



A Pixel-Wise Distance Regression Approach for Joint Retinal Optical Disc and Fovea Detection

Maria Ines Meyer^{1(✉)}, Adrian Galdran^{1(✉)}, Ana Maria Mendonça^{1,2},
and Aurélio Campilho^{1,2}

¹ INESC-TEC - Institute for Systems and Computer Engineering,
Technology and Science, Porto, Portugal

{`maria.i.meyer, adrian.galdran`}@inesctec.pt

² Faculdade de Engenharia da Universidade do Porto, Porto, Portugal
{`amendon, campilho`}@fe.up.pt

Abstract. This paper introduces a novel strategy for the task of simultaneously locating two key anatomical landmarks in retinal images of the eye fundus, namely the optic disc and the fovea. For that, instead of attempting to classify each pixel as belonging to the background, the optic disc, or the fovea center, which would lead to a highly class-imbalanced setting, the problem is reformulated as a pixelwise regression task. The regressed quantity consists of the distance from the closest landmark of interest. A Fully-Convolutional Deep Neural Network is optimized to predict this distance for each image location, implicitly casting the problem into a per-pixel Multi-Task Learning approach by which a globally consistent distribution of distances across the entire image can be learned. Once trained, the two minimal distances predicted by the model are selected as the locations of the optic disc and the fovea. The joint learning of every pixel position relative to the optic disc and the fovea favors an automatic understanding of the overall anatomical distribution. This results in an effective technique that can detect both locations simultaneously, as opposed to previous methods that handle both tasks separately. Comprehensive experimental results on a large public dataset validate the proposed approach.

Keywords: Optic Disk Detection · Fovea detection · Retinal Image Analysis

1 Introduction

Among the landmarks of interest in the human retina, the fovea (a small depression in the macula center) and the optic disc (the location where the optic nerve and blood vessels leave the retina) are key for diagnostic purposes. Consequently, plenty of research has addressed their automatic location in the past [3, 6, 10].

Most previous work approaches either the problem of detecting the optic disc (OD) or the fovea. In the OD case, several techniques leverage the knowledge that

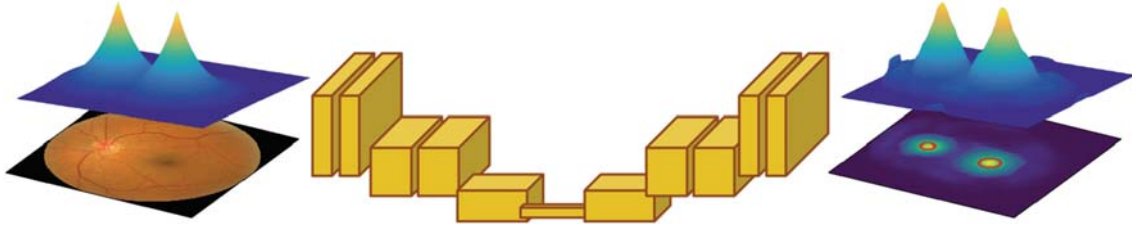


Fig. 1. Proposed approach for joint OD and fovea detection via regressing a distance map from both landmarks to every other pixel in the image.

retinal vessels originate on it in order to find its center. This is typically achieved by extracting geometrical and orientation information from the vascular tree [5]. Fuzzy convergence algorithms [7], Hough transforms [9], and matched-filtering approaches [8] have also been reported for the task of OD location.

Regarding fovea detection, in [6] authors proposed to pre-detect a region likely to contain the fovea based on constraints on its anatomical position relative to OD and blood vessels, with a subsequent thresholding stage to refine its location.

Several other works have reported results on both tasks, although solved in separate stages. The method introduced in [14] analyzed the intensity differences around the OD center to identify it, followed by a template matching technique to locate the fovea. Template matching was also proposed in [16], coupled with directional matched filters. The method in [11] devised a cost function based on a set of global and local anatomical features that was then minimized to yield the most likely locations of OD, fovea and vascular arch. The same authors re-defined in [10] both tasks as regression problems, applying a k-NN strategy to estimate OD and fovea location.

Fewer work has addressed the location of both anatomical landmarks simultaneously. In [3] a super-elliptical convergence index filter was applied for this purpose, while in [2] two Convolutional Neural Networks (CNN) were built, the first one aiming at locating a region of interest around both the OD and the fovea, and the second one refining these predictions.

The main contribution of this work is a new strategy for jointly detecting the OD and the fovea. In contrast with previous techniques, the proposed method does not attempt to directly detect only OD and fovea centers. Instead, the distance to both locations is regressed for every pixel in a retinal image, as illustrated in Fig. 1. This regression problem can be effectively solved by means of a Fully-Convolutional Neural Network. This strategy poses a multi-task-like problem, on which information of every pixel contributes to generate a globally consistent prediction map where likelihood of OD and fovea locations are maximized.

