Optic Disc Segmentation from Retinal Fundus Images via Deep Object Detection Networks

Xu Sun[1], Yanwu Xu[1], Wei Zhao[1], Tianyuan You[1], Jiang Liu[2]

***Abstract*— Accurate optic disc (OD) segmentation is a fundamental step in computer-aided ocular disease diagnosis. In this paper, we propose a new pipeline to segment OD from retinal fundus images based on deep object detection networks. The fundus image segmentation problem is redefined as a relatively more straightforward object detection task. This then allows us to determine the OD boundary simply by transforming the predicted bounding box into a vertical and non-rotated ellipse. Using Faster R-CNN as the object detector, our method achieves state-of-the-art OD segmentation results on ORIGA dataset, outperforming existing methods in this field.**

I. INTRODUCTION

The optic disc (OD) is the exit point where ganglion cell axons leave the eye. Reliable OD segmentation is a necessary step in the diagnosis of various retinal diseases such as diabetic retinopathy and glaucoma.

In the literature, a plenty of research has been conducted to solve this challenging problem. Since the traditional segmentation techniques are usually difficult to achieve good performance, especially in illness cases, some machine learning based approaches have been investigated. However, comparing to the latest developed deep neural networks (DNN) based methods, the performance of these conventional machine learning approaches mostly replies on hand-crafted features, thus the performance is expected to be further improved by introducing DNN methods to learn more discriminative features automatically.

Recently, DNN based segmentation approaches, like M-Net [1], is shown to outperform segmentation and conventional machine learning based methods. However, like most of the existing approaches, this work is still a two-step approach. It performs coarse boundary detection first and then applies an ellipse fitting to generate a smooth ellipse shape boundary.

Actually, there are two typical methods for the optic cup (OC) segmentation, which do not need performing ellipse fitting and can be similarly introduced into OD segmentation. Xu *et al.* [2] proposed a sliding window approach to localize the optic cup directly with a bunch of rectangles. In [3], OC ellipse parameters are directly calculated from OD reconstruction. These work are based on a simple yet effective assumption that the optic disc and cup are in non-rotated ellipse shape, and this is well accepted by many ophthalmologists and researchers to simplify the OD/OC analysis procedure.

Inspired by [2], a natural question arises as if we could lend impressive object detection results of deep learning from computer vision community to investigate the challenging OD segmentation problem. Follow this basic concept, in this paper, we propose a simple yet effective one-step end-to-end method to segment optic disc from a retinal fundus image using deep object detection architectures.

II. RELATED WORK

A. Deep Object Detection Networks

With the resurgence of deep learning, computer vision community has significantly improved object detection results over a short period of time. Modern object detection systems can mainly be divided into two groups: one-stage detectors and two-stage detectors. OverFeat [4] was one of the pioneered modern one-stage object detector based on deep networks. More recent works like RetinaNet [5], have demonstrated their promising results. Generally, these approaches are applied over regularly sampled candidate object locations across an image. In contrast, two-stage detectors are based on a proposal-driven mechanism, where a classifier is applied to a sparse set of candidate object locations. Following the R-CNN work [6], recent progresses on two-stage detectors have focused on processing all regions with only one shared feature map, and on eliminating explicit region proposal methods by directly predicting the bounding boxes. Various extensions to this framework have been presented, *e.g.*, Faster R-CNN [7], and Mask-R-CNN [8].

B. Optic Disc Segmentation

As mentioned in the introduction, a plenty of research have been conducted for the OD segmentation. These methods can be roughly categorized into three types:

- **Traditional segmentation methods:** Pallawala *et al.* proposed an ellipse fitting and wavelet transform method [9] to detect the OD contour. Liu *et al.* applied a level-set [10] based approach to segment OD and OC. Similarly, in [11], active shape model was employed to detect the boundaries of OD and OC sequentially. Cheng *et al.* developed a template matching method [12] to segment OD with additional peripapillary atrophy elimination.
- **Conventional machine learning based methods with hand-crafted features:** Generally, these methods deal with classification issues, *i.e.*, identifying whether a pixel or small region (*e.g.* super-pixels) on, inside of or

[1]X. Sun, Y. Xu, W. Zhao, and T. You are with the Central Research Institute of Guangzhou Shiyuan Electronics Co., Ltd, 510530, Guangzhou, China. {sunxu, xuyanwu, zhaowei, youtianyuan}@cvte.com

[2]J. Liu is with Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences, 315201, China. jimmyliu@nimte.ac.cn

