2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)
Venice, Italy, April 8-11, 2019

ELLIPSE DETECTION OF OPTIC DISC-AND-CUP BOUNDARY IN FUNDUS IMAGES

Zeya Wang, Nanqing Dong, Sean D Rosario, Min Xu, Pengtao Xie, Eric P. Xing

Petuum Inc., Pittsburgh, PA 15222, USA

ABSTRACT

Glaucoma is an eye disease that damages the optic nerve and leads to loss of vision. The diagnosis of glaucoma involves measurement of cup-to-disc ratio from retinal fundus images, which necessitates the detection of the optic disc-and-cup boundary as a crucial task for glaucoma screening. Most existing computer-aided diagnosis (CAD) systems focus on the segmentation approaches but ignore the localization approaches, which requires less human annotation cost. In this paper, we propose a deep learning-based framework to jointly localize the ellipse for the optic disc (OD) and optic cup (OC) regions. Instead of detecting a bounding box like in most object detection approaches, we directly estimate the parameters of an ellipse that suffices to capture the morphology of each OD and OC region for calculating the cup-to-disc ratio. We use two modules to detect the ellipses for OD and OC regions, where the OD region serves as attention to the OC region. The proposed framework achieves competitive results against the state-of-the-art segmentation methods with less supervision. We empirically evaluate our framework with the recent state-of-the-art segmentation models on two scenarios where the training data and test data come from the same and different domains.

Index Terms— Ellipse Detection, Optic Disc-and-Cup Boundary, Deep Learning

1. INTRODUCTION

Glaucoma is one of the leading causes of irreversible blindness in the world that brings about damage to the optic nerve. Early screening can help patients receive better treatment options. One common method to identify glaucoma is through analysis of the optical nerve head (ONH). Optic nerve examination requires localization of the optic disc-and-cup and their borders [1]. An increase in the size of the optic cup may indicate a presence of glaucoma. The cup-to-disc ratio (CDR), which compares the size of the cup to that of the disc on fundus images, is a significant indicator of glaucoma [2]. A large CDR may be indicative of glaucoma, and CDR upwards of 0.65 is considered glaucomatous [3]. The segmentation of optic disc-and-cup can be used to estimate the CDR. The full segmentation for skilled graders takes on average eight minutes per eye [4], which makes a fast and accurate CAD system desirable for large-scale clinical diagnosis. Traditional knowledge-based segmentation methods rely on the selection of hand-crafted features, while the selected features can be biased during the feature engineering process [1]. With the success of deep learning, convolutional neural networks (CNNs) have played an important role in glaucoma screening [1, 4, 5].

For optic disc-and-cup segmentation, Fu et al. propose M-Net with a polar transformation (PT) to ensemble multi-scale information, which is robust in the test phase and achieves the state-of-the-art performance [1]. The segmentation masks predicted by these methods will be post-processed through an ellipse fitting procedure to generate two ellipses for OD and OC regions [1]. Given the oval shape of OD and OC, CDR can be calculated for glaucoma screening based on the mathematical diameters. Considering that the final objective is to have an accurate estimate of the diameters for both OD and OC, instead of acquiring a segmentation mask, it is intuitive to find an alternative solution to bypass the segmentation step.

Overwhelmed by the recent segmentation-based methods, the regression-based approach is less developed. We also notice that there is few regression-based method which can handle non-rectangle objects. Inspired by the recent success of region proposal networks (RPN) [6], we propose a novel framework that leverages the geometric property of OD and OC. Instead of detecting a bounding box in the conventional object detection methods, we are motivated to design a network to detect rotated ellipses. It is worth mentioning that our method can be viewed as counterpart to the traditional methods, which usually consist of both segmentation and ellipse fitting steps. The contribution is twofold. First, we reformulate the OD and OC segmentation problem from a perspective of ellipse detection. We propose a region proposal-based framework to jointly learn the ellipse parameters for OD and OC regions. Given the OC region is contained in OD, we design two ellipse proposal networks to respectively detect OD and OC. The detection of OC is regularized by an attention module from the detection of OD, which mimics a zoom-in process from OD to OC. Second, we empirically evaluate the proposed method and the recent state-of-the-art segmentation models on two public datasets, the REFUGE dataset [1] and Drishti-GS1 dataset [2] [7, 8].

