DIABETIC RETINOPATHY DETECTION WITH NEURAL NETWORKS

Peizheng Li
Institute of Signal Processing and System Theory
University Stuttgart
st169530@stud.uni-stuttgart.de

Fangwen Liao
Institute of Signal Processing and System Theory
University Stuttgart
st169178@stud.uni-stuttgart.de

February 16, 2021

ABSTRACT

Diabetic retinopathy (DR) is an eye disease that seriously impairs the vision of diabetic patients. In this paper, we mainly developed a detection model based on deep convolutional neural networks. This model is based on IDRID and Kaggle EyePACS datasets, which are also augmented through Sample Pairing. After analysis using deep visualization, we selected various architectures, and the model reached an binary accuracy of nearly 90% through ensemble learning. In the end, we concluded that the size of the dataset is currently the most critical factor that determines the detection ability.

1 Introduction

Diabetic retinopathy (DR) is the most important manifestation of diabetic microangiopathy. The current detection of DR is time-consuming in most cases and difficult to spread to areas with insufficient medical resources. Therefore, we have developed a set of DR detection models based on deep neural networks, to diagnose the disease and determine the stage of disease progression.

2 Baseline Model

2.1 Inputpipeline and Preprocessing (accomplished by Fangwen Liao)

We use the Indian Diabetic Retinopathy Image Dataset (IDRID)[1] which contains 516 color fundus images with the size of 4288×2848. Data are converted into TFRecord files as the input pipeline, where the original training set is divided into training set and validation set at a ratio of 3:1.

In spite of high resolution and location uniformity, these images still suffer from problems such as black borders, too many pixels, and unbalanced distribution, etc. In response to these, the following preprocessing is performed:

- cut off excess black borders of images and convert them to a standard size of 256×256;
- oversample the images with labels 1, 3, and 4.

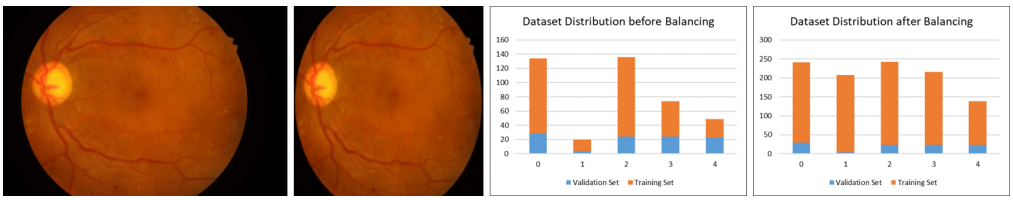

Figure 1: Comparison of Data before and after Preprocessing

As shown in the figure 1, the images after preprocessing are more concise and basically maintains a uniform distribution.

2.2 Model Architecture (accomplished by Fangwen Liao)

We first choose the standard VGG16 as the baseline model, in order to establish a benchmark for the comparison later.

2.3 Selection of Output Type (accomplished by Peizheng Li)

There are three optional output types: binary classification, 5-class classification, and regression, in which the latter two can be converted to the first. After comparison, we can see that the performance of regression and 5-class classification models are obviously superior. So we first choose regression as the output type.

Table 1: Comparison of Model with Different Output Types

Output Types of Models	Regression	Binary Classification	5-class Classification
Binary Accuracy	83.50%	71.84%	85.44%

2.4 Metrics (accomplished by Peizheng Li)

The accuracy and confusion matrix are the general metrics during our training and validation. For the binary classification problem, that is, whether the patient has diabetic retinopathy, we also consider precision, recall, F1 score, ROC as well as PRC as evaluation metrics. These are mainly used in the evaluation of the subsequent ensemble learning method.

2.5 Evaluation and Preliminary Detection Performance (accomplished by Fangwen Liao)

On the baseline model, we have achieved 71.78% binary accuracy. The corresponding confusion matrices are shown in the figure 2. Obviously, the baseline model is insufficient to complete the binary classification task.

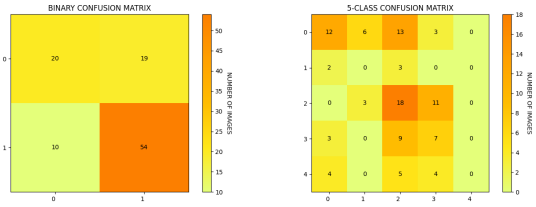

Figure 2: Confusion Matrices of the Baseline Model

3 Analysis and Optimization

3.1 Deep Visualization (accomplished by Peizheng Li)

In order to analyze the model, we adopt four deep visualization methods, as shown in the figure 3. Among them, since the feature size of the last CNN layer is 8×8, the resolution of Grad-CAM is relatively low. In contrast, Integrated Gradients can better highlight the important regions of the image without being affected by resolution.

