PM-Net: Pyramid Multi-label Network for Joint Optic Disc and Cup Segmentation

Pengshuai Yin[1], Qingyao Wu[1], Yanwu Xu[4(✉)], Huaqing Min[1], Ming Yang[2], Yubing Zhang[2], and Mingkui Tan[1,3(✉)]

[1] South China University of Technology, Guangzhou, China
mingkuitan@scut.edu.cn
[2] Guangzhou Shiyuan Electronic Technology Company Limited, Guangzhou, China
[3] Peng Cheng Laboratory, Shenzhen, China
[4] Cixi Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China
ywxu@ieee.org

Abstract. Accurate segmentation of optic disc (OD) and optic cup (OC) is a fundamental task for fundus image analysis. Most existing methods focus on segmenting OD and OC inside the optic nerve head (ONH) area but paying little attention to accurate ONH localization. In this paper, we propose a Mask-RCNN based paradigm to localize ONH and jointly segment OD and OC in a whole fundus image. However, directly using Mask-RCNN faces some critical issues: First, for some glaucoma cases, the highly overlapping of OD and OC may lead to the missing of OC proposals. Second, some proposals may not fully surround the object, and thus the segmentation can be incomplete. Last, the instance head in Mask-RCNN cannot well incorporate the prior such as the OC is inside the OD. To address these issues, we first propose a segmentation based region proposal network (RPN) to improve the accuracy of proposals and then propose a pyramid RoIAlign module to aggregate the multi-level information to get a better feature representation. Furthermore, we employ a multi-label head strategy to incorporate the prior for better performance. Extensive experiments verify our method.

Keywords: Medical image process · Fundus image · Optic disc · Segmentation

1 Introduction

Fundus images assist doctors to diagnose many eye diseases such as glaucoma, which is one of the leading causes of blindness. The early detection and treatment for glaucoma often protect the eyes against serious vision loss. Clinically, the

P. Yin and Q. Wu—Equally contribution to this work.

D. Shen et al. (Eds.): MICCAI 2019, LNCS 11764, pp. 129–137, 2019.
https://doi.org/10.1007/978-3-030-32239-7_15

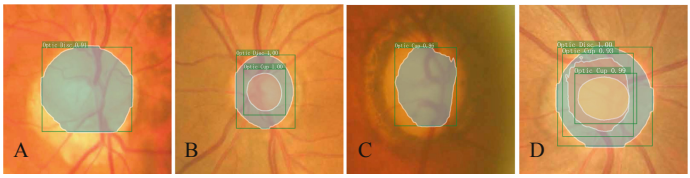

Fig. 1. Issues when adopting Mask-RCNN framework for OD and OC segmentation. (A): No proposals for OC and the proposal does not fully surround the OD. (B): The proposal does not fully surround the OD but the objectness score of this proposal is 1. (C): No proposals for OD. (D): Multiple instances of OC.

vertical cup to disc ratio (CDR) is a popular optic nerve head (ONH) assessment that is widely adopted by trained glaucoma specialists to screen glaucoma. The CDR is the comparison of the diameter of the cup to disc [11]. A larger CDR may indicate glaucoma or other diseases such as neuro-ophthalmic diseases. Accurate optic disc and cup segmentation are essential for CDR measurement. However, manual CDR assessment is time-consuming and costly, so an automatic glaucoma screening method is necessary.

OD and OC segmentation is a fundamental task in fundus image analysis [1]. Initially, many methods are based on hand-craft features, such as template based methods [2,5], deformable based methods [10,15,21], pixel classification based methods [16,20], label transfer based methods [18] and superpixel based methods [6,7,19]. However, these methods are limited in performance and can be easily affected by pathological regions. Recently, deep learning based methods show promising performance on OD and OC segmentation [3,4,8,9,23]. Most methods employ a two-step paradigm: first locate ONH area, and then segment OD and OC within the ONH area to avoid the influence from other fundus regions. In practice, accurate ONH localization is essential for accurate OD and OC segmentation. However, most methods focus on the second step and pay little attention to the accurate ONH localization.

In this paper, we propose a Mask-RCNN based method to jointly localize ONH and segment OD and OC in a whole fundus image. Unfortunately, directly using Mask-RCNN for OD and OC segmentation [14] may suffer from several issues, as shown in Fig. 1. Essentially, the performance of Mask-RCNN highly depends on the accuracy and compactness of bounding boxes. However, the proposals generated from the region proposal network may not completely enclose the object. More critically, in some glaucoma cases, the highly overlapping of OD and OC leads to the missing of OC proposals. Moreover, the instance segmentation may produce multiple instances of OC, but each fundus image in fact has only one OD and OC. Last but not least, the instance head is hard to model the prior that the OD contains the OC.

