of-the-art approaches, such as pOSAL framework [19], GL-Net [20], M-Net [16], Stack-U-Net [21], WGAN [22], two-stage Mask R-CNN [23], multi-modal self-supervised pretraining network [24], Shuang Yu [25] and A. Sevastopolsky [10]. Additionally, we compare with the Fully convolutional network U-Net [15]. In these methods, pOSAL framework, GL-Net, and WGAN used Generative Adversarial networks(GAN) [26] to segment the fundus pictures, and both achieved good segmentation results. M-Net, Stack-U-Net, A. Sevastopolsky [10] and Shuang Yu [25] are modified based on U-Net structure. M-Net used multi-input, multi-output and polar transformation method to segment the optic disc and optic. Stack-U-Net stack more than one U-NET and achieved impressive results on DRISHTI-GS dataset. A. Sevastopolsky [10] modified a faster U-Net segmentation network and Shuang Yu [25] using a modified U-Net architecture, which combines the pre-trained ResNet-34 model as encoding layers with classical U-Net decoding layers. Two-stage Mask-RCNN uses Mask-RCNN-based method to locate and then segment OD and OC. While A'lvaro S. Hervella etl. [24] propose a novel self-supervised pre-training method to improve the segmentation of the optic disc and cup which also achieved a good result. All of the above methods have achieved good segmentation results, but through experimental comparison, our proposed method achieves a higher score than them, especially in OC segmentation task.

The final experimental results are shown in Table 3. On Drishti-GS dataset, our method obtained DC score, JAC score, sensitivity and precision of 0.9780, 0.9570, 0.9784 and 0.9778 for the OD segmentation and 0.9123, 0.8442, 0.9220 and 0.9149 for OC segmentation, respectively. On REFUGE dataset, we achieve DC score, JAC score, sensitivity and precision of 0.9601, 0.9239, 0.9814 and 0.9412 for the OD segmentation and 0.8903, 0.8065, 0.9209 and 0.8749 for OC segmentation, respectively.

Ablation experiment results

Table 4 shows the ablation results scores of the OD and OC segmentation on Drishti-GS dataset and REFUGE dataset. And the visual examples is showed in Figure 6. Table 4 shown that utilize Multi input and deeper network both can improve the performance of the model, and reprocess can also improve the segmentation results. When using DDSC-Net, the segmentation performance of the model has been significantly improved on two public datasets. On Drishti-GS dataset, the OD and OC

segmentation results on DC are 0.33% and 1.8% higher than the simple DDSC-Net, and 0.36% and 0.47% higher than the DDSC-Net without multi input, respectively. When post processing is utilized, the segmentation performance is further improved. On REFUGE dataset, the OD and OC segmentation results on DC are increased by 0.09% and 0.09%, respectively.

It can be seen from the experiment results that Image pyramid multi-scale input, deeper network and reprocess can improve the performance of the model. As can be seen from Figure 6, compared with simple DDSC-Net and DDSC-Net without multi input, our DDSC-Net can significantly improve the accuracy of OD and OC, especially on OC segmentation results.

Compared with U-shaped networks

Since our network structure is modified based on the U-Net architecture, we firstly compare our model with other methods that also use the U-shaped network. Firstly, we compare with the original U-net on the Drishti-GS dataset. From Table 5, We have achieved 2.8% and 9.23% higher than the original U-net in the Dice coefficients of the optic disc and the optic cup, respectively. When compared with the same improved method modified based on the U-net structure [16], [21], [10], [25], our method is better than the best improved method proposed by Shuang Yu [25], which is 0.42% higher in OD, and 2.46% in the more difficult OC segmentation task. On REFUGE dataset, our method is also better than that of M-net [16] using this dataset. From Table 6 we can see that our method is 1.65% higher in the optic disc and 5.88% in the optic cup compared to M-Net.

