Udacity ML Engineer Nanodegree

Capstone Proposal

Predicting Blindess with Deep Learning

Nikita Kozodoi

May 2020

Contents

1 Domain Background 2

2 Problem Statement 2

3 Data Sets and Inputs 2

4 Solution Statement 3

5 Evaluation Metrics 3

6 Benchmarks 4

7 Project Design 4

1 Domain Background

Diabetic retinopathy (DR) is one of the leading causes of vision loss. According to a recent study from International Diabetes Federation, the global prevalence of DR among the individuals with diabetes for the period from 2015 to 2019 was at more than 25% [5]. According to the World Health Organization, more than 300 million people worldwide have diabetes, and the disease prevalence has been rising rapidly in developing countries [6].

Early detection and treatment are crucial steps towards preventing DR [2]. Currently, detecting DR is a time-consuming process. The screening procedure requires a trained clinical expert to examine the fundus photographs of the patient's retina. This creates delays in diagnosis and treatment of the disease. Automated evaluation of retina photographs can speed up the efficiency and coverage of the DR screening programs. This is especially relevant for developing countries, which often lack qualified medical stuff to perform the diagnosis.

Previous research has explored the usage of deep learning for detecting DR and concluded that convolutional neural networks (CNNs) have high potential in this task [2]. The Asia Pacific Tele-Ophthalmology Society (APTOS) has launched two Kaggle competitions with a goal of promoting the use of deep learning for DR detection and boosting the development of automated detection systems.

2 Problem Statement

The project aims at developing a deep learning model for predicting the severity of the DR disease based on the patient's retina photograph. Severity of the DR determines the necessary actions for prevention and/or required treatment. Therefore, we consider the multiple classification task, in which five disease stages are distinguished based on its severity:

- 0 – no disease
- 1 – mild stage
- 2 – moderate stage
- 3 – severe stage
- 4 – proliferative stage

3 Data Sets and Inputs

The project leverages two data sets: (i) main data set used for modeling and performance evaluation; (ii) supplementary data set used for pre-training the model.

The main data set is provided by APTOS. The data set has been employed in the APTOS 2019 Blindness Detection competition on Kaggle and is available for the download at the competition website: `https://www.kaggle.com/c/aptos2019-blindness-detection/data`.

The data set includes 3,662 labeled retina images of clinical patients. The images are taken using a fundus photography technique. The data set is collected from multiple clinics in India using a variety of different camera models, which creates discrepancies in the image resolution, aspect ratio, exposure and other parameters.

The images are labeled by a clinical expert. The labels indicate the severity of DR from 0 to 4 in accordance with the scale provided in Section 2. The data also include a test set with 1,928 images whose labels are not explicitly disclosed. The test set can be used for performance evaluation using a «late submission» functionality at the Kaggle competition website.

In addition to the main data set described above, the project uses a supplementary data set with the retina images to pre-train the developed models. The supplementary data set features 35,126 retina images labeled by a clinician using the same scale as the main data set. The data set has been used in the 2015 Diabetic Retinopathy Detection competition and is available for the download at the corresponding competition website: `https://www.kaggle.com/c/diabetic-retinopathy-detection`.

4 Solution Statement

The proposed solution uses historical retina image data to develop a CNN-based model that is able to achieve high predictive performance in the DR classification task.

The project applies different image preprocessing techniques to standardize retina photographs. We adapt one of the state-of-the-art CNN model architectures to the considered problem using transfer learning. We test multiple model specifications and use different techniques such as test-time augmentation and ensembling to increase the predictive performance of the classification model. The effectiveness of our solution is validated on the test data set from the 2019 APTOS Kaggle competition.

5 Evaluation Metrics

The project considers DR detection as an ordinal classification task. In such a setting, a suitable evaluation metric is the Cohen's Kappa, which measures the agreement between the actual and predicted labels [1]. The metric varies from 0 (random agreement) to 1 (perfect agreement).

The Kappa is computed as follows. First, one constructs three matrices: (i) the matrix of observed scores X; (ii) the matrix of expected scores based on chance agreement M; (iii) the weight matrix W. Next, the Kappa is computed as:

$$\kappa_w = 1 - \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} x_{ij}}{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} m_{ij}}, \tag{1}$$

where x_{ij} and m_{ij} are elements in the observed and predicted matrices, respectively, whereas the w_{ij} are weights. We use Kappa with quadratic weights to stronger penalize larger errors.

6 Benchmarks

The project uses multiple benchmark models to evaluate the effectiveness of the proposed solution. We compare performance of the developed solution with the publicly available solutions derived by different teams in the APTOS 2019 Blindness Detection competition. We evaluate the model performance on the same test sample to enable direct comparison.

The particular benchmarks include:

- Fine-tuned Resnet-50 model [3] ($\kappa_w = 0.8747$)
- Optimized Resnet-50 model with train and test-time image augmentations ($\kappa_w = 0.8832$)
- EfficientNet B5 model [4] with advanced image preprocessing ($\kappa_w = 0.9098$)

7 Project Design

The project pipeline consists of five stages outlined below.

1. **Image prepossessing.** This step is important because images were taken in different conditions and exhibit differences that need to be taken into account. We consider different ways to crop and resize the images to remove as many discrepancies as possible.
2. **Selecting the model architecture.** The project uses CNNs for image classification. We use one of the EfficientNet neural network architectures, which is one of the recent state-of-the-art CNN models for computer vision [4].
3. **Pre-training the model on the 2015 data set.** The main 2019 data set has a limited number of images (N = 3,662). We pre-train the model on a larger data set from the 2015 competition (N = 35,126).
4. **Fine-tuning the model on the 2019 data set.** We fine-tune the pre-trained model on the main data set. We also consider multiple variants of image augmentations to improve the model performance. We use cross-validation for data partitioning and make modeling decisions based on the performance observed on the out-of-fold predictions.
5. **Producing predictions**. We predict labels of images in the test sample and evaluate the performance. We average predictions of base models trained on different cross-validation folds and use test-time augmentation to further improve the performance.

References

[1] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.

[2] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.

[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.

[4] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*, 2019.

[5] RL Thomas, S Halim, S Gurudas, S Sivaprasad, and DR Owens. Idf diabetes atlas: A review of studies utilising retinal photography on the global prevalence of diabetes related retinopathy between 2015 and 2018. *Diabetes research and clinical practice*, page 107840, 2019.

[6] Tien Yin Wong and Neil M Bressler. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. *Jama*, 316(22):2366–2367, 2016.