

Optic Disc Segmentation in Fundus Images with Deep Learning Object Detector

Manal AlGhamdi

Department of Computer Science, University of Umm AL-Qura, Makkah, Saudi Arabia

Article history

Received: 20-03-2020

Revised: 09-04-2020

Accepted: 08-05-2020

Email: maalghamdi@uqu.edu.sa

Abstract: The Optic Disc (OD) is an important anatomical landmark in the fundus image to diagnose a myriad of diseases, such as glaucoma and Diabetic Retinopathy (DR) and to locate structures such as the macula and the main vascular arcade. However, locating and segmenting the OD are not easy tasks. Previous methods have employed a deep Convolutional Neural Network (CNN) without any need for hand-crafted features. Among these methods, RetinaNet has recently attracted attention as a simple one-stage object detector that performs quickly and efficiently while achieving state-of-the-art results. RetinaNet has proven its efficiency in multiple conventional object detection tasks with a larger training set that contains a sufficient number of diverse cases which are beyond reach in medical tasks. Thus, we propose an OD segmentation model from fundus images based on RetinaNet extension with DenseNet that addresses the vanishing gradient problem, enhances feature propagation, performs deep supervision, strengthens feature reuse and reduces the number of parameters. The experimental results using three publicly available databases show the efficacy of deep object detection network and the dense connectivity when applied to fundus images, which is a promising step in providing a segmentation to detect patients in the early stages of the disease.

Keywords: DenseNet, RetinaNet, Deep Learning, Image Segmentation, Detection

Introduction

In fundus images, the Optic Disc (OD) is characterized by the bright, yellow and approximately elliptical area where the vessels are thick and dense (Sopaharak *et al.*, 2008). An accurate OD localization and segmentation is an important step in diagnosing a variety of retinal diseases, such as glaucoma and optic disc pit and to check for any neovascularization at the optic disc as this is a manifestation of diabetic retinopathy (Jonas *et al.*, 1988; Joshi *et al.*, 2011). The OD is also used to locate other structures such as the fovea and the blood vessels, as shown in Fig. 1. In a healthy retina, the OD has clear and well-defined edges, while blurry and elevated in pathological one as presented in Fig. 2. The majority of the recent studies have addressed this challenging task by using the traditional segmentation techniques. However, reliable results are difficult to achieve using such methods, especially in relation to illness. As an alternative, machine learning-based approaches using Convolutional Neural Networks (CNNs) have been presented.

CNNs have achieved the highest performance for many medical image segmentation tasks. The standard CNN framework begins by applying pre-trained networks to large datasets as feature extractors, followed by a classifier to detect the object (Asiri *et al.*, 2019). However, these methods use a CNN as a component with other pre-processing and classification steps. Thus, it is a slow-speed model which prevents it from being used in real-time systems and, in addition, its architecture is too complicated, so it could not be learned as a whole one. To eliminate these shortcomings, end-to-end object detectors were proposed, combining separate stages of previous work into one CNN and converging this task with the regression problem, which enables this network to be learned as a whole from beginning to end.

For object detectors, two methodologies can be applied to generate object predictions. The first is known as the two-stage object detector, where the detection is performed in two stages. First, a regional proposal network is applied to subtract the objects from the background regardless of their classes, followed by bounding-box regression to process the extracted regions.

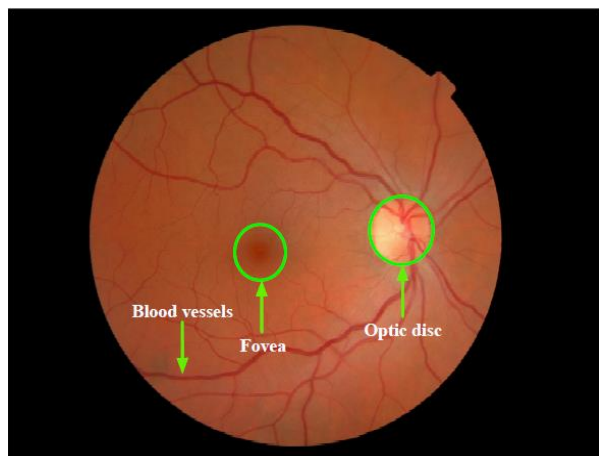


Fig. 1: The blood vessels, optic disc and fovea in a fundus image

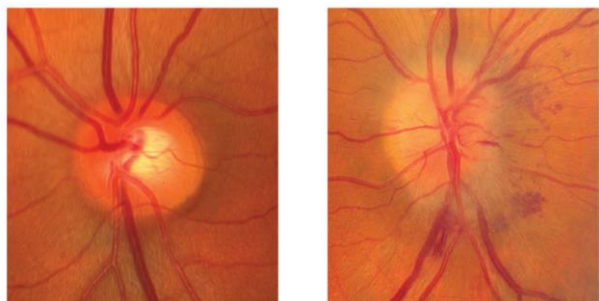


Fig. 2: Screening for the Optic Disc (OD) where disc margins are sharp and clear in (left) a healthy eye and blurry and elevated in (right) pathological eye (Newman and Biousse, 2014)

