Detection of Retinal Anatomical Structures and its Application to Image Quality Assessment

Rodrigo de Castro Michelassi and Nina S. T. Hirata
Instituto de Matemática, Estatística e Ciência da Computação
Universidade de São Paulo

***Abstract*—Given the general shortage of ophthalmologists in Brazil, automated retinal image screening emerges as a fundamental tool to promote early diagnosis of retinal diseases and prevent, for example, blindness. Motivated by this, we propose the use of YOLO-based object detection models to detect two retinal structures – the optic disc and the macular region – which play a crucial role in ophthalmological examinations. As a case study, based on detections performed on BRSET images, we determined the distances between the structures and the retinal edges. We found that the calculated distances do not always meet the minimum distance established in the Image Quality Assessment Image Field protocol. The results highlight the ability of YOLO-based models to reliably detect anatomical structures in retinal images and their potential to contribute to improving analyses that rely on these detections.**

I. Introduction

Retinal imaging is a common procedure in ophthalmological examinations and in monitoring eye conditions. It helps detect eye diseases such as diabetic retinopathy (DR), glaucoma, and others. Furthermore, as it provides a unique window into vascular networks, it is also important for assessing cardiovascular risks or abnormalities potentially related to systemic diseases such as hypertension, diabetes, and cerebrovascular disease [1].

This work is part of a broader project led by a team from the Faculty of Medicine of the University of São Paulo. The project aims to promote early diagnosis of DR in primary healthcare in Brazil, to reduce preventable blindness. Given the general shortage of eye specialists, the development of computational tools to assist in screening retinal images is of great importance. To advance in that direction, we focus on the detection of two anatomical structures, the optic disc and the macular region, using deep learning approaches. These two structures play crucial roles in eye examinations.

The detection of anatomical structures in retinal images has been widely studied in previous works. In [2], the authors propose a distance regression network, that calculates the distance from each pixel of the image to the optic disc and macular center jointly. More recent approaches, such as [3], use deep learning architectures for optic disc segmentation and detection. Other works focus on analyzing the angular relation between the macular center and the optic disc [4] or apply handcrafted image processing, exploring its visual features for detection [5], [6].

Since the retinal images collected in the aforementioned project so far have no image-level annotations, we decided to use publicly available datasets, repurposing their annotations to our goals. We chose the state-of-the-art YOLO model [7] for the detection tasks. After training and validation, we qualitatively evaluate the model in the Brazilian Multilabel Ophthalmological Dataset (BRSET) [8]. BRSET is a multilabel retinal dataset, containing annotations regarding several retinal pathologies, quality assessment parameters, anatomical classifications and other useful patient information, for 16,266 retinal images.

Additionally, considering that one of the first steps in assessing DR in retinal images is quality evaluation, we analyze one of the Image Quality Assessment (IQA) criteria in BRSET. The criterion we analyze is named "Image Field" and is described as "This parameter is graded as adequate when all the following requirements are met: 1) The optic disc is at least 1 disc diameter (DD) from the nasal edge; 2) The macular center is at least 2 DD from the temporal edge; 3) The superior and inferior temporal arcades are visible in a length of at least 2 DD" [8]. In particular, we compute the distances in (1) and (2) and see how well they agree with the IQA labels available on BRSET. The retinal structures mentioned are presented in Fig. 1.

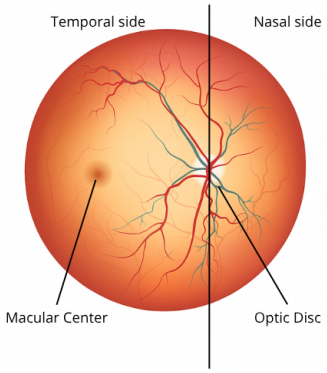

Fig. 1. The retinal structure, with the optic disc located on the nasal side, and the macular center located on the temporal side.

Our main contributions are summarized as follows:

- Adaptation of several datasets for optic disc and macular region detection.
- A state of the art performance on optic disc and macular region detections.
- Distance determination from the optic disc to the nasal edge, and from the macular center to the temporal edge.

- A statistical analysis of BRSET IQA annotations, regarding the image field criteria, with respect to the computed distances.

