Joint Localization of Optic Disc and Fovea in Ultra-widefield Fundus Images

Zhuoya Yang[1,2,3], Xirong Li[1,2,3](✉), Xixi He[3], Dayong Ding[3], Yanting Wang[4], Fangfang Dai[4], and Xuemin Jin[4]

[1] MOE Key Lab of DEKE, Renmin University of China, Beijing, China
[2] AI and Media Computing Lab, School of Information, Renmin University of China, Beijing, China
xirong@ruc.edu.cn
[3] Vistel AI Lab, Visionary Intelligence Ltd., Beijing, China
[4] Henan Provincal Peoples' Hospital, Zhengzhou, China

Abstract. Automated localization of optic disc and fovea is important for computer-aided retinal disease screening and diagnosis. Compared to previous works, this paper makes two novelties. First, we study the localization problem in the new context of ultra-widefield (UWF) fundus images, which has not been considered before. Second, we propose a spatially constrained Faster R-CNN for the task. Extensive experiments on a set of 2,182 UWF fundus images acquired from a local eye center justify the viability of the proposed model. For more than 99% of the test images, the improved Faster R-CNN localizes the fovea within one optic disc diameter to the ground truth, meanwhile detecting the optic disc with a high IoU of 0.82. The new model works reasonably well even in challenging cases where the fovea is occluded due to severe retinopathy or surgical treatments.

Keywords: Object localization · UWF fundus image · Faster R-CNN

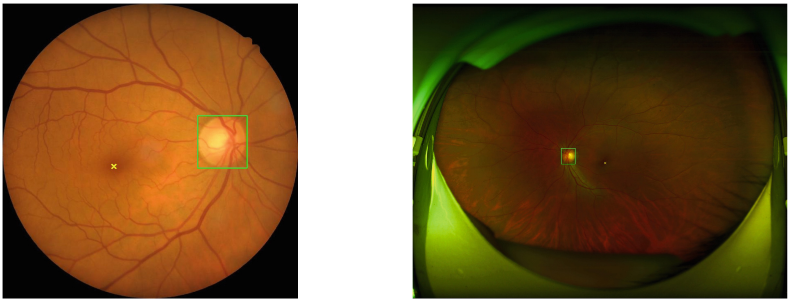
1 Introduction

This paper studies joint localization of the optic disc (OD) and the fovea in ultra-widefield (UWF) fundus images. Localizing the OD is a prerequisite for computer-aided diagnosis of optic nerve diseases such as glaucoma, the progression of which is assessed by the optic cup-to-disc ratio. The fovea is responsible for sharp central vision, so any retinal lesions observed in its surrounding area shall be taken seriously. Automated localization of the two objects is thus crucial for fundus image analysis.

Existing works on localizing either the OD [6,10], the fovea [4], or their combination [1,7,8] deal with normal fundus images. Different from normal fundus

This work was supported by NSFC (No. 61672523), the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (No. 18XNLG19).

H.-I. Suk et al. (Eds.): MLMI 2019, LNCS 11861, pp. 453–460, 2019.
https://doi.org/10.1007/978-3-030-32692-0_52



(a) Normal, 45° field of view

(b) Ultra-Widefield, 200° field of view

Fig. 1. Two types of fundus images. (a) A normal fundus image captures up to 15% of the retina, showing main structures such as optic disc (green bounding box), fovea (yellow cross) and vessel. (b) A UWF fundus image, which reaches the far peripheral retina and covers approximately 82% of the retina. (Color figure online)

photography, a UWF fundus image provides a much larger field of view. It covers more retinal surface, as shown in Fig. 1(b), and thus peripheral retinal lesions can be spotted. Despite the increasing use of UWF fundus images in varied clinical scenarios [2,3], automated localization of both OD and fovea in a UWF fundus image has not been touched, to the best of our knowledge.

In a fundus image, the OD is an oval bright object at the nasal side of the fovea. Meanwhile, the fovea is a small pit located at the center of the darkest area known as macula, see Fig. 1. As the visual appearance of the two objects seems to be vivid, a natural idea is to use a present-day object detection network, *e.g.,* Faster R-CNN [9]. Although there have been few CNN-based solutions for OD and fovea localization in normal fundus images [1], we see no attempt for exploiting any deep object detection network.

