Combination of supervised learning and unsupervised learning to detect ungradable images in the AIROGS Challenge

Tien-Dung Le[1]

[1]KBC, Belgiumn

This paper presents a solution that combines a supervised deep learning classification and an unsupervised autoencoder network in order to detect upgradable images in the AIROGS challenge. In this challenge, there are only gradable images in the train set. I.e., it is not possible to build a supervised learning to detect upgradable images. However, using only an unsupervised learning techniques does not yield a good result. The gap is then reduced by using a classification learned through gradable images. The performance in the hidden test set shows potential of this approach.

Correspondence: ***tien.le@kbc.be***

Introduction

Glaucoma is a leading cause of irreversible blindness and impaired vision [1], [2]. Early detection of this disease can avoid visual impairment, which could be facilitated through glaucoma screening. Glaucomatous patients can be identified with the use of color fundus photography (CFP). The analysis of CFP images performed by human experts, however, is a highly costly procedure. Artificial intelligence (AI) could increase the cost-effectiveness of glaucoma screening, by reducing the need for this manual labor. AI approaches for glaucoma detection from CFP have been proposed and promising at-the-lab performances have been reported. However, large performance drops often occur when AI solutions are applied in real-world settings. Unexpected out-of-distribution data and bad quality images are major causes for this performance drop [3]. This paper presents an effort towards to the development of solutions that are robust to real-world scenarios. The paper is structured as follow: section 2 introduces the challenge, section 3 describes our solution, section 4 shows the results and discussion.

The AIROGS challenge

The AIROGS challenge looks for solutions for glaucoma screening from CFP [3]. These solutions should be robust to real-world scenarios where unexpected out-of-distribution data and bad quality images are major causes for this performance drop.

Dataset. The Rotterdam EyePACS AIROGS dataset (in full, so including train and test) contains 113,893 color fundus images from 60,357 subjects and approximately 500 different sites with a heterogeneous ethnicity. All images were assigned by human experts with the labels referable glaucoma, no referable glaucoma, or ungradable.

To encourage participants to develop technologies that are equipped with inherent robustness mechanisms, the training set set is an in-the-lab set where only gradable images are considered and ungradable images excluded. The test set, however, includes all image types acquired during screening, simulating a real-world scenario. The training set contains approximately 102,000 gradable images. The test set contains about 11,000 gradable and ungradable images (both gradable and ungradable).

For each input image during evaluation, the desired output is a likelihood score for referable glaucoma (O1), a binary decision on referable glaucoma presence (O2), a binary decision on whether an image is ungradable (O3, true if ungradable, false if gradable), and a non-thresholded scalar value that is positively correlated with the likelihood for ungradability (e.g. the entropy of a probability vector produced by a machine learning model or the variance of an ensemble) (O4).

Evaluation. The evaluation will be based on two aspects: screening performance and robustness. The screening performance will be evaluated using the partial area under the receiver operator characteristic curve (90-100% specificity) for referable glaucoma (α) and sensitivity at 95% specificity (β). The screening performance metrics are based on these specificity ranges, since a high specificity is generally desired in screening settings. To calculate α and β, we compare output O1 to the referable glaucoma reference provided by human experts.

AIROGS Combination Algorithm

Our solution contains 4 components: a pre-processing box, a binary classification, an autoencoder network and a blending engine.

Pre-processing box. This pre-processing component remove the black wrapper around the color fundus images and resize them to the size 1024x1024 (Fig. 1).

Binary classification. This binary classification is a deep learning network that learns from the labels: referable glaucoma presence / no referable glaucoma presence. The

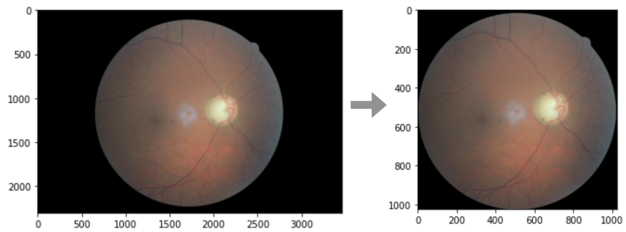

Fig. 1. Images before and after pre-processing.

MONAI framework [4] is used to train 2 networks: efficient net and densenet. The predicted probability is an average of the 2 networks outputs

Autoencoder network. An autoencoder is a neural network that is trained to reconstruct its input [5]. Their main purpose is learning in an unsupervised manner an "informative" representation of the data. The reconstruction error is the score to measure how likely an image is ungradable. Fig. 2 shows a typical example. The higher the reconstruction error is the more likely that the image is ungradable.

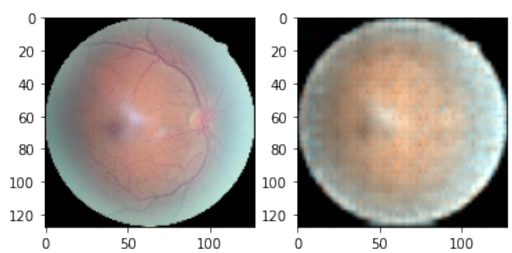

Fig. 2. Input image and reconstructed image.

Blending engine. The blending engine is there to profit all information from the dataset. The blending engine considers the probability output from the binary classification as a weight factor to the reconstruction error. The highest weight is 1 when the probability is 0.5, which is unable to predict, and the weight is lowest when the probability is certain, which is either 0 or 1.

Results and Discussion

The final results are shown in the leaderboad at https://airogs.grand-challenge.org/evaluation/final-test-phase/leaderboard/.

Bibliography

1. Mariotti SP. Global data on visual impairment. *Geneva, Austria: World Health Organisation*, 2010.
2. Aubreville M. Wilm F., Breininger K. Domain adversarial retinanet as a reference algorithm for the mitosis domain generalization (midog) challenge. *arXiv preprint arXiv:2108.11269*, 2021.
3. Coen de Vente et al. Airogs: Artificial intelligence for robust glaucoma screening challenge, 2022.
4. Working Groups. Medical open network for ai, 2021.
5. Raja Giryes Dor Bank, Noam Koenigstein. Autoencoders. *arXiv preprint arXiv:2003.05991*, 2021.