978-1-5386-3646-6/18/$31.00 ©2018 IEEE

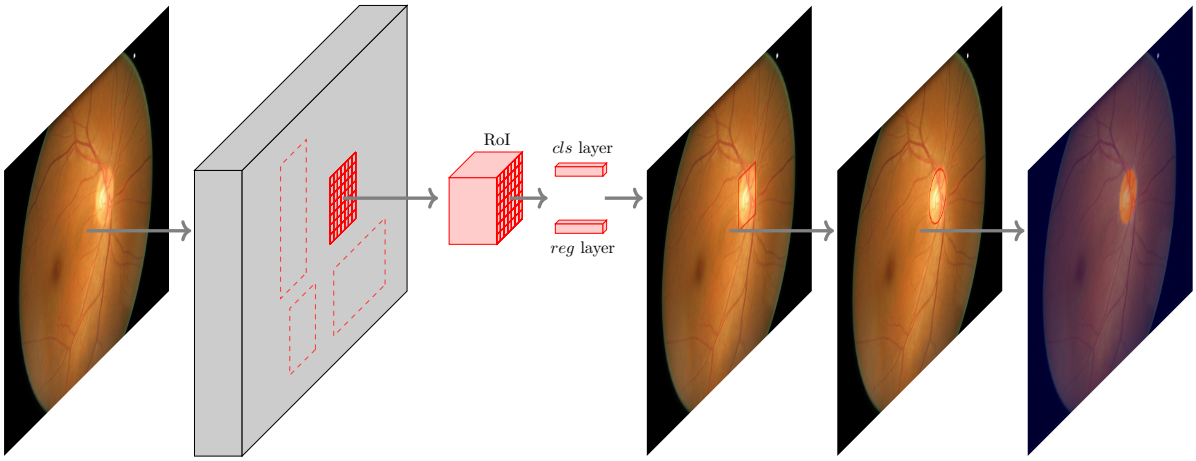

Fig. 1. Illustration of the proposed pipeline for OD segmentation from color fundus images, where red region denotes the detected OD.

outside of the OD boundary. In [13], the local texture features around each point of interest are utilized to classify this point is on the OD boundary or not. In [14], Cheng *et al.* utilizes contrast enhanced histogram and center surround statistics features to classify a super-pixel is within the OD or not. However, reliance on hand-crafted features make these methods susceptible to low quality images and pathological regions.

- **Deep neural networks based methods:** To our knowledge, very few DNN based approaches have been studied for OD segmentation/localization. However, this is a promising direction to develop reliable and efficient automated OD segmentation and computer-aided ocular disease diagnosis systems. The latest work of this kind is probably [1], in which a M-net deep architecture is proposed for OD boundary segmentation after polar transformation.

III. METHOD

In this paper, we propose a new pipeline for OD segmentation from retinal fundus images. The pipeline of our basic algorithm is shown in Fig. 1.

- We first feed the color fundus image into a deep object detection networks to get the most confident bounding box for OD.
- The parameters of this bounding box are then used to generate an ellipse-like boundary which well approximates the OD appearance.
- If necessary, segmentation masks can be obtained based on the generated boundary.

A. Network Architecture

In this paper, we adopt *Faster R-CNN* [7] as the object detector due to its flexibility and robustness. Faster R-CNN consists of two stages. The first stage, called region proposal network (RPN), processes images with a deep convolutional network *(e.g.*, VGG-16 [17]), and predicts a set of rectangular candidate object locations using features at some selected intermediate layer (*e.g.*, "conv5"). During training, the loss for this first stage is defined as

$$L = L_{cls} + L_{reg} \quad (1)$$

where L_{cls} and L_{reg} denote the classification loss and bounding box regression loss, respectively. We refer readers to [7] for the more details of these two quantities.

In the second stage, these (*e.g.*, 300) candidate bounding boxes are mapped to the same intermediate feature space, and then fed to the subsequent layers of the convolutional network (*e.g.*, "fc6" followed by "fc7") to output a class label and a bounding box offset for each proposal. The loss function for this second stage box classifier also takes the form of (1) using proposals produced from the RPN as anchors.

B. Segmentation Generation

The object detector predicts a probability score for each candidate bounding box in the input image. Non-maximum suppression (NMS) is often required to reduce redundancy. However, for OD detection, a retinal fundus image contains one and only one object. Therefore, NMS is no longer necessary in our pipeline since we only need to retain the bounding box with the highest confidence score.

