XIRAN ZHANG·XICHENG LIU· ALEXANDRE LI HAO YAO

ResNet-50 based grading of diabetes retinopathy

**Abstract** Diabetes retinopathy is the main cause of permanent blindness. diabetes is a globally recognized problem that affects eye health. It is estimated that by 2050, more than 700 million people in the world will be affected by this problem. Diabetes retinopathy is a common complication of diabetes, which is mainly caused by retinal vascular injury. The early symptoms of DR are swelling of the micronerves and blood entering the retina after capillary damage; The late stage symptoms of DR are more pronounced, manifested as the proliferation of new blood vessels, ultimately leading to detachment of the retinal layer and permanent blindness. The use of fundus images captured by fundus cameras is an effective diagnostic tool for detecting early symptoms of DR lesions, making it highly suitable for automatic diagnostic systems.

Although deep learning has proposed a series of algorithms in recent years, it is mainly aimed at solving some screening tasks, such as distinguishing healthy individuals and patients. Few studies have involved DR grading tasks.

This article is based on the ResNet-50 deep learning model, customizing the FC layer, and using image rotation to solve the problem of 2 labels being too sparse.

**Keywords** DR Grading · ResNet-50 · Deep learning

1 Introduction

Diabetic retinopathy (DR) is one of the leading causes of blindness. DR Is diagnosed by examining retinal fundus images and OCTA for the presence of retinopathy.In DR Analysis, the image quality of ultra-wide field optical coherence tomography(OCTA) needs to be evaluated firstly, and the image with better imaging quality is selected, and then DR Analysis, such as lesion segmentation and DR Grading, is performed.It is crucial to construct a robust model for automatic image quality assessment, lesion segmentation and DR Grading for DR Screening and diagnosis.

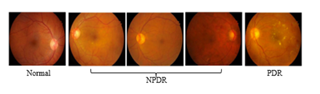


Fig 1 Classification of DR

Fig1 shows five types of DR lesions in the dataset that are gradually becoming more severe, which can be summarized into three types.

Grade1: Normal

Grade2 (NPDR):presence of multiple micro aneurysms, spotted bleeding, intravenous injection, and cotton wool pad;

Grade3 (PDR):neovascularization and vitreous hemorrhage.

This problem can be defined as a clasification problem.

Traditional image classification algorithms are based on machine learning techniques such as Support Vector Machines (SVM) and K-Nearest Neighbors (KNN).

One advantage of traditional machine learning algorithms is that they can have a better solution effect on small sample data sets. However, one disadvantage is that they may not perform as well on large sample data sets compared to deep learning techniques such as Convolutional Neural Networks (CNN).

In this project, we use deep learning to solve this problem. ResNet50 is a deep learning architecture for image classification. It is a convolutional neural network (CNN) that is 50 layers deep and was developed by researchers at Microsoft. It is a variant of the ResNet architecture and is commonly used for image classification tasks.

# 2 Method

## 2.1 ResNet50

The advantage of the deep learning model ResNet-50 used in this article is that：Its depth and residual learning methods enable it to learn more complex features, thereby improving its accuracy. Its global average pooling layer can reduce the number of model parameters, thus reducing the risk of overfitting.

The ResNet50 model consists of 5 stages each with a convolution and Identity block. Each convolution block has 3 convolution layers and each identity block also has 3 convolution layers. The architecture is based on the concept of residual learning, where shortcut connections are used to “skip over” some layers, converting a regular network to a residual network .

The ResNet50 model has over 23 million trainable parameters and uses a bottleneck design for the building block, which reduces the number of parameters and matrix multiplications, enabling much faster training of each layer.

A residual network block was proposed: artificially causing certain layers of the neural network to skip the connections of neurons in the next layer, connecting each other in layers, and weakening the strong connections between each layer.

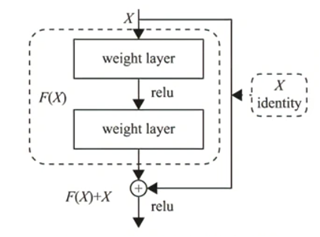


Fig 2 Residual network structure unit

The input image format is [image length, image width, number of image channels].