[1] https://refuge.grand-challenge.org

[2] Provided by Medical Image Processing (MIP) group, IIIT Hyderabad.

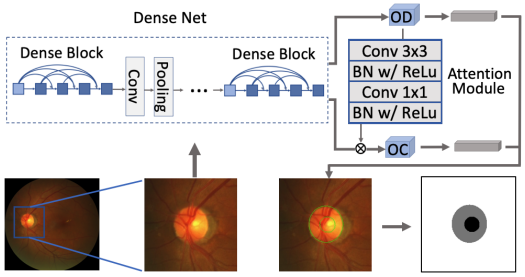

Fig. 1. Illustration of our ellipse detection framework [9].

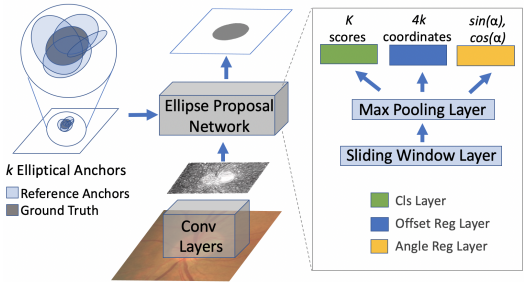

Fig. 2. Ellipse Proposal Network. The elliptical anchors are generated in different ratios and scales.

2. METHODS

Our overall flowchart for detecting OD and OC regions consists of a step for cropping rough OD regions, and our ellipse detection network (see Figure 1). In order to save computational costs for detection network, we first extract rough OD regions based on a simple automatic detection method described in [1] to occlude most of the fundus surface with no annotated information. The regions are cropped and resized to a small size (e.g. 224×224) and fed to our ellipse detection network. Our network shares a similar region proposal network as in baseline object detection systems, such as Faster R-CNN [6]. However, our detection network replaces the bounding boxes and rectangular anchors (see details about anchor in [6]) used in the conventional object detection tasks with ellipse hoops and elliptical anchors, to solve our reformulated problem. We also incorporate an attention mechanism to tailor our network to our target.

2.1. Network Architecture

As demonstrated in Figure 2, the cropped input images are first fed into a DenseNet-based feature extractor [9], which is truncated at the last convolution layers to generate shared feature maps. The feature extractor is followed by two different branches of ellipse proposal networks respectively, for OD and OC detection. Each proposal network will generate probability scores and offsets of the true ellipse for a set of default anchors, as well as two logits for the rotation angle. The use of two branches for detecting OD and OC can help us model each detection task individually, thereby having a weighted loss for unbalanced positive and negative anchors. In the inference phase, we will decode the estimates according to the anchor with the highest likelihood and output the decoded shape and position parameters with the estimated angles.

2.2. Ellipse Proposal Network

The RPN is first proposed in Faster R-CNN [6] to generate rectangular object proposals through associating a set of default rectangular anchor boxes. The benefit of using anchors is to predict the offsets relative to the default boxes to which the ground truth bounding boxes are anchored, instead of directly predicting the coordinates. Our ellipse proposal network (EPN) is similar to RPN but is designed for detecting elliptical regions. An ellipse can be parameterized with center coordinate (X_0, Y_0), major axis $(F1)$, minor axis $(F2)$ and rotation angle $(\alpha \in [0, \pi))$. To estimate $\{X_0, Y_0, F1, F2\}$, we generate k elliptical anchors with parameters $\{X_0', Y_0', F1', F2'\}$ and consider the estimation of α separately. Without setting angle in the default anchors, our k anchors are actually assumed to be aligned with any elliptical region that will be encoded (i.e. share the same α with encoded regions). When calculating the intersection of truth region and each anchor, we map the coordinates of these ellipses with $(x, y) \Rightarrow (x\cos\alpha + y\sin\alpha, -x\sin\alpha + y\cos\alpha)$ and then approximate it by calculating the intersection of the bounding boxes that surround these ellipses. Given that the input image to EPN has been preprocessed to encompass the rough OD regions, we generate k anchors which are centered close to the center of the input rough regions with specific scales and ratios of $F1$ to $F2$, thereby reducing the high computational cost of encoding and decoding work for anchors with low overlap with the disc and cup regions.