This type of detector is more widely utilized; however, its complex network architecture causes the training and inference to be less efficient (He *et al.*, 2017; Dai *et al.*, 2016). The second methodology is known as the one-stage object detector and this skips the region extraction stage and applies the detection model directly over possible locations. OverFeat (Sermanet *et al.*, 2013), Single Shot multiBox Detector (SSD) (Liu *et al.*, 2016) and You Only see Once (YOLO) (Redmon *et al.*, 2016) are one-stage detectors that have been well studied due to their fast processing; however, they are limited in terms of accuracy. Recently, RetinaNet (Lin *et al.*, 2017b) has been proposed as a one-stage object detector and this method demonstrates high performance by utilizing the focal loss function. It solves the problem of the traditional cross-entropy loss function, while maintaining the efficiency of the processing, which is the main feature of one-stage object detectors. However, achieving the desirable results with

RetinaNet is restricted by the availability of large annotated training sets which is beyond reach in medical tasks. With small training set, the model suffers from the problem of vanishing gradients where the gradients of the loss function approach zero, making the network hard to train (Glorot and Bengio, 2010). An alternative solution, the DenseNet (Huang *et al.*, 2017) was presented with a dense block that performs an iterative concatenation of all previous feature maps. This facilitates the reuse of computation and improves gradient flow, leading to improved accuracy and easier training of deep network.

Inspired by the promising results achieved using the DenseNet (Huang *et al.*, 2017) in many biomedical segmentation problems and the RetinaNet (Lin *et al.*, 2017b) in object detection tasks, this paper presented the DestinaNet that extended the RetinaNet with dense blocks to automatically detect OD in fundus images. Using both models help to facilitate the reuse of computation and improves gradient flow, leading to improved accuracy and easier training of deep networks. The presented model efficiently segments the OD from a retinal fundus image in a simple one-stage, end-to-end manner based on the deep object detection architectures. Unlike most of the previous studies in this field, the model uses a whole image as an input and outputs a segmentation result as a final result without any pre- or post-processing steps to finalize the optic disc area. The remainder of this paper is organized as follows: The proposed framework is introduced in section 2, datasets, experimental set-ups and results are presented in section 3, we discuss our results in section 4 and conclude the paper in section 5.

Methods

The proposed method comprises two parts: OD localization and OD segmentation. Fig. 3 shows the flowchart for the proposed method.

DetinaNet Architecture

RetinaNet (Lin *et al.*, 2017b) is a one-stage object detector that outperforms the state-of-the-art methods in both accuracy and running time. It advances the other one-stage object detectors by utilizing a simple and efficient novel loss function that enables the detector to focus more on difficult samples. This loss function, called focal loss, updates the standard cross entropy loss to over-weight the losses from harder classes over easier classes. This addresses the class imbalance problem that causes the performances of one-stage detectors to fail against the two-stage detectors.

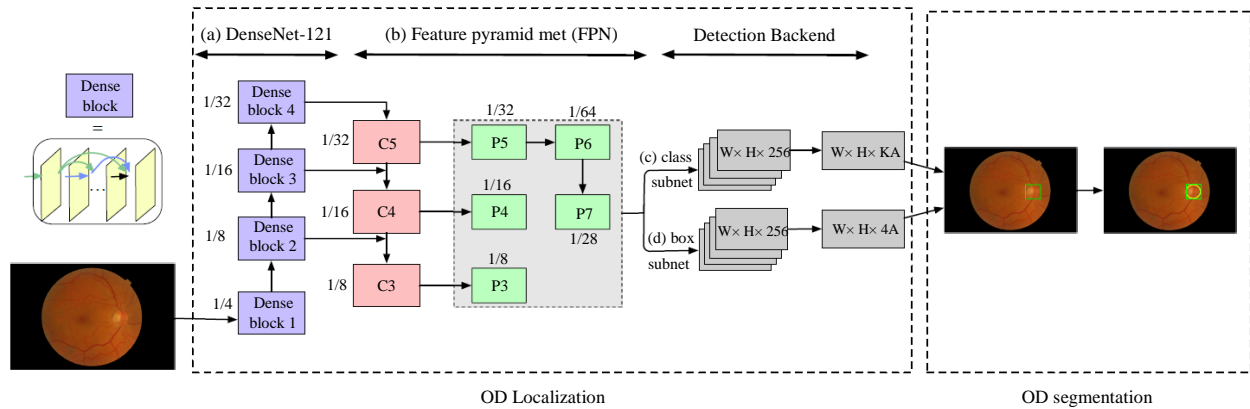


Fig. 3: Flowchart of the proposed DetinaNet model based on the architecture of RetinaNet Lin *et al.* (2017b), the Feature Pyramid Network (FPN) Lin *et al.* (2017a) and the DenseNet Huang *et al.* (2017). (a) The backbone network was chosen to be the DenseNet-121. (b) The multi-scale convolutional feature pyramid generated from the backbone. (c) The first subnetwork used for anchor boxes classification. (d) The second subnetwork used for regression between the anchor boxes and the ground-truth boxes

The model consists of a backbone network and two subnetworks. The backbone network employs the ResNet (He *et al.*, 2016) to compute the convolutional feature maps from the input image. Using the output from the backbone network, the first subnetwork is an object classification while the second is a convolutional bounding-box regression. RetinaNet’s architecture is simpler than those of the two-stage object detectors that contain separate networks for classification and regression. However, with small datasets it is easy to encounter the vanishing gradient problem. Thus, we employ the DenseNet (Huang *et al.*, 2017) as the backbone network which solves this problem leading to better performance. The architecture of DetinaNet is presented in Fig. 3 and each component is described below.