II. Methodology

A. Datasets

We use multiple publicly available retinal image datasets, listed next. Note that only one of the datasets has the annotations in the format needed for YOLO. Thus, we preprocessed them in order to generate bounding box annotations (center coordinates, width and height) from existing annotations, as detailed below.

a) REFUGE2 [9], Drishti-GS [10] and IDRiD for optic disc detection [11]: Since these datasets only contain segmentation masks for the optic disc and the mask is approximately circular, we extracted the extreme points on top, bottom, left and right sides of the mask, and generated the bounding box from those points.

b) G1020 [12]: We extracted only annotations regarding the optic disc bounding box and formatted according to the YOLO standards.

c) RIGA [13]: This dataset contains image annotations made by 6 different doctors, on top of the images. We randomly chose the annotations made by one of them and computed the difference between the original image and the annotated image. From that we just applied the same method used for REFUGE2. However, we discarded the BinRushed subset of RIGA because it resulted in noisy masks.

d) IDRiD for macular center detection: For this specific task, the IDRiD dataset has only annotations concerning the macular center coordinates. We followed the procedure described in [14] to generate heatmaps around the macular region, then binarized the heatmap mask and generated bounding box annotations following the procedure applied on REFUGE2. In fact, since we are only interested in the center of the macular region, we judged this would be an adequate approach to frame the problem as a detection problem. The authors of [14] use a Gaussian function, $H(x, y) = exp(-\frac{(x-\alpha)^2+(y-\beta)^2}{2\sigma^2})$, where (α, β) is the coordinate of the macular center, (x, y) is an image pixel coordinate, and σ is a constant used to adjust the Gaussian blob, to generate the heatmap. We adjusted σ to fit our image dimensions, based on the value chosen by the authors.

Table I summarizes the used datasets. For each dataset, the table shows the number of images after preprocessing, the type of annotation we used, and the detection task for which the dataset was used.

B. Training Procedure

We trained two YOLO detection models: one for optic disc and another for macular center detection. We trained both models for 40 epochs, using the Adam optimizer and starting with YOLO 11 weights [15]. All images were resized to 224 × 224. We also limited the number of detections in each image to only one.

TABLE I
Datasets used for Optic Disc and Macular Center detection

Dataset	Images	Annotation	Detection Target
Refuge2	1200	Segmentation mask	Optic disc
G1020	1020	Bounding boxes	Optic disc
Drishti-GS	101	Segmentation mask	Optic disc
RIGA	555	Optic disc boundaries	Optic disc
IDRiD	81	Segmentation mask	Optic disc
IDRiD	516	Fovea keypoints	Macular center

For optic disc detection, we considered two training scenarios:

- Scenario 1: We separated 70% of the available data from the 5 datasets for training, and the remaining 30% was equally separated into validation and test datasets.
- Scenario 2: To evaluate a cross-dataset situation, we separated the entire MESSIDOR dataset, a subset of RIGA, from our training set, and used it entirely for testing. For the rest of the data, we used 80% for training and 20% for validation, keeping approximately the same amount of images for each set as in Scenario 1.

For macular center detection, we just used the original IDRiD dataset training/testing split.

C. Distances Determination

Given the locations of the optic disc and the macular region, we compute three measures: the angle between the centers of the optic disc and the macular region, the distance from the optic disc to the nasal edge, and the distance from the macular center to the temporal edge. The nasal and temporal edges are, respectively, the retinal edges on the nasal and temporal sides.

The angle was calculated adapting the implementation made available in [4]. For the distances to the edges, since we observe some high angles between the centers of the retinal structures in BRSET images, we decided to consider the distance along the line segment that passes through the centers of both structures. See Fig. 2.

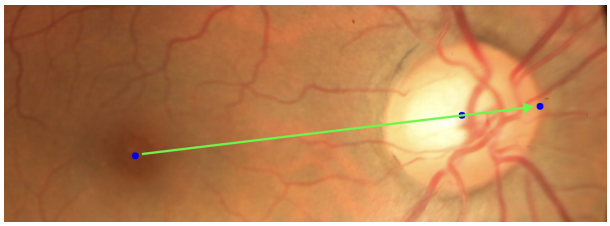

Fig. 2. Extended unit vector $\vec{U}$ and points representing, from left to right, the macular center, the optic disc center and the nasal point.