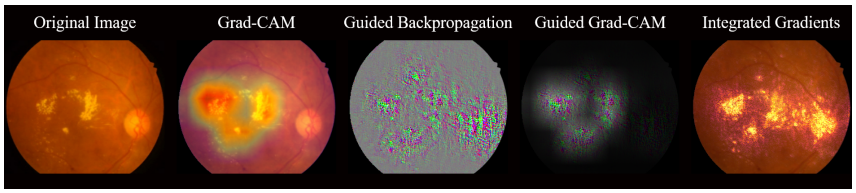

Figure 3: Four Kinds of Deep Visualization Based on the Baseline Model

It can be analyzed from figure 4: the baseline model can roughly detect and locate pathological features. However, some regions are alse misrecognized, especially in the optic disc area which has high brightness similar to bleeding points.

Based on the analysis, we focus our further optimization on 2 aspects: data augment and model architecture optimization.

3.2 Traditional Data Augment (accomplished by Peizheng Li)

We first used the traditional augment method for images:

- First rotate and cut the original image randomly
- Then randomly crop a 256×256 sub-image
- Finally randomly change its contrast, saturation and hue

Note: The amplitude of above operations should not be too large, so as not to change the original distribution.

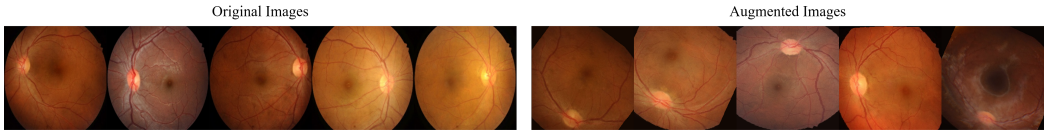

Figure 4: Images with and without Augment

3.3 Model Architecture Optimization (accomplished by Peizheng Li)

The upper bound of model performance is also a major factor restricting the classification ability. In this regard, we add batch normalization (BN) layers to all convolutional layers and try different model structures for optimization tasks.

3.3.1 Inception and SEResNeXt (accomplished by Peizheng Li)

The width of the receptive field and the model depth are 2 major factors that affect network performance. Inception[2] and Resnet[3] have taken corresponding measures for these 2 aspects, namely, the combination of receptive fields of different sizes and the shortcut of identity mapping. Therefore, we build self-made Inception as well as Resnet, which is added with the group convolution and SE module (attention) and becomes the SEResNeXt[4].

3.3.2 DenseNet and EfficientNet through Transfer Learning (accomplished by Peizheng Li)

DenseNet is similar to the extension of Resnet[5]. In addition, it expands the shortcut between adjacent layers to the connection between each layer. EfficientNet is derived from a paper published by Mingxing Tan in ICML in 2019[6]. It differ from previous heuristic conceptions of network architecture, and provides a systematic model scaling method that integrates network depth, network width and Input image resolution.

In view of the model complexity, we use transfer learning: pre-training, and then fine tuning by using ray tuning.

3.3.3 RepVGG (accomplished by Peizheng Li)

The RepVGG proposed by the team from Tsinghua University in 2021 provides a novel perspective[7]. It simplifies all convolution layer to the most basic 3*3 convolution, supplemented by 1*1 and identity branches. Then, through reparameterization it can achieve higher speed and lower memory usage, while retaining the high performance.

3.3.4 Comparison (accomplished by Peizheng Li)

The binary accuracy of all models can exceed 80%, and the performance of EfficientNet is the best, which can reach 87.38%. In terms of 5-class classification, DenseNet is particularly outstanding with the highest accuracy of 62.14%.

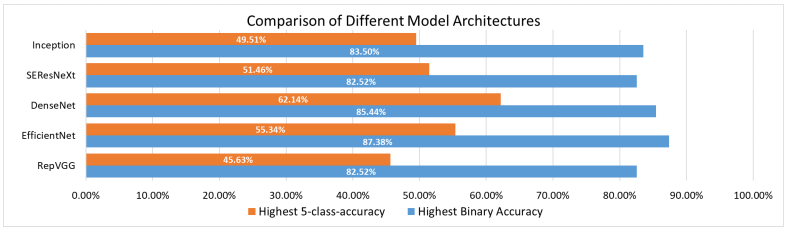

Figure 5: Comparison of Different Model Architectures

3.4 Sample Pairing (accomplished by Peizheng Li)

Compared with millions of parameters, the dataset of less than one thousand images is obviously too small. Therefore, over-fitting has now become the primary obstacle to the optimization.

