Constrastive Learning for the Airogs Challenge

Wei Tang[1], **Mingcai Chen**[2], **and Baoqiang Ma**[3]

[1]UMCG, University of Groningen, the Netherlands
[2]Nanjing University, China
[3]UMCG, University of Groningen, the Netherlands

Abstract

Glaucoma is an eye illness that causes vision loss and can potentially result in blindness. Artificial intelligence might make glaucoma screening more effective by decreasing the need for human labor. This paper provides a method that uses convolution neural networks for detecting glaucoma and unsupervised networks for differentiating low-quality fundus images. We use Mahalanobis distance for out-of-distribution detection, and the results on the test set demonstrate the approach's promise.

Introduction

Glaucoma is a vision-impairing eye disease that can be discovered and prevented with early diagnosis and color fundus photography (CFP) screening. Human professionals undertake CFP picture analysis, which is a very expensive operation. By decreasing the requirement for physical labor, artificial intelligence (AI) can increase the cost-effectiveness of glaucoma screening. AI methods for identifying glaucoma via color fundus photography (CFP) have been proposed, with promising laboratory results. When AI systems are used in real-world situations, however, significant performance loss is frequently found. The main reasons for this include unexpected out-of-distribution data and poor image quality.

The AIROGS challenge

The AIROGS competition was organized to exploit solutions based on CFP glaucoma screening tools to help clinicians with accurate and robust computationally aided diagnosis. The Rotterdam EyePACS AIROGS dataset contains 113,893 color fundus images from 60,357 subjects in about 500 different locales, including a wide range of ethnicities. Referable glaucoma, no referable glaucoma, and ungradable were all assigned to all of the photographs by human specialists.

Preprocessing

The images vary in size, but the eyeballs are all the same size. So we tried to remove the black wrap around the color fundus image. R channel < 10 was used to filter the peripheral during this process. The images were then resized to 256x256 pixels.

Classification

We first applied shape-based and color-based data augmentation, depending on the dataset's characteristics. The former consists of horizontal and vertical flips. The HSV channel and random brightness contrasting augmentation are included in the latter. Finally, we used the Resnext 30x4d deep learning network to train 80 of the data for 300 epochs, using balanced sampling (1:1). In order to save training time, only 2000 samples were used for training and validation each time. A cosine annealing scheduler was applied with a maximum number of 30 epochs. The initial learning rate was set at 1e-3 and the loss function was binary cross-entropy loss.

Color fundus ungradable image detection

Color fundus ungradable image detection, i.e., identifying whether a given color fundus image is drawn from outside the training distribution, is changeling and essential for robust Glaucoma Screening. We borrow the latest development in the field of contrastive learning [1] and novelty detection [2] to accomplish this task.

Generally, we use a contrastive learning algorithm named SimCLR [1] to train a deep network. For the detection of ungradable images, we consider the maximum similarity between the representation of any given image and that of the training images. Furthermore, the norm of extracted features is also used to form the critria.

Using f, g represent the network feature extractor and a linear projection head, respectively. We train the model following the SimCLR [1]. We first draw two augmentation functions t, t' from a predefined augmentation method: $t \sim \mathcal{T}$, $t' \sim \mathcal{T}$. Then, we extract representation on both of them:

$$\begin{aligned} \tilde{x}_{2k-1} &= t(x_k) \\ h_{2k-1} &= f(\tilde{x}_{2k-1}) \\ z_{2k-1} &= g(h_{2k-1}) \\ \tilde{x}_{2k} &= t'(x_k) \\ bmh_{2k} &= f(\tilde{x}_{2k}) \\ z_{2k} &= g(h_{2k}) \end{aligned} \tag{1}$$

Having the representations, we calculate the pairwise similarity:

$$s_{i,j} = z_i^\top z_j / (\|z_i\| \|z_j\|) \tag{2}$$

The final loss function for training is:

$$\begin{aligned} \ell(i,j) &= -\log \frac{\exp(s_{i,j}/\tau)}{\sum_{k=1}^{2N} k \neq i \exp(s_{i,k}/\tau)} \\ \mathcal{L} &= \frac{1}{2N} \sum_{k=1}^{N} [\ell(2k-1,2k) + \ell(2k,2k-1)] \end{aligned} \tag{3}$$

For detection, we calculate the maximum similarity between the representation of any given image x and that of the training

images x_m.

$$score = \max_m \cdot \|f(x_m)\| f(x)^\top f(x_m) / (\|f(x)\| \|f(x_m)\|) \tag{4}$$

The $score$ is to be compared with the $score$ of training set. If it is smaller than 95% of the training examples, this example is regarded as out of distribution. Otherwise, we classify it as in distribution, i.e. gradable. We show the distribution of in-distribution samples' scores in Fig. 1.

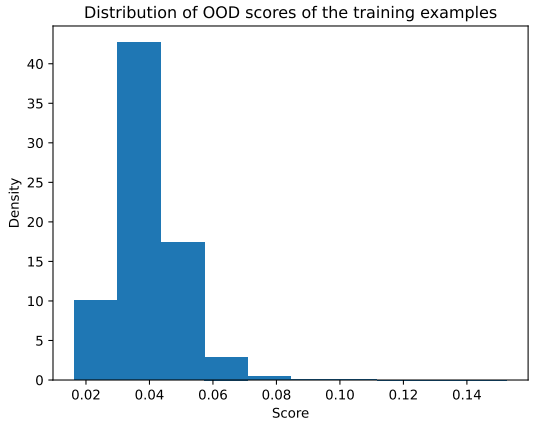

Fig. 1. Distribution of in-distribution samples' scores

More specificly, we use a common backbone model: 18 layers resnet [3] with 128 dimension output. The model is pretrained on STL-10, so the model can extract some common shape and texture features. Then we feed all cropped and resized color fundus images (96×96) into the model for contrastive learning. The mini-batch SGD is used with a learning rate of 0.0003, a weight decay of 0.0001, a batch size of 256, a cosine annealing learning rate, for 100 rounds.

Bibliography

1. Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.
2. Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. *Advances in neural information processing systems*, 33:11839–11852, 2020.
3. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.