We first calculate the unit vector $\vec{U}$ between both center coordinates. We also assume that the optic disc can be approximated by an ellipse, to define a point we refer to as *nasal point* — the point on the optic disc boundary closest to the nasal edge, as shown in Fig. 2. Although not a standard anatomical term, we use nasal point throughout this work to

denote this geometrically defined location, calculated using the parametric equation of the ellipse:

$$x_{nasal} = c_x + \frac{optic\ disc\ width}{2} \times cos(\theta) \quad (1)$$

$$y_{nasal} = c_y + \frac{optic\ disc\ height}{2} \times sin(\theta) \quad (2)$$

in which (c_x, c_y) are the center coordinates of the optic disc, and θ is the angle of the unit vector $\vec{U}$ and the x axis. The optic disc width and height, as well as the center coordinates, are given by the bounding box generated by the models.

To compute the distances, we consider the intersection points N and T between the line that extends the vector $\vec{U}$ in both directions and the retinal edges, respectively, on the nasal and temporal sides. Then, the distance between the optic disc and the nasal edge is computed as the Euclidean distance between N and the nasal point. Similarly, the distance between the macular center and the temporal edge is computed as the Euclidean distance between the macular center and T.

III. Experiments and Results

A. Optic Disc and Macular Region Detection

We perform detections using the models trained according to the hyperparameter settings described in the previous section. Recall (see Table I) that for optic disc we considered two training scenarios, using images from five datasets, while for the macular region we trained only one model using images from just one dataset. We considered a detection when the model confidence was at least 50%, i.e., $c \geq 0.5$. The test results for precision, recall and mean average precision (mAP50 and mAP50-95) are given in Table II.

TABLE II
Test results for optic disc and macular region detections (%)

Training task	Precision	Recall	mAP50	mAP50-95
Scenario 1	98.7	99.5	99.0	88.2
Scenario 2	99.3	99.3	98.9	87.1
Macular Region	93.5	93.2	96	75.2

As can be seen in Table II, even though the Scenario 2 model is tested on a dataset that was not used for training, it still achieves high performance. This indicates that the model is generalizing well. We show in Fig. 3 some detection results on BRSET. As can be seen, the bounding boxes are placed properly on the optic disc and the macular region.

B. BRSET Analysis

Based on visual inspection, we assessed that the optic disc and macular region were properly detected on BRSET by the corresponding trained models (Scenario 2 and Macular Region detection models). Therefore, we considered it appropriate to calculate some distance statistics from the detection results.

First we computed the angle between the line segment that links the centers of the optic disc and the macular region whenever the model confidence level c satisfied $c \geq 0.5$, as previously mentioned. The angle distribution is shown in Fig. 4. As can be seen, there is a peak around 8 degrees, which is considered normal. Instances with very large angles could hint violation of the image field protocol.

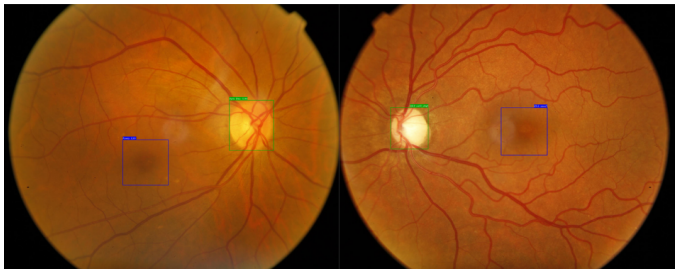

Fig. 3. Macular center (blue) and optic disc (green) detections by the YOLO models, in BRSET images.

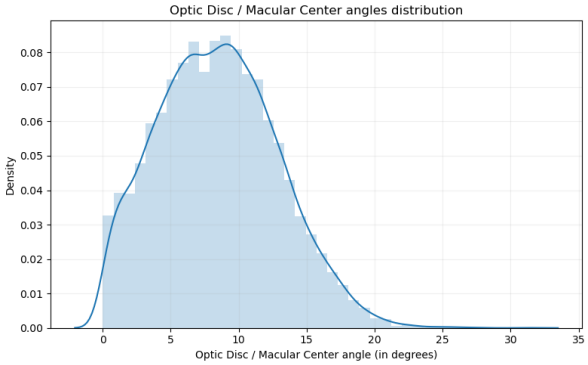

Fig. 4. Optic Disc/Macular Center angle distribution on BRSET.