Compared with other networks

Because GAN-based methods and other deep learning networks also achieve satisfactory segmentation results in OD and OC segmentation task. Therefore, we then compared the use of generative adversarial ideas, WGAN [19], GL-Net [20] and pOSAL framework [22] and other deep learning methods [23], [24]. From Table 5 and Table 6, we can see that our proposed method still performs better than these methods. Particularly in the optic disc and cup segmentation results on the Drishti-GS dataset, our model outperforms the state-of-the-art method GL-Net by 0.7% and 0.73% in term of dice coefficients, respectively. In the optic cup segmentation results on the REFUGE dataset our model outperforms the pOSAL by 0.77%. These comparative results demonstrate that our proposed method can effectively improve the segmentation accuracy of the OC and obtain competitive results in the segmentation of the OD.

Compared with REFUGE challenge

We also compared our segmentation results with the results of the REFUGE challenge. The segmentation results of the 12 participating teams are shown in Table 7. It is obvious that in the DC metric our method achieved the best segmentation result in the segmentation of the cup, which is 0.66%. higher than the best result of 0.8837. In the segmentation of the optic disc, we achieved the second-best segmentation result, showing strong competitiveness.

From the above, it can be concluded that our method can effectively segment the optic disc and the optic cup, especially in the segmentation task of OC, and achieved

the start-of-the-art segmentation performance on the Drishti-GS and REFUGE dataset.

Discussion

In this paper, a densely connected depthwise separable convolution deep network for joint OD and OC segmentation method was proposed. As experimental results and comparative experiments above, we can draw a conclusion that, Multi-scale image pyramid input and densely connected deep separable network can effectively perform OD and OC segmentation. The correct segmentation of optic disc and optic cup is essential for calculating CDR, which helps clinicians diagnose glaucoma more effectively. Because OC is located in OD and there is no obvious boundary like OD, OC segmentation has always been a more challenging task in the segmentation task of OD and OC. In prior segmentation work, the proposed method could not obtain satisfactory OC segmentation results, whereas the network we designed can effectively improve the OC segmentation results. According to our experimental results, using DDSC blocks to deepen the network and enhance the fusion of contextual semantic information through dense connections and skip connections can effectively improve the segmentation effect. Compared with the current optimal OD and OC segmentation network, our method obtains the state-of-the-art segmentation results both in two public datasets. Our DDSC module is formed by densely connected deep separable convolution, which have 8 to 9 times fewer parameters than standard convolution. Therefore, the advantage of using the DDSC module is that it can deepen the network depth and learn deeper semantic information for feature fusion without increasing the network parameters.

Although our method has achieved the best OD and OC segmentation results on the testing images of two public datasets. However, our proposed method still has limitations. The limitation of our research is that for fundus images used by different equipment and institutions, the model parameters we trained cannot perform stable segmentation on fundus images of these different shooting devices. To use our method to perform segmentation on fundus images of a new dataset, the network parameters need to be retrained. Improving the generalization of the model and stably segmenting fundus images from different data sets are also our future research directions.

Conclusion

In this paper, we have developed a dense depthwise separable convolution based network named DDSA-Net for joint optic disc and optic cup segmentation. The proposed network uses a dense connection of deep separable convolution network as the backbone network, and adds multi-scale image input at the input end to widen network and prevent the information loss of deep network. Finally, image morphology is employed to post-process the segmentation results. We also utilize data augmentation to improve the segmentation of the model and also help to prevent the network from overfitting. The DDSA-Net is verified on the DRISHTI-GS and REFUGE dataset. The experimental results show that our model produces better segmentation results than the state-of-the-art method on the DRISHTI-GS and REFUGE dataset, which show more better potential for application. Especially

in the more challenging OC segmentation task, the optimal segmentation result was achieved. In future research work, we can try to apply our method to other medical segmentation tasks, and improve network segmentation performance by combining self-supervision mechanisms. We hope that our research methods can assist clinicians in better screening and diagnosis of glaucoma, so that patients can receive effective treatment as soon as possible to prevent vision loss.

Method

The method proposed in this paper use a two-stage approach to implement the segmentation of the optic disc and cup. In the first stage, CNN and Hough circle detection are used to obtain the center coordinates of the optic disc and extract the ROI. In the second stage, ROI is fed into the model to train a high-precision segmentation network to obtain the accurate segmentation results of optic disc and cup. The overall flowchart of our proposed method is shown in Figure 1.

