



Retinal Disease Analysis based on Retinal Fundus Photographs via Deep Learning

Xuezhou Wen

New York University, New York, 10012, U.S.

Email: xw2447@nyu.edu

Yufang Tang and Firstname C. Lastname

Name of Institution/Department, City, Zip Code, Country

Email: second.author, third.author@hostname2.org

Abstract—本文的贡献：使用数据增强的方法，解决了医学图像数据量不足，以及数据分布不均衡等问题，提升了基于深度学习的眼底疾病诊断准确率。As deep learning methods are being used in the field of retinal image analysis increasingly, the challenges of insufficient and imbalanced dataset begin to affect the performance of the retinal image analysis models. In this article, we introduced a training method which use data augmentation to solve the insufficiency and imbalance of retinal image dataset, particularly the retinal image for the classification of Diabetic Retinopathy (DR). By using the data augmentation method, we were able to obtain a balanced augmented retinal image dataset from an imbalanced original retinal image dataset, expand the size of the original retinal dataset about 4 times, and lead to better classification accuracy.

Index Terms— Retina fundus photographs, Retinal image, Deep learning, Feature map, Data augmentation

I INTRODUCTION

该部分内容主要介绍：（该部分主要是查阅其他人的论文，查阅过程中，将引述其他论文的地方，加

参考文献)

- 介绍眼底疾病对人的健康的严重危害 First of all, it is necessary for us to learn about the damage to human health caused by diabetic retinopathy. Diabetic Retinopathy (DR) is a major microvascular complication of diabetes, and it is the leading cause of blindness of working-age populations around the globe [1]. DR has become an urgent issue to the global health system as the numbers of people with DR and VTDR (Visual-Threatening Diabetic Retinopathy) are both growing [1]. Thus, it is an important task to develop an intelligent system for retinal disease detection.
- 眼底图像在眼底疾病诊断中的重要性 To diagnose DR, the physicians need to analyze the fundus images of the patients. The fundus photographs are visual records which demonstrate the current ophthalmoscopic appearance of patients' retina as well as the characteristics of diabetic retinopathy [2]. Thus, it is very important

for the physicians to analyse the fundus images when diagnosing DR.

- 目前，眼底图像分类的研究情况，主要介绍其他论文在眼底诊断中使用的方法，包括：非深度学习方法、深度学习方法，以及这两种方法取得的效果，即深度学习方法比非深度学习方法好。Nowadays, there are two major methods to diagnose DR. The first one is to manually inspect the morphological changes in microaneurysms, exudates, blood vessels, hemorrhages, and macula in the retinal image, which is very time-consuming and tedious [3]. The other method is to employ CAD (computer-aided diagnosis), powered by deep learning, and combined with human intervention, which has better performance and efficiency in detecting DR.
- 基于深度学习方法的眼底图像分类存在的问题，即训练样本数量不足、样本类别分布不均衡，以及这两个问题对基于深度学习方法的眼底图像分类方法的性能影响。Although the deep learning techniques for retinal images have surpassed most of the state-of-the-art-methods [4], there are still challenges remained. Due to the existence of the hidden layers in deep neural networks (DNN), it is hard for the researchers to come out a theoretical explanation on how to exactly extract the features of a retinal image. On the other hand, training a good model may require a massive and balanced dataset of labeled retinal images [4], which is difficult to acquire. Thus, it becomes necessary to develop a method to generate more retinal images by using data augmentation methods, so that we can obtain more and balanced dataset used for training.

通过阐述以上内容，引出本文的写作意图

- 基于眼底图像的眼底疾病诊断对人的健康很重

要。The diagnosis of retinal disease based on retinal image analysis is very important to people's health, especially in preserving people's vision, since DR is one of the most common cause for blindness [5].

- 基于深度学习的方法在眼底疾病诊断中具有很好的应用潜在价值。The retinal disease diagnosis powered by deep learning has great potential in applications including DR detection, since the deep network is able to exploit both simple and complex compositional features of retinal disease images [4]. Moreover, the deep network does not rely on hard-coded knowledge, but its self-learning ability to recognize raw data [4], using deep learning network can save a tremendous amount of human resources and achieve a higher diagnosis accuracy.
- 基于深度学习的方法在眼底疾病诊断中，训练样本数量不足、样本类别分布不均衡两个问题，必须要解决。In order to let deep learning better power the retinal diagnosis, we have to recognize that some problems need to be solved, such as insufficient labeled retinal images for the training of deep learning model, and the imbalance of the retinal image dataset. To solve these two issues, we choose to employ data augmentation to expand and balance the insufficient dataset.