To analyze the distances to the retinal edges, we considered only images whose structures were detected with a high confidence level (greater than 75% and 80%, respectively, for the macular region and the optic disc). A total of 11,384 images met this criterion. We normalize the distances based on disc diameter (DD) and compute the distance distribution separated into two subgroups of images: one whose "Image Field" is labeled *Adequate* and the other which is labeled *Inadequate*.

With respect to the *Adequate* images, contrary to the specification of the IQA protocol, the histogram in Fig. 5 shows that there is a considerable number of images with the distance between the optic disc and the nasal edge lower than 1DD. In contrast, Fig. 6 shows that most of the images labeled *Adequate* present the distance between the macular center and the temporal edge at least 2DD, as expected.

Regarding the group of *Inadequate* images, in Fig. 5 we see that most of them do not satisfy the 1DD distance requirement, as expected. Those few that have distance above 1DD might have been labeled *Inadequate* because they do not meet the second criterion (distance at least 2DD).

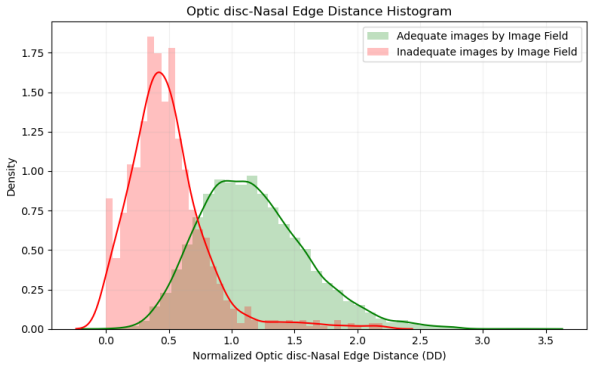

Fig. 5. Distribution of the normalized distances between the optic disc and the nasal edge in the adequate/inadequate "Image Field" subgroups.

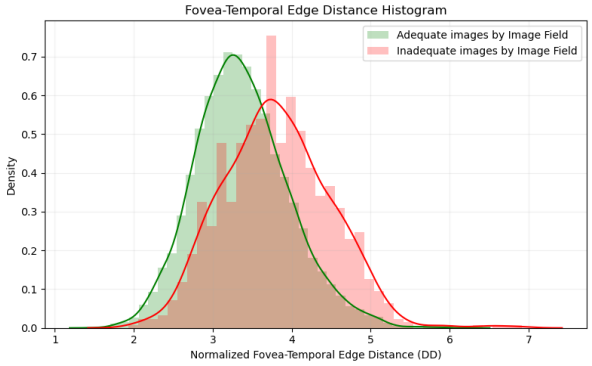

Fig. 6. Distribution of adequate/inadequate images considering the distance between macular center and temporal edge.

IV. Conclusion

In this work we trained and evaluated models to detect the optic disc and the macular region in retinal images. Results indicate that YOLO-based object detection models, trained with a diverse set of datasets, lead to a high performance for anatomical retinal structure detection, allowing such structures to be accurately identified even in cross-dataset evaluation scenarios. Furthermore, we computed the distances between these structures and retinal edges for BRSET images. We observed some disagreement between the computed distances and the distance requirements regarding the optic disc and the nasal edge in images labeled as "Adequate". Nevertheless, according to experts, the crucial information for the diagnosis of diabetic retinopathy is in the macular region and, therefore, it is not uncommon to consider some criteria less strictly. As future work, we would like to generate annotations of structures such as optic disc and macular region locations and segmentation masks for the images in BRSET. We would also like to extend the coverage of image quality assessment to other quality criteria.

Acknowledgment

The authors thank São Paulo Research Foundation (FAPESP) – grant #2022/15304-4, and MCTI (Ministério da Ciência, Tecnologia e Inovações, Brazil), law 8.248, PPI-Softex - TIC 13 - 01245.010222/2022-44.

