Fundus Image Analysis for Glaucoma: REFUGE-2020 Challenge Report

Sharath M Shankaranarayana

Zasti.AI
sharath@zasti.ai

Abstract. Glaucoma is one of the commonly occurring ocular diseases which could potentially lead to permanent loss in vision. Color fundus imaging is one the most common modalities used for the imaging the retina. Using the color fundus images, the assessment of glaucoma is performed by the examination of the optic nerve head and the region around it. For the computer aided diagnosis of glaucoma, we propose a deep learning based framework for three major tasks - localization of optic nerve head, image based classification for glaucoma, semantic segmentation of optic disc and cup.

Keywords: Fundus image analysis· Glaucoma · deep learning · classification · detection · segmentation.

1 Methodology

1.1 Task 1: Glaucoma Classification

The first task of the challenge requires us to predict the probability of glaucoma for a given retinal fundus image. We follow a similar process followed in [5]. The dataset consists of 1200 training images of which 120 have glaucoma and the rest 1180 do not have glaucoma. To overcome the inherent imbalance in the dataset, we resort to data augmentation along with oversampling of the glaucoma fundus images. Since glaucoma is characterized by abnormalities in optic nerve head (ONH) region, we also crop out the ONH region. For data augmentation, we employ the following techniques:

1. Random flipping and rotation
2. Photometric distortion
3. Specific histogram based image processing techniques such as histogram equalization, adaptive histogram equalization, intensity rescaling at different levels, histogram matching by randomly selecting a few cannonical images from the validation set.

With the augmented dataset, we then employ multiple pretrained deep convolutional neural networks (CNNs) for the task of binary classification. The classification networks employed are:

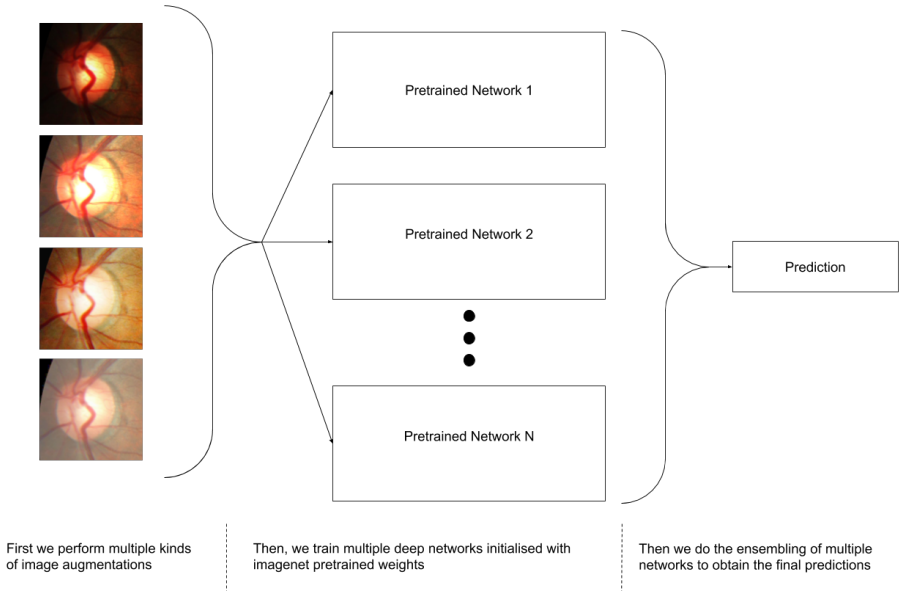

Fig. 1. Classification pipeline

1. EfficientNet [9]: EfficientNets are recently introduced class of networks and they employ a model scaling method to scale up CNNs in a more structured way and they have been shown to surpass the performance of other deep networks on ImageNet dataset and with better efficiency. We use multiple classes of EfficientNets- EfficientNet-B4, EfficientNet-B5, EfficientNet-B6, EfficientNet-B7. We use them because they are optimized for training at larger resolutions (380, 456, 528 and 600, respectively) when compared to standard 224x224 resolution of other Imagenet pretrained networks.
2. Inception-Resnet [8]: We use the Inception-Resnet architecture since it combines the two most commonly used blocks- inception block, which helps in multiscale feature extraction and residual block, which helps in faster convergence and alleviating vanishing gradients. These two blocks help in better feature extraction.
3. Resnext [10]: Resnext is another highly modularized network architecture for image classification tasks, which has also proved to be state-of-the-art in ImageNet classification task. Along with ImageNet pretrained Resnext, we also use pretrained weights obtained by weakly supervised learning on the Instagram dataset [3].
4. Squeeze and Excitation networks [2]: We use this class of network architectures since they consist of "Squeeze-and-Excitation" (SE) blocks that tend to generalize well across different datasets.