Backbone Network

The architecture of the FPN (Lin *et al.*, 2017a) was adopted to develop the backbone network on top of the DenseNet-121 (Huang *et al.*, 2017). In the Retina, the output of the l th layer is calculated by a non-linear activation function as:

$$x_l = H_l([x_{l-1} + x_{l-2}]) \quad (1)$$

where, x_l is the output of l^{th} layer, x_{l-1} is the output of $(l-1)^{th}$ layer, x_{l-2} is the output of $(l-2)^{th}$ layer and H is defined as the summation of the two preceding feature maps followed by Rectified Linear Unit (ReLU). While, the DenseNet (Huang *et al.*, 2017) contains dense blocks that use all the preceding feature maps in its output as:

$$x_l = H_l([x_{l-1}, x_{l-2}, \dots, x_1, x_0]) \quad (2)$$

where, $[\dots]$ represents the concatenation operation and H_l is defined as a function of three consecutive operations: Batch Normalization (BN), followed by a Rectified Linear Unit (ReLU) and a 3×3 convolution (conv). The FPN constructs a multi-scale convolutional feature pyramid by combining the standard convolutional network using a top-down architecture and lateral connections (Fig. 3a and 3b). The developed pyramid consists of 5 levels from P_3 to P_7 computed using the output of the corresponding dense blocks C_3 to C_5 . Each level of the pyramid detects an object with a different scale.

Anchor

At each level of the pyramid, $A = 9$ translation-invariant anchors were used. Each anchor was associated with two vectors. The first one is a K length vector of classification targets, where K represents the number of classes. Meanwhile, the second vector is a 4-dimensional vector of box regression targets. For the assignment of the anchors to the object boxes, the same procedure from the Region Proposal Network (RPN) of Faster R-CNN (Ren *et al.*, 2015) was adopted.

Class Subnet

Each level of the FPN is connected with a Fully Convolutional Network (FCN) to compute the probability of object presence at each spatial position of each anchor and each object class. Given a feature map with 256 channels from one pyramid level, the class subnet utilizes four 3×3 convolution layers with 256 filters each and ReLU activation and another 3×3 convolution layer with $A \times K$ filters. Finally, the binary predictions are produced using a sigmoid activation function (Fig. 3c).

Box Subnet

Similar to the structure of the class subnet, the box subnet is defined for each pyramid level to compute regressions of the existing offset between each anchor box and its neighboring ground-truth box and to generate $4 \times A$ linear outputs for each spatial location, as shown in Fig. 3d.

Focal Loss

The goal of the Focal Loss (FL) is to focus on the gap between classes that represents the foreground and background during the training of the one-stage object detectors. Given the ground-truth class $y \in \{\pm 1\}$ and the model's prediction $p \in [0, 1]$ for the class with label $y = 1$ as:

$$p_i = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \quad (3)$$

The focal loss function extended the Cross Entropy (CE) loss function as:

$$CE(p_i) = -\log(p_i) \quad (4)$$

One of the CE loss function properties is that there is a loss value for all the samples including the ones which are easy to classify, which would be successful when the training dataset is balanced (Lin *et al.*, 2017b). However, the RetinaNet model explicitly produces more negative samples from the background to be trained and to be more able to distinguish actual objects and the background. Therefore, the CE loss function is unsuitable for the one-stage detectors.

In the one-stage methods, the loss is computed using all produced samples and actual objects, which might output a larger value for the easily classified samples of background than for the more difficult classified samples of the objects (Lin *et al.*, 2017b). This means that the majority of the loss value is caused by the easy common samples for the background, which are less significant.

To overcome this issue, the focal loss function introduces a modulating factor $\alpha(1-p_i)^\gamma$ into the CE loss function, as follows:

$$FL(p_i) = -\alpha(1-p_i)^\gamma \log(p_i) \quad (5)$$

where, α is the weight factor defined according to the class and γ is the modulating factor. Adding this modulating factor allows the one-stage object detectors to calculate the loss with more concentration on the samples that receive less loss.

Training and Inference Procedure

RetinaNet was developed as an object detector which needed to be classified into one of 80 object classes

following the task of the COCO challenge (Lin *et al.*, 2014). However, the DetinaNet for OD detection task classified the ROI into one of only two binary classes: OD and background. Thus, the number of the classes k used by the class subnet is a set of two – that is $k = 2$. The backbone network is chosen to be the DenseNet-121 (He *et al.*, 2016) which has a depth of 121.

In our model, the data augmentation is used to artificially expand the training samples and prevent overfitting with respect for obvious invariances for the detection task at hand (e.g., flipping an image should not affect the existence of the disease) (Pratt *et al.*, 2016). Four types of data augmentation were employed: Vertical shift, which randomly shifts the images up and down using the nearest pixel to fill the blank; horizontal shift, which randomly shifts the images left and right using the nearest pixel to fill the blank; horizontal flip, which randomly flips the image horizontally; and random zoom-in of the image. These affine transformations are the most effective when applied to fundus images since they do not affect the signs of the disease in the image (Chen *et al.*, 2015). In addition, they help the model to have a better understanding of the input image since it views the images in many transformed views including scaling. All the images, including the augmented data, are used as an input to our OD detector for a single inference.

Segmentation Generation

The DetinaNet model predicts a score per bounding box in the input image. In the non-medical image processing, the Non-Maximum Suppression (NMS) (Rosenfeld and Kak, 1982) is an integral part of the object detection pipeline to reduce redundancy by finding the local maxima in an image. Since the OD detection task from the retinal fundus image will produce one and only one object, NMS is not employed and the model will produce the bounding box with the highest confidence score.