Different from objects in natural photos, the OD and the fovea are spatially correlated. Using the optic disc diameter (DD) as a unit, the horizontal distance between the OD and the fovea is around 2.5 DD, with the latter located slightly below the former. For normal fundus images, some initial efforts have been made to leverage such spatial constraints, implemented either as a two-step approach [7] or as post-processing to refine the localization [8]. How to supervise an object detection network with the spatial constraints remains open.

The contributions of this paper are as follows:

- We present the first study on joint localization of both OD and fovea in UWF fundus images.
- We propose *spatially constrained* Faster R-CNN that effectively leverage the spatial constraints between the OD and the fovea.
- We provide an extensive evaluation, justifying the effectiveness of the proposed solutions on a real-world dataset.

Table 1. State-of-the-art for fovea and OD localization in fundus images. This paper goes one step further w.r.t. the target domain and the localization technique.

Fundus image	Target	Paper	Localization method
Normal	OD	Zou *et al.* [10]	Intensity-based ROIs proposal + Vessel-based verification
		Meng *et al.* [6]	Sliding window + Patch-based CNN classification
	Fovea	Gegundez *et al.* [4]	OD & Vessel-based ROIs proposal + Contour finding
	OD & Fovea	Niemeijer *et al.* [7]	KNN based regression
		Qureshi *et al.* [8]	Ensemble of multiple low-level OD/fovea detectors
		Al-Bander *et al.* [1]	Two-step CNN based regression
UWF	OD & Fovea	*This work*	Proposed *spatially constrained* Faster R-CNN

2 Related Work

Good efforts have been made for automated localization of the OD [6,10], the fovea [4], and both [1,7,8]. However, all target at normal fundus images, as we summarize in Table 1.

For OD localization, Zou *et al.* [10] generate regions of interest (ROIs) by simple intensity thresholding, and subsequently identify the OD by vessel-based verification. Meng *et al.* [6] train a patch-based convolutional neural network (CNN) to locate the OD, using sliding window to generate proposals. It predicts only the OD center, without precise boundary.

For fovea localization, Gegundez *et al.* [4] assume the availability of the OD and the vascular tree. A pixel within the fovea region is estimated, which is set to be 2.5 optic disc diameters (DD) away from the OD center. The orientation of the vasculature is used to determine if the pixel is on the OD's left or right side. Then, contour finding is performed on a 2DD×2DD sub-image centered on the pixel to localize the fovea.

For joint localization, Niemeijer *et al.* [7] develop a regression model. For each pixel in a given image, its distance to the (unknown) OD center is predicted based on intensity and vascular features. The pixel with the smallest distance is chosen as the OD center. In a similar vein, the fovea is localized, with the search area enforced to be 2DD away from the detected OD center. Qureshi *et al.* [8] take an ensemble approach to locate OD (fovea) by combining several low-level OD (fovea) detectors. Among the multiple OD and fovea candidates, the pair that best match spatial constraint is selected. More recently, Al-Bander *et al.* [1] build three CNN based regression models. One CNN is used to predict the initial centers of the OD and the fovea. With a sub-image cropped around the predicted OD center as input, the second CNN is used to re-predict the OD center. In a similar way, the third CNN is used to re-predict the fovea center.

Note that the majority of the previous efforts rely heavily on intensity and vascular features [4,7,8,10]. However, such features are unreliable due to varied factors including changes in imaging conditions, retinal disorders and surgical treatments on the eye. The few CNN based attempts are cumbersome, as they require either pixel-by-pixel classification [6] or a triplet of CNNs [1]. Spatial constraints between the two objects are largely unexplored.

3 Spatially-Constrained Joint Localization

We aim to automatically localize both OD and fovea in a given UWF fundus image. The task differs from conventional object detection in the following two aspects. First, there is only one OD and one fovea. Second, due to the physiological structure of the retina, constraints on spatial locations of the two objects exit. We hypothesize that such constraints are helpful for localizing the two objects, especially the fovea, in challenging cases where UWF images are presented with severe retinopathy. In order to exploit the spatial constraints, we propose two strategies, one is an OD-guided two-step approach and the other is to directly incorporate the constraints into the loss function of Faster R-CNN, and consequently achieve a simultaneous localization in one forward computation.