Finally, with all the networks trained for the glaucoma classification task, we ensemble the network predictions using simple averaging of posterior probabilities (as shown in Fig. 1)

1.2 Task 2: Optic Disc and Cup Segmentation

For the second task, we train a deep network for the task of semantic segmentation of OD and OC. We employ the same methodology proposed in our previous works [6] for training. But instead of using the RGB color channels, we use inverted green channel images as proposed in [7] since the inverted green channel images provides better contrast for the ONH region and the background. We use adversarial training setting and use ResUnet as our base architecture as mentioned in the paper [6]. The reader is advised to refer [6] for details. Once we train the network and predict the segmentation maps for the given retinal images, we perform some post-processing operations on the predicted maps. We first keep only the largest connected component in the segmentation maps and remove the other smaller components. And later we apply convex-hull operator to obtain the final segmentation map.

1.3 Task 3: Fovea Localization

This task requires us to predict the point coordinates of the location of fovea. We follow the same procedure followed in [5]. Instead of treating it as a standard coordinates regression problem, we convert this into an image to image translation problem. For this, from the information of ground truth point coordinates, we create distance maps having the same size as the fundus images. Distance maps are using the euclidean distance transforms computed from the fovea location i.e., the distances farther from the fovea have higher values than the distances closer to the fovea. For the purpose of easier training, we normalize the distance map and then invert so that the distances nearer to fovea has higher values. We later truncate it such that it only contains a specific radius around the fovea. This is done to improve training and is shown in figure Fig. 2.We then treat it as an paired image to image translation problem [11] as we did with od segmentation task similar to our work in [6]. The overall block diagram of generative adversarial network (GAN) framework used for image translation task is shown in figure Fig. 3

We train the generator to learn a mapping from the input fundus image x to the fovea distance map $y : G : x \rightarrow y$. We train the discriminator to distinguish between the generated distance map and real distance map:

$$L_{GAN}(G, D) = \mathbb{E}_{x,y}[log(D(y))] + \mathbb{E}_x[log(1 - D(G(x))] \quad (1)$$

where $\mathbb{E}_{x,y}$ represents the expectation of the log-likelihood of the pair (x, y) being sampled from the underlying probability distribution of real pairs of input fundus images and groundtruth distance maps.

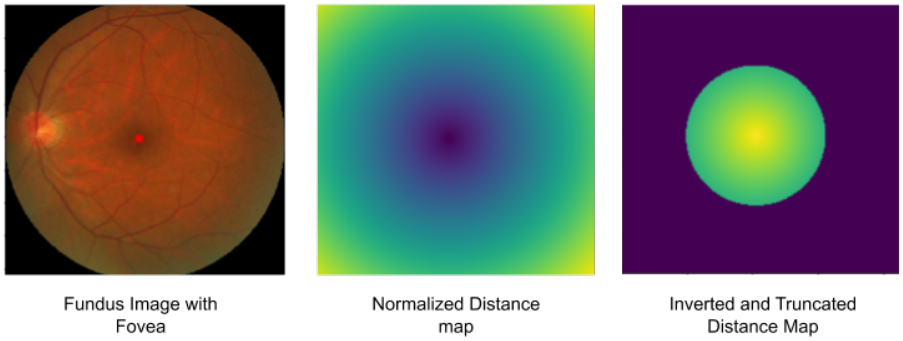

Fig. 2. Process of creating distance maps

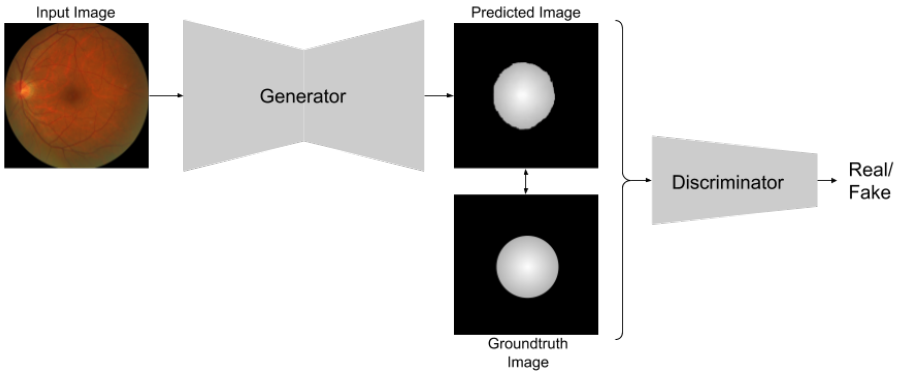

Fig. 3. Overall block diagram of fovea distance map regression

Additionally, we also use L_1 loss between the generator predicted distance map and groundtruth distance map and therefore the final objective becomes