ROI extraction network

Since the resolution of a complete fundus picture taken by professional camera is generally relatively large, and the area of interest is only a small area in the fundus picture, locating and cropping out the region of interest can reduce the interference of unnecessary background information on the segmentation result, which improve the segmentation accuracy and reduce the amount of calculation. However, the methods which employ green channel images [8] or morphological operations [27] to detect optic disc is susceptible to the effects of images taken by different devices, fundus image quality, brightness, internal blood vessels, and lesions in fundus images, resulting in reduced location accuracy. In our work, we utilize a method based on CNN network to extract features to solve this problem. The model and segmentation process are shown in the Figure 2. At this stage, we design a simple convolutional neural network to segment the optic disc simply, then use Circular Hough Transform (CHT) [28] to calculate the center of the optic disc. With this method, we can locate the optic disc with 100% accuracy and crop the ROI area. The location result is shown in Table 1.

CHT is an extension of the Hough transform [29], which is mainly used to detect the circle object in the image. For circle detection, the HT is based on the equation of circle, defined as:

$$(x - a)^2 + (y - b)^2 = r^2$$

Where (a, b) represents the coordinates of circle center and r is radius. Center coordinates can be obtained by performing the CHT on the image. CHT can be defined as:

$$(P_c, r) = CHT(I, r_{\min}, r_{\max})$$

Where $p_c = (i_c, j_c)$ and r represents the center position and the radius respectively which define the circle with the highest punctuation in the Circular Hough Transform implemented by CHT. I is the input image. The radius r is restricted to be between $r_{\min}$ and $r_{\max}$. In our method we set $r_{\min}$ and $r_{\max}$ as 40, 160 respectively. After obtaining the coordinates of the center of the disc, we use it as the center

point to cut the original image into a small picture with a resolution of 480×480 on REFUGE dataset and 560×560 on DRISHTI-GS dataset. The image contains the optic disc, optic cup and some background information. The visual result examples of ROI extraction are shown in Figure 3

DDSC network Architecture

In the object detection network [30], [31], [32], features extracted from shallow network can be used to detect small objects, while features extracted from deep network can be used to detect large objects. In the segmentation task of optic disc and optic cup, this idea was adopted to design our network structure. Considering the prior knowledge that the optic cup is located in the optic disk, we use dense and skip connection to make full use of the context semantics of the shallow layers and deep layers. The proposed network structure is detachedly shown in the Figure 4. The proposed deep network, named DDSC-Net, is consists of three main parts. The first part is the image pyramid [33], which is used as the multi-scale input of the network so that the network can receive image information of different scales. Multi-scale input can solve the problem of losing part of the image information with the depth of the network. The second part is a U-shaped convolutional neural network which includes encoder module on the left and decoder module on the right. The output map is activated by the softmax activation function, and then the cross entropy loss function is introduced to calculate the difference between the segmentation result and the real ground truth.

Image pyramid multi-scale input

The input of the DDSC-Net is an image pyramid, which can effectively improve the segmentation quality of the network. This method employs the average pooling layer to build an image pyramid, which is then introduced into different layers of the encoder module. The advantages of this are as follow: 1). To avoid a large increase in network parameters; 2). increase the network width of the decoder depth; 3). and reduce the loss of information caused by the deepening of the network.

DDSC network

In our proposed network, we use depthwise separable convolution layers to replace most of the standard convolutional layers in the network, which can significantly reduce the amount of computation. Therefore, we design a deeper network to learn more feature information from input data, especially the semantics of the optic cup. In addition, we execute more skip connections between encoder and decoder to enhance the transfer of contextual feature information in our model. The DDSC network is composed of three parts: densely connected depthwise separable convolution blocks, subsampled layers and upsampling layers. A dense depthwise separable convolution (DDSC) block contains five densely connected layers which consist of a batch normalization layer, a rectified linear unit (Relu) activation function, and a depthwise separable convolution layer with kernel size of 3×3 . The subsampled layer is a max pooling layer with kernel size of 2 and stride of 2. And the upsampling layer is a 3×3 transposed convolution layer.

