Test task: Eye state classification

Pavel Tikhomirov

April 2024

1 Introduction

The task is to train an algorithm that allows to classify the state of the eye - open or closed, moreover the predicted value should be float from 0 to 1 to estimate confidence of the method. It's an image binary classification task. The dataset consists of 4000 24×24px gray scale images of eyes. The specificity of the task is that the dataset does not contain class labels of open or closed eyes. Also this set contains images that are which are difficult to classify even by humans or contains defects (Figure 1)

Figure 1: One of images in dataset. The image is very dark, and it's impossible for human to detect the eye

The target metric is equal error rate (EER). To calculate EER we need to calculate false positive rates (fpr) and false negative rates (fnr) for different threshold of classification. $\text{EER} = \text{fpr}_j$, $j = \text{argmin}_i(|\text{fpr}_i - \text{fnr}_i|)$. To evaluate quality of methods the first 400 examples were labeled. This examples were excluded from training set.

2 Experiment setup

2.1 Self-supervised learning

The manual data labeling is very expensive. So firstly we consider methods that requires minimum number of labeled examples. As we know that data contains two general classes we use clustering algorithms that uses this prior knowledge. But we can't apply clustering directly on image data, and use convolution networks (CNN)(Krizhevsky et al. 2012) to extract meaningful features to build clusters on them. Extracting such features is a separate task in machine learning called representation learning (Bengio et al. 2014).

The goal of this experiment is compare different techniques to extract useful representations of image. We train variational autoencoder (VAE) Kingma and Welling 2022) based on CNN with KL-divergence and MSE loss, and only encoder with triplet loss (Hoffer and Ailon 2018). Tuned hyper-parameter is the size representations. For VAE the resulting representation is a concatenation of mean and variance vectors. The second step is the Gaussian mixtures clustering algorithm over representations. The motivation behind choosing this clustering algorithms is the following: the number of classes is known and the output is probabilistic. Images are augmented during training. The resulting EER on test set for both feature extractors ~ 0.41, accuracy ~ 0.61.

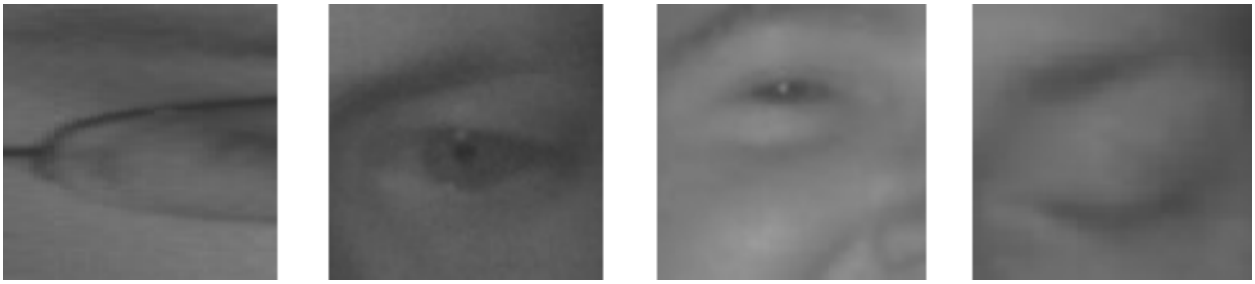

Figure 2: Examples of images from open datasets

2.2 Pre-trained networks

One more approach to go away from manual labeling is to use pre-trained networks. There are two open datasets with similar domain on Kaggle: (Shah 2020) $\sim$ 7.1k, and *Eyes-Image-Dataset-For-Machine-Learning* $\sim$ 14.5k images. Examples are presented at Figure 2.

We splits data in train and validation parts with ratio 4:1. During training images are scaled to the target size and augmented. We train ResNet18 network (He et al. 2015). Achieved accuracy on validation part is 0.92 for the first dataset and 0.95 for the second, however the performance of network on test set is very low and close to random classifier.

2.3 Semi-supervised learning

The next experiment is to use half of labeled data as training set for semi-supervised task. We train ResNet18 in setup of noisy student (Xie et al. 2020) with augmentations and noise injection. Performance of the model trained in such a way is better: EER $\sim$ 0.26, accuracy $\sim$ 0.71. The low performance can be caused by not increasing the size of the model, however bigger models behave worse because of overfitting. Increasing the size of unlabeled part of dataset is a way to solving this problem.

2.4 Supervised learning

After this we train ResNet18 in a standard way with 206 train and 206 test examples. The resulting test quality is the following: EER $\sim$ 0.09, accuracy $\sim$ 0.91. Metrics are improved to $\sim$ 0.05, and $\sim$ 0.95 correspondingly via multiple retraining and tuning augmentations.