To solve these issues, we improve Mask-RCNN framework in three aspects. First, we introduce a segmentation branch on the region proposal network (RPN) to segment the ONH area from the whole image. Second, we propose a pyramid

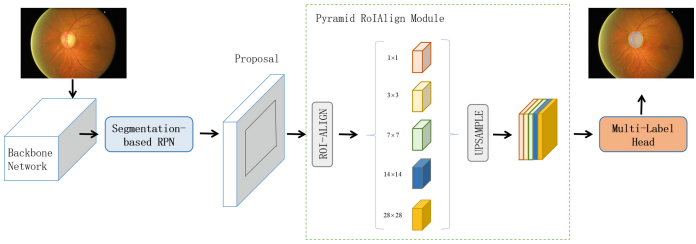

Fig. 2. The flowchart of our proposed network. The whole network can be trained end-to-end. A fundus image is fed into a convolutional neural network to extract features for segmentation based region proposal network to localize the ONH area. The pyramid RoIAlign module is developed to aggregate multi-level context information in proposals. Last, a multi-label segmentation head is used to jointly segment OD and OC.

RoIAlign module to aggregate multi-level information. The pyramid RoIAlign module helps to incorporate global features and learn stronger feature representations. Last, we employ a multi-label segmentation head instead of the instance segmentation head to better incorporate the prior that the OD is inside the OC.

The main contributions of this paper are listed as follows:

- We propose a segmentation-based RPN to generate more accurate and complete proposals for localizing ONH area.
- We propose a pyramid RoIAlign module to aggregate multi-level context information within proposals, which makes the final prediction more reliable.
- We propose a multi-label head to segment OD and OC jointly by better modeling the relation of OD and OC.

2 Proposed Method

The flowchart of our pyramid multi-label network (PM-Net) is shown in Fig. 2. A fundus image is first fed into a backbone network to extract features. Here, to increase the accuracy of proposals, we propose a **segmentation-based RPN** to segment the ONH area and produce proposals from the segmentation. Which helps to avoid proposal missing and accurately localize the ONH area based on the extracted features. We then propose a pyramid RoIAlign module to aggregate multi-level information in proposals to learn stronger representations. Last, we employ a multi-label head to jointly segment OD and OC by considering the relations between OD and OC. The whole network can be trained end-to-end.

2.1 Segmentation-Based Region Proposal Network

The RPN takes an image as input and produces rectangle proposals for OC and OD, each with an objectness score [12]. Unfortunately, as shown in Fig. 1, the rectangle proposals commonly do not fully surround the object. More seriously, the anchor based method may miss the proposals for OD or OC. Especially

for some glaucoma cases, the proposals of OD is highly overlapped with the proposals of OC. Therefore, the non-maximum suppression (NMS) filters out some OC proposals. RPN also may detect multiple proposals due to the similarity appearance of OC and OD.

To solve these issues, we add a sibling segmentation branch (seg) in parallel with classification branch (cls) and bounding box regression branch (reg) in the RPN. The seg branch segments the rough ONH area and generates proposals from the bounding box of the segmentation. For each image, the segmentation branch generates only one proposal for the ONH area and does not generate proposals for OC. Therefore, our segmentation-based RPN avoids generating multiple instances of an object and we enlarge the proposal by 20 pixels in order to let the proposal completely enclose the ONH area.

Our RPN is implemented with a 3×3 convolution followed by three sibling 1×1 convolution layers (for cls, reg and seg respectively). The seg branch segments the ONH area from the whole image and it is trained simultaneously with cls and reg branches. The segmentation-based region proposal network is trained by minimizing the objective function following the multi-task loss defined in [12]:

$$Loss_{segrpn} = Loss(\{p_i\}, \{t_i\}) + Loss_{seg}. \tag{1}$$

$$Loss(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*). \tag{2}$$

$$Loss_{seg} = g_i log c_i + (1 - g_i) log(1 - c_i). \tag{3}$$

Here, i is the index of an anchor, $L_{cls}(p_i, p_i^*)$ is the log loss between the predicted probability p_i of anchor i and the corresponding ground-truth label p_i^*. $L_{reg}(t_i, t_i^*) = R(t_i - t_i^*)$ where R is the smooth L_1 loss, t_i is a vector representing the 4 parameterized coordinates of the predicted bounding box and t_i^* is the corresponding ground-truth bounding box. $loss_{seg}$ is a binary cross-entropy loss between ground-truth class g_i at pixel i and the corresponding prediction c_i.

