Localizing Optic Disc and Cup for Glaucoma Screening via Deep Object Detection Networks

Xu Sun[1], Yanwu Xu[1,4(✉)], Mingkui Tan[2], Huazhu Fu[3], Wei Zhao[1], Tianyuan You[1], and Jiang Liu[4]

[1] Guangzhou Shiyuan Electronic Technology Company Limited, Guangzhou, China
ywxu@ieee.org
[2] Institute for Infocomm Research, A*STAR, Singapore, Singapore
[3] South China University of Technology, Guangzhou, China
[4] Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences, Cixi, China

Abstract. Segmentation of the optic disc (OD) and optic cup (OC) from a retinal fundus image plays an important role for glaucoma screening and diagnosis. However, most existing methods only focus on pixel-level representations, and ignore the high level representations. In this work, we consider the high level concept, *i.e.*, objectness constraint, for fundus structure analysis. Specifically, we introduce a deep object detection network to localize OD and OC simultaneously. The end-to-end architecture guarantees to learn more discriminative representations. Moreover, data from a similar domain can further contributes to our algorithm through transfer learning techniques. Experimental results show that our method achieves state-of-the-art OD and OC segmentation/localization results on ORIGA dataset. Moreover, the proposed method also obtains satisfactory glaucoma screening performance with the calculated vertical cup-to-disc ratio (CDR).

1 Introduction

As the second leading cause of blindness, glaucoma is predicted to affect about 80 million people by 2020 [7]. Since damage to optic nerves cannot be reversed, early detection of glaucoma is critical in preventing further deterioration. The vertical cup-to-disc ratio (CDR) is a commonly-used metric for glaucoma screening. Thus, accurate segmentation of the optic disc (OD) and optic cup (OC) is essential for developing practical automated glaucoma screening systems. Most existing methods tackle this challenging problem by using traditional segmentation techniques like thresholding, edge-based and region-based methods [5,13]. While these solutions work well with images of healthy retina, they tend to be misleading in illness cases where retinas suffer from different types of retinal lesions (*e.g.*, drusen, exudates, hemorrhage, *etc.*). Alternatively, some methods based on conventional machine learning pipelines have been proposed [6]. However, since these approaches rely too heavily on handcrafted features, their applicability is limited. A promising way to improve the performance is to employ

D. Stoyanov et al. (Eds.): COMPAY 2018/OMIA 2018, LNCS 11039, pp. 236–244, 2018.
https://doi.org/10.1007/978-3-030-00949-6_28

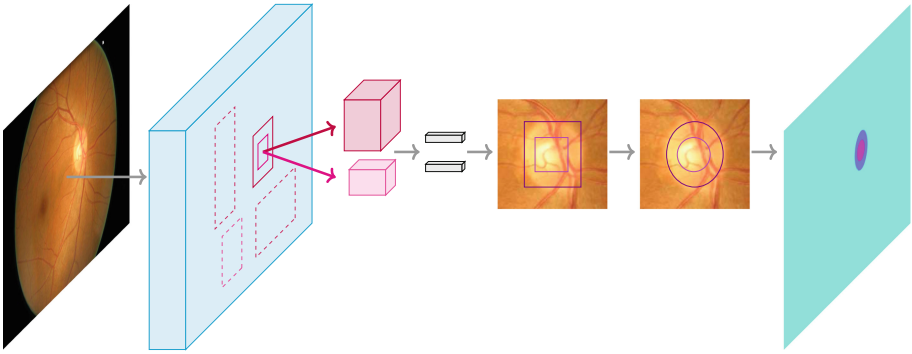


Fig. 1. Architecture of the proposed method for OD and OC segmentation/localization, where purple and magenta regions denote OD and OC respectively. (Color figure online)

the deep neural networks (DNN) architectures as they are capable of learning more discriminating features.

The effectiveness of DNN structure, indeed, has been well demonstrated in a recent state-of-the-art work termed M-Net [3]. Nevertheless, like most of the existing algorithms, M-Net is still a pixel-wise classification based approach, which first classifies each pixel as one of the three classes, *i.e.*, OD, OC and non-target, and then uses ellipse-fitting to approximate the smooth boundaries of OD and OC. In fact, the ellipse fitting step can be easily bypassed if OD and OC are assumed to be in a non-rotated ellipse shape. Therefore, the OD and OC can be treated as a whole object instead of a bunch of pixels without objectness constraint, which enables tackling the segmentation task from an object detection perspective. Follow this basic idea, two typical methods are presented in literature [12,14]. Unfortunately, these two methods are initially developed for OC localization only and not easy to adapt to OD localization.