In order to generate a more accurate OD boundary from the bounding box, the following points were considered. Firstly, the structure of the OD has a bright, yellow and elliptical region in the color fundus images; thus, an ellipse shape can be used to approximate the shape of the OD. Secondly, the generated bounding box used parameters including width, height and central point coordinates, which are exactly the same as the ones needed for the vertical ellipse. Therefore, the vertical ellipse around the OD boundary is simply redrawn from the predicted bounding box without any post-processing steps.

Experiments and Results

The proposed method is evaluated on three public eye fundus datasets with both healthy and pathological images. These datasets were released for OD segmentation evaluation and thus all the images have ground-truth for OD segmentation. The datasets include:

- Messidor image dataset (Decenciere *et al.*, 2014), which consists of 1, 200 fundus images acquired with 45° Field-Of-View (FOV). The images are 1440×960, 2240×1488 or 2304×1536 pixels in size and 8 bits per color plane. The OD boundary was manually delimited by two experts and used as a gold standard for the evaluation
- Drishti-GS dataset (Sivaswamy *et al.*, 2014), which consists of a total of 101 images centered on the OD with a FOV of 30° and 2896×1944 pixels in size and PNG uncompressed image format. The OD boundaries were annotated using four experts
- DRIONS-DB (Carmona *et al.*, 2008) consists of 110 colour digital retinal images, with 600×400 pixels in size and 8 bits per color plane. The OD boundaries were annotated by two experts using a software tool provided for image annotation

Evaluation Criteria

The performance metrics used to evaluate the proposed OD segmentation model are chosen to target the area of overlap between the ground-truth segmented OD (G) and the automated segmented OD (S) and the distance in pixels between them, as illustrated in Fig. 4. Given a test image, the following values are calculated: The True Positive (TP) is the correctly detected region of OD by the automated algorithm, the False Positive (FP) is the incorrectly detected region of OD by the automated algorithm, the False Negative (FN) is the region of the ground-truth OD missed by the automated segmentation and finally the True Negative (TN) is the region within the retinal region that is not the OD. After that, the performance metrics to evaluate the effectiveness of the proposed OD segmentation model are calculated as follows:

- Sensitivity (SEN) which is the portion of the ground- truth OD area detected by the automated algorithm, where the higher value means better segmentation It is calculated as:

$$SEN = \frac{TP}{TP + FN} \quad (6)$$

- Accuracy (Acc), which is the ratio of truly identified pixels to the total number of pixels in the retinal region, where the higher value means better segmentation. It is calculated as:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

- Intersection-over-Union (IoU), which is the portion of overlapping region between the ground-truth and automated segmented OD, where the higher value means better segmentation. It is calculated as:

$$IoU = \frac{G \cap S}{G \cup S} \quad (8)$$

Moreover, a tradeoff between the true positive rate and false positive rate, producing ROC (Swets, 1988) curves with AUC, is also used to compare the performances of both RetinaNet and DetinaNet.

Experiment Set-Up

We used a NVIDIA GTX980M 8GB GPU card with a 1536 CUDA parallel-processing core for implementation. The model was developed using the open source framework Tensor Flow (Abadi *et al.*, 2016).

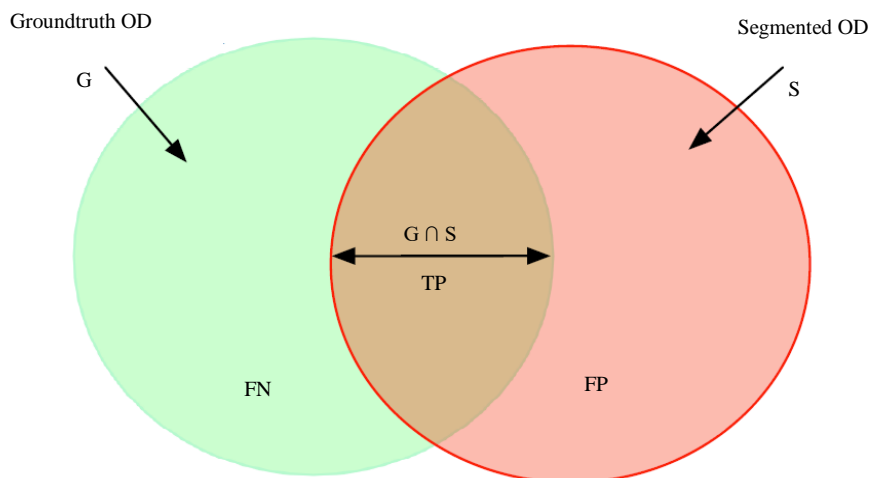


Fig. 4: The performance metrics used to evaluate the proposed OD segmentation model

Our whole framework was implemented with Python based on Keras API (Chollet, 2015) with Tensorflow backend. For data augmentation, a range between [-20, 20] was used for vertical and horizontal shift and a range between [0, 0.5] was used for the zoom-in. For each dataset, the classifier was trained in a 5-fold cross-validation using 80% and evaluated on 20% of the images with batch size 20 using the Adam optimizer (Kingma and Ba, 2014) at a small learning rate of 0.001. The values for focal loss parameters in Equation 4 are set similarly to Lin *et al.* (2017b), i.e., $\alpha = 0.25$ and the $\gamma = 2$.