During the training, we use both anchor and segmentation based proposals to train the mask branch. In the testing phase, we use only segmentation based proposals.

2.2 Pyramid RoIAlign Module

The network tends to misclassify OC and OD due to the similar appearance of these objects. Motivated by [22], we solve this issue by incorporating suitable global features. We extend the RoIAlign layer to pyramid RoIAlign module. RoIAlign layer converts the feature inside any valid region into a feature map with a fixed spatial content of $H \times W$ by using bilinear interpolation to avoid quantization. Different output spatial sizes of the RoIAlign layer represent different level of context information. Small output spatial sizes represent the global context. Our pyramid RoIAlign module is a five-level RoiAlign layer with output spatial sizes of $\{1 \times 1\}$, $\{3 \times 3\}$, $\{7 \times 7\}$, $\{14 \times 14\}$ and $\{28 \times 28\}$ respectively. These output features are upsampled to $\{28 \times 28\}$ and concatenated together to form

the final feature representation, which carries different level context information. The local and global context together make the final prediction more reliable. The representation is sent to multi-label head to predict the final segmentation.

2.3 Multi-label Head for Joint OD and OC Segmentation

Note that the OD contains OC so the pixels within OC has the same labels to OD. The instance head assigns one label to each class and predicts a binary mask. Instead, multi-label head learns an independent binary classifier for each class and assigns each pixel to multiple binary labels. The multi-label head can better use the relative relation of OD and OC. Moreover, for some glaucoma cases, the OC occupy the most area of OD. Using instance head leads to imbalance pixel number for OD and OC. The multi-label head solves the imbalance problem since the classifier is independent for OD and OC. For the above reasons, we treat the OD and OC segmentation problem as a multi-label problem. Our multi-label head divide this multi-label problem into two binary classification problem with single label: {OD,¬OD}, {OC,¬OC} (¬ represents negative examples), our multi-label loss is defined as:

$$Loss_m = -\frac{1}{N}\sum_{n=1}^{N}[g_{n,i}logp_{n,i} + (1 - g_{n,i})log(1 - p_{n,i})]. \tag{4}$$

Here, N is the class number. $p_{n,i}$ represents the predicted probability when assign pixel i to class n. $g_{n,i}$ represents the ground truth label for pixel i.

3 Experiments

We compare our method with several state-of-the-art methods, including R-Bend [10], ASM [21], LRR [17], U-Net [13], Superpixel [7], M-Net [8] and Faster-RCNN based method [14].

Datasets. We test our method on two datasets: ORIGA and REFUGE. ORIGA has 650 images with 168 glaucomatous eyes and 482 normal eyes. Following [8], we use 325 images for training (including 73 glaucoma cases) and 325 images for testing (including 95 glaucoma cases). For REFUGE, we use 400 images for training and 400 images for validation.

Implementation Details. We train our model with 15000 iterations using the initial learning rate 0.002. Then, we decay the learning rate to 0.0002 and fine-tune the model with another 15000 iterations. For architecture, we adopt the ResNet-50 with feature pyramid network (FPN) as the backbone and pretrain the model on MS COCO dataset to avoid overfitting.

Evaluation Metrics. We adopt the overlapping error (E) and balanced accuracy (A) as the evaluation metrix for OD, OC and rim regions:

$$OE = 1 - \frac{Area(S \bigcap G)}{Area(S \bigcup G)}, A = \frac{1}{2}(Sen + Spe), \tag{5}$$

with $Sen = \frac{TP}{TP+FN}$, and $Spe = \frac{TN}{TN+FP}$. Here, S and G denote the predicted mask and corresponding ground-truth. TP and FP denote true and false positives, respectivel. TN and FN denote true and false negatives, respectively. Moreover, we also calculate CDR and adopt absolute CDR error δ_E as an evaluation metric: $\delta_E = |CDR_S - CDR_G|$, where CDR_G is the ground-truth CDR from trained clinician and CDR_S is calculated on segmentation result.

Table 1. Performance evaluation on ORIGA dataset.

