doi:10.3233/SHTI230575

A Deep Learning-Based Approach Towards Simultaneous Localization of Optic Disc and Fovea from Retinal Fundus Images

Mohammad Tariqul ISLAM^a, Ferdaus AHMED^b, Mowafa HOUSEH^c and Tanvir ALAM^{c,1}

^aSouthern Connecticut State University, Connecticut, USA ^bInfosys Ltd., Texas, USA

^cCollege of Science and Engineering, Hamad bin Khalifa University, Doha, Qatar

Abstract. In this work, we propose a multi-task learning-based approach towards the localization of optic disc and fovea from human retinal fundus images using a deep learning-based approach. Formulating the task as an image-based regression problem, we propose a Densenet121-based architecture through an extensive set of experiments with a variety of CNN architectures. Our proposed approach achieved an average mean absolute error of only 13pixels (0.04%), mean squared error of 11 pixels (0.005%), and a root mean square error of only 0.02 (13%) on the IDRiD dataset.

Keywords. Retina, Fovea, Optic Disc, Qatar Biobank

1. Introduction

The human retina plays an important role in the diagnosis of a number of diseases such as diabetes [1], diabetic retinopathy (DR), cardiovascular disease [2], and other diseases. Optic disc (OD) and Fovea are integral parts of the human eye and their localization is a central part of computer aided screening tools in ophthalmology clinics [3,4]. Accurately localizing the optic disc (OD) and fovea is a non-trivial task, as they may not always be easily distinguishable due to low contrast with the surrounding background. Additionally, the fovea exhibits a high degree of variability in terms of its shape, structure, and boundaries, and is often quite small and difficult to discern with the naked eye. Consequently, deep learning techniques have been proposed to improve the precision of OD and fovea localization. In this present article, we propose a deep learning-based approach to locate the fovea and the OD in a retinal fundus image.

¹ Corresponding Author: Tanvir Alam, E-mail: talam@hbku.edu.qa.

2. Materials and Methods

2.1. Dataset

We model the problem as a supervised learning — regression task that accepts a digital retinal fundus image as the input and predicts the coordinates of the optic disc and fovea centers. We considered the IDRiD dataset that contains labeled data for fovea and optic disc center locations. It contains retinal fundus images and three types of corresponding ground truth labels: (a) multi-class segmentation masks for retinal phenomena and bodies, (b) diabetic retinopathy and diabetic macular edema severity grade, and (c) optic disc center and fovea locations as (x,y) coordinates. We used the third category of labels in our task. The IDRiD dataset contains 516 RGB retinal images in total, each of size 4288 pixels (width) by 2848 pixels (height).

2.2. Model Development and Evaluation

Given an RGB retinal fundus image I, we use a convolutional neural network to perform regression on I in predicting the location of the fovea, F and the optic disc, D, each of which are represented in (x, y) coordinates in image-space. Our proposed network architecture is based on the DenseNet architecture, specifically, the 121-layer variant of the network. The network is divided into two segments - the stem, which contains all but the last classification layer from the network pretrained on ImageNet-1k and the head, which has the classifier segment containing the following layers: a batch normalization (1024), dropout (0.25), linear (1024), ReLU, batch normalization (512), dropout (0.5), linear (512), and sigmoid (-1, 1) where the information within the parenthesis indicate the hyper-parameters for the layers. Using a pre-trained network ensured that we could leverage transfer learning as (i) training such a deep neural network would require a much larger dataset to prevent overfitting, and (ii) benefit from the knowledge the network gained by being trained on a dataset containing more than a million images. To compare our proposed network architecture for the task with other widely-used alternatives, we experimented with six other CNNs. Specifically, we finetuned AlexNet, VGG, GoogleNet, ResNet, SqueezeNet, and EfficientNet. Table 1 shows the variants we used of these networks, where applicable. The networks were fine-tuned for 50 epochs, with a maximum learning rate of 3e-2 that changed according to the cosine annealing scheduler. To prevent the networks from overfitting, we used L2 regularization with lambda=1e-2 and also applied early stopping with a minimum improvement of 0.001 for the monitoring metric (MSE) and a patience of 20 iterations. We used the Mish activation function in the inner layers of the network due to its superior performance to ReLU and Leaky ReLU. The outcome of the experiments were statistically stabilized by making use of nested K-fold cross validation with K = 5 for both the inner and the outer folds. We fine-tuned these networks on a workstation with a AMD Ryzen 5800X CPU, 64 GB of DDR4 memory, an Nvidia RTX 3090 GPU, and 1TB SSD. For faster turnaround time, we also utilized two other less powerful computers that run older GPUs (TitanX Pascal and RTX 2080) to tune the hyperparameters. The runtime of each fold of the experiment ranged from 30 minutes to several hours depending on the candidate model. We considered mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) as evaluation metrics. In addition, since we proposed a joint estimation of two independent quantities, namely the optic disc and the fovea centers, we also computed these metrics on these independent targets to understand how the

marginal error distribution behaves. Finally, we went a level deeper and also estimated the performance per coordinate (x and y) of the locations to further investigate the outcome.