Given the bounding box, the question now is how to generate a satisfactory OD boundary. Since the OD appears as a bright yellowish elliptical region in color fundus images, it is promising to use an ellipse to approximate the shape of OD. Furthermore, we also observe that localizing a bounding box requires exactly the same parameters as a vertical ellipse, *i.e.*, its width, height, and central point (horizontal and vertical coordinates). With this in mind, we propose to generate the OD boundary by simply redrawn the predicted bounding box as a vertical ellipse.

IV. EXPERIMENTS

A. Dataset and Evaluation Criteria

We use the ORIGA dataset [15] for our experiments, and evaluate the proposed disc segmentation method using the manual masks as *ground truth*. The ORIGA dataset contains 650 images with a uniform size of $3,072 \times 2,048$, where 168 images are with glaucoma and the other 482 images are normal. In every image, OD is labeled by ophthalmologists using a vertical and non-rotated ellipse. These images are

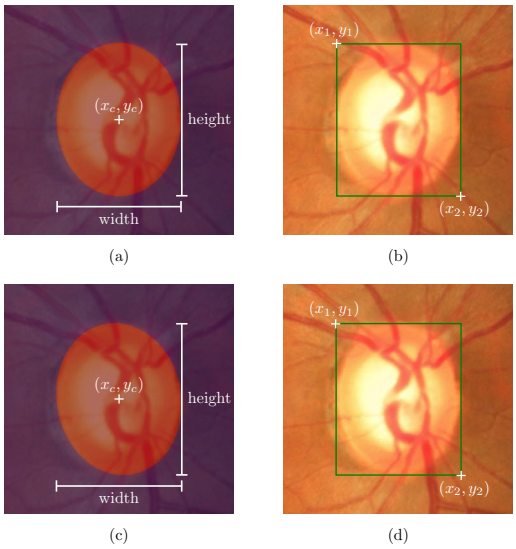

Fig. 2. One example from the training set to illustrate how we generate "ground truth" bounding box with the manual segmentation mask.

equally divided into 325 images for training and 325 images for testing.

Various evaluation criteria have been applied to OD detection/segmentation, among which the overlapping ratio is the most widely used one. Due to space limitation, the proposed technique is only evaluated based on the overlapping ratio here. Let R denote the overlapping ratio between the manually labeled OD region and the estimated one as follows:

$$R = \frac{A_{\text{GT}} \cap A_{\text{ET}}}{A_{\text{GT}} \cup A_{\text{ET}}} \tag{2}$$

where A_{GT} and A_{ET} denote the area of the image region enclosed by the reference OD boundary (*i.e.* ground truth) and the area of the image region enclosed by the OD boundary determined by a tested method, respectively.

B. Implementation Details

Preprocessing: The manual "ground truth" provided by the ORIGA dataset is the segmentation masks, while what we need to train the object detector is the bounding box for OD. Therefore, we have to preprocess the dataset to enable training of Faster R-CNN. As mentioned in the previous subsection, the OD are labeled by vertical ellipse. It means that we can determine the OD boundary by only four parameters, *i.e.*, its width w, height h and center location (x_c, y_c), as illustrated in Fig. 2 (a). These parameters, on their own, can then be easily converted into coordinates of the upper left corner point (x_1, y_1) and the lower right corner point (x_2, y_2) of the bounding box, which is exactly what we want. But we found that there is another simpler way to accomplish that which allows us to bypass the tedious processing of fitting an ellipse first and then converting it parameters. To enable the bounding box to tightly localize the labeled OD segmentation mask, we simply let x_1 and x_2 denote the minimum and maximum horizontal coordinates of OD mask, and similarly, y_1 and y_2 denote minimum and maximum vertical coordinates.

Data argumentation: The training set of ORIGA dataset contains only 325 images, which is insufficient to learn so many parameters of the deep neural networks without overfitting. The easiest and most common method to reduce overfitting on image data is to artificially enlarge the dataset using label-preserving transformations. We employ two ways to argument our data. The first way is to rotate images by a set of angles over -10(2)10 degrees, where the notation -10(2)10 stands for a list starting from -10 to 10 with an increment of 2. Since image rotation also changes the location of OD in it, we have to modify the "ground truth" bounding box accordingly. This can be easily accomplished using the same method as what we do for the original dataset, provided that the manual segmentation masks are also rotated (as illustrated in bottom row of Fig. 2). The second way to argument our data is to use horizontal reflection, on both the original training set and it rotated counterparts. This increases the size of our training set by a factor of 20.