2 Joint Optic Disc and Fovea Detection Methodology

Next, the motivation for considering joint OD and fovea detection as a pixel-based distance regression problem and the methodology to solve it are detailed.

2.1 Casting the Problem into a Pixel-Wise Regression Task

In order to solve the joint detection of the OD and the fovea on retinal images, we implicitly adopt a Multi-Task Learning (MTL) approach. MTL is known to be an effective regularizing technique for machine learning models with a large quantity of parameters to be learned, which are typically prone to overfitting.

MTL can be described as a strategy to improve generalization of machine learning models based on solving simultaneously two or more related tasks. It has been observed that this approach can improve learning of abstract representations, which can be mutually useful for the range of tasks to be solved [13].

Jointly detecting the OD and the fovea can be considered as a pixelwise classification problem, where there are three classes: background, OD location, and fovea location. However, under this configuration, the resulting problem becomes highly skewed in terms of examples for each class. A straightforward MTL-like solution would consist on designing two sub-tasks, which would be defined by binary classification problems, and solve them jointly. Nevertheless, this approach would still suffer from a class imbalance issue that would complicate substantially the optimization of a model in this setting.

In this paper, a pixel-wise MTL-like strategy is adopted. Instead of attempting to classify each pixel into two or three classes, we reformulate the problem as regressing the distance from each image location to the closest of both retinal landmarks of interest. For this, we first define the *Bi-Distance Map* $\mathcal{B}(x, y)$ for each pixel location $(x, y) \in \Omega$, being $\Omega \subset \mathbb{R}^2$ the image domain on which a retinal image $I(x, y)$ is defined. Given the location of the OD (x_{od}, y_{od}) and the fovea (x_{fov}, y_{fov}) , $\mathcal{B}(x, y)$ is defined as follows:

$$\mathcal{B}(x, y) = \min \left(\sqrt{(x - x_{od})^2 + (y - y_{od})^2}, \sqrt{(x - x_{fov})^2 + (y - y_{fov})^2} \right). \quad (1)$$

From the Bi-Distance Map definition, a normalized form, bounded in $[0, 1]$, can be easily built:

$$\mathcal{B}^N(x, y) = \left(1 - \frac{\mathcal{B}(x, y)}{\max_{\Omega} \mathcal{B}(x, y)} \right)^{\gamma}, \quad (2)$$

where γ is a decay parameter governing the spread of \mathcal{B}^N across the image domain. The effect of modifying the decay parameter γ on the normalized Bi-Distance Map is shown in Fig. 2.

The goal of the model designed in the next sub-section will be to produce an accurate estimate of $\mathcal{B}^N(x, y)$ simultaneously for every pixel. By casting the initial classification problem into a pixel-wise regression task, the model is required

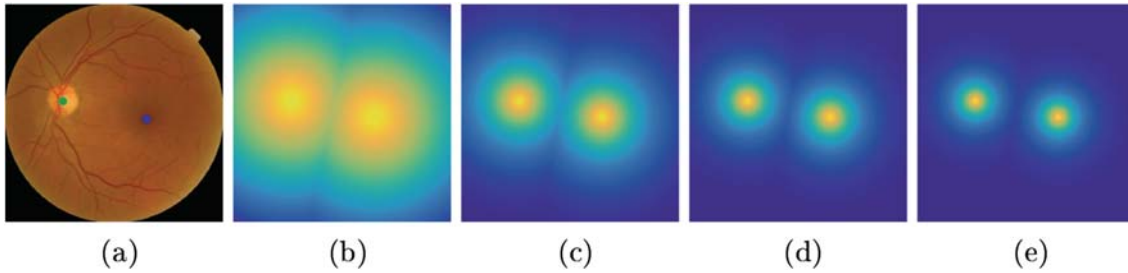


Fig. 2. (a) Retinal Image with **O** or **F** marked green and blue respectively. (b)–(f) Normalized Bi-Distance Maps for different decays parameters: $\gamma \in \{1, 3, 5, 7\}$.

to solve a different task on each location, implicitly turning the regression of Eq. (1) into a hierarchical multi-task technique: (1) at a high-level, for each pixel location the model needs to predict which landmark, the OD or the fovea, is closest to it, and (2) at a low-level, for each pixel location the model is required to produce an estimate of the distance to the closest landmark.

Hence, each pixel poses two independent but deeply related goals, effectively regularizing the initial problem, and favoring a globally consistent solution.