There is a common stem layer that performs convolution operations on the input image, followed by a pooling layer. Afterwards, there is the stacking of layers.

In the specific residual block, taking ResNet50 as an example, the formula is calculated in the residual network, and the data structure of is [56, 56, 64]; The data structure of is the result of three convolution operations;

In the figure 3, the data structure of F (x) is [56, 56, 256], which is different from x, so we transform the data structure of and perform convolution operations. The input data structure of is [56, 56, 64], and the output is [56, 56, 256]

The input data and operation results of this residual block are added, ensuring that shallow features can be continuously transmitted to deep layers.

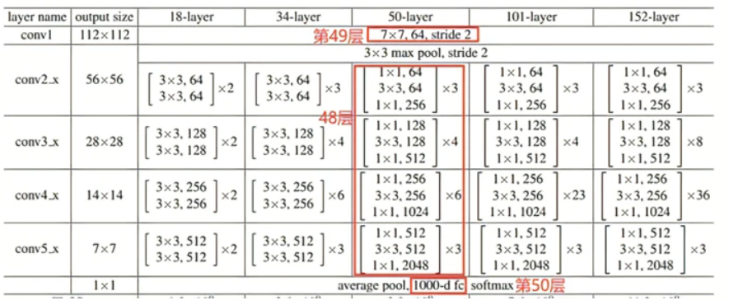


Fig 3 ResNet-50 model network structure

To adapt to the dataset, the last layer of ResNet-50 is replaced. The input of the original last fully connected layer is fed to a linear layer with 256 output units, followed by a ReLU layer and a Dropout layer, and then a 256 x 6 linear layer. The output is a 6-channel softmax layer.

Additionaly, in the training dataset for image classification tasks, there is a problem of sparse labels.

For label 0, there are 329 images out of 611. 212 label 1images out of 611 and only 70 label 2 images.

If there is a problem of sparse labels in an image classification task, it means that there are very few training examples for some of the classes. This can make it difficult for the model to learn to accurately classify those classes, as it has less information to work with. As a result, the model may have lower overall accuracy and may struggle to correctly classify images from the underrepresented classes.

There are several techniques that can be used to address the problem of sparse labels, such as data augmentation, oversampling, and transfer learning. These techniques can help to increase the amount of training data available for the underrepresented classes and improve the model’s performance.

In this project, we used transfer learning to train a pre-trained model. To address the problem of sparse labels, we used data augmentation. We rotated the images in label 2 and added them as new samples to the training set to increase the number of label 2 samples.

# 3 Results and Conclusion

In this task, we use AUC and quadratic-weighted-kappa to evaluate the model.

AUC stands for “Area under the ROC Curve.” The ROC curve is a graphical representation of the performance of a binary classifier as the discrimination threshold is varied. It plots the true positive rate (TPR) against the false positive rate (FPR) at different threshold values. The AUC measures the entire two-dimensional area underneath the entire ROC curve, with a value ranging from 0 to 1. A model with perfect classification ability will have an AUC of 1, while a model with no classification ability will have an AUC of 0.5.

AUC is commonly used to evaluate the performance of binary classification models, as it provides a single number that summarizes the model’s ability to discriminate between positive and negative classes.

Quadratic-weighted kappa (QWK), also known as Cohen’s kappa, is a statistical measure of inter-rater reliability. It is used to assess the level of agreement between two or more raters when assigning categorical ratings to a set of items. Quadratic-weighted kappa takes into account the possibility of agreement occurring by chance and adjusts for this, providing a more accurate measure of agreement than simple percent agreement.

Quadratic-weighted kappa is calculated by comparing the observed agreement between raters to the expected agreement if the ratings were assigned randomly. The weights used in the calculation are based on the squared difference between the ratings, with larger differences receiving greater weight. A quadratic-weighted kappa value of 1 indicates perfect agreement between raters, while a value of 0 indicates no agreement beyond what would be expected by chance.

To solve the problem of 2 labels being too sparse, we expanded the dataset by rotating the images 90 ° and 180 ° before re adding them to the dataset.