3.1 Strategy 1. OD-guided Fovea Localization

The OD in a fundus image appears as a bright area where blood vessels converge. Such a visual pattern is unique and relatively stable, making the object more easily to be localized than the fovea. This observation motivates us to localize the OD first, and accordingly use it as a guidance to localize the fovea. We term this strategy OD-guided fovea localization. Note that the OD-guided strategy conceptually resembles [7] to some extent, as both works use the OD to narrow down the search space for fovea localization. Nonetheless, our module is end-to-end and requires no vascular information, making the overall solution simplified.

We first train a Faster R-CNN for OD localization. While the OD is known to be placed to the nasal side of the fovea, no laterality information is available in our study. So at each side, a squared candidate area is heuristically estimated, with its location and size relatively determined in the unit of the DD. The region is cropped and fed into another Faster R-CNN trained for fovea localization. OD-guided fovea localization implements the spatial constraints by enforcing the object detection network to search for the fovea in the two sub-images.

This strategy is effective for reducing false alarms in the peripheral area of the retina. However, both the training and the execution of the fovea localization network remain independent of the OD. Moreover, two Faster R-CNNs are required. To overcome these downsides, we consider another strategy as follows.

3.2 Strategy 2. Spatially-Constrained Loss

We train the Faster R-CNN network with a new loss which takes into account the spatial constraints between the OD and the fovea. In order to make the paper more self-contained, we describe briefly how Faster R-CNN works in this new context. As an end-to-end object detection network, Faster R-CNN is composed of a Region Proposal Network (RPN) for ROI generation and a Fast R-CNN [5] for ROI refinement and classification. Given a UWF fundus image, the RPN generates many bounding-box proposals and classifies them either as foreground or as background. Proposals classified as foreground, after bounding-box regression and non-maximum suppression, are preserved as candidate ROIs. The Fast

R-CNN then classifies these ROIs into one of the three classes, *i.e.,* OD, fovea and background. In order to get a reliable set of ROIs to represent the OD, we let OD-ROIs consist of ROIs predicted as OD with scores larger than 0.5. In a similar vein we obtain fovea-ROIs. For better localization, these chosen OD-ROIs and fovea-ROIs are further fed into a bounding-box regression subnetwork.

With the OD-ROIs and fovea-ROIs identified, we exploit the spatial constraint in two aspects, one is distance-based and the other is direction-based. Let $d(OD, fovea)$ be the Euclidean distance between the averaged center of OD-ROIs and fovea-ROIs. We define a distance-based loss $loss_d$ as

$$loss_d := \max(0, d_{min} - d(OD, fovea)) + \max(0, d(OD, fovea) - d_{max}), \quad (1)$$

where d_{min} and d_{max} are the 2nd and the 98th percentiles of the OD-fovea distances calculated using the ground truth of our training data. Whenever the distance is smaller than d_{min} or larger than d_{max}, a loss occurs.

To consider the direction-based constraint, we compute the angle between the averaged center of OD-ROIs and that of fovea-ROIs, denoted by $\theta(OD, fovea)$. The value is positive if the fovea is below the OD, and negative otherwise. Note that in a well-positioned fundus image, the fovea shall be placed slightly below the OD. However, minor rotations might occur in practice. We therefore consider the upper bound only, defining a direction-based loss $loss_\theta$ as

$$loss_\theta := \max(0, abs(\theta(OD, fovea)) - \theta_{max}), \quad (2)$$

where the upper bound θ_{max} is the 98th percentile of the angles, computed using the same ground truth as used for obtaining d_{min} and d_{max}.

By adding the above two losses to the original loss of Faster R-CNN (denoted as $loss_{fr}$), we obtain the new spatially-constrained loss $loss_{sc}$ as

$$Loss_{sc} := loss_{fr} + \lambda_1 \cdot loss_d + \lambda_2 \cdot loss_\theta, \quad (3)$$

where λ_1 and λ_2 are two positive weights controlling the influence of the distance-based and direction-based losses, respectively. Based on a hold-out validation set, we empirically set $\lambda_1 = 0.002$ and $\lambda_2 = 0.1$.

4 Evaluation

4.1 Experimental Setup

Data. As no public dataset is available, we acquire 2,182 UWF images from our collaborating eye center, with manually labeled bounding boxes of the OD and the coordinate of the fovea. The dataset is divided at random into three disjoint subsets for training, validation and test, respectively, with a ratio of 4:1:1.5. To avoid over-fitting, the data split is based on patients s so images from a specific patient appear only in one subset.