II 介绍眼底图像和眼底疾病的相关知识

通过这部分内容，说明可以通过分类眼底图像来诊断眼底疾病

A 眼底图像的相关介绍

如何采集、眼底图像中各部分对应的眼底区域（例如：眼底血管等）The retina can be photographed directly by a fundus camera, which is a specialized microscope with a camera attached [2]. When the

patient is taken the fundus photograph, he will be asked to sit in front of the fundus camera, place his chin on the chin rest and his forehead against the bar of the fundus camera [2]. A retinal image contains a large amount of information, including characteristics of DR, such as macular edema and microaneurysms [2]. In Figure 1 and Figure 2, the basic structure of a normal human retina is shown. In Figure 1, the fovea is the dark round shape in the left side of image, macula being the round margin surrounding the fovea, and optic disc being the bright round shape covered under the blood vessels. The thick blue and red vessels in the center respectively refer to retinal vein and artery, and the thin blue and red vessels that spread across the entire retina are referred respectively to retinal venules and arterioles.

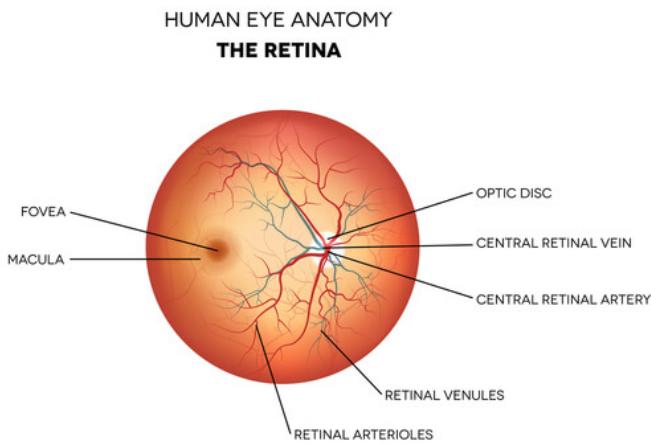


Fig. 1: The structure and components of a human retina,

refer to

<https://www.visioncareofmaine.com/retina-bangor/>

B 眼底疾病的介绍

There are several diseases that manifest as retinal impairments, including DR, AMD (age related macular degeneration), Cardiovascular diseases, Glaucoma and so on [4]. In this paper, we mainly focus on improving the deep learning methods for DR diagnosis. DR is a disease serving as

one of the prevalent consequences of diabetes mellitus (DM), caused by fluctuating level of glycemia. When DR is severe, it will become microaneurysms, neovascularization, hemorrhages, cotton wool spots, and exudates in the retinal region. From Figure 2, we can see four zoomed picture of lesions. The first one, the sac-like structure [4], enclosed by the yellow box is called microaneurysms, which refers to unusual dilation of retinal capillaries caused by lack of oxygen supply [4]. The emanation of fluffy white pitches [4] enclosed by blue box is called soft exudates or cotton wool spots. The lesion enclosed by the black box is called hemorrhages, which refers to the burst of vessels [4] caused by the blockade in arterioles. The part enclosed by the green box is called hard exudates, which refers to a lesion caused by leakage of fat, proteins and water from walls of retinal vessels [4].

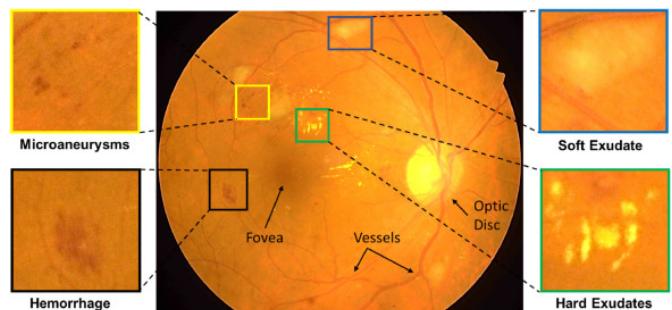


Fig. 2: Example of a retinal disease image

眼底疾病有哪些，每种眼底疾病的形成原因，以及在眼底图像中的表现（病灶）

III 基于深度学习的眼底图像分类方法

该部分内容，主要阐述基于深度学习的眼底图像分类方法的相关知识

A 深度学习方法的原理

Deep learning is a subfield of machine learning, and they both employ neural network. In a maiden