Results

Fig. 5 presents the ROC curves calculated using our DetinaNet and the RetinaNet (Lin *et al.*, 2017b). Comparing these curves, it is clear that the DetinaNet outperform the RetinaNet by fusing the advantages of both densely connections and end-to-end object detector. The dense connectivity mitigates the problem of overfitting in such small dataset and encourages the model to learn a more discriminative set of features. As shown, our DetinaNet model achieved FPR = 0.2 per image at a 80% TPR detection rate and FPR = 0.4 per image at an 90% TPR detection rate. While at the same FPR = 0.2 per image, RetinaNet (Lin *et al.*, 2017b) achieved only 60% TPR.

The comparison between the proposed OD boundary segmentation algorithm and the related methods is presented in Table 1. The results show that the proposed model outperformed the state-of-the-art OD segmentation models in all three datasets. Unlike the existing work, the presented model is simple to extend to

different applications since it does not require pre-processing steps, hand-crafted features or an extensive parameter for tuning. The existing methods based on traditional segmentation techniques such as morphological operations (Marin *et al.*, 2015; Roychowdhury *et al.*, 2016), Sliding Band Filters (SBF) Dashtbozorg *et al.* (2015) and a variational model (Dai *et al.*, 2017) achieved lower results than the other studies which are based on deep learning techniques. This is because these methods are based on the shape and color features of the bright area around OD, which do not work well on patients with uncommon and damaged structure.

Examples of the OD segmentation performance using the presented method on the images for DR and glaucoma patients from the Messidor datasets are shown in Fig. 6. The ground-truth annotations are represented in green boundaries and the automated predictions in blue boundaries. Fig. 6a to 6c show successful examples with a low number of false positives. Accurate OD boundaries were detected at the exact position of the annotated area. However, Fig. 6d to 6f show fail images that have an excessive number of false positives. The shape of the OD in Fig. 6d is expanded and with an uncommon and un-round shape due to the glaucoma disease. Bright spots which look similar to the OD are also misidentified as the OD in Fig. 6e and 6f. Images of the patients who have severe DR are likely to have this type of error. An insufficient number of abnormal ODs in the dataset is a limitation of our OD detection model.

Table 1: Comparison of the proposed segmentation method results with related methods, '-' means not available

Method	Accuracy	Sensitivity	IoU
Messidor Decenciere <i>et al.</i> (2014)			
Our Work	99.6	0.98	0.95
Xu <i>et al.</i> (2017)	99.43	-	-
Alghamdi <i>et al.</i> (2016)	99.20	-	-
Niu <i>et al.</i> (2017)	98.75	-	-
Kim <i>et al.</i> (2019)	-	0.97	0.94
Marin <i>et al.</i> (2015)	-	0.92	0.87
Roychowdhury <i>et al.</i> (2016)	-	0.92	0.84
Dai <i>et al.</i> (2017)	-	-	0.91
Dashtbozorg <i>et al.</i> (2015)	-	-	0.89
Teresa <i>et al.</i> Araujo <i>et al.</i> (2018)	-	-	0.88
Drishti-GS Sivaswamy <i>et al.</i> (2014)			
Our Work	99.8	0.94	0.92
Al-Bander <i>et al.</i> (2018)	99.6	0.92	0.90
Zilly <i>et al.</i> (2017)	-	-	0.91
Sevastopolsky (2017)	-	-	0.75
DRIONS-DB Carmona <i>et al.</i> (2008)			
OurWork	99.7	0.93	0.91
Al-Bander <i>et al.</i> (2018)	99.6	0.92	0.89
Sevastopolsky (2017)	-	-	0.89

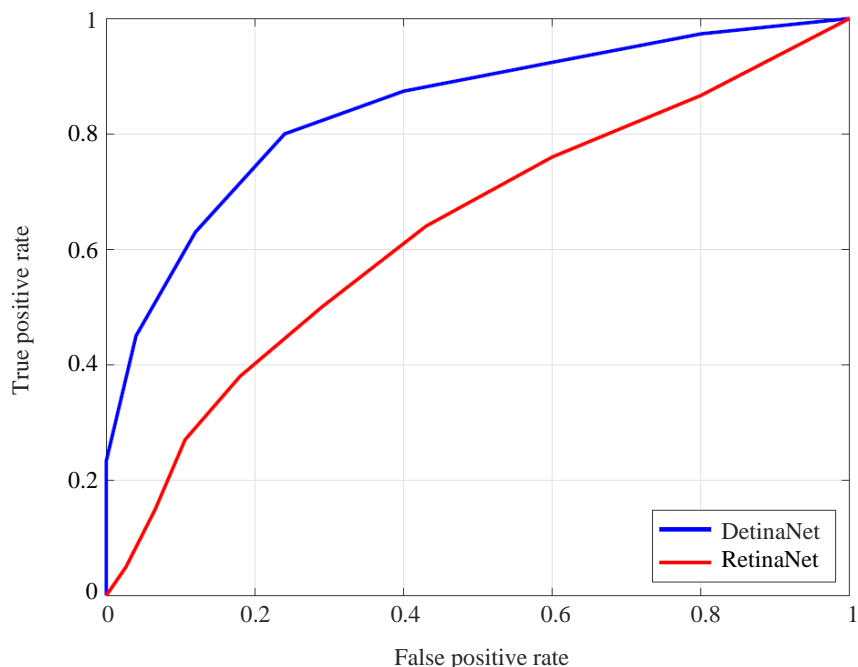


Fig. 5: Evaluating the DetinaNet model performance for OD segmentation in terms of ROC curves

