Glaucoma Detection Algorithm for the Artificial Intelligence for RObust Glaucoma Screening Challenge

Temirgali Aimyshev[1] and Kanat Eleusiz[1]

CMC Technologies LLP, Nur-Sultan, Kazakhstan

Abstract. Artificial Intelligence for Robust Glaucoma Screening (AIROGS) Challenge is aimed at the development of solutions for glaucoma screening from color fundus photographs that are robust to real-world scenarios. In this work we describe our approach for the AIROGS Challenge. We constructed a two-stage pipeline: a) optic disk detection stage; b) glaucoma classification stage. The problem of ungradability was addressed during the first stage of the pipeline.

Keywords: AIROGS · Glaucoma

1 Introduction

Glaucoma is a group of eye diseases which result in damage to the optic nerve and cause vision loss. If treated early, it is possible to slow or stop the progression of disease and save vision. Screenings using fundus imaging are potentially a cost-effective solution for glaucoma detection.

When assessing fundus image for presence of glaucoma, ophthalmologists focus their attention on the optical disk and cup, or, to be more specific, on the cup-to-disk ratio (CDR). When CDR is more than 0.7, glaucoma is suspected [4]. The ongoing research shows that other areas of the retina might also show some signs of the glaucoma [2], but most of the works are concentrated around CDR estimation through optic disk and cup segmentation.

2 Proposed Method

2.1 Overview

We propose a two-stage approach for glaucoma classification. The first step is to detect and crop optic disk area of the image. The second step is a convolutional neural network that classifies cropped image from the first step.

2.2 Optic Disk Detection

As previously outlined, optic disk area is the most important part of the fundus image in terms of glaucoma detection. Therefore, we decided to train our

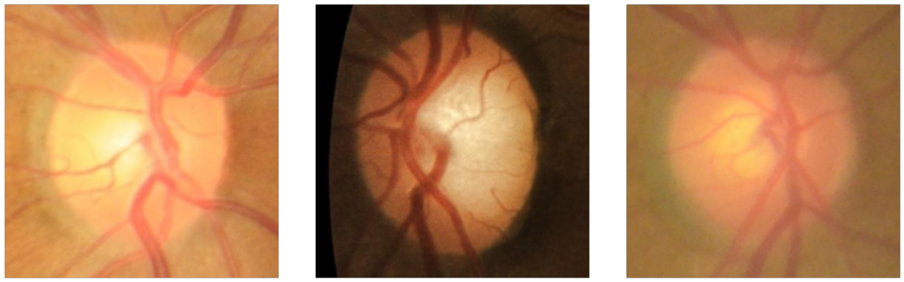

Fig. 1: Exemplary cropped images from the first stage.

glaucoma classificator only on parts of the images that contained optic disks to maximize valuable input. At first, we wanted to segment optic disk, but as the use of external data was not allowed in this competition, we decided that manually segmenting disks would require too much labor and would introduce additional errors. So we decided to use a much more simple approach - to use only bounding box labels.

We manually labelled 500 images from the provided dataset and trained YOLO5 [3] based model to detect optic disks. Then we checked the results and manually fixed errors (about 20 errors among 1000 images). After that we retrained the model on 1500 images. We ran the model on the entire dataset and got square sized cropped images that contained only optic disks. See example crops on Figure 1.

Provided dataset did not contain any images with "ungradable" label, however, it was stated that the test set will have some images with this label and that our models should be able to account to that. At first, we considered a two-step algorithm for identification of ungradable images: 1) whether or not we can detect an optic disk on the image; 2) whether or not we can recognize a cup on the disk using different out of distribution calculations or classic image processing techniques. But after a careful examination of the provided data, we found images that our second step would classify as "ungradable" (see Figure 2). So we decided to use only the first step (optic disc presence detection) as our estimator of gradability.

2.3 Classification

The second stage of our proposed pipeline is a vision transformer (as described in [1]) with the input image size as 384x384. The dataset was split into train set, test set and validation set randomly in a 60-20-20 proportion. We trained a convolutional neural network, based on EfficientNet architecture, with the 512x512 inputs for 100 epochs, 5-fold validation. However, a vision transformer achieved 0.875 score on local validation set, while a convolutional neural network - 0.852. So we decided to use a transformer for the final submission.

Fig. 2: Examples of cropped images that have low visual quality of optical disk.

3 Experimental Results and Discussion

At the time of writing (28 February 2022), our ungradability kappa score (=0.78) on Preliminary Test Phase 2 Leaderboard is top-1 among all submissions. This suggests that our approach, though simple, fits better to the given dataset. Our hypothesis is that annotators were very aggressive in assessing image gradability and provided labels even when optic discs were barely visible. More detailed analysis of the full dataset is needed to check this hypothesis.

Another interesting point for us was better performance (and faster training) of a vision transformer over a convolutional neural network.

References

1. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale (2021)
2. Hemelings, R., Elen, B., Barbosa-Breda, J., Blaschko, M.B., De Boever, P., Stalmans, I.: Deep learning on fundus images detects glaucoma beyond the optic disc. Scientific Reports **11**(1) (Oct 2021)
3. Jocher, G.: ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements (Oct 2020). https://doi.org/10.5281/zenodo.4154370
4. Virk, J.K., Singh, M., Singh, M.: Cup-to-disk ratio (cdr) determination for glaucoma screening. In: 2015 1st International Conference on Next Generation Computing Technologies (NGCT). pp. 504–507 (2015). https://doi.org/10.1109/NGCT.2015.7375170