2.2 A Fully-Convolutional Deep Neural Network for Distance Regression

In recent years, Convolutional Neural Networks (CNNs) have attained a remarkable success in medical image analysis problems. Although CNNs were initially employed for image classification tasks, subsequent advances allowed the application of convolutional architectures for detection and segmentation tasks. For the latter, a CNN is reformulated to produce pixel-wise classification based on the ideas of *Fully-Convolutional Neural Networks* (F-CNN) and *skip connections*, which enable the coupling of coarse and fine layers of a CNN. In our case, the goal is also to assign a single prediction to each pixel, similarly to a segmentation problem. However, this prediction should not match a discrete set of categories, but rather be a continuous value in $[0, 1]$.

According to the above considerations, our approach is to build on successful F-CNN architectures tailored for segmentation problems, but modifying the loss function to perform distance regression. A popular F-CNN architecture is U-net, introduced in [12]. In U-Net, a contracting sub-network is coupled with a symmetric upsampling sub-network in such a way that the representation produced by the final layer of the upsampling path matches the dimensions of the second last layer. This representation is then fused with the one coming from the corresponding layer in the contracting section. This process is iterated until the output of the upsampling path shares dimensionality with the initial input. This results in a U-shaped architecture, whereby the output feature maps of the contracting sub-network are effectively combined with the output from the upsampling sub-network, providing multi-scale context to the model. The U-Net based architecture employed in this paper is illustrated in Fig. 3.

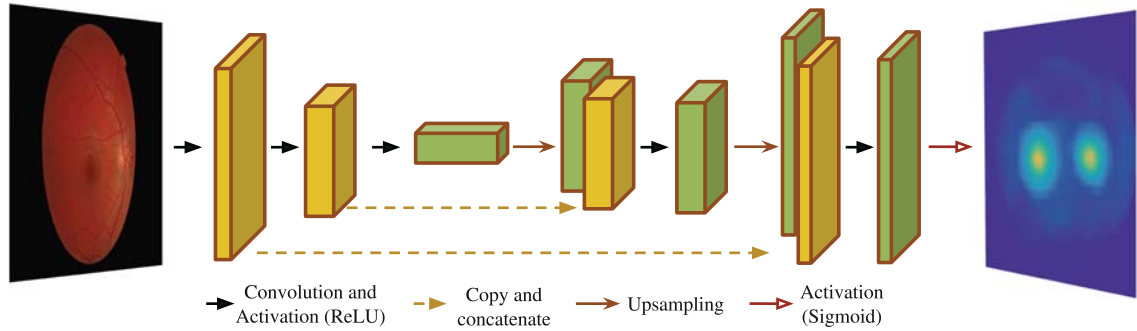


Fig. 3. The proposed F-CNN architecture for Normalized Bi-Distance Map Regression.

Regarding the objective function employed to optimize the model \mathcal{U}_θ , the goal is to predict, for each pixel (x, y) , $\mathcal{B}^N(x, y)$ as defined in Eq. (2). Given the smooth nature of $\mathcal{B}^N(x, y)$, we select a standard \mathcal{L}^2 loss for this task:

$$\mathcal{L}_{reg}(\theta) = \frac{1}{M} \sum_{x,y \in \Omega} \|\mathcal{U}_\theta(x, y) - \mathcal{B}^N(x, y)\|^2, \quad (3)$$

where M is the number of pixels within a retinal image.

The above loss is backpropagated to update iteratively the weights θ , with mini-batch stochastic gradient descent using the Adam optimizer. The initial learning rate was $\alpha = 0.01$, and it was halved whenever the loss monitored on the validation set was not observed to be improving¹.

2.3 Optic Disc and the Fovea Assignment

The output of the above model is a smooth prediction on distances to both landmarks of interest. Out of this, a Laplacian-of-Gaussian operator was applied in order to extract the two most prominent maxima.

Even if the outlined technique can find the locations of both landmarks, it is unable to specify which of them corresponds to the OD and which one to the fovea. Fortunately, this can be easily solved by a simple local intensity analysis around the two detected maxima produced by \mathcal{U}_θ . Specifically, the region around the OD is typically brighter, whereas the fovea is generally darker than the rest of the image. As such, the predicted locations were examined in a local neighborhood of the original images, and the mean red intensity extracted. The region with higher average intensity was selected as belonging to the OD.

3 Experimental Evaluation

3.1 Data

The model was trained and evaluated on the Messidor dataset [4]. Messidor comprises 1200 retinal images of varying resolutions (2240×1488 , 1440×960 ,

¹ An implementation can be found at github.com/minesmeyer/od-fovea-regression.

and 2304×1536 pixels). The locations of the fovea and the OD for 1136 of these images were provided in [6]. The OD centroids were extracted from these segmentations and used to define their location.