Method	E_{disc}	E_{cup}	E_{rim}	δ_E	A_{disc}	A_{cup}	A_{rim}
R-Bend [10]	0.129	0.395	-	0.154	-	-	-
ASM [21]	0.148	0.313	-	0.107	-	-	-
LRR [17]	-	0.244	-	0.078	-	-	-
U-Net [13]	0.115	0.287	0.303	0.102	0.959	0.901	0.921
Superpixel [7]	0.102	0.264	0.299	0.077	0.964	0.918	0.905
M-Net [8]	0.083	0.256	0.265	0.078	0.972	0.914	0.921
M-Net+PT [8]	0.071	0.230	0.233	0.071	0.983	0.930	0.941
Sun's [14]	0.069	0.213	-	0.067	-	-	
Mask-RCNN(Baseline)	0.074	0.231	0.260	0.079	0.985	0.941	0.929
PM-Net(Ours)	**0.066**	**0.208**	**0.224**	**0.065**	**0.986**	**0.942**	**0.949**

3.1 Comparison with State-of-the-arts

We compare our method with several baselines on ORIGA dataset and record results in Table 1. Our method does not rely on post-processing such as ellipse fitting. From Table 1, our PM-Net method outperforms all other methods. There are several reasons accounting for this. First, our segmentation based region proposal network produces more accurate proposals for the ONH area. Conversely, faster-RCNN based methods [14] may miss OC proposals or generate proposals that do not fully surround the ONH area. Second, the multi-label head considers the relation of OD and OC for joint segmentation while the instance head segment OD and OC separately. For M-Net, it is performed on the ONH area; While our method is performed on the whole image. Although the background accounts for a large content of the image compared to the OD, our method is still better than M-Net. One possible reason is that our proposed pyramid RoIAlign module helps to learn stronger representations than M-Net.

Table 2. Segmentation results on REFUGE validation set.

Method	E_{disc}	E_{cup}	E_{rim}	δ_E	A_{disc}	A_{cup}	A_{rim}
Mask-RCNN	0.092	0.228	0.211	0.055	0.973	0.976	0.923
PM-Net	**0.088**	**0.223**	**0.204**	**0.048**	**0.979**	**0.980**	**0.936**

We have the same observations on the experiments on REFUGE validation set in Table 2, which further verifies our method. We also show example segmentation results and discussions in Fig. 3.

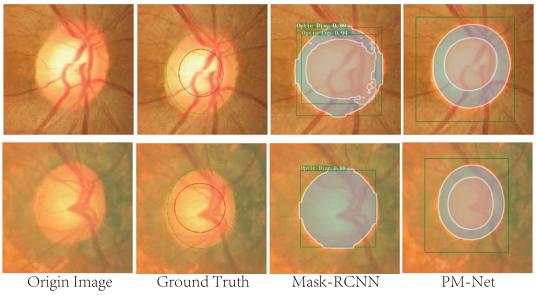

Fig. 3. Example segmentation results. The first row is a normal eye while the second row is a glaucoma eye. Our method achieve more accurate and reliable results.

Table 3. Effect of the pyramid RoIAlign module on ORIGA dataset.

Method	E_{disc}	E_{cup}	E_{rim}	δ_E	A_{disc}	A_{cup}	A_{rim}
Baseline	0.074	0.231	0.260	0.079	0.985	0.941	0.929
Max-pooling	0.073	0.210	0.234	**0.067**	0.982	0.938	0.944
Sum-average	**0.069**	0.211	0.237	0.068	**0.987**	**0.942**	0.942
Concatenation	0.070	**0.207**	**0.228**	0.069	0.986	0.936	**0.950**

3.2 Ablative Studies

In this part, we conduct ablative studies on ORIGA dataset. First, we evaluate the different feature map fusion strategy for the pyramid RoIAlign module. Then, we evaluate each module of our proposed method. Table 3 shows the performance

Table 4. Effect of different components of our method on ORIGA dataset. ML is for multi-label head. SegRPN is for segmentation based region proposal network.