machine learning, we use a simple neural network, including an input layer, a hidden layer and an output layer, to train the machine. Data in a neural network is transmitted from one neural to another one, and this kind of non-deep machine learning requires manual data representation in the form of the features [4]. This means if we were to use a maiden machine learning, then we need to "teach" the machine about the features of all possible lesions which lead to DR detection. A deep learning machine, with its ability to conduct hierarchical feature extraction [4], has self-learning ability and does not require handcrafted features to be fed. 通过阐述深度学习方法的原理，说明深度学习方法与非深度学习方法（比如：神经网络）有着很大的不同，从而说明在理论上，深度学习方法是一种眼底疾病诊断的有效方法。

B 深度学习中的特征提取方法

In deep learning, particularly in retinal image analysis, a CNN (convolutional neural network) is usually used to better extract the features of the input image. A convolution operation uses dozens of random filters to "slide" across the input image, and creates the same number of filtered images called feature map, and passes them to the processing layers like maxpooling or dropout. The convolution operation repeats for several times, and eventually forms a very deep neural network. By performing multiple convolution operations, CNN's performance can surpass a simple neural network. With the power of CNN, a deep learning model can extract high level features from the images, and thus detect the DR more accurately. 阐述卷积算子提取眼底图像特征的原理、以及深度学习方法中核心的问题，即能够提取眼底图像的高级特征，从而能够比其他方法，更准确地诊断眼底疾病。

C 训练数据对深度学习的重要性

Researches have indicated that, the increase of data imbalance will have negative effect on the performance of classification models [6]. In the experiment mentioned above [6], the data with different class prevalence indexes were fed into the same group of neural network models, and the result shows that as the class prevalence index gets closer to 0.5, which represents a balanced class prevalence, the performance of the model increases, and eventually leads to the conclusion about the impact of imbalance dataset on the performance of neural network models. Moreover, researches have also shown that increasing the size of dataset can lead to better accuracy and improved generalization capability [7]. In another experiment [7], three datasets with the size of 400, 800, 1200 were fed into the same training model with the same data partition ratio. The result show that the error of the models to process unknown data decreases as the size of the dataset increases, which lead to the conclusion of the impact of dataset size referred above. Therefore, by analysing these two experiments conducted, we can learn that it is very beneficial to create a balanced dataset with a considerable size if we want to achieve any satisfying performance. 通过引述相关论文，阐述训练数据数量和类别数据的均衡对模型训练的重要性。

IV 本文提出的诊断眼底疾病的深度学习方法

这部分主要介绍本文使用的深度学习网络结构、以及数据增强方法

A 网络结构

We use four models [8][9][10][11] with different complexities to verify our hypothesis about the impact of balancing the dataset and expanding the size of the dataset. The model that we focus on most

is the model 1 [8], which is described as a mimic of the VGG16 neural network but with a smaller size. The model 1 contains five main structures: the first one is the input module, the second and third are convolution modules, the fourth is a flattening module, and the last one is the softmax output module. The first module contains a Con2D layer, a "relu" activation layer, a batch normalization layer, a Max-pooling layer and a dropout layer. The second and third module both have two groups of Con2D, "relu" activation, batch normalization, then a maxpooling and a dropout. The fourth module has a flattening layer, a dense layer, a "relu" activation layer, a batch normalization layer and a dropout layer. The last module has a dense layer and a "softmax" activation layer. For other three models, they all have a simpler structure than the model 1, and use simple layers, including Con2D, "relu" activation, maxpooling and dropout layers.

介绍本文的网络结构，阐述其中关键的计算，例如：卷积核提取特征、pooling 数据降维避免过拟合、及其他

B 数据增强方法

In an effort to expand and balance our dataset, we employ several augmentation methods to produce augmented images for the purpose of training. These augmentation methods include rotation, width and height shifting, rescaling, zooming, horizontally and vertically flipping, brightness and filling. We compare with two pairs of original and augmented images, from Figure 3 to Figure 6.