3. Results

Table 1 shows the performances of the candidate models on the test set in pixel-space, respectively. We can see that the proposed Densenet121-based network outperforms the other models by a significant margin across all modes of the metrics. We used RMSE as the metric for selecting the best candidate model based on their validation set performances due to this metric's ability to measure error in the same space as the input, unlike other metrics such as MSE that represents error in the quadratic space.

Network	MAE	MSE	RMSE	RMSE (FOV)	RMSE (OD)	RMSE (F_X)	RMSE (F_Y)	RMSE (O_X)	RMSE (O_Y)
Efficientnet_ b3a	8.11 ± 0.15	0.66 ± 0.14	12.17 ± 1.29	12.47 ± 1.17	11.87 ± 1.42	$\begin{array}{c} 12.02 \pm \\ 1.54 \end{array}$	$\begin{array}{c} 12.89 \pm \\ 0.83 \end{array}$	14.46 ± 2.94	$\begin{array}{c} 8.18 \pm \\ 1.08 \end{array}$
Alexnet	11.70 ± 1.57	1.26 ± 0.45	$\begin{array}{c} 16.52 \pm \\ 3.06 \end{array}$	17.02 ± 2.92	15.99 ± 3.22	15.43 ± 1.8	18.44 ± 3.89	14.43 ± 2.65	17.41 ± 3.71
VGG16	26.69± 15.44	7.80 ± 6.57	36.65± 20.09	28.91± 11.95	42.60 ± 26.46	28.05 ± 12.95	29.68 ± 11.04	51.12 ± 35.51	30.88 ± 14.22
DenseNet121	7.60 ± 0.77	0.463 ± 0.13	10.06 ± 1.52	11.64 ± 1.92	8.19 ± 1.00	10.91 ± 0.97	12.27 ± 2.78	7.66 ± 0.25	8.53 ± 2.16
SqueezeNet1 _0	24.46 ± 3.30	5.95 ± 0.61	36.46 ± 1.89	27.08 ± 2.99	43.83 ± 1.29	26.94 ± 5.09	27.07 ± 0.92	49.79 ± 5.55	36.28 ± 4.49
ResNet34	9.10 ± 0.01	0.73 ± 0.14	12.77 ± 1.23	13.82 ± 1.83	11.61 ± 0.52	12.15 ± 0.57	15.04 ± 3.84	9.44 ± 1.31	13.26± 1.86
GoogleNet	11.38 ± 2.08	1.10 ± 0.45	$\begin{array}{c}15.37\pm\\3.31\end{array}$	14.71 ± 2.71	15.99 ± 3.87	13.45 ± 1.46	15.82 ± 3.80	16.32 ± 4.26	15.65 ± 3.46

Table 1. Pixel-space (in range [0, 448]) performance of the candidate models.

A visual inspection of the predictions from our proposed network showed that it was able to accurately locate the fovea and the optic disc centers in most cases. Figure 1 shows a subset of the test set where the predicted and the ground truth locations were overlaid on the corresponding input image.



Figure 1. Predictions (green marks) and ground truth (red marks) locations overlaid on retinal images. Circles and Plus represent the location of the optic disc and the fovea, respectively.

4. Discussion

Figure 2 shows the distribution and the histogram of the RMSE for the centers of the fovea, optic disc, and their individual (x/y) coordinate. We can see from Figure 2a that vertical error for estimating the fovea center has a wider and slightly different distribution, indicating our model faced a challenge estimating this quantity accurately. This has contributed directly to the distribution of the distance from fovea, seen from the same plot — it has a higher mean and variance than the optic disc distance distribution. However, the means for all the distances, both overall for each center and their horizontal and vertical components were less than 15 pixels. It is to be noted that for an input image of size 448x448, this accounts for less than 3.5% error. The histogram in Figure 2b also corroborates our observation on the error distribution — almost 80% of the errors in estimating the fovea location were below 20 pixels, whereas 90% predictions on the optic disc centers also exhibit similar characteristics. The right skewed distribution of the errors in the histogram shows that most of the errors were low, indicating that our proposed approach was effective in the task.



Figure 2. The distribution of RMSE (distance from ground truth) on the left and a pixel-space histogram of the same on the right.

5. Conclusion

Using an extensive set of experiments, we showed that proposed DenseNet based model is superior to other candidate models by evaluating across multiple metrics for regression tasks. Our future plan includes integrating this method as a diagnostic tool for diseases such as diabetes and DR, using retinal image data gathered from Qatar Biobank (QBB).

References

- Islam MT, Al-Absi HRH, Ruagh EA, Alam T. DiaNet: A Deep Learning Based Architecture to Diagnose Diabetes Using Retinal Images Only. IEEE Access, 2021. (accessed April 13, 2023).
- [2] Al-Absi HRH, Islam MT, Refaee MA, Chowdhury MEH, Alam T. Cardiovascular Disease Diagnosis from DXA Scan and Retinal Images Using Deep Learning. Sensors 2022;22.
- [3] Hasan MK, Alam MA, Elahi MTE, Roy S, Martí R. DRNet: Segmentation and localization of optic disc and Fovea from diabetic retinopathy image. Artif Intell Med 2021;111:102001.
- [4] Niemeijer M, Abràmoff MD, van Ginneken B. Fast detection of the optic disc and fovea in color fundus photographs. Med Image Anal 2009;13:859–70.