3 Results & Conclusion

The supervised mode shows the best results even with little amount of labels. Model makes more false positive mistakes then false negative (7 vs 4). Despite good overall performance the critical errors happen (when the model is overconfident about false classification). The inference examples are presented in Figure 3. Images with a well-viewed sclera are classified with high confidence and correctly. At the same time model isn't confident about some images of low contrast, or squint, that's can be interpreted as close to human behaviour of model. Overconfident mistakes can be caused by lack of data. To get rid of such mistakes without additional labeling one can use different adversarial training techniques (Bai et al. 2021).

Reported weights obtained as a result of retraining on the whole labeled data (412 images). Inference speed of the whole pipeline on NVIDIA RTX 4070 and Intel Core i5-13400F is 207 fps (estimated on 20100 images)

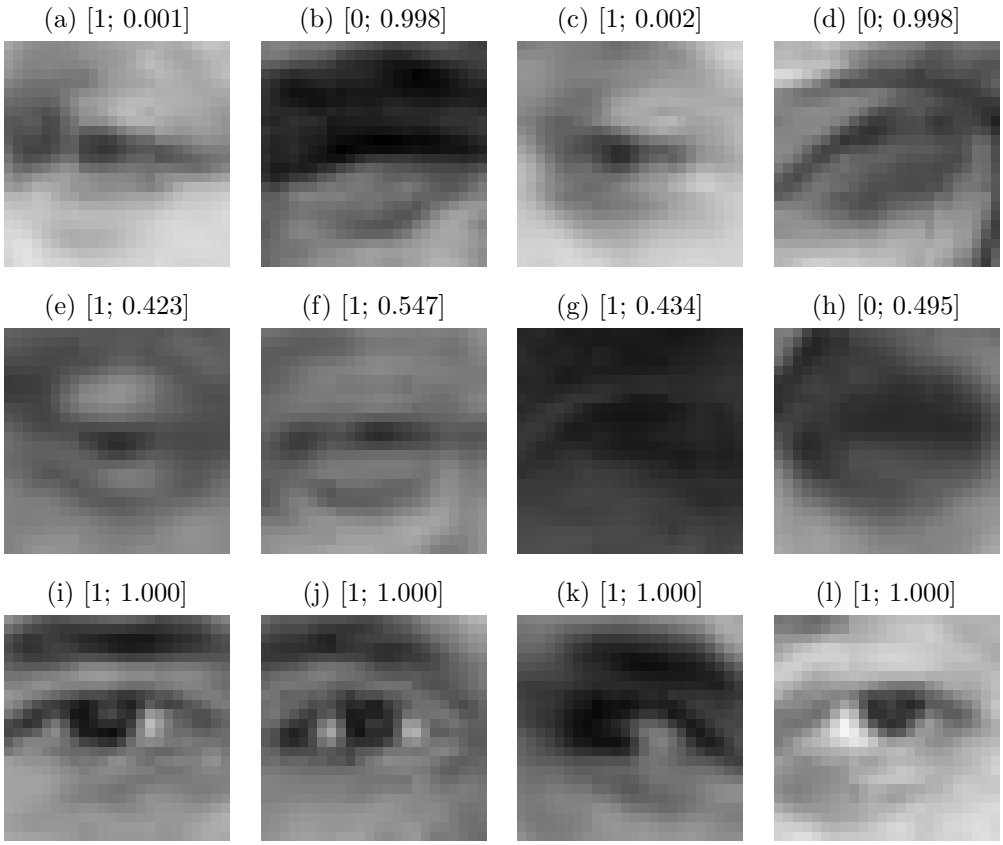

Figure 3: Results on test set. The fist number is ground truth label, the second – predicted probability of the 1st class. (a)-(d) – top4 "critical" mistakes, where the model is farthest from the correct answer. (e)-(h) – top4 "difficult" examples, where model's prediction is close to the 0.5. (i)-(l) – top4 "easy" examples, where model is closest to the correct answer

References

[1] T. Bai, J. Luo, J. Zhao, B. Wen, and Q. Wang. Recent advances in adversarial training for adversarial robustness, 2021.

[2] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives, 2014.

[3] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.

[4] E. Hoffer and N. Ailon. Deep metric learning using triplet network, 2018.

[5] D. P. Kingma and M. Welling. Auto-encoding variational bayes, 2022.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL `https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf`.

[7] K. Shah. Eye-dataset, 2020. URL `https://www.kaggle.com/dsv/1093317`.

[8] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le. Self-training with noisy student improves imagenet classification, 2020.