介绍本文用到了哪些数据增强方法，以及这些方法对图像做了哪些处理，用处理前后的图片做对比说明。



Fig. 3: Example of an original retinal image (degree 0)

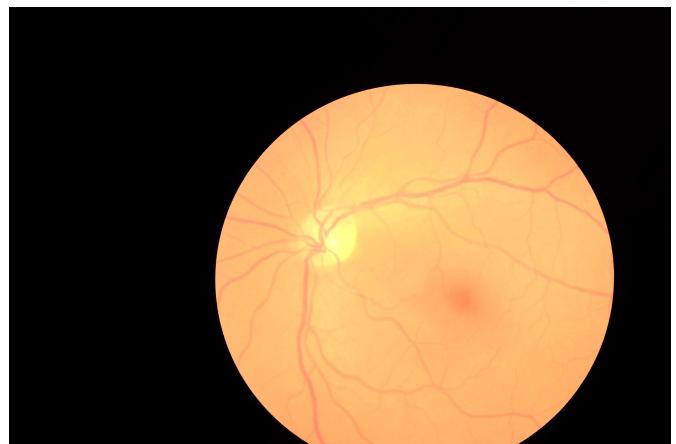


Fig. 4: Example of an augmented retinal image (degree 0)

V 对比实验

阐述使用的数据增强方法后，同一种网络结构，训练得到的模型，其性能得到了提升。

A 数据集介绍

本实验所用的数据集原图全部来自于世界上最大的数据科学社区 kaggle, <https://www.kaggle.com/c/diabetic-retinopathy-detection/data>。目前任何人都可以通过浏览以上的网址进行眼底图片的下载。kaggle 的眼底图片分为五对数据包，每对分别有训练和测试两个数据包，所以数据标签都存储在了一个 csv 文件中。本实验主要使用了全部五个训练数据包，我们将所有标签的图片都放在了同一个文件夹中，并将与每个图片对应的



Fig. 5: Example of an original retinal image (degree 4)

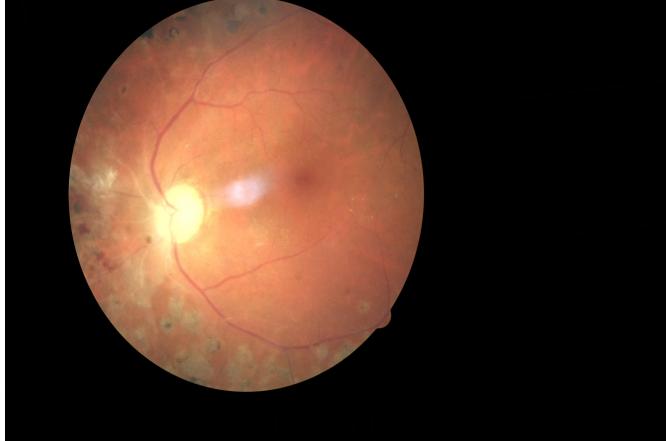


Fig. 6: Example of an augmented retinal image (degree 4)

标签都加在了它们的文件名中，以便之后的数据处理。

Kaggle 眼底图片的标签分为五类，0 类为阴性，1 类为轻度，2 类为中度，3 类为重度，4 类为糖尿病视网膜病变增殖。由于客观上这些类别发生的纪律性不同，所以每个数据包中的每类所含的图片数量也相差较大。在我们将五个训练数据包合并之后，各个类别的图片的数量为：0 类 13423 张，1 类 1312 张，2 类 2755 张，3 类 449 张，4 类 384 张，同时这个数据集也是用于训练 1 号模型的数据集。在我们使用了数据增强后的数据的每个类别的数量：0 类 13306 张（原图），1 类 13422 张（增强），2 类 13244 张（增强），3 类 13235 张（增强），4 类 13216 张（增强），同时这个数据集也是用来训练 2 号模型的数据集。

All the data used by this experiment come from Kaggle, which is one of the largest data science communities in the world. To obtain these data, you simply need to visit their website: <https://www.kaggle.com/c/diabetic-retinopathy-detection/data> and register a free account. The diabetic retinopathy detection dataset in Kaggle has five zip packages for training, five zip packages for testing, and a csv file for storing the ground truth of all ten packages, and we used the five training packages in our experiment. Before beginning to train, we reorganize the file structure of the Kaggle dataset, by checking the csv file, we find out each image's label, write it on its filename, and divide all training and testing images into two separate folders. After that, we convert all of these images into tfrecords.