TABLE I

OPTIC DISC SEGMENTATION PERFORMANCE COMPARISON OF DIFFERENT METHODS ON ORIGA DATASET.

Method	*R(%)*
MCV [13]	87.1
ASM [11]	88.7
EHT [12]	89.7
MDM [18]	89.2
SP+ASM [14]	90.5
SDM [19]	91.1
U-Net [20]	88.5
M-Net [1]	92.9
Proposed	**93.1**

Training details: The join-training scheme is adopted to train the Faster RCNN detection framework. The network is implemented using `Tensorflow` based on the publicly available code provided by Chen *et.al.* [16]. Several minor revisions are made in this implementation, which give potential improvements. Interested readers may refer to the technical report [16] for more details about the modifications. We trained and test the proposed method on a single-scale image using a single model. We rescale the images such that their shorter side is $s = 600$ pixels before feeding them to detector. VGG-16 [17] is used as the backbone of the Faster R-CNN and is initialized with an ImageNet pre-trained model. For anchors, we use 3 scales with box areas of 128^2, 256^2, and 512^2 pixels, and 3 aspect ratios of $1:1$, $1:2$, and $2:1$. The entire network is fine-tuned end-to-end with a training set of 7,150 images for 200,000 iterations (about 28 epochs) on a single NVIDIA TITAN XP GPU. The learning rate is set to 0.001 at the beginning of the training process and then changed to 0.0001 after 100,000 iterations.

Testing: The pixels inside the predicted bounding box are labeled as OD. This region is then used to calculate the overlapping ratio with the manual "ground truth" from ORIGA dataset.

C. Segmentation Results

In Table I, we compare the proposed method with the modified ChanVese method (MCV) [13], active shape model (ASM) [11], elliptical Hough transform (EHT) [12], modified deformable models (MDM) [18], superpixel-based method with ASM post-processing (SP+ASM) [14], supervised de-

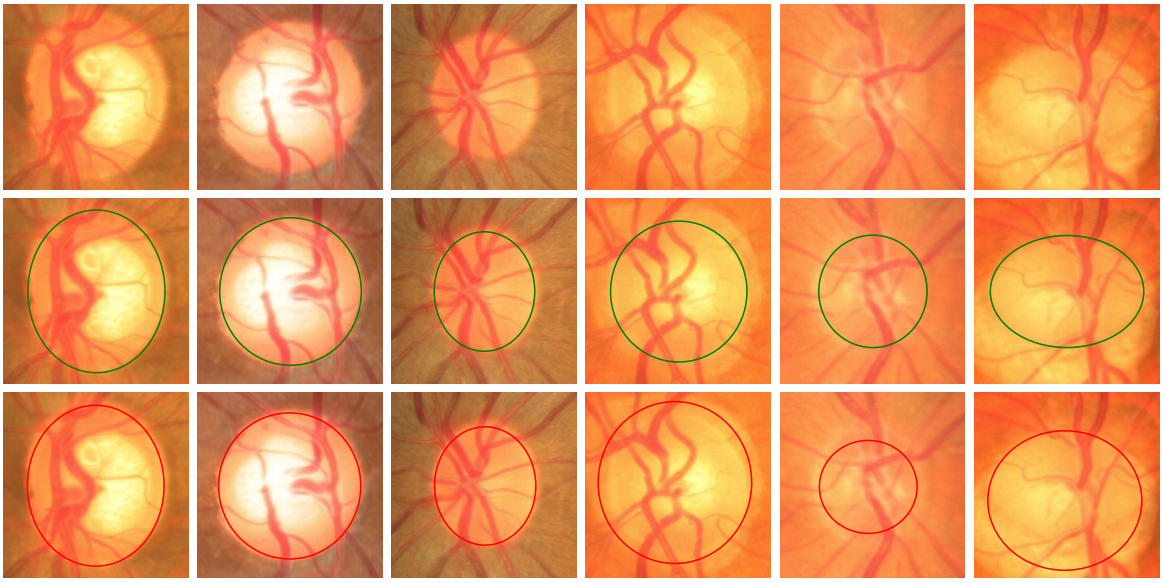

Fig. 3. Sample results. From top to bottom: the cropped original images, the manual "ground truth" and outlines by the proposed method. From left to right, the overlapping ratios by the proposed method are 98.23%, 98.20%, 98.06%, 77.01%, 75.00% and 73.68%, respectively.

scent method (SDM) [19], and two deep learning based methods, *i.e.*, U-Net [20] and M-Net [1]. As shown in Table I, our detection based method achieves state-of-the-art segmentation results on ORIGA dataset, with the average overlapping ratio of 93.1% for OD segmentation. Fig. 3 shows some visual examples of the segmentation results, where the first three columns are images from which the proposed method obtains the highest overlapping ratios and the rest columns are lowest ones.