For standard convolution, the output feature map F for standard convolutional when assuming stride and padding as one is computed as

$$F_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot I_{k+i-1,l+j-1,m}$$

The parameters and computational cost of the standard convolutions are respectively computed as:

$$k * k * M * N$$

and

$$k * k * M * N * H * W$$

Where I is the input feature map or input image, K is the convolution kernel size with $k \times k$, M is the number of input channers, N is the number of output channel, H and W are the height and width of the input feature map or input image respectively. While depthwise separable convolution is made of depthwise and pointwise convolutions [34]. The output feature map F for depthwise separable convolutional is computed as:

$$F'_{k,l,n} = \sum_{i,j} K'_{i,j,m} \cdot I_{k+i-1,l+j-1,m}$$

And the parameters and computational cost of the depthwise separable convolutions are respectively computed as:

$$k * k * M + M * N$$

and

$$k * k * M * H * W + M * N * H * W$$

Comparing the parameters of the depthwise separable convolution with the standard convolution can be obtained as follows:

$$\frac{k * k * M + M * N}{k * k * M * N} = \frac{1}{N} + \frac{1}{k^2}$$

It can be seen that the depth separable convolution uses about 8 to 9 times less parameter than the standard convolution. Therefore, we can deepen and widen the network without causing an explosive increase in the number of parameters, and also enable the network to learn more contextual information.

post-processing

The output of the network is a map with resolution of 240×240 . We used cubic interpolation to restore it to 480×480 and 560×560 . Then adopted morphological operations to smooth the edges. There are four kinds of operation methods of image morphology: erode, dilate, open and close. Based on the prior knowledge that most of the optic disc and cup are elliptical structure, we use the closed operation in the

image morphology to fuse the pixel points with fine boundary connection and fill the concave angle of the image, so as to make the boundary of the segmented image smoother slippery. The closed operation can be expressed as follows:

$$F = (f \oplus s) \ominus s$$

Where f is the image, s the Structure element, $\oplus, \ominus$ represent dilate and erode respectively. In our work, s is a $7 * 7$ circular structure element.

Loss function

In our work, we regard the optic disc and optic cup segmentation as a multi-category segmentation task and use One-Hot encoding to process the data. Let $x \in R^{C*H*W}$ be the input image, and $y \in \{y_o, \dots, y_i\}^{i*H*W}$ is the One-Hot representation of the ground truth label, when the pixel belong to category i , $y_i = 1$, otherwise, $y_i = 0$. We treat the output as 3 categories $i = 3$ and the output of our model is a map of $f_i(x, v) = y' \in \{p_o, \dots, p_i\}^{i*C*W}$. In our work, we use Multi-class cross-entropy loss function to measure the difference between the output of the model and the ground truth label. The loss function Loss is defined as:

$$Loss = - \sum_{i=0}^2 y_i \log(f_i(x, v))$$

The output map $f_i(x, v)$ is a probability distribution, and each element $\{p_o, \dots, p_i\}^{i*C*W}$ represents the probability that the pixel belongs to the i -th category.

List of abbreviations

IOP: Intraocular Pressure assessment;
ONH Optic Nerve Head;
CDR: Cup to Disc Ratio;
OD: Optic Disc;
OC: Optic Cup;
CAD: Computer Aided Diagnosis;
CNN: Convolution Neural Network;
ROI Region of Interest;
CHT: Circular Hough Transform;
DC: Dice Coefficients;
JAC: Jaccard;
SEN: Sensitivity;
PRE: Precision;

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets are publicly available at the follow links:

1. DRISHTI-GS dataset: <http://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Home.php>;
2. REFUGE dataset: <https://refuge.grand-challenge.org>.

Competing interests

The declare that they have no competing interests.

Funding

This study was supported by the National Natural Science Foundation of China under Grant NO.61771206 and the National Youth Fund of China under Grant no.61801184.