First we will briefly introduce the classification of the Kaggle dataset images. The Kaggle dataset images provided were classified into five severity degrees of diabetic retinopathy. Degree 0 corresponds to no DR, degree 1 corresponds to mild, degree 2 corresponds to moderate, degree 3 corresponds to severe, degree 4 corresponds to proliferative DR. Due to the objective conditions, such as the actual unbalance distribution of the number of patients in each degree, the Kaggle dataset is also very unbalanced. We merge the images from the five training packages, obtain 13423 images for degree 0, 1312 images for degree 1, 2755 images for degree 2, 449 images for degree 3 and 384 images for degree 4, and we use this image dataset to train our model A. After the augmentation process, we get 13422 original images for degree 0, 13092 augmented images for degree 1, 13244 augmented images for degree 2, 13235 augmented images for degree 3, 13216 augmented images for degree 4, and this image dataset is used to train

model B. Note that the degree 0 images are original since there are already sufficient number to make the entire dataset generally balanced.

介绍采用的数据集来自哪里，数据的情况：数据量、数据类别、每种类别的数据量。实验中使用了哪几个类别，每个类别用的数据量（数据增强前后的数据量）。

B 实验数据

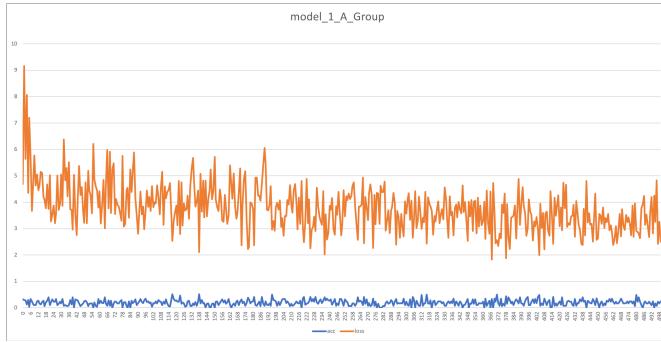


Fig. 7: Training data of model 1 using original dataset

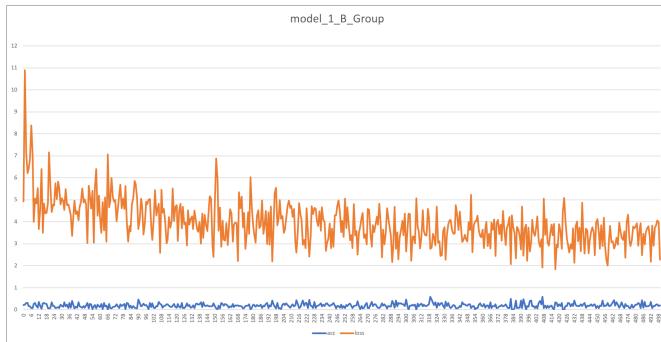


Fig. 8: Training data of model 1 using augmented dataset

From the 4 pairs of images, Figure 7 and 8, Figure 9 and 10, Figure 11 and 12, Figure 13 and 14, showing the four models' performance when training with original dataset and augmented dataset, it can be seen that, when considering the scaling of y-axis, model 2, model 3 and model 4 displayed an improvement in accuracy when training with augmented dataset. The accuracy of model 2, model 3 and model 4 were normally below 0.5 when training

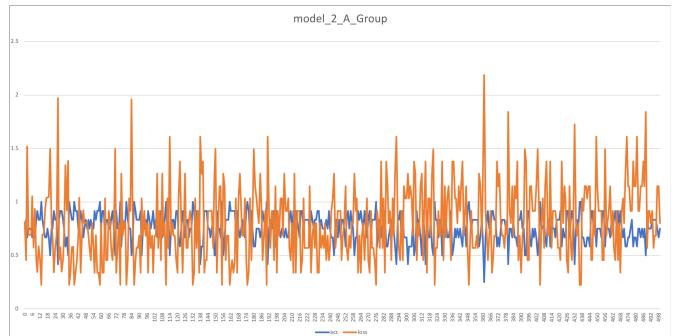


Fig. 9: Training data of model 2 using original dataset

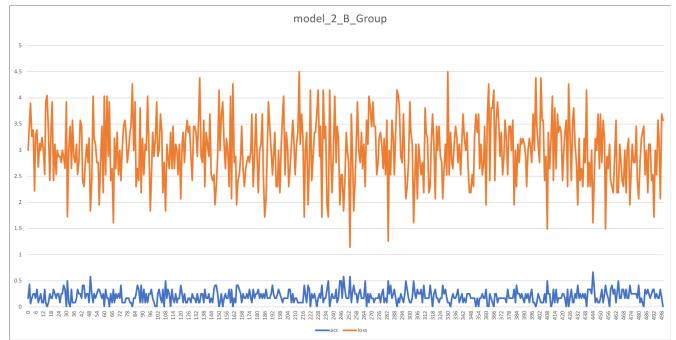


Fig. 10: Training data of model 2 using augmented dataset

with original dataset, while the accuracy of the same model was above 0.5 when training with augmented dataset. For model 1, training with augmented data does not have an accuracy lower than training with original data. Considering all above proof, we can come to conclusion that data augmentation can bring a certain degree of improvement on the accuracy of retinal disease diagnosis models.