In this paper, we formulate the OD and OC segmentation as a multiple object detection problem, with the introduction of the objectness constraints to improve the accuracy. Different from tradition pixel-wise based two-step approaches, we propose a simple yet effective method to jointly localize/segment OD and OC in a retinal fundus image based on deep object detection networks. The proposed method inherently holds four desirable features: (1) the multi-object network involves the OD and OC relationship and localizes them simultaneously; (2) the object detection network contains the objectness property, which presents the high-level discriminate representation; (3) the end-to-end architecture guarantees learning image features automatically, and also allows for transfer learning to address the challenging of small scale data; (4) by simply using Faster R-CNN [8] as the deep object detector, our method outperforms state-of-the-art OC and/or OD segmentation/localization methods on ORIGA dataset, and obtains satisfactory glaucoma screening performances with calculated CDR on ORIGA and SCES datasets.

2 Methodology

2.1 Architecture Overview

As shown from Fig. 1, in our detection driven method, the retinal fundus image is first fed into a deep convolutional network (*e.g.*, ResNet [4]) to produce a shared feature map at the last convolutional layer (*e.g.*, outputs of the 5th convolution block in ResNet [4]). Then a sparse set of rectangular candidate object locations are generated based on the feature map. This stage is commonly known as a Region Proposal Network (RPN). Then, the proposals are processed by the fully connected layers (*e.g.*, "fc6" and "fc7" in Faster R-CNN [8]) of the networks to predict class-specific scores and regressed bounds (*e.g.*, bounding box offset). For each foreground class (*i.e.*, OD and OC), we keep the bounding box with the highest confidence score as the final output of the detector.

Provided these two detected bounding boxes, the next stage is how to generate satisfactory OD and OC boundaries. It is widely accepted by many ophthalmologists and researchers that the shape of OD and OC can be well approximated by a vertical ellipse. Inspired by this concept, we propose to obtain the OD and OC boundaries by simply redrawing the predicted bounding boxes as vertical ellipses. The fundus image segmentation problem thus reduces to a relatively more straightforward localization task in our setting.

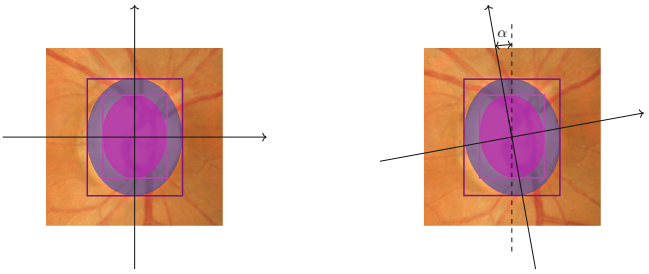
2.2 Implementation

In this paper, we adopt *Faster R-CNN* [8] as the object detector due to its flexibility and robustness comparing to many follow-up architectures. Faster R-CNN consists of two stages. During training, the loss for the first stage RPN is defined as

$$L(\{p_i, t_i\}) = \beta \sum_i L_{\text{cls}}(p_i, p_i^*) + \gamma \sum_i p_i^* L_{\text{reg}}(b_i, b_i^*) \tag{1}$$

where β, γ are weights balancing localization and classification losses. i is the index of an anchor in a training mini-batch. p_i is the predicted probability of the ith anchor being OD/OC. The ground truth label p_i^* indicates if the overlapping ratio between the anchor and the manual OD/OC mask is either larger than an given threshold (*e.g.*, 0.3) or the largest among all anchors. b_i is a vector standing for the 4 coordinates of the predicted bounding box, and b_i^* is that of the ground-truth box associated with a positive anchor (*i.e.*, with $p_i^* = 1$). The classification loss L_{cls} is the log loss over target and non-target classes, and the regression loss is a robust loss function (*e.g.*, the smooth L_1 loss). We refer readers [8] to for the more details of these entries. Meanwhile, the loss function for the second stage box classifier also takes a similar form of (1) using proposals produced from the RPN as anchors.