Implementation. Original images are downsized from 3900×3072, to 762×600. We implement Faster R-CNN with VGGNet-16 as their backbone. Per model, the top-ranked OD-ROI is used as the final OD region, while the center of the top-ranked fovea-ROI is used as the predicted coordinate of the fovea.

Evaluation Criteria. As the OD is an area and the fovea is a point to be localized, we use Intersection over Union (IoU) for OD and the Euclidean distance between the center of predicted box and the center of ground true for fovea. Overall performance is obtained by averaging scores of all test images.

Table 2. Performance of different models for optic disc/fovea localization. The operator $[\![\cdot]\!]$ computes accuracy, *i.e.,* the rate of test images satisfying a given criterion. Per test image, DD is the vertical disc diameter of the ground truth. The proposed spatially constrained Faster R-CNN models achieve the best joint localization.

Model	Optic Disc (OD)		Fovea				
	IoU	$[\![IoU \geq 0.5]\!]$	Distance	Std.	$[\![< \frac{1}{5}DD]\!]$	$[\![< \frac{1}{4}DD]\!]$	$[\![< DD]\!]$
Baselines:							
OD Faster R-CNN	**0.8445**	**0.9980**	–	–	–	–	–
Fovea Faster R-CNN	–	–	35.90	140.15	0.8039	0.8495	0.9841
Joint Faster R-CNN	0.8131	0.9881	31.24	92.59	0.8039	0.8475	0.9841
Spatially Constrained:							
OD-guided	**0.8445**	**0.9980**	27.25	82.55	0.8019	0.8514	0.9881
$loss_{sc}$	0.8174	0.9960	**25.29**	**52.47**	**0.8059**	**0.8594**	**0.9920**

4.2 Experiment 1. Joint Localization Versus Separate Localization

The first uncertainty we need to address is the necessity of joint localization. We train two Faster R-CNN models separately, one for the OD and the other for the fovea. We then train another model that localizes the two objects simultaneously.

The performance of the three models is reported in the top part of Table 2. Comparing with the two individual models, Joint Faster R-CNN performs worse than the OD model for OD localization, while providing more precise fovea localization than the fovea model. Note that IoU exceeding 0.80 is sufficient, as shown in Fig. 2. So we focus our discussion on fovea localization.

Joint Faster R-CNN obtains a noticeably smaller standard deviation for fovea localization. The result justifies the necessity of joint localization, and also suggests the spatial relations between the OD and the fovea are implicitly modeled.

4.3 Experiment 2. Spatially-Constrained Joint Localization

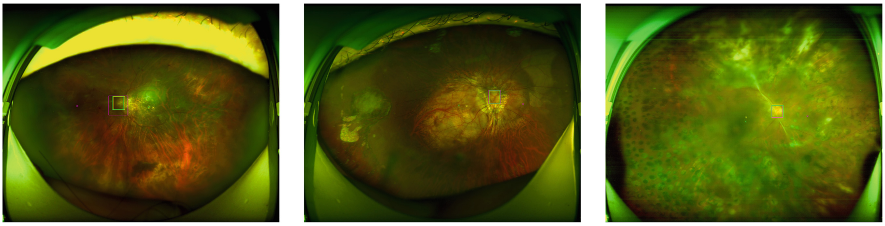
With Joint Faster R-CNN as our baseline, we now evaluate the two proposed strategies for spatially-constrained joint localization. Recall that the OD-guided strategy uses the OD model in its first step, so the strategy scores the same IoU as the baseline in Table 2. Both strategies give better fovea localization than the baseline, suggesting the importance of explicitly modeling the spatial constraints. Moreover, Faster R-CNN trained with the proposed loss gives the best fovea localization with the smallest deviation (25.29 ± 52.47). For more than

99% of the test images, the fovea is localized within one DD to the ground truth. We further conduct an ablation study concerning distinct spatial constraints, *i.e.,* distance-based, direction-based and their combination. As shown in Table 3, the joint loss is the best.