The available 1136 images were divided into two sets containing 568 images, and the proposed model was trained alternately on one subset and tested on the other one, reporting average performance. During training, 20% of the data was separated and employed to monitor the value of the loss in Eq. (3).

3.2 Evaluation Approach

In order to provide a standardized and fair comparison to prior work, we adopt the evaluation approach from [6]. Accordingly, we calculated the Euclidean distance between the ground-truth OD and fovea coordinates and their predicted positions. Predictions falling within a certain distance of the ground-truth are considered successful. This distance is set as a multiple of the OD radius R , and different distances are considered: $(1/8)R$, $(1/4)R$, $(1/2)R$ and $1R$.

Since R varies between images with different resolutions, a separate R was set for each of the three resolutions in the dataset. Namely, $R = 68$ for resolution 1440×960 , $R = 103$ for resolution 2240×1488 and $R = 109$ for resolution 2304×1536 . Following [6,9], we also compute the mean euclidean distance (\bar{D}) between predicted and ground-truth positions normalized by the FOV diameter ($\bar{D}_{FOV} = (D(p_{exp}, p_{real})/d_{FOV} \cdot 100)$) and by the OD radius ($\bar{D}_R = (D(p_{exp}, p_{real})/R \cdot 100)$).

3.3 Quantitative Evaluation

Initially, the decay parameter γ defined in Eq. (2) was varied in the range of $\{1, 2, \dots, 9\}$ with the goal of estimating the most appropriate value. In order to avoid contaminating the training data, we performed this experiment on an independent dataset provided in [1], which contains 413 retinal fundus images and corresponding annotations for OD and fovea centers. The resulting mean distances for both landmarks are presented in Fig. 4. From this experiment, we selected $\gamma = 7$ as a good decay parameter, since it achieved a low error in combination with the lowest standard deviations in both tasks.

After fixing $\gamma = 7$, we proceed to train on the Messidor dataset as specified in the previous section. Table 1 presents the obtained results in comparison to the state-of-the-art in terms of R criteria and average normalized Euclidean distances. When considering the median values of \bar{D}_{FOV} and \bar{D}_R , OD and fovea detection reached a mean/median of $\bar{D}_R = 15.01/9.87$ and $\bar{D}_{FOV} = 0.94/0.70$ respectively, which comes close to human observer performance. The proposed technique achieves a high detection performance in the two tasks, and surpasses other state of the art methods in fovea detection for $1/2R$ and $1R$ criteria.

The generalization ability of the model was evaluated by means of a cross-dataset experiment on the DRIVE dataset [15]. Results for OD detection in this case were well-aligned with current techniques, meeting the $1R$ criterion in 97.5% of the cases. It is worth noting that the method failed only in one image,

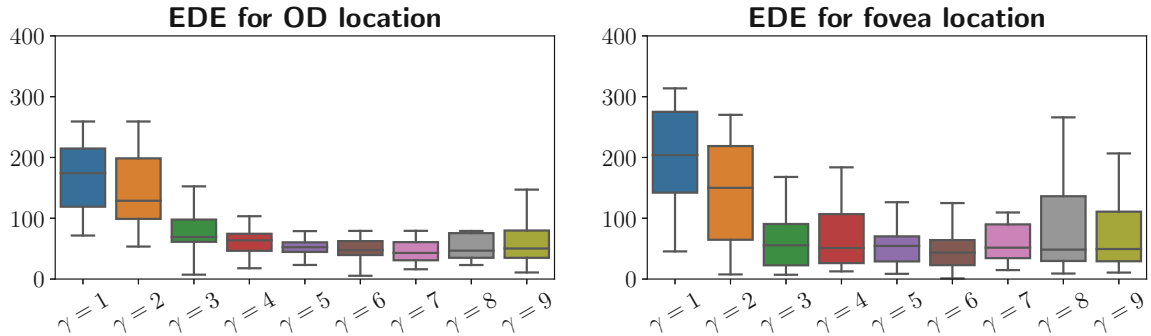


Fig. 4. Euclidean Distance Error (EDE) distributions in pixels for varying decay parameters.

Table 1. Performance comparison of the proposed model on OD and fovea detection.