Method	E_{disc}	E_{cup}	E_{rim}	δ_E	A_{disc}	A_{cup}	A_{rim}
Mask-RCNN(Baseline)	0.074	0.231	0.260	0.079	0.985	0.941	0.929
ML	0.071	0.217	0.232	0.071	0.984	0.940	0.947
RoIAlign	0.069	0.211	0.237	0.068	**0.987**	**0.942**	0.942
ML + RoiAlign	0.068	0.217	0.237	0.073	0.986	**0.942**	0.941
ML + SegRPN	0.069	0.219	0.230	0.070	0.986	0.933	**0.950**
ML + RoiAlign + SegRPN	**0.066**	**0.208**	**0.224**	**0.065**	0.986	**0.942**	0.949

using different feature fusion strategy for pyramid RoIAlign module. All three strategies improve E_{cup} and δ_E significantly. Table 4 shows the effect of different components on our PM-Net. The multi-label head and pyramid RoiAlign module improve A_{cup} significantly. Moreover, the segmentation based RPN improve A_{disc} since the proposal of OC is more accurate.

4 Conclusions

In this paper, we have proposed a pyramid multi-label network (PM-Net) for simultaneously ONH localization and joint OD/OC segmentation. PM-Net produces more accurate proposals and avoids missing object proposals by using a segmentation-based RPN to locate the ONH area. Furthermore, PM-Net adopts pyramid RoIAlign module to incorporate suitable global features and employs a multi-label head to model the relationship of OD and OC. Extensive experiments verify the effectiveness of our method.

Acknowledgement. This work was supported by National Natural Science Foundation of China (NSFC) 61602185, 61876208, Guangdong Introducing Innovative and Entrepreneurial Teams 2017ZT07X183, Guangdong Provincial Scientific and Technological Fund 2018B010107001, 2017B090901008, 2018B010108002, Pearl River S&T Nova Program of Guangzhou 201806010081, CCF-Tencent Open Research Fund RAGR20190103.

References

1. Almazroa, A., Burman, R., et al.: Optic disc and optic cup segmentation methodologies for glaucoma image detection: a survey. J. Ophthalmol. (2015)
2. Aquino, A., Gegúndez-Arias, M.E., et al.: Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques. IEEE TMI **29**(11), 1860–1869 (2010)
3. Chen, X., Xu, Y., Yan, S., Wong, D.W.K., Wong, T.Y., Liu, J.: Automatic feature learning for glaucoma detection based on deep learning. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 669–677. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_80
4. Chen, X., Xu, Y., et al.: Glaucoma detection based on deep convolutional neural network. In: EMBC, pp. 715–718. IEEE (2015)
5. Cheng, J., Liu, J., et al.: Automatic optic disc segmentation with peripapillary atrophy elimination. In: EMBC, pp. 6224–6227. IEEE (2011)
6. Cheng, J., Liu, J., et al.: Superpixel classification for initialization in model based optic disc segmentation. In: EMBC, pp. 1450–1453. IEEE (2012)
7. Cheng, J., Liu, J., et al.: Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE TMI **32**(6), 1019–1032 (2013)
8. Fu, H., Cheng, J., et al.: Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE TMI **37**, 1597–1605 (2018)
9. Fu, H., Cheng, J., et al.: Disc-aware ensemble network for glaucoma screening from fundus image. IEEE TMI **30**, 2493–2501 (2018)

10. Joshi, G.D., Sivaswamy, J., et al.: Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. IEEE TMI **30**(6), 1192–1205 (2011)
11. Kaufman, P.L., Levin, L.A., Adler, F.H., Alm, A.: Adler's Physiology of the Eye. Elsevier Health Sciences (2011)
12. Ren, S., He, K., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. In: NeurIPS, pp. 91–99 (2015)
13. 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
14. Sun, X., Xu, Y., et al.: Optic disc segmentation from retinal fundus images via deep object detection networks. In: EMBC, pp. 5954–5957, July 2018
15. Tang, L., Garvin, M.K., et al.: Segmentation of optic nerve head rim in color fundus photographs by probability based active shape model. IOVS **53**(14), 2144 (2012)
16. Wong, D.W.K., Liu, J., et al.: Learning-based approach for the automatic detection of the optic disc in digital retinal fundus photographs. In: EMBC. IEEE (2010)
17. Xu, Y., Duan, L., Lin, S., Chen, X., Wong, D.W.K., Wong, T.Y., Liu, J.: 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
18. 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
19. 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
20. 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
21. Yin, F., Liu, J., et al.: Model-based optic nerve head segmentation on retinal fundus images. In: EMBC, pp. 2626–2629. IEEE (2011)
22. Zhao, H., Shi, J., et al.: Pyramid scene parsing network. In: CVPR (2017)
23. Zilly, J., Buhmann, J.M., et al.: Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. Comput. Med. Imaging Graph. **55**, 28–41 (2017)