Our EPN module also starts with a similar design of 3×3 spatial sliding window (i.e. 3×3 convolutional layers) over the 7×7 feature maps (given input size is 224×224) after the last shared convolution layers. Given that we only detect one region in each EPN, we add a max pooling layer to capture the strongest activation, which are followed by two sibling FC layers: an ellipse-classification layer and an ellipse-regression layer. For the two branches of EPNs for detecting OD and OC regions, we integrate an attention mechanism with 3×3 and 1×1 convolutional layers with batch normalization from OD to OC which mimics a zoom-in process from the OD region to its surrounded OC region.

2.3. Training Objective

Our training objective is extended from the multi-task loss in object detection network. Our loss function for images with a

batch size B is:

$$L = \frac{1}{N_{cls}^{OD}} L_{cls}^{OD}(s,p,w) + \lambda_1 \frac{1}{N_{reg}^{OD}} L_{reg}^{OD}(s,g,t) + \lambda_2 \Delta^{OD}(\alpha,\hat{\alpha}) + \gamma\{\frac{1}{N_{cls}^{OC}} L_{cls}^{OC}(s,p,w) + \lambda_1 \frac{1}{N_{reg}^{OC}} L_{reg}^{OC}(s,g,t) + \lambda_2 \Delta^{OC}(\alpha,\hat{\alpha})\} \quad (1)$$

where N_{cls} is the number of anchors and N_{reg} is the number of positive anchors (e.g., $\frac{|A \cap G|}{|A \cup G|} > 0.8$ for OD and 0.6 for OC, where $|A|$ and $|G|$ are the areas of the bounding boxes that encompass anchors and ground truth ellipse) in a batch. λ_1 and λ_2 are balancing parameters to weight between L_{cls}, L_{reg} and Δ. γ is used to weight between the loss of OD and OC. $L_{cls}(s,p)$ is a weighted cross entropy loss over positive anchor versus negative anchor, so $s \in \{0,1\}$ is an indicator of positive anchors and p is the estimated confidence for anchors. Considering that positive anchors are often less than negative ones, the errors for the positive anchors are up-weighted by w, which can be the inverse of ratio of positive anchor to the total number of anchors. $L_{reg}(s,g,t)$ is a regression loss for positive anchors. $L_{reg}(s,g,t) = \sum_{s_{ij}=1} H(\hat{g}_{ij}, \hat{t}_{ij})$. $H(\cdot)$ is a $L1$ smooth loss between the estimates $\hat{t}$ and encoded truth $\hat{g}$. i and j is the sample index and anchor index. $\hat{g}_{ij}$ is encoded from the parameters of the ground truth g_i for positive anchor j, and for inference $\hat{t}_{ij}$ with the highest likelihood will be decoded to get the estimated ellipse parameters t_i (see [6] for more details). Δ is an error function for rotation angle. Since the rotation angle α is defined in $[0,\pi)$, we first transform α to α' such that $\alpha' \in [-\frac{\pi}{2}, \frac{\pi}{2})$. α' will be encoded by $\{\sin\alpha', \cos\alpha'\}$. Given $\sin\alpha' \in [-1,1)$ and $\cos\alpha' \in [0,1)$, we respectively compute the hyperbolic tangent value u and sigmoid value v of the two logits generated in the last layer of EPN for angle. Δ is given by calculating the $L1$ smooth loss between $\{\sin\alpha', \cos\alpha'\}$ and $\{u,v\}$. For inference, we will use $\hat{\alpha} = \arctan\frac{u}{v} + \frac{\pi}{2}$ as the estimate of α.