Author's contributions

B.L. designed research, performed research, collected data, analysed data, and wrote the paper. D.P. and H.S. supervised the theoretical development and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Mr. zhongshuo Zhuo for his valuable help in the revising this manuscript.

References

1. 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* (2014)
2. H, A, Quigley, A, T, Broman: The number of people with glaucoma worldwide in 2010 and 2020. *British Journal of Ophthalmology* (2006)
3. Michelson, G., Wäntges, S., Hornegger, J., Lausen, B.: The papilla as screening parameter for early diagnosis of glaucoma. *Deutsches Rztblatt International* **105**(34-35), 583 (2008)
4. Aquino, A., Gegundez-Arias, M.E., Marin, D.: Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques. *IEEE Transactions on Medical Imaging* **29**(11), 1860–1869 (2010)
5. Dashtbozorg, B., Mendonça, A.M., Campilho, A.: Optic disc segmentation using the sliding band filter. *Computers in Biology & Medicine* **56**, 1–12 (2015)
6. Chakravarty, A., Sivaswamy, J.: Joint optic disc and cup boundary extraction from monocular fundus images. *Comput Methods Programs Biomed* **147**, 51–61 (2017)
7. Zheng, Y., Stambolian, D., O'Brien, J., Gee, J.C.: Optic disc and cup segmentation from color fundus photograph using graph cut with priors. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2013)
8. Zilly, J., Buhmann, J.M., Mahapatra, D.: Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. *Computerized Medical Imaging and Graphics* (2017). doi:10.1016/j.compmedimag.2016.07.012
9. Maninis, K.K., Pont-Tuset, J., Arbeláez, P., Gool, L.V.: Deep retinal image understanding (2016)
10. Sevastopolsky, A.: Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network. *Pattern Recognition & Image Analysis* **27**(3), 618–624 (2017). doi:10.1134/S1054661817030269
11. Shankaranarayana, S.M., Ram, K., Mitra, K., Sivaprakasam, M.: Joint optic disc and cup segmentation using fully convolutional and adversarial networks. In: *Fetal, Infant and Ophthalmic Medical Image Analysis*, pp. 168–176. Springer, ??? (2017)
12. 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)
13. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440 (2015)
14. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition (2014). arXiv:1409.1556
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 234–241 (2015). Springer
16. Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X.: Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE transactions on medical imaging* **37**(7), 1597–1605 (2018). doi:10.1109/TMI.2018.2791488
17. Sivaswamy, J., Krishnadas, S., Joshi, G.D., Jain, M., Tabish, A.U.S.: Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation. In: *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pp. 53–56 (2014). IEEE. doi:10.1109/ISBI.2014.6867807
18. Orlando, J.I., Fu, H., Breda, J.B., van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.-A., Kim, J., Lee, J., et al.: Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Medical image analysis* **59**, 101570 (2020). doi:10.1016/j.media.2019.101570
19. Wang, S., Yu, L., Yang, X., Fu, C.-W., Heng, P.-A.: Patch-based output space adversarial learning for joint optic disc and cup segmentation. *IEEE transactions on medical imaging* **38**(11), 2485–2495 (2019). doi:10.1109/TMI.2019.2899910
20. Jiang, Y., Tan, N., Peng, T.: Optic disc and cup segmentation based on deep convolutional generative adversarial networks. *IEEE Access* **7**, 64483–64493 (2019). doi:10.1109/ACCESS.2019.2917508
21. Sevastopolsky, A., Drapak, S., Kiselev, K., Snyder, B.M., Keenan, J.D., Georgievskaya, A.: Stack-u-net: refinement network for improved optic disc and cup image segmentation. In: *Medical Imaging 2019: Image Processing*, vol. 10949, p. 1094928 (2019). International Society for Optics and Photonics. doi:10.1117/12.2511572
22. Kadambi, S., Wang, Z., Xing, E.: Wgan domain adaptation for the joint optic disc-and-cup segmentation in fundus images. *International Journal of Computer Assisted Radiology and Surgery* (2020). doi:10.1007/s11548-020-02144-9