References

[1] P. Burlina, A. Galdran, P. Costa, A. Cohen, and A. Campilho, "Chapter 18 - artificial intelligence and deep learning in retinal image analysis," in *Computational Retinal Image Analysis*, E. Trucco, T. MacGillivray, and Y. Xu, Eds. Academic Press, 2019, pp. 379–404. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780081028162000198

[2] M. I. Meyer, A. Galdran, A. M. Mendonça, and A. Campilho, "A Pixel-Wise Distance Regression Approach for Joint Retinal Optical Disc and Fovea Detection," in *Medical Image Computing and Computer Assisted Intervention – MICCAI*, A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, Eds. Springer International Publishing, 2018, pp. 39–47.

[3] T. Araújo, G. Aresta, A. Galdran, P. Costa, A. M. Mendonça, and A. Campilho, "UOLO - Automatic Object Detection and Segmentation in Biomedical Images," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer International Publishing, 2018, pp. 165–173.

[4] H. Kang, S. Lee, and A. Lee, "Measuring ocular torsion and its variations using different nonmydriatic fundus photographic methods," *PLOS ONE*, vol. 15, p. e0244230, 12 2020.

[5] M. A. P. Da Silva, M. S. M. Mafalda, A. B. Alvarez, and R. F. L. Chavez, "Optic Disc Localization from Retinal Fundus Image Using Discrete Cosine and Hough Transforms," in *IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE)*, 2023, pp. 134–139.

[6] J. Singh, G. D. Joshi, and J. Sivaswamy, "Appearance-based object detection in colour retinal images," in *15th IEEE International Conference on Image Processing*, 2008, pp. 1432–1435.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, 2016, pp. 779–788.

[8] L. F. Nakayama, D. Restrepo, J. Matos, L. Z. Ribeiro, F. K. Malerbi, L. A. Celi, and C. S. Regatieri, "BRSET: A Brazilian Multilabel Ophthalmological Dataset of Retina Fundus Photos," *PLOS Digital Health*, vol. 3, no. 7, pp. 1–16, 07 2024. [Online]. Available: https://doi.org/10.1371/journal.pdig.0000454

[9] H. Fang, F. Li, J. Wu, H. Fu, X. Sun, J. Son, S. Yu, M. Zhang, C. Yuan, C. Bian, B. Lei, B. Zhao, X. Xu, S. Li, F. Fumero, J. Sigut, H. Almubarak, Y. Bazi, Y. Guo, Y. Zhou, U. Baid, S. Innani, T. Guo, J. Yang, J. I. Orlando, H. Bogunović, X. Zhang, and Y. Xu, "REFUGE2 Challenge: A Treasure Trove for Multi-Dimension Analysis and Evaluation in Glaucoma Screening," 2022. [Online]. Available: https://arxiv.org/abs/2202.08994

[10] J. Sivaswamy, S. R. Krishnadas, G. Datt Joshi, M. Jain, and A. U. Syed Tabish, "Drishti-GS: Retinal image dataset for optic nerve head (ONH) segmentation," in *IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, 2014, pp. 53–56.

[11] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian Diabetic Retinopathy Image Dataset (IDRiD)," 2018. [Online]. Available: https://dx.doi.org/10.21227/H25W98

[12] M. N. Bajwa, G. A. P. Singh, W. Neumeier, M. I. Malik, A. Dengel, and S. Ahmed, "G1020: A Benchmark Retinal Fundus Image Dataset for Computer-Aided Glaucoma Detection," 2020. [Online]. Available: https://arxiv.org/abs/2006.09158

[13] A. Almazroa, S. Alodhayb, E. Osman, E. Ramadan, M. Hummadi, M. Dlaim, M. Alkatee, K. Raahemifar, and V. Lakshminarayanan, "Retinal fundus images for glaucoma analysis: the RIGA dataset," in *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*, J. Zhang and P.-H. Chen, Eds., vol. 10579, International Society for Optics and Photonics. SPIE, 2018, p. 105790B. [Online]. Available: https://doi.org/10.1117/12.2293584

[14] B. Bhargav Bhatkalkar, V. Nayak S, S. Shenoy, and R. Arjunan, "FundusPosNet: A Deep Learning Driven Heatmap Regression Model for the Joint Localization of Optic Disc and Fovea Centers in Color Fundus Images," *IEEE Access*, vol. PP, pp. 1–1, 11 2021.

[15] G. Jocher and J. Qiu, "Ultralytics YOLO11," 2024. [Online]. Available: https://github.com/ultralytics/ultralytics