3. EXPERIMENTS AND RESULTS

3.1. Dataset

The REFUGE dataset contains the 400 retinal fundus images, which are captured with the Zeiss Visucam 500 fundus camera at a resolution of 2124×2056 pixels. The pixel-wise disc-and-cup gray-scale annotations are provided. The DrishtiGS1 dataset consists of a total of 101 images which were collected at Aravind eye hospital, Madurai, India. All images were taken centered on OD with a field-of-view of 30 degrees and of dimensions 2896×1944 pixels. The images have been marked by 4 eye experts with varying clinical experience, so the pixel-wise labels are provided.

Method	I_{Disc}	D_{Disc}	I_{Cup}	D_{Cup}	δ_{CDR}
SP [3] [11]	0.756	0.861	0.385	0.556	0.181
M-Net + PT [4] [1]	0.913	0.954	0.770	0.870	0.048
Ellipse Detection	0.911	0.953	0.773	0.872	0.047

Table 1. Performance comparison under supervised setting.

Method	Test Time (Unit: 0.01 Seconds)
M-Net + PT + EF	22.6
Ellipse Detection	2.3

Table 2. Computational cost comparison.

3.2. Supervised Learning

In our experiment on supervised learning, we augment training images with these following steps: 1) rotation by $\{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$; 2) flipping; 3) converting the RGB image to LAB format, performing a contrast-limited adaptive histogram equalization (CLAHE) on the lightness component with different thresholds for contrast limiting, and then converting the images back to RGB for enhancing the local contrast [10]. Before training, we perform a mean subtraction to subtract the average value from the R, G and B channels separately.

The performance of our method is evaluated for supervised OD and OC localization. We randomly split the REFUGE dataset into 280 training images, 60 validation images and 60 test images. We report the model performance on the test set in terms of Intersection-Over-Union (IOU) score I and Dice coefficient D between the true elliptical regions and estimated regions. We also calculate the CDR by taking square root of the cup to disc area ratio with the fitted ellipses respectively for the true masks and detection results [2]. We report the mean absolute error δ between estimated CDR and true CDR. We compare the proposed approach with superpixel classification-based segmentation (SP) [11], and M-Net with PT [1]. All the performances are given in Table 1.

Our method gives higher IOU and Dice scores in comparison to SP and close scores with M-Net for both OD and OC localization. With accurate estimation of these elliptical regions, our method can generate CDR close to that between truth OD and OC regions. The ellipse detection method can provide competitive performance with cutting-edge segmentation methods.

We also report the test time of the proposed method and M-Net with PT and ellipse fitting (EF) under the same setting. The test environment is TensorFlow 1.3.0 on top of Intel® Core™ i7-6800K CPU @ 3.40GHz × 12 and NIVIDIA GeForce®GTX 1080 Ti graphic card. We report the single image processing time over an average of 60 test images in Table 2. Our framework is efficient at test time, as it can di-

[3] Reimplementation. No data augmentation.
[4] https://github.com/HzFu/MNet_DeepCDR

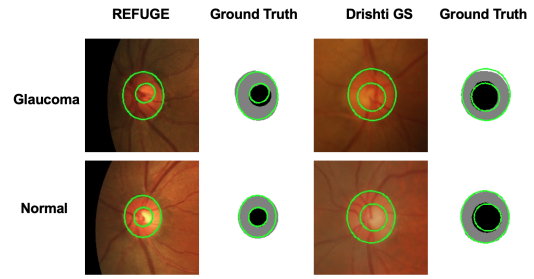

Fig. 3. The visualization of ellipses detection (marked with green contours) for optic disc-and-cup boundary.