On the other side, we write a python script to manually calculate the accuracy of our models' predictions. 505 images from the validation set, which have not been seen by the models when they are training, are used as prediction sample. The models provide five probability value corresponding to the five labels of severity of DR, and the label with largest value would be regarded as a prediction result of a single test case. The accuracy is calculated as total number of predictions that match the ground

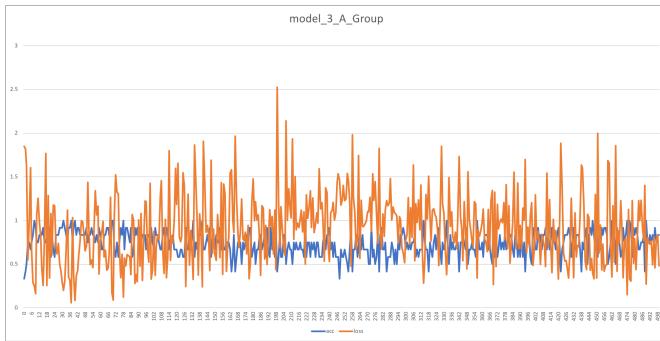


Fig. 11: Training data of model 3 using original dataset

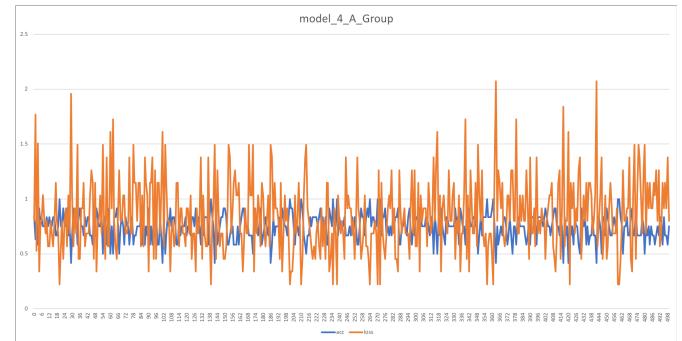


Fig. 13: Training data of model 4 using original dataset

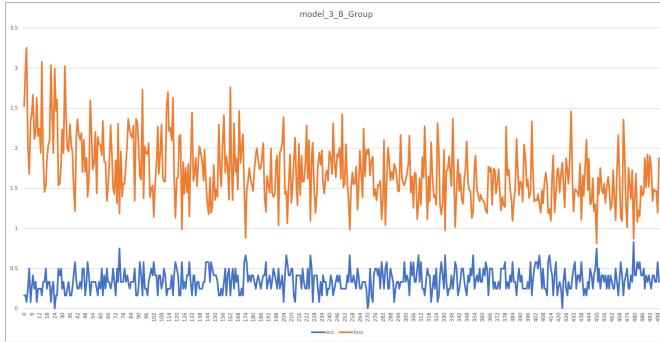


Fig. 12: Training data of model 3 using augmented dataset

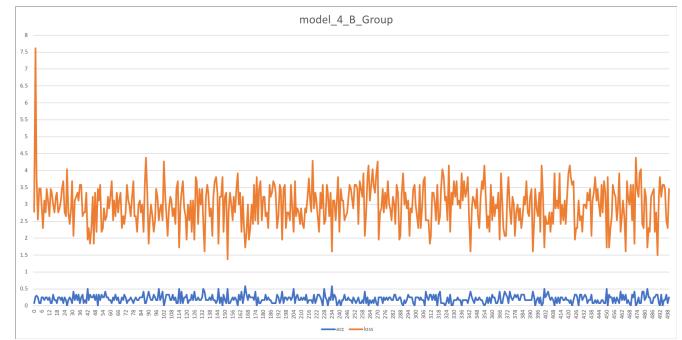


Fig. 14: Training data of model 4 using augmented dataset

truth divided by total number of images tested. In Figure 15, we can observe that the Group B using the augmented dataset displays an improved prediction accuracy. 以表格的形式，分别列出几个模型（5个），在使用原始数据，和使用增强后的数据，得到的模型，其前后分类准确率。即使用增强后的数据，训练得到的模型，其准确率高于使用原始数据训练得到的模型。