During the training process, we adopted different strategies. The first was to only train the last FC layer for the pre-trained network. The second was to involve the entire network in training. Another influencing factor was whether the data was augmented. After multiple experiments, we obtained the following results.

**Table 1** ExperimentalResult

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Net | Data | AUC | QWK |
| 1 | Last layer | Normal | 0.836999695742681 | 0.7080253366951601 |
| 2 | Last layer | **Augmented** | 0.8528934377312402 | 0.7160168913665019 |
| 3 | Full | Normal | 0.8214358229975333 | 0.5541224889460077 |
| 4 | Full | **Augmented** | 0.8512585946521067 | 0.6974806911982201 |

Epoch=30

In the table above, we can obviously see concluded that the outcome is better after the data is augmented. But regarding to the net, we find that trainning only the FC Layer is better than full training. Reasons for it may be that the number of epoches is too small so the ResNet50 is not learning enough. Or that the pre-trained ResNet50 is already capable for image clasification. So we did another experiment to investigate if we increase the number of epoches, whether the outcome will change.

**Table 1** ExperimentalResult

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Net | AUC | QWK |
| 1 | Last layer | | 0.8175994344393365 | 0.645419266100113 |
| 2 | Full | | 0.8715757937073662 | 0.7273896930284596 |

Epoch=200

We can see that after the network has been fully trained, the effect of full training is better than training only the FC layer.

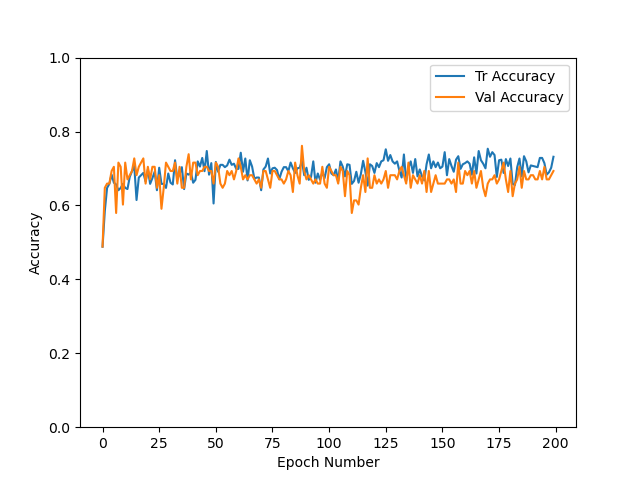


Fig 4 Only FC Layer

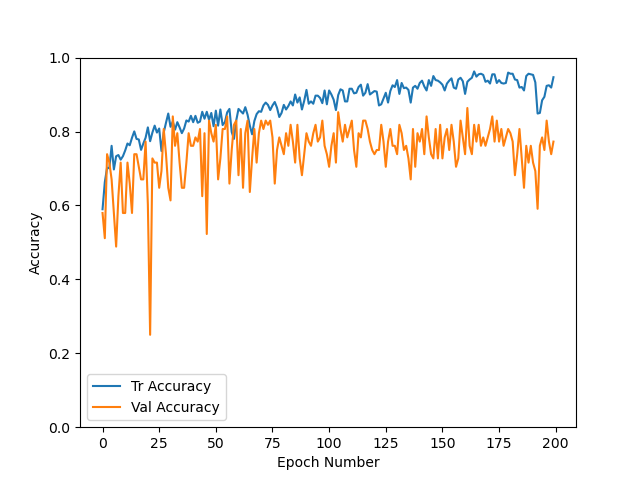


Fig 5 Full Trained

As we can see from Fig 4 and Fig 5, only FC layer method can quickly reach its optimal but it stays stable after

Overall, this project implemented a Diabetic Retinopathy Grading system using deep learning methods, which improved accuracy and speed compared to traditional methods. At the same time, this project used a data augmentation method to address the problem of sparse labels in medical images. However, there is still much room for improvement in this project. For example, different networks could be used to refine tasks such as image segmentation, feature extraction, and classification filtering.

ZHANGXIRAN

E-mail: 2697953125@sjtu.edu.cn

XiCheng Liu S. Author

ALEXANDRE LI HAO YAO S. Author