23. Almubarak, H., Bazi, Y., Alajlan, N.: Two-stage mask-rcnn approach for detecting and segmenting the optic nerve head, optic disc, and optic cup in fundus images. *Applied Sciences* **10**(11), 3833 (2020). doi10.3390/app10113833
24. Hervella, Á.S., Ramos, L., Rouco, J., Novo, J., Ortega, M.: Multi-modal self-supervised pre-training for joint optic disc and cup segmentation in eye fundus images. In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 961–965 (2020). IEEE. doi10.1109/ICASSP40776.2020.9053551
25. Shuang, Xiao, Shaun, Frost, Yogesan, Kanagasigam: Robust optic disc and cup segmentation with deep learning for glaucoma detection. *Computerized Medical Imaging & Graphics the Official Journal of the Computerized Medical Imaging Society* (2019). doi10.1016/j.compmedimag.2019.02.005
26. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. *Advances in Neural Information Processing Systems* **3**, 2672–2680 (2014)
27. Abdullah, M., Fraz, M.M., Barman, S.A.: Localization and segmentation of optic disc in retinal images using circular hough transform and grow-cut algorithm. *PeerJ* **4**, 2003 (2016)
28. Hough, B.P.: Methods and means for recognizing complex pattern (1962)
29. Duda, R.O., Hart, P.E.: Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM* **15**(1), 11–15 (1972)
30. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: Ssd: Single shot multibox detector. In: *European Conference on Computer Vision*, pp. 21–37 (2016). Springer
31. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement (2018)
32. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: Yolov4: Optimal speed and accuracy of object detection (2020)
33. Adelson, E.H., Anderson, C.H., Bergen, J.R., Burt, P.J., Ogden, J.M.: Pyramid methods in image processing. *RCA engineer* **29**(6), 33–41 (1984)
34. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (2017). [arXiv:1704.04861](https://arxiv.org/abs/1704.04861)

Figures

Figure 3 Location and cropping results of ROI. The first column of images is the original images, the second column is the position of the optic disc located, the third and fourth is the position of the optic disc after cropping and the cropping result.

Figure 4 Network architecture of the DDSC-Net. Our DDSC-Net architecture includes multi-scale input of image pyramid and DDSC-blocks. As shown in the figure, each DDSC-block is composed of five "BN + Relu + depthwise separable convolution" dense connection layers, the convolution kernel is 3×3 . The down sample is the maximum pooling layer, and the upsample is transposed convolution. Finally, the output of the network uses a softmax activation function to classify the output map into three categories: 0, 1 and 2. 0 is the background, category 1 is the optic disc, and 2 is the optic cup.

Tables

Table 1 Localization accuracy results

Dataset	Images Accurately Located	Total	Accuracy
DRISHTI-GS	101	101	1.0
REFUGE	1200	1200	1.0

Table 2 Statistics of the dataset used in evaluation the proposed method

DATASET	Train and Val	Test	Total	Image size	Release year
DRISHTI-GS	50	51	101	2047×1759	2014
REFUGE	800	400	1200	2124×2056 and 1634×1634	2018

Table 3 Experiment Results

DATASET	OD				OC			
	<i>DC</i>	<i>JAC</i>	<i>SEN</i>	<i>PRE</i>	<i>DC</i>	<i>JAC</i>	<i>SEN</i>	<i>PRE</i>
DRISHTI-GS	0.9780	0.9570	0.9784	0.9778	0.9123	0.8442	0.9220	0.9149
REFUGE	0.9601	0.9239	0.9814	0.9412	0.8903	0.8065	0.9209	0.8749

Table 4 The model segmentation results with and without Multi input, deeper network and post-processing

Method	DRISHTI-GS Dataset		REFUGE Dataset	
	$DC_{disc}(mean)$	$DC_{cup}(mean)$	$DC_{disc}(mean)$	$DC_{cup}(mean)$
Simple DDSC-Net	0.9746	0.8939	0.9575	0.8837
without Multi input	0.9743	0.9076	0.9589	0.8880
DDSC-Net	0.9779	0.9123	0.9592	0.8894
DDSC-NET+Post-Processing	0.9780	0.9123	0.9601	0.8903