V. CONCLUSION

In this paper, we redefine the optic disc segmentation problem as an object detection problem, and then propose a new pipeline to segment OD from retinal fundus images using Faster R-CNN as the object detector. Tested on the widely used and publicly available ORIGA dataset, our method outperforms existing methods, achieving state-of-the-art OD segmentation accuracy. In the future, we plan to investigate other deep object detectors and extend the application to optic cup segmentation as well.

REFERENCES

[1] H. Fu, J. Cheng, Y. Xu, D. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Trans. Medical Imaging*, 2018.

[2] Y. Xu, D. Xu, S. Lin, J. Liu, J. Cheng, C.Y. Cheung, T. Aung, and T.Y. Wong, "Sliding window and regression based cup detection in digital fundus images for glaucoma diagnosis," in *MICCAI*, 2011, vol. 6893, pp. 1–8.

[3] Y. Xu, S. Lin, D. Wong, J. Liu, and Dong Xu, "Efficient reconstruction-based optic cup localization for glaucoma screening," in *MICCAI*, Kensaku Mori, Ichiro Sakuma, Yoshinobu Sato, Christian Barillot, and Nassir Navab, Eds., 2013, pp. 445–452.

[4] P. Sermanet, X. Eigen, D.and Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," in *ICLR*, 2014.

[5] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *ICCV*, 2017, pp. 2980–2988.

[6] R. Girshick, J. Donahue, Tr. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *CVPR*, 2014, pp. 580–587.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *NIPS*, 2015, pp. 91–99.

[8] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *ICCV*. IEEE, 2017, pp. 2980–2988.

[9] P. Pallawala, W. Hsu, M. Lee, and K. Eong, "Automated optic disc localization and contour detection using ellipse fitting and wavelet transform," in *ECCV*, 2004.

[10] J. Liu, D. Wong, J. Lim, H. Li, N. Tan, Z. Zhang, T. Wong, and R. Lavanya, "Argali: an automatic cup-to-disc ratio measurement system for glaucoma analysis using level-set image processing," in *EMBC*, 2008.

[11] F. Yin, J. Liu, S. Ong, Y. Sun, D. Wong, N. Tan, C. Cheung, M. Baskaran, T. Aung, and T. Wong, "Model-based optic nerve head segmentation on retinal fundus images," in *EMBC*. IEEE, 2011, pp. 2626–2629.

[12] J. Cheng, J. Liu, D. Wong, F. Yin, C. Cheung, M. Baskaran, T. Aung, and T. Wong, "Automatic optic disc segmentation with peripapillary atrophy elimination," in *EMBC*. IEEE, 2011, pp. 6224–6227.

[13] G. Joshi, J. Sivaswamy, and S. Krishnadas, "Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment," *IEEE Trans. medical imaging*, vol. 30, no. 6, pp. 1192–1205, 2011.

[14] J. Cheng, J. Liu, Y. Xu, F. Yin, D. 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. Medical Imaging*, vol. 32, no. 6, pp. 1019–1032, 2013.

[15] Z. Zhang, F Yin, J. Liu, W. Wong, N. Tan, B. Lee, J. Cheng, and T. Wong, "Origa-light: An online retinal fundus image database for glaucoma analysis and research," in *EMBC*. IEEE, 2010, pp. 3065–3068.

[16] X. Chen and A. Gupta, "An implementation of faster rcnn with study for region sampling," *arXiv preprint arXiv:1702.02138*, 2017.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015.

[18] J. Xu, O. Chutatape, E. Sung, C. Zheng, and P. Kuan, "Optic disk feature extraction via modified deformable model technique for glaucoma analysis," *Pattern recognition*, vol. 40, no. 7, pp. 2063–2076, 2007.

[19] A. Li, Z. Niu, J. Cheng, F. Yin, D. Wong, S. Yan, and J. Liu, "Learning supervised descent directions for optic disc segmentation," *Neurocomputing*, vol. 275, pp. 350–357, 2018.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.