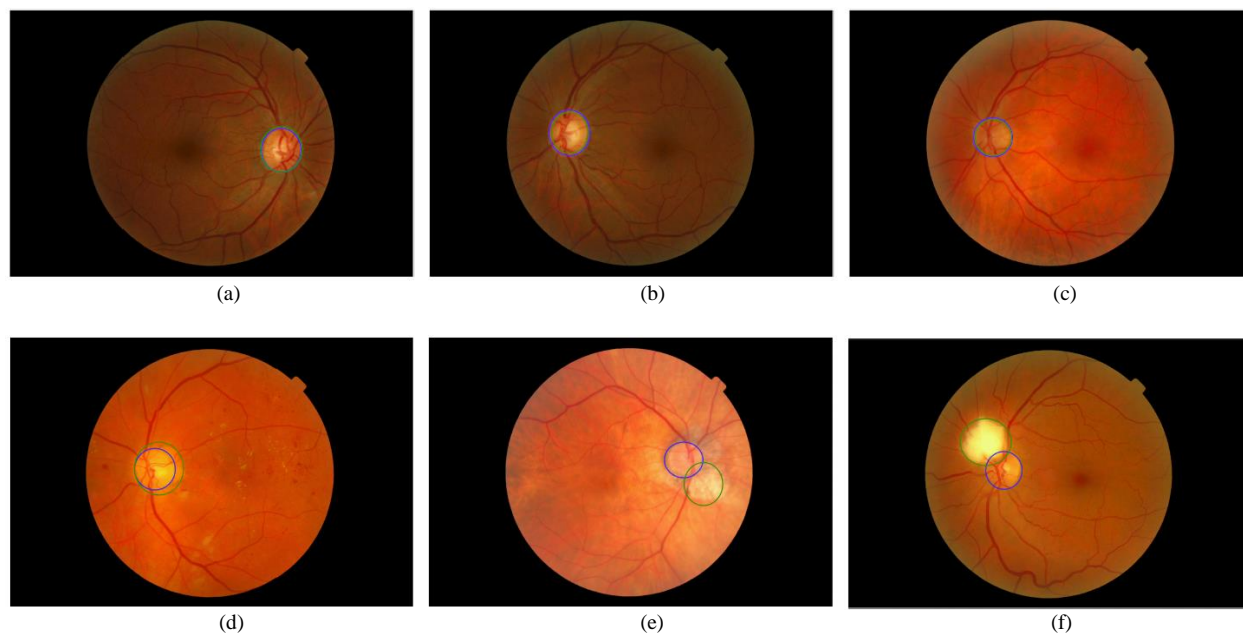


Fig. 6: Examples of the OD segmentation performance using the presented method on the images for DR and glaucoma patients from the Messidor datasets: (a,b,c) are samples of best results, (d,e,f) are samples from the worst ones. The green boundaries represent the ground-truth while the blue boundaries represent the automated predictions by DetinaNet

Discussion

We proposed DetinaNet as an OD detection and segmentation model based on RetinaNet, which is the current state-of-the-art region-based deep learning object detector. It efficiently segments the OD from a retinal

fundus image in a simple one-stage, end-to-end manner based on the deep object detection architectures. The model was evaluated on three public eye fundus datasets, including Messidor, Drishti-GS and DRIONS-DB and as demonstrated above it outperforms the stat-of-the-art results with considerable margins. Experiments on three

datasets with reasonable diversity demonstrates that the presented model is invariant to variations such as population, camera, operators, disease and image. Among these datasets the presented model achieves the best results on the Messidor dataset. This is due to the availability of a large number of annotated images in this dataset, compared with the other two, which helps to improve the performance of the model.

As presented in Table 1, the presented model achieves a high sensitivity values comparable with that of the recent U-Net-based OD detection models (Zilly *et al.*, 2017; Al-Bander *et al.*, 2018; Kim *et al.*, 2019), because our method is a simple end-to-end one-stage object detector. In addition, these models have depended on pre-processing steps to extract the OD-centered region under the assumptions that the algorithm can accurately detect the ROI (Al-Bander *et al.*, 2018; Kim *et al.*, 2019), or detect the OD and the OC separately Zilly *et al.* (2017). It also achieves an overwhelmingly higher sensitivity compared with the conventional models with deep learning-based techniques (Srivastava *et al.*, 2015; Niu *et al.*, 2017; Alghamdi *et al.*, 2016; Xu *et al.*, 2017). This type of method works better in image classification tasks, which require learning the deepest features of the scene, than in object detection tasks. Compared with the other object detector-based models such as YOLOv2 Araujo *et al.* (2018) and RCNN Zhang *et al.* (2018), we also performed better thanks to the dense connections that addresses the problem of limited size training sample. This is an extremely promising outcome for the deployment of such model in clinical environments and its uses in assessing the early detection of diseases.

Conclusion

In this study, the task of OD segmentation was presented as an object detection task and thus a DetinaNet was presented to segment the OD from retinal fundus images. The model was developed based on promising results achieved by the RetinaNet and the DenseNet in many object detection problems. Combining both models facilitates the reuse of computation through dense connections and improves gradient flow. Experimental results using three different publicly available datasets with reasonable diversity show the effectiveness of the presented method which outperforms existing methods and achieves state-of-the-art OD segmentation accuracy. In the future, investigating other deep object detectors will be considered as well as extending the application to optic cup segmentation.