Besides the traditional data augmentation, we also try the Sample Pairing (SP) method, which superimposes two original images to a new image (What we used is averaging)[8]. Since the fundus images have high similarity, after superimposition, there will not be problems such as ghosting. On the contrary, features that are not related to diabetic retinopathy are halved, while highly related features are strengthened.

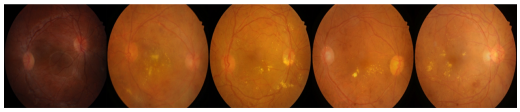

Figure 6: Images after Sample Pairing

However, due to partial change of distribution, these images are not suitable for pre-training, but for fine tuning.

Our comparison also proves this statement: Using SP data in the entire training will not improve performance, whether it is binary accuracy or 5-class-accuracy. But using the initial data for pre-training before fine tuning with SP images can help the model perform better in 5-class classification (the reason why there is no significant difference in binary accuracy is mainly due to the small and imperfect test set. There are many images that even experts may not be able to make correct judgments)

Table 2: Comparison between Different Training Methods in Terms of Using Sample Pairing

Training Method	Average Binary Accuracy	Average 5-class Accuracy
Training with Original Images	83.69%	47.57%
Training with SP Images	83.50%	48.16%
Training with Original Images and then Fine Tuning with SP Images	83.11%	49.32%

Generally speaking, using Sample Pairing for fine tuning can improve performance of models based on small datasets.

3.5 Ensemble Learning (accomplished by Peizheng Li)

In order to improve the generalization ability of the model, we adopt ensemble learning to achieve functional complementarity between models. The specific strategy is as follows:

- Regression model: take the average of the prediction of each corresponding image
- 5-class classification model: take the prediction with the most occurrences of each corresponding image

In contrast, the ensemble learning result of the regression model is better than the 5-class classification model. So, we finally choose the three best-performing regression models for ensemble learning, which achieves 88.35% binary accuracy and 53.40% 5-class-accuracy. And the balanced accuracy scores of binary and 5-class classification are 89.62% and 44.67% respectively.

Table 3: Evaluation of Ensemble Learning

Binary Accuracy	Binary Balanced Accuracy Score	Precision	Recall	F1 Score	5-class Accuracy	5-class Balanced Accuracy Score
88.35%	89.62%	96.43%	84.38%	89.90%	55.40%	44.67%

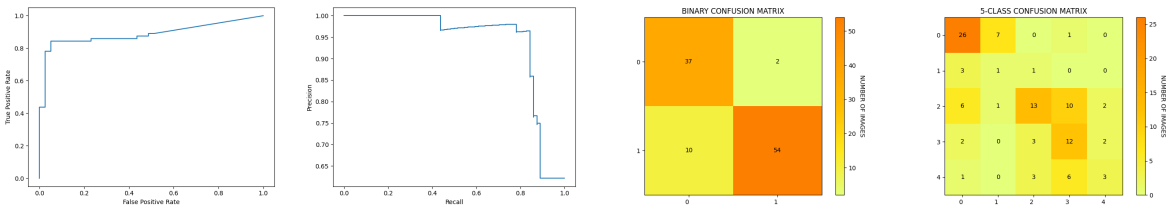

Figure 7: ROC, PRC and Confusion Matrices of Ensemble Learning

3.6 Other Attempts (accomplished by Peizheng Li)

- We use Tensorflow Profiler to analyze the training process and partially change the preprocessing to offline, saving 14% of training time.
- We try the preprocessing of Ben Graham and find that it can not help performance improvement.
- We train an EfficientNet model on the Kaggle's EyePACS dataset[9], which achieves 87.00% binary accuracy and 69.23% 5-class-accuracy.

4 Conclusion

In Conclusion, the use of convolutional neural networks can accomplish the detection task of diabetic retinopathy.

On the basis of small dataset, by selecting appropriate data augment and preprocessing methods, as well as matching with appropriate model architectures, the overall performance can be effectively improved.

However, compared to the millions of model parameters, hundreds of samples are seriously insufficient. In other words, insufficient data is currently the most critical factor restricting the detection effect.

Therefore, the future researches should mainly focus on expanding the dataset and mining deeper information in it.

References

[1] https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid

[2] Szegedy, Christian, et al. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[3] He, Kaiming, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[4] Hu, Jie, Li Shen, and Gang Sun. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[5] Huang, Gao, et al. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[6] Tan, Mingxing, and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. International Conference on Machine Learning. PMLR, 2019.

[7] Ding, Xiaohan, et al. RepVGG: Making VGG-style ConvNets Great Again. arXiv preprint arXiv:2101.03697 (2021).

[8] Inoue, Hiroshi. Data augmentation by pairing samples for images classification. arXiv preprint arXiv:1801.02929 (2018).

[9] https://www.kaggle.com/c/diabetic-retinopathy-detection/data