$$G^* = arg\min_G \max_D (L_{GAN}(G, D) + \lambda(L_{L1}(G)) \tag{2}$$

where λ balances the contribution of two losses. In equation (2), the discriminator tries to maximize the expression by classifying whether the distance map is real or generated. The generator tries to minimize both adversarial loss and $L1$ loss in equation (2).

The architecture for generator part of GAN is described in the next section since the base architecture is same for this task as well as the lesions segmentation task. The discriminator uses a conventional CNN architecture used for classification. Let Ln denote a Convolution-BatchNorm-ReLU layer with n filters. The discriminator uses the following architecture
L64-L128-L256-L512-L512-L512

Finally, with the predicted distance maps from the generator, we extract the fovea point coordinates by performing a post-processing operation. Ideally, we could just take the pixel with highest intensity as the fovea coordinate, but doing so results in other erroneous regions as well. Therefore we cluster one percent of the highest intensities and segment out the largest cluster. The centroid of this largest cluster gives us the fovea coordinate.

2 Experiments and Results

For the first task of Glaucoma classification, we employ only the images provided by the challenge organisers for training, and a subset of RIMONE dataset [1] for validation. All the networks are trained for 20 epochs and stochastic gradient descent (SGD) optimizer except the EfficientNets, which are trained using ADAM optimizer, with an initial learning rate of $1e^{-3}$ for all models. We save the model giving the highest accuracy on the validation set. While testing, we perform ensembling of all models by taking the mean of posterior probabilities for glaucoma. The final ensembling gave us an AUC score of 0.959 in the validation stage of the competition.

For the second and third tasks, along with the challenge data, we employ 1200 images from the REFUGE challenge [4]. We train the GAN models from scratch with initialization from a Gaussian distribution with mean 0 and standard deviation 0.02 and do the training for 200 epochs. We keep the initial learning rate to be 10^{-4} and halve every 50 epochs. We train the models at a high resolution of 640x640. We save the model which gives the best validation score. For the task of OD and OC segmentation, we obtained dice scores of 0.9599 and 0.8495 respectively in the validation stage of the competition. For the third task, we obtain *Euclidean Distance* difference of 15.26 pixels.

3 Acknowledgement

We thank the organizers of the 2nd version of the Retinal Fundus Glaucoma Challenge (REFUGE2) (https://refuge.grand-challenge.org/) for hosting the challenge and kindly providing the dataset.

References

1. Fumero, F., Alayón, S., Sanchez, J., Sigut, J., Gonzalez-Hernandez, M.: Rim-one: An open retinal image database for optic nerve evaluation. In: Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on. pp. 1–6. IEEE (2011)
2. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
3. Mahajan, D., Girshick, R., Ramanathan, V., He, K., Paluri, M., Li, Y., Bharambe, A., van der Maaten, L.: Exploring the limits of weakly supervised pretraining. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 181–196 (2018)

4. Orlando, J.I., Fu, H., Breda, J.B., van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.A., Kim, J., Lee, J., et al.: Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical image analysis **59**, 101570 (2020)
5. Shankaranarayana, S.M.: Fundus image analysis for age related macular degeneration: Adam-2020 challenge report (2020)
6. Shankaranarayana, S.M., Ram, K., Mitra, K., Sivaprakasam, M.: Joint optic disc and cup segmentation using fully convolutional and adversarial networks. In: Fetal, Infant and Ophthalmic Medical Image Analysis, pp. 168–176. Springer (2017)
7. Shankaranarayana, S.M., Ram, K., Mitra, K., Sivaprakasam, M.: Fully convolutional networks for monocular retinal depth estimation and optic disc-cup segmentation. IEEE journal of biomedical and health informatics **23**(4), 1417–1426 (2019)
8. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, inception-resnet and the impact of residual connections on learning (2016). arXiv preprint arXiv:1602.07261 (2016)
9. Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946 (2019)
10. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1492–1500 (2017)
11. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)