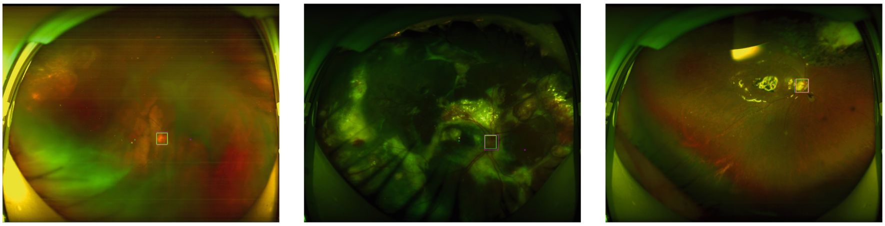
Some qualitative results are provided in Fig. 2. From Fig. 2(a) to (e), while the baseline incorrectly predicts the fovea on the opposite side, our improved Faster R-CNN localizes the fovea on the correct side. For Fig. 2(f) where the macular

Table 3. The influence of distinct spatial constraints, *i.e.,* distance-based (d_{min} and d_{max} in Eq. 1), direction-based (θ_{max} in Eq. 2) and their combination, on fovea localization. Faster R-CNN trained with the joint loss (the last row) performs the best.

d_{min}	d_{max}	θ_{max}	Distance	Std.
✓	✗	✗	27.04	75.29
✓	✓	✗	25.75	62.76
✓	✓	✓	**25.29**	**52.47**



(a)	*IoU*	*Distance*	(b)	*IoU*	*Distance*	(c)	*IoU*	*Distance*
Baseline	0.410	1005.14	Baseline	0.829	903.17	Baseline	**0.881**	874.22
This paper	**0.818**	**67.26**	*This paper*	**0.864**	**109.37**	*This paper*	0.852	**59.30**



(d)	*IoU*	*Distance*	(e)	*IoU*	*Distance*	(f)	*IoU*	*Distance*
Baseline	**0.874**	785.00	Baseline	0.720	935.93	Baseline	0.842	**55.44**
This paper	0.863	**58.69**	*This paper*	**0.949**	**39.11**	*This paper*	**0.844**	61.02

Fig. 2. Some localization results by the baseline (Joint Faster R-CNN) and the improved Faster R-CNN. OD and fovea are indicated by bounding boxes and crosses, respectively. Ground truth/baseline/our results are shown in green/purple/yellow. Below each image are IoU of the detected OD and distance of the predicted fovea to the ground truth. Better numbers are shown in bold font. Best viewed digitally in close-up. (Color figure online)

area appears to be filled with silicone oil, the baseline gives a better localization. At the cost of slightly dropping the performance of OD localization, the new Faster R-CNN noticeably improves fovea localization.

5 Conclusions

For joint localization of OD and fovea in UWF fundus images, we recommend to train Faster R-CNN with the proposed joint loss. As experiments on a set of 2,182 real-world images show, for more than 99% of the test images, the improved Faster R-CNN localizes the fovea within one DD to the ground truth, meanwhile localizing the OD with a high IoU of 0.82.

References

1. Al-Bander, B., Al-Nuaimy, W., Williams, B., Zheng, Y.: Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc. Biomed. Sig. Process. Control **40**, 91–101 (2018)
2. Assil, K.K., Batra, V.N.: The role of ultra-widefield retinal imaging as a standard assessment tool in the cataract practice. Am. J. Ophthalmol. **10**(1), 31–34 (2017)
3. Falavarjani, K.G., Wang, K., Khadamy, J., Sadda, S.R.: Ultra-wide-field imaging in diabetic retinopathy; an overview. J. Curr. Ophthalmol. **28**(2), 57–60 (2016)
4. Gegundez-Arias, M.E., Marin, D., Bravo, J.M., Suero, A.: Locating the fovea center position in digital fundus images using thresholding and feature extraction techniques. Comput. Med. Imaging Graph. **37**(5–6), 386–393 (2013)
5. Girshick, R.: Fast R-CNN. In: Proceedings of ICCV (2015)
6. Meng, X., Xi, X., Yang, L., Zhang, G., Yin, Y., Chen, X.: Fast and effective optic disk localization based on convolutional neural network. Neurocomputing **312**, 285–295 (2018)
7. Niemeijer, M., Abrámoff, M.D., Ginnekena, B.V.: Fast detection of the optic disc and fovea in color fundus photographs. Med. Image Anal. **13**(6), 859–870 (2009)
8. Qureshi, R.J., Kovacs, L., Harangi, B., Nagy, B., Peto, T., Hajdu, A.: Combining algorithms for automatic detection of optic disc and macula in fundus images. Comput. Vis. Image Underst. **116**(1), 138–145 (2012)
9. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017)
10. Zou, B., Chen, C., Zhu, C., Duan, X., Chen, Z.: Classified optic disc localization algorithm based on verification model. Comput. Graph. **70**, 281–287 (2017)