Data Augmentation: We employ two distinct forms of data augmentation in our experiment. The first form is to rotate fundus images from the training set using a set of angles over $-10(2)10$ degrees, where the notation $N_1(\Delta)N_2$



(a) Original image (b) Rotated image

Fig. 2. The generated "ground truth" OD and OC bounding boxes for the augmented image (*right*) from the original manual segmentation masks (*left*), where purple and magenta regions denote OD and OC respectively. The right image is obtained by rotating the whole original fundus image with an angle α regarding to its center. (Color figure online)

represents a list ranging from N_1 to N_2 with an increment of Δ. We limit the degree of rotation into such a small interval because of the assumption that the OD and OC are in a vertical ellipse shape. The second form is to generate image horizontal reflections on both the original training set and its rotated counterparts. With this transformation operation, a left eye image is artificially turned into the "right eye" image, and vice versa. This is desirable as we now have a balanced training set that consists of equal number of images from the left eye and the right eye. These two augmentation schemes increase the amount of our training set by a factor of 20.

Training Details: To enable training of the deep object detector, we first need to transform the manual segmentation masks into the "ground truth" bounding boxes. As illustrated in Fig. 2, this can be simply achieved by finding a vertical rectangle whose bounds lie exactly on the edge of the provided mask for each type of targets. Faster R-CNN [8] is implemented using `Tensorflow` based on a publicly available code [1].

We train the detection networks on a single-scale image using a single model. Before feeding images to the detector, we rescale their shorter side to 600 pixels. A 101-layer *ResNet* [4] is used as the backbone of Faster R-CNN. For anchors, we use 5 naive scales with box areas of $32^2, 64^2, 128^2, 256^2$, and 512^2 pixels, and 3 naive aspect ratios of $1:1$, $1:2$, and $2:1$. Instead of training all parameters from scratch, we fine-tune the network end-to-end from an ImageNet pre-trained model on a single NVIDIA TITAN XP GPU. We use a weight decay of 0.0001 and momentum of 0.9 for optimization. We start with a learning rate of 0.001, divide it by 10 at 100k iterations, and terminate training at 200k iterations.

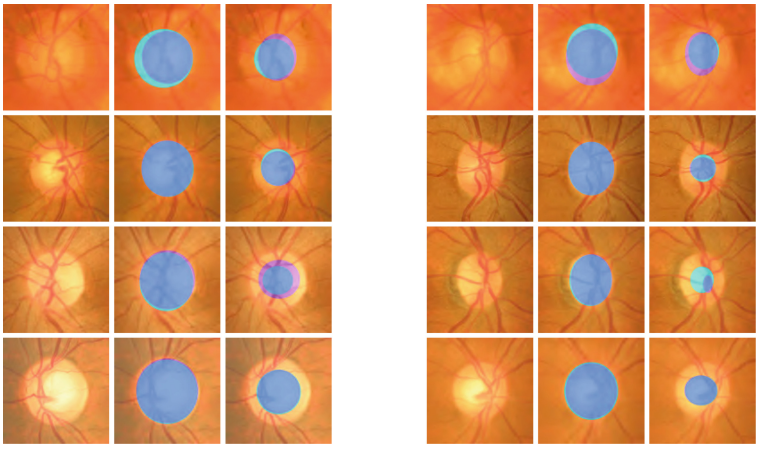
3 Experimental Results

3.1 OD and OC Segmentation

Following previous work in the literature, we evaluate and compare the OD and OC segmentation performance on ORIGA dataset [14]. In each image, OD and OC are labelled as vertical ellipses by experienced ophthalmologists. These images are divided into 325 training images (including 73 glaucoma cases) and 325 testing images (including 95 glaucoma cases). We employ two measurements to evaluate the performance, the overlapping error (E) and the absolute CDR error (δ) defined as:

$$E = 1 - \frac{A_{\mathrm{GT}} \cap A_{\mathrm{SR}}}{A_{\mathrm{GT}} \cup A_{\mathrm{SR}}}, \text{ and } \delta = |d_{\mathrm{GT}} - d_{\mathrm{SR}}| \tag{2}$$

where A_{GT} and A_{SR} denote the areas of the ground truth and segmented mask, respectively. d_{GT} denotes the manual CDR provided by ophthalmologists, and d_{SR} denotes the CDR that is calculated by the ratio of vertical cup diameter to vertical disc diameter from the segmentation results.