Table 5 Comparison with the deep learning methods on DRISHTI-GS dataset

Method	$DC_{disc}(mean)$	$DC_{cup}(mean)$	$JC_{disc}(mean)$	$JC_{cup}(mean)$
U-Net [15]	0.950	0.820	-	-
M-Net [16]	0.959	0.866	-	-
Stack-U-Net [21]	0.970	0.890	-	-
A. Sevastopolsky [10]	-	0.850	-	0.750
Shuang Yu [25]	0.973	0.887	-	-
MULTI-MODEL-PRE-TRAINING [24]	0.960	0.902	0.924	0.823
WGAN [22]	0.954	0.840	0.912	0.724
pOSAL [22]	0.965	0.858	-	-
GL-Net [20]	0.971	0.905	-	-
Proposed	0.978	0.912	0.957	0.844

Table 6 Comparison with the deep learning methods on REFUGE dataset

Method	$DC_{disc}(mean)$	$DC_{cup}(mean)$	$JC_{disc}(mean)$	$JC_{cup}(mean)$
M-Net [16]	0.9436	0.8315	-	-
MULTI-MODEL-PRE-TRAINING [24]	-	-	0.9225	0.7902
Two-stage Mask R-CNN [23]	0.9504	0.8546	-	-
pOSAL [22]	0.9602	0.8826	-	-
Proposed	0.9601	0.8903	0.9239	0.8065

Table 7 Result of OD and OC segmentation on the REFUGE dataset. Top Three Results are marked in bold

TEAM	$DC_{disc}(mean)$	$DC_{cup}(mean)$
CHUKMED	0.9602	0.8826
Masker	0.9496	0.8837
BUCT	0.9525	0.8728
NKSG	0.9488	0.8643
VRT	0.9532	0.8600
AIML	0.9505	0.8519
Mammoth	0.9361	0.8667
SMILEDeepDR	0.9386	0.8367
NightOwl	0.9487	0.8257
SDSAIRC	0.9436	0.8315
Cvblab	0.9077	0.7728
WinterFell	0.8772	0.6861
Proposed	0.9601	0.8903

Figures

Figure 1

Overview of our proposed method. Firstly, simple DDSC-Net without multi input is used to rough segment the optic disc, and located the optic disc by CHT. Then fed cropped image into segmentation network to joint segment OD and OC.

Figure 2

ROI extraction model. Simple DDSC-Net architecture is a simple version of DDSC-Net. After obtain the OD segmentation result, CHT is employed to calculate the center and radius of OD. ROI is a small image cut out from the center of the calculated circle.

Figure 3

Location and cropping results of ROI. The first column of images is the original images, the second column is the position of the optic disc located, the third and fourth is the position of the optic disc after cropping and the cropping result.

Figure 4

Network architecture of the DDSC-Net. Our DDSC-Net architecture includes multi-scale input of image pyramid and DDSC-blocks. As shown in the figure, each DDSC-block is composed of five "BN + Relu + depthwise separable convolution" dense connection layers, the convolution kernel is 3×3 . The down sample is the maximum pooling layer, and the upsample is transposed convolution. Finally, the output of the network uses a softmax activation function to classify the output map into three categories: 0, 1 and 2. 0 is the background, category 1 is the optic disc, and 2 is the optic cup.

Figure 5

he visual examples of OD and OC segmentation on DRISHTI-GS dataset. The first row is the input image, the second and fourth rows are the ground truth of the OD and the OC respectively, and the third and fifth rows are the OD,OC segmentation results of our model, where the black denote the background, and the white part denote the OD and OC segmentations.

Figure 6

Some ablation experiment results on REFUGE dataset. The first row is the input image, the second is the ground truth, the third rows is the simple DDSC-Net segmentation results, fourth row is the DDSC-Net without multi input, fifth row is the results of DDSC-Net and sixth row is the reprocess result, where the white region denote the background, and the gray and black region denote the OD and OC segmentations.