Computer Vision for Glaucoma Detection

Anne-Louise Kopff[1] Anh Hoang Nguyen [1] Quiazhong Chen [1]

[1]Stanford University, ICME, Computer Science, and Mechanical Engineering

Overview

Glaucoma is an asymptomatic eye disease and a leading cause of irreversible blindness. It is detected by specialists using eye fundus images. In this project, we used computer vision neural networks to predict glaucoma using a small dataset of 650 eye fundus images. With the Vision Transformer model, we got precision = 0.73 and F-1 score = 0.66.

Background and Dataset

Our publicly available dataset is composed of 650 pre-processed eye fundus images. 25% of them corresponds to patients with glaucoma. Previous work on glaucoma prediction used different cropping (as shown below) and ran ResNet models on them. Best score was obtained by keeping the optic nerve head region.

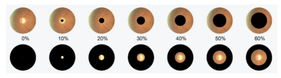

Figure 1. Example of eye fundus image with glaucoma (2)

The damage caused by glaucoma can be quantified in fundus photos by measuring the cup-to-disc ratio (CDR) in the optic nerve head region. As shown in the histogram below (our dataset), an elevated CDR is considered suspicious for the glaucoma diagnosis.

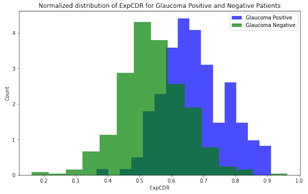

Figure 2. Histogram of CDR values for positive and negative patients

Methods

Trained the baseline CNNs with a split of 520 training images and 130 test images. Then improved the performance using different methods and models.

1. Baseline CNNs: With our initial split, our best scores were obtained for MobileNet and InceptionV3. The training curves are presented in the next section
2. Dealing with small dataset: Change training/test splits to increase the training size and reduce overfitting:

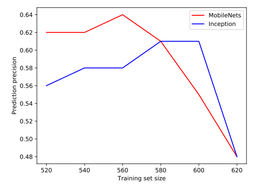

Figure 3. Precision vs training set size

To generate more training images using our existing ones, we applied some basic **data augmentation** (resizing to 72x72 pixels, flipping, rotating and zooming)

3. Vision Transformer (ViT): ViT splits each image into sub-image patches, embeds each patch with a linear projection and then the sequence of these embedded patches is fed into the transformer model.

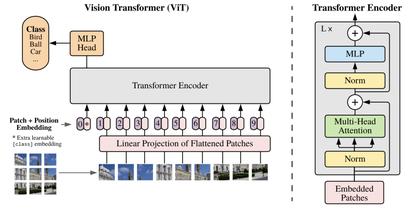

Figure 4. Vision transformer model overview

Experiments

Training curves and model evaluation table for image input

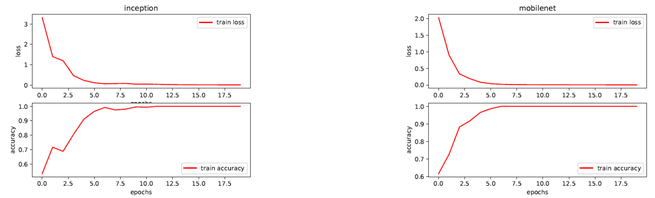

Figure 5. Training curves of baseline CNN models

Model	Precision	Recall	F1-score
Inception	0.56	0.57	0.57
MobileNet	0.60	0.61	0.60
ViT	**0.73**	**0.75**	**0.66**

Table 1. Test results for image inputs only

Model	Precision	Recall	F1-score
Inception	0.59	0.60	0.59
MobileNet	0.62	0.64	0.63

Table 2. Test results for image inputs and feature table inputs

The results show that utilizing the CDR features improves the CNN models' detection accuracy. Transformer is a powerful architecture and it can be applied to Glaucoma detection task. A promising next step is to explore a multi-modal Transformer architecture utilizing both images and CDR features as inputs.

References

[1] Lawrence Zitnick and Piotr Doll´ar. Edge boxes: Locating object proposals from edges. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, Computer Vision – ECCV 2014, pages 391–405, Cham, 2014. Springer International Publishing. poster

[2] Deep learning on fundus images detects glaucoma beyond the optic disc Ruben Hemelings, Bart Elen, João Barbosa-Breda, Matthew B. Blaschko, Patrick De Boever and Ingeborg Stalmans Scientifc Reports, Article 20313 (2021)