(a) Glaucoma images (b) Non-glaucoma images

Fig. 3. The segmentation results of the proposed method, where the purple, cyan and blue regions denote the manual masks, the segmentation outputs and their overlapping regions, respectively. From top to bottom rows are images with highest disc overlapping error, lowest disc overlapping error, highest cup overlapping error and lowest cup overlapping error, for cases with and without glaucoma, respectively. The overlapping errors from top to bottom rows, left to right are 0.219, 0.021, 0.096, 0.044, 0.247, 0.119, 0.471, 0.038, 0.264, 0.008, 0.045, 0.062, 0.293, 0.175, 0.752, and 0.035, respectively. (Color figure online)

We compare the proposed method to the state-of-the-art methods in OD and OC segmentation, including the relevant-vessel bends method (R-bend) [5],

Table 1. OD and OC Segmentation Performance Comparison of Different Methods on ORIGA Dataset.

Method	E_{disc}	E_{cup}	δ
R-Bend [5]	0.129	0.395	0.154
ASM [15]	0.148	0.313	0.107
SP [2]	0.102	0.264	0.077
LRR [11]	-	0.244	0.078
SW [14]	-	0.284	0.096
Reconstruction [12]	-	0.225	0.071
U-Net [9]	0.115	0.287	0.102
M-Net [3]	0.083	0.256	0.078
M-Net + PT [3]	0.071	0.230	0.071
Proposed	**0.069**	**0.213**	**0.067**

active shape model (ASM) [15], superpixel-based classification method (SP) [2], low-rank superpixel representation method (LRR) [11], sliding-window based method (SW) [14], reconstruction based method (Reconstruction) [12] and three deep learning based methods, *i.e.*, U-Net [9], M-Net [3] and M-Net with polar transformation (M-Net + PT). As shown in Table 1, our proposed deep object detection based method outperforms all state-of-the-art OD and OC segmentation algorithms on ORIGA dataset in terms of all aforementioned three evaluation criteria. Figure 3 shows some visual outputs of our method.

3.2 Glaucoma Screening/Classification Based on CDR

Following clinical convention, we evaluate the proposed method for glaucoma screening by using the calculated CDR value. Generally, the larger CDR value indicates the higher risk of glaucoma. We train our model using 7,150 images augmented from ORIGA training set, and then test it on ORIGA testing set and the whole SCES dataset [3] individually. We evaluate glaucoma screening/classification performance using the area under Receiver Operating Characteristic curve (AUC). As illustrated in Fig. 4, the AUC values of our method on ORIGA and SCES are 0.845 and 0.898, respectively, which are slightly lower than M-Net. Here we justify that: 1) the major objective of this work is to minimize OD and OC segmentation errors, which are not directly associated to glaucoma classification accuracy; 2) the state-of-the-art method M-Net [3] has no significant difference from our proposed method ($p >> 0.05$ on ORIGA and $p >> 0.05$ on SCES using DeLong's test [10]); 3) on the independent test dataset SCES, our proposed object detection method with objectness constraint achieves consistent higher sensitivity (*i.e.*, true positive rate) than other two competitive methods when false positive rate (*i.e.*, 1-Specificity) is lower than 0.2, which indicates that our approach is promising for practical glaucoma screening.

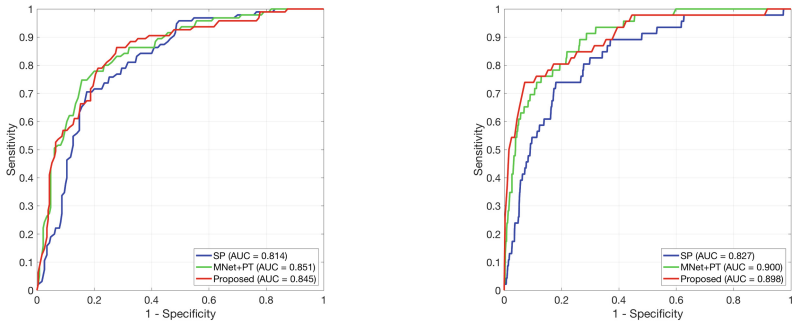


Fig. 4. Glaucoma screening performance on the ORIGA (*left*) and SCES (*right*) datasets.