Optic disc detection	No. images	1/8R	1/4R	1/2R	1R	\bar{D}_{FOV}	\bar{D}_R
Al-Bander et al. [2]	1200	—	83.6	95.00	97.00	—	—
Yu et al. [16]	1200	—	—	99.08	98.24	—	—
Marin et al. [9]	1200	87.33	97.75	99.50	99.75	—	7.03
Proposed Approach	1136	65.58	93.57	97.10	98.94	1.12	15.01
Fovea detection							
Gegundez-Arias et al. [6]	800	82.00	94.25	95.88	96.50	1.41	—
Yu et al. [16]	800	23.63	64.88	94.00	98.00	2.34	—
Niemeijer et al. [10]	800	76.88	93.25	96.00	97.38	1.87	—
Dashtbozorg et al. [3]	1200	—	66.50	93.75	98.87	—	—
Al-Bander et al. [2]	1200	—	66.80	91.40	96.60	—	—
Proposed Approach	1136	70.33	94.01	97.71	99.74	0.94	12.55
<i>Human observer</i> [6]	800	94.38	98.50	99.88	99.88	0.52	—

which contains a highly degenerated OD. Excluding this fail case, the method achieved a mean normalized distance \bar{D}_{FOV} from the OD centroid of 2.99.

4 Conclusions and Future Work

A novel approach for jointly detecting the OD and fovea centers has been presented. The proposed technique is based on regressing the distance from each pixel to the closest of both landmarks. A F-CNN is optimized to solve this task, achieving competitive results in OD detection and surpassing the current state-of-the-art in fovea location.

The idea of replacing single-point landmark location by distance map regression is not limited to the task of OD and fovea detection. Therefore, in the future, different medical image analysis problems involving similar detection challenges may benefit from an analogous approach.

Acknowledgments. This work is funded by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and the European Regional Development Fund (ERDF), within the project “NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016”. The Titan Xp used for this research was donated by the NVIDIA Corporation.

References

1. IDRiD - Indian Diabetic Retinopathy Image Dataset (2018). <https://idrid.grand-challenge.org/>
2. Al-Bander, B., Al-Nuaimy, W., Williams, B.M., Zheng, Y.: Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc. *Biomed. Signal Process. Control.* **40**, 91–101 (2018)
3. Dashtbozorg, B., Zhang, J., Huang, F., ter Haar Romeny, B.M.: Automatic optic disc and fovea detection in retinal images using super-elliptical convergence index filters. In: Campilho, A., Karray, F. (eds.) *ICIAR 2016*. LNCS, vol. 9730, pp. 697–706. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41501-7_78
4. Decencière, E., et al.: Feedback on a publicly distributed database: the Messidor database. *Image Anal. Ster.* **33**(3), 231–234 (2014)
5. Foracchia, M., Grisan, E., Ruggeri, A.: Detection of optic disc in retinal images by means of a geometrical model of vessel structure. *IEEE Trans. Med. Imaging* **23**(10), 1189–1195 (2004)
6. 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**, 386–393 (2013)
7. Hoover, A., Goldbaum, M.: Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. *IEEE Trans. Med. Imaging* **22**(8), 951–958 (2003)
8. Lalonde, M., Beaulieu, M., Gagnon, L.: Fast and robust optic disc detection using pyramidal decomposition and hausdorff-based template matching. *IEEE Trans. Med. Imaging* **20**(11), 1193–1200 (2001)
9. Marin, D., Gegundez-Arias, M.E., Suero, A., Bravo, J.M.: Obtaining optic disc center and pixel region by automatic thresholding methods on morphologically processed fundus images. *Comput. Methods Programs Biomed.* **118**(2), 173–185 (2015)
10. Niemeijer, M., Abràmoff, M.D., van Ginneken, B.: Fast detection of the optic disc and fovea in color fundus photographs. *Med. Image Anal.* **13**(6), 859–870 (2009)
11. Niemeijer, M., Abràmoff, M.D., Van Ginneken, B.: Segmentation of the optic disc, macula and vascular arch in fundus photographs. *IEEE Trans. Med. Imaging* **26**(1), 116–127 (2007)
12. 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. [arXiv:1505.04597](https://arxiv.org/abs/1505.04597)

13. Ruder, S.: An overview of multi-task learning in deep neural networks, June 2017. <http://arxiv.org/abs/1706.05098>
14. Sinthanayothin, C., Boyce, J.F., Cook, H.L., Williamson, T.H.: Automated localisation of the optic disc, fovea and retinal blood vessels from digital color fundus images. *Br. J. Ophthalmol.* **4**(83), 902–910 (1999)
15. Staal, J., Abramoff, M., Niemeijer, M., Viergever, M., van Ginneken, B.: Ridge based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **23**(4), 501–509 (2004)
16. Yu, H., et al.: Fast localization of optic disc and fovea in retinal images for eye disease screening. In: *SPIE Medical Imaging*, p. 796317 (2011)