Acknowledgement

The author would like to thank the University of Umm AlQura (UQU) for their support.

Ethics

This article is original and contains unpublished material.

References

- Abadi, M., P. Barham, J. Chen, Z. Chen and A. Davis *et al.*, 2016. Tensorflow: A system for large-scale machine learning. Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation, (SDI' 16), pp: 265-283.
- Al-Bander, B., B. Williams, W. Al-Nuaimy, M. Al-Tae and H. Pratt *et al.*, 2018. Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis. *Symmetry*, 10: 87-87. DOI: 10.3390/sym10040087
- Alghamdi, H., H.A. Lilian Tang, S. Waheeb and T. Peto, 2016. Automatic optic disc abnormality detection in fundus images: A deep learning approach. Proceedings of the Medical Image Computing and Computer-Assisted Interventions (CAI' 16), OMIA, Greece, pp: 17- 24. DOI: 10.17077/omia.1042
- Araujo, T., G. Aresta, A. Galdran, P. Costa and A.M. Mendonca *et al.*, 2018. UOLO - Automatic Object Detection and Segmentation in Biomedical Images. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Stoyanov, D., Z. Taylor, G. Carneiro, T. Syeda-Mahmood and A. Martel *et al.* (Eds.), Springer International Publishing, pp: 165-173. DOI: 10.1007/978-3-030-00889-5_19
- Asiri, N., M. Hussain, F.A. Adel and N. Alzaidi, 2019. Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. *Arti. Intell. Med.*, 99: 101701-101701. DOI: 10.1016/j.artmed.2019.07.009
- Carmona, E.J., M. Rincon, J.G. Feijoo and J. Casa, 2008. Identification of the optic nerve head with genetic algorithms. *Arti. Intell. Med.*, 43: 243-259. DOI: 10.1016/j.artmed.2008.04.005
- Chen, X., Y. Xu, D.W.K. Wong, T.Y. Wong and J. Liu, 2015. Glaucoma detection based on deep convolutional neural network. Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Aug. 25-29, IEEE Xplore Press, Milan, Italy, pp: 715-718. DOI: 10.1109/EMBC.2015.7318462
- Chollet, F., 2015. Keras. <https://keras.io>.
- Dai, B., X. Wu and W. Bu, 2017. Optic disc segmentation based on variational model with multiple energies. *Patt. Recogn.*, 64: 226-235. DOI: 10.1016/j.patcog.2016.11.017

- Dai, J., Y. Li K. He and J. Sun, 2016. R-FCN: Object Detection Via Region-Based Fully Convolutional Networks. In: *Advances in Neural Information Processing Systems*, Lee, D.D., M. Sugiyama, U.V. Luxburg, I. Guyon and R. Garnett (Eds.), pp: 379-387.
- Dashtbozorg, B., M. Mendonca and A. Campilho, 2015. Optic disc segmentation using the sliding band filter. *Comput. Biol. Med.*, 56: 1-12.
DOI: 10.1016/j.compbiomed.2014.10.009
- Decenciere, E., X. Zhang, G. Cazuguel, B. Lay and B. Cochener *et al.*, 2014. Feedback on a publicly distributed database: The messidor database. *Image Analy. Stereol.*, 33: 231-234.
DOI: 10.5566/ias.1155
- Glorot, X. and Y. Bengio, 2010. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, (AIS' 10), pp: 249-256.
- He, K., G. Gkioxari, P. Dollár and R. Girshick, 2017. Mask R-CNN. *Proceedings of the International Conference on Computer Vision*, Oct. 22-29, IEEE Xplore Press, Venice, Italy, pp: 2980-2988.
DOI: 10.1109/ICCV.2017.322
- He, K., X. Zhang, S. Ren and J. Sun, 2016. Deep residual learning for image recognition. *Proceedings of the Conference on Computer Vision and Pattern Recognition*, Jun. 27-30, IEEE Xplore Press, Las Vegas, NV, USA, pp: 770-778.
DOI: 10.1109/CVPR.2016.90
- Huang, G., Z. Liu, L.V.D. Maaten and K.Q. Weinberger, 2017. Densely connected convolutional networks. *Proceedings of the Conference on Computer Vision and Pattern Recognition*, Jul. 21-26, IEEE Xplore Press, Honolulu, HI, USA, pp: 2261-2269.
DOI: 10.1109/CVPR.2017.243
- Jonas, J.B., G.C. Gusek and G.O.H. Naumann, 1988. Optic disk morphometry in high myopia. *Graefe's Archive Clin. Exp. Ophthalmol.*, 226: 587-590.
DOI: 10.1007/BF02169209
- Joshi, G.D., J. Sivaswamy and S.R. Krishnadas, 2011. Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. *IEEE Tran. Med. Imag.*, 30: 1192-1205.
DOI: 10.1109/TMI.2011.2106509
- Kim, J., L. Tran, E.Y. Chew, S. Antani and G.R. Thoma, 2019. Optic disc segmentation in fundus images using deep learning. *SPIE Med. Imag.*, 10954: 103-108. DOI: 10.1117/12.2512798
- Kingma, D.P. and J. Ba, 2014. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.
- Lin, T., P. Dollár, R. Girshick, K. He and B. Hariharan *et al.*, 2017a. Feature pyramid networks for object detection. *Proceedings of the Confer. Computer Vision and Pattern Recognition*, (VPR' 17), IEEE Xplore Press, Belongie, pp: 936-944.
DOI: 10.1109/CVPR.2017.106
- Lin, T., P. Goyal, R. Girshick, K. He and P. Dollár, 2017b. Focal loss for dense object detection. *Proceedings of the International Conference on Computer Vision*, (CCV' 17), IEEE Xplore Press, pp: 2999-3007. DOI: 10.1109/ICCV.2017.324
- Lin, T.Y., M. Maire, S. Belongie, J. Hays and P. Perona *et al.*, 2014. Microsoft COCO: Common objects in context. *Proceedings of the European Conference on Computer Vision*, Springer, pp: 740-755.
DOI: 10.1007/978-3-319-10602-1_48
- Liu, W., D. Anguelov, D. Erhan, C. Szegedy and S. Reed *et al.*, 2016. SSD: Single shot multibox detector. *Proceedings of the European Conference on Computer Vision*, Springer, pp: 21-37.
DOI: 10.1007/978-3-319-46448-0_2
- Marin, D., M.E. Gegundez-Arias, A. Suero and J.M. Bravo, 2015. Obtaining optic disc center and pixel region by automatic thresholding methods on morphologically processed fundus images. *Comput. Meth. Prog. Biomed.*, 118: 173-185.
DOI: 10.1016/j.cmpb.2014.11.003
- Newman, N.J. and V. Biousse, 2014. Diagnostic approach to vision loss. *Continuum*, 20: 785-815.
DOI: 10.1212/01.CON.0000453317.67637.46
- Niu, D., P. Xu, C. Wan, J. Cheng and J. Liu, 2017. Automatic localization of optic disc based on deep learning in fundus images. *Proceedings of the 2nd International Conference on Signal and Image Processing*, Aug. 4-6, IEEE Xplore Press, Singapore, p: 208-212.
DOI: 10.1109/SIPROCESS.2017.8124534
- Pratt, H., F. Coenen, D.M. Broadbent, S.P. Harding and Y. Zheng, 2016. Convolutional neural networks for diabetic retinopathy. *Proc. Comput. Sci.*, 90: 200-205.
DOI: 10.1016/j.procs.2016.07.014
- Redmon, J., S. Divvala, R. Girshick and A. Farhadi, 2016. You only look once: Unified, real-time object detection. *Proceedings of the Conference on Computer Vision and Pattern Recognition*, (VPR' 16), IEEE Xplore Press, pp: 779-788.
DOI: 10.1109/CVPR.2016.91
- Ren, S., K. He, R. Girshick and J. Sun, 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems*, (IPS' 15), pp: 91-99.
- Rosenfeld, A. and A.C. Kak, 1982. *Digital Picture Processing*. 2nd Edn., Academic Press, Inc.
- Roychowdhury, S., D.D. Koozekanani, S.N. Kuchinka and K.K. Parhi, 2016. Optic disc boundary and vessel origin segmentation of fundus images. *J. Biomed. Health Inform.*, 20: 1562-1574.
DOI: 10.1109/JBHI.2015.2473159