Method	I_{Disc}	D_{Disc}	I_{Cup}	D_{Cup}	δ_{CDR}
M-Net + PT	0.882	0.937	0.621	0.766	0.166
Ellipse Detection	0.864	0.927	0.635	0.777	0.159

Table 3. Performance comparison for transfer learning.

rectly estimate the parameters of the elliptical OD and OC regions, thereby bypassing the steps of polar transformation and ellipse fitting.

3.3. Transfer Learning

We use the whole images in REFUGE dataset as the training set and test the trained model on the Drishti-GS1 dataset. Given the scales of the cropped region for these two datasets are different, for data augmentation in the training process we replace the rotation and flipping with a step of re-scaling the training images to different scales. We evaluate the test performance in terms of the same metrics. All the performances are given in Table 3. The proposed method can achieve similarly robust performance as the state-of-the-art segmentation method, when the target domain is different from the source domain. The visual examples are provided in Figure 3.

4. DISCUSSION

We present a robust and computationally efficient deep learning framework that is able to directly fit ellipses for OD and OC regions, which shows performance competitive with the state-of-the-art. In our future work, we will further improve our model by integrating novel features, (e.g. domain adaption [12]) and optimize the architecture from our current design.

5. REFERENCES

[1] 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 TMI*, 2018.

[2] S.B. Syc, C.V. Warner, S. Saidha, S.K. Farrell, A. Conger, E.R. Bisker, J. Wilson, T.C. Frohman, E.M. Frohman, L.J. Balcer, et al., "Cup to disc ratio by optical coherence tomography is abnormal in multiple sclerosis," *Journal of the neurological sciences*, vol. 302, no. 1-2, pp. 19–24, 2011.

[3] M.U. Akram, A. Tariq, S. Khalid, M.Y. Javed, S. Abbas, and U.U. Yasin, "Glaucoma detection using novel optic disc localization, hybrid feature set and classification techniques," *Australasian Physical & Engineering Sciences in Medicine*, vol. 38, no. 4, pp. 643–655, 2015.

[4] G. Lim, Y. Cheng, W. Hsu, and M.L. Lee, "Integrated optic disc and cup segmentation with deep learning," in *ICTAI*. IEEE, 2015, pp. 162–169.

[5] H. Fu, J. Cheng, Y. Xu, C. Zhang, D.W.K. Wong, J. Liu, and X. Cao, "Disc-aware ensemble network for glaucoma screening from fundus image," *IEEE TMI*, 2018.

[6] 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.

[7] J. Sivaswamy, S. Krishnadas, A. Chakravarty, G.D. Joshi, and A.S. Tabish, "A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis," *JSM Biomedical Imaging Data Papers*, vol. 2, no. 1, pp. 1004, 2015.

[8] J. Sivaswamy, S.R. Krishnadas, G.D. Joshi, M. Jain, and A.U.S. Tabish, "Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation," in *ISBI*. IEEE, 2014, pp. 53–56.

[9] G. Huang, Z. Liu, L. Van Der Maaten, and K.Q. Weinberger, "Densely connected convolutional networks," in *CVPR*. IEEE, 2017, pp. 4700–4708.

[10] Z. Wang, N. Dong, W. Dai, S.D. Rosario, and E.P. Xing, "Classification of breast cancer histopathological images using convolutional neural networks with hierarchical loss and global pooling," in *ICIAR*. Springer, 2018, pp. 745–753.

[11] J. Cheng, J. Liu, Y. Xu, F. Yin, D.W.K. Wong, N.M. Tan, D. Tao, C.Y. Cheng, T. Aung, and T.Y. Wong, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE TMI*, vol. 32, no. 6, pp. 1019–1032, 2013.

[12] N. Dong, M. Kampffmeyer, X. Liang, Z. Wang, W. Dai, and E. Xing, "Unsupervised domain adaptation for automatic estimation of cardiothoracic ratio," in *MICCAI*. Springer, 2018, pp. 544–552.