从而证明，数据增强方法，能够提升基于深度学习方法的眼底疾病诊断准确率。

C 实验分析

We observe that the reason why data augmentation can improve retinal disease model's accuracy. It is because that by using data augmentation, more training examples can be provided, which solves the issue of unbalanced dataset. As mentioned above, the original dataset is extremely unbalanced, with

13423 images for degree 0, 1312 images for degree 1, 2755 for degree 2, 449 for degree 3 and 384 for degree 4. By using augmentation method, we can create an expanded and balanced retinal image dataset, with 13422 original images for degree 0, 13092 augmented images for degree 1, 13244 augmented images for degree 2, 13235 augmented images for degree 3, 13216 augmented images for degree 4. The augmentation method greatly increased the number of images for label 1, label 2, label 3 and label 4, as well as the overall size of the retinal image dataset. 分析为什么数据增强方法，能够提升模型诊断眼底疾病的准确率。1、增加了训练样本数量。2、解决了各类别样本不均衡问题。

通过表格，罗列数据增强前后，每个类别图片数量，以及整个数据集图片总数量的对比。

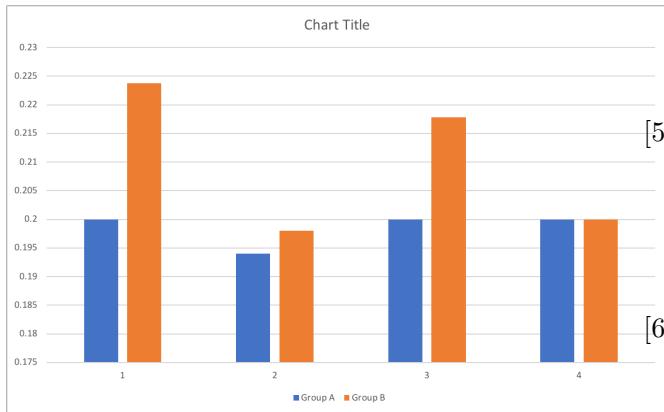


Fig. 15: Prediction results: Group A uses the original dataset, Group B uses the augmented dataset.

VI CONCLUSION

进一步说明，使用数据增强方法，能够提升基于深度学习的眼底疾病诊断准确率。From the experiment and analysis above, we can conclude that by using the method of data augmentation, we can improve the accuracy of retinal disease diagnosis powered by deep learning.

REFERENCES

- [1] Yingfeng Zheng, Mingguang He, and Nathan Congdon. The worldwide epidemic of diabetic retinopathy. *Indian journal of ophthalmology*, 60(5):428, 2012.
- [2] Jonathan Shankle. Ophthalmic photography: retinal photography, angiography, and electronic imaging: by patrick j. saine and marshall e. tyler, london, elsevier science, 2002, 398 pp., illus., hardcover. price us 150.00. *Survey of Ophthalmology*, 49(2) : 264, 2004.
- [3] Javeria Amin, Muhammad Sharif, and Mussarat Yasmin. A review on recent developments for detection of diabetic retinopathy. *Scientifica*, 2016, 2016.
- [4] Maryam Badar, Muhammad Haris, and Anam Fatima. Application of deep learning for retinal im-
- [5] Yinan Zhang and Mingqiang An. An active learning classifier for further reducing diabetic retinopathy screening system cost. *Computational and mathematical methods in medicine*, 2016, 2016.
- [6] Maciej A Mazurowski, Piotr A Habas, Jacek M Zurada, Joseph Y Lo, Jay A Baker, and Georgia D Tourassi. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural networks*, 21(2-3):427–436, 2008.
- [7] AR Ajiboye, Ruzaini Abdullah-Arshah, and Q Hongwu. Evaluating the effect of dataset size on predictive model using supervised learning technique. 2015.
- [8] Adrian Rosebrock. Multi-label classification with keras. 2018.
- [9] Jason Brownlee. How to classify photos of dogs and cats (with 97%). 2019.
- [10] Connor Shorten. Image classification keras tutorial: Kaggle dog breed challenge. 2018.
- [11] Francois Chollet. Building powerful image classification models using very little data. *Keras Blog*, 2016.