4 Discussion

To illustrate why the proposed method is more preferable, below we highlight its main features by comparing it with two most related work in literature. The first one is the sliding-window based method [14], which first introduces the concept to segment OC via object detection technique. However, it is only developed for detecting OC after obtaining OD from another individual procedure. Our method, instead, incorporates these two separate tasks into a joint framework. Additionally, the sliding window method relies on handcrafted features. In contrast, our method learns deep representation directly from data. It should be pointed out that a fairly large amount of annotated data is usually required for training a highly accurate deep model, while in practice, such annotated data are expensive to acquire, especially in the field of medical imaging. One typical way of addressing a lack of data problem is by using a technique known as transfer learning and fortunately, this can be easily performed in DNN-based frameworks including our method. We also highlight that the training takes much longer time to converge and can hardly get satisfactory results, when the pre-trained model on ImgaeNet is not used to initialize the networks.

The second work to be compared is M-Net [3], which also trains a DNN for extracting image features and shares some aforementioned advantages of our method. To deploy M-Net, besides the end-to-end U-shape segmentation network, we also require an OD detector for detecting the disc center, a polar transformation method for mapping the disc image from the Cartesian coordinate system to polar coordinate system, an inverse polar transformation operation for recovering the segmentation result back to the Cartesian coordinate system, and an ellipse-fitting for generating smooth boundaries of OD and OC. In contrast, our method requires only a deep object detector.

5 Conclusion

In this paper, we tackle the fundus image segmentation problem from an object detection perspective, based on the circumstance that OD/OC can be well

approximated with vertical ellipse shape. The proposed method is not only conceptually simpler but also easier to deploy comparing to other multi-step based approaches such as M-Net [3]. Evaluated on the ORIGA dataset, our method outperforms all existing methods, achieving state-of-the-art segmentation results. Moreover, the proposed method also obtains satisfactory glaucoma screening performance with CDR calculated on the ORIGA and SCES datasets. In the future, we plan to investigate other deep object detectors and to explore more diagnostic indicators for glaucoma screening.

References

1. Chen, X., Gupta, A.: An implementation of faster RCNN with study for region sampling. arXiv preprint arXiv:1702.02138 (2017)
2. Cheng, J., et al.: Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE Trans. Med. Imaging **32**(6), 1019–1032 (2013)
3. 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 Trans. Med. Imaging **37**(7), 1597–1605 (2018)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778. IEEE (2016)
5. Joshi, G.D., Sivaswamy, J., Krishnadas, S.: Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. IEEE Trans. Med. Imaging **30**(6), 1192–1205 (2011)
6. Li, A., et al.: Learning supervised descent directions for optic disc segmentation. Neurocomputing **275**, 350–357 (2018)
7. Quigley, H., Broman, A.: The number of people with glaucoma worldwide in 2010 and 2020. Br. J. Ophthalmol. **90**(3), 262–267 (2006)
8. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: NIPS, pp. 91–99 (2015)
9. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
10. Sun, X., Xu, W.: Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. IEEE Sig. Process. Lett. **21**(11), 1389–1393 (2014)
11. Xu, Y., et al.: Optic cup segmentation for glaucoma detection using low-rank superpixel representation. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8673, pp. 788–795. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10404-1_98
12. Xu, Y., Lin, S., Wong, D.W.K., Liu, J., Xu, D.: Efficient reconstruction-based optic cup localization for glaucoma screening. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8151, pp. 445–452. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40760-4_56
13. Xu, Y., et al.: Efficient optic cup detection from intra-image learning with retinal structure priors. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7510, pp. 58–65. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33415-3_8

14. Xu, Y., et al.: Sliding window and regression based cup detection in digital fundus images for glaucoma diagnosis. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6893, pp. 1–8. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23626-6_1
15. Yin, F., et al.: Model-based optic nerve head segmentation on retinal fundus images. In: EMBC, pp. 2626–2629. IEEE (2011)