- Sermanet, P., D. Eigen, X. Zhang, M. Mathieu and R. Fergus *et al.*, 2013. Overfeat: Integrated recognition, localization and detection using convolutional networks. Proceedings of the International Conference on Learning Representations, (CLR' 13).
- Sevastopolsky, A., 2017. Optic disc and cup segmentation methods for glaucoma detection with modification of U- Net convolutional neural network. *Patt. Recognit. Image Anal.*, 27: 618-624.
DOI: 10.1134/S1054661817030269
- Sivaswamy, J., S.R. Krishnadas, G.D. Joshi, M. Jain and A.U.S. Tabish, 2014. Drishti-GS: Retinal image dataset for Optic Nerve Head (ONH) segmentation. Proceedings of the 11th International Symposium on Biomedical Imaging, Apr. 29-May 2, IEEE Xplore Press, Beijing, China, pp: 53-56.
DOI: 10.1109/ISBI.2014.6867807
- Sopharak, A., B. Uyyanonvara, S. Barman and T.H. Williamson, 2008. Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods. *Comput. Med. Imag. Graph.*, 32: 720-727.
DOI: 10.1016/j.compmedimag.2008.08.009
- Srivastava, R., J. Cheng, D.W.K. Wong and J. Liu, 2015. Using deep learning for robustness to parapapillary atrophy in optic disc segmentation. Proceedings of the 12th International Symposium on Biomedical Imaging, Apr. 16-19, IEEE Xplore Press, New York, USA, pp: 768-771.
DOI: 10.1109/ISBI.2015.7163985
- Swets, J., 1988. Measuring the accuracy of diagnostic systems. *Science*, 240: 1285-1293.
DOI: 10.1126/science.3287615
- Xu, P., C. Wan, J. Cheng, D. Niu and J. Liu, 2017. Optic disc detection via deep learning in fundus images. Proceedings of the International Workshop on Ophthalmic Medical Image Analysis, (MIA' 17), Springer, pp: 134-141.
DOI: 10.1007/978-3-319-67561-9_15
- Zhang, D., W. Zhu, H. Zhao, F. Shi and X. Chen, 2018. Automatic localization and segmentation of optical disk based on faster R-CNN and level set in fundus image. *Medical Imaging*.
- Zilly, J., J.M. Buhmann and D. Mahapatra, 2017. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. *Comput. Med. Imag. Graph.*, 55: 28-41.
DOI: 10.1016/j.compmedimag.2016.07.012