International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V68I12P210

Volume 68 Issue 12, 54-57, December 2020

Application of Convolutional Neural Network in Classification of Autofluorescence Image of Diabetic Retina Fundus

Suzhe Ye[#1], Daguan Ke[#2]

 # School of Ophthalmology & Optometry, School of Biomedical Engineering, Wenzhou Medical University, China

[1] 2540988415@qq.com, [2] 43774064@qq.com

Abstract *— Diabetic retinopathy is one of the complications of diabetes. The common medical diagnosis method is fluorescein fundus angiography, but the diagnosis process requires fluorescein sodium injection. The fundus autofluorescence technology used in this study can be harmless and has a better application prospect for patients who cannot be angiographic examinations. However, the naked eye cannot recognize the early fundus images and need to introduce computer-aided diagnosis. This paper's research object is 190 fundus autofluorescence images, and the accuracy of the 10-fold cross-check is used as the evaluation index. Compare the effects of convolutional neural network algorithms on classification performance under different image resolutions and image enhancement operations. The optimal image resolution is 64*64, the image enhancement operation is horizontal flip, and the optimal accuracy rate is 0.92105. After exploring the network structure, it is found that there is a better result without modifying the network. This article summarizes the following training steps: first, use the basic model to select the appropriate image resolution and image enhancement operation, and secondly, modify the network layer and explore the network through trial and error.*

Keywords *— Autofluorescence image of fundus; Convolutional neural network; Computer-aided diagnosis*

I. INTRODUCTION

Machines obtain knowledge and generate data models from data through big data and deep learning algorithms. In 1948, Shannon [1] proposed information entropy in the mathematical principles of communication. The introduction of information can eliminate the system's uncertainty and turn many intelligent problems into information processing problems. The deep mind deep learning tool developed by Google in 2010 solved the problem of multi-model parallelization. To stabilize the algorithm, a dedicated processor TPU for deep learning is also specially developed.

In 1995, Yann LeCun [2] proposed a convolutional neural network for handwritten digit recognition. In 2006, Hinton [3] and others first proposed the concept of deep learning. Use pre-training and fine-tuning techniques to find the approximate optimal solution, greatly reducing network training time. The convolutional neural network structure ranges from simply stacking convolutional layers to shrinking the convolution kernel to a more streamlined structure design. In 2012, AlexNet [4] won the ImageNet large-scale visual recognition competition and was significantly ahead of the traditional machine learning feature engineering methods in the same period. The structure of AlexNet is to pile up the convolutional layer, the pooling layer, the fully connected layer, and finally add the Softmax layer. The VGGnet [5] proposed in 2014 is deeper and can learn more complex representations of data. The main change is to use multiple 3*3 small convolution kernels to replace the previous 5*5 and 7*7 convolution kernels.

In contrast, the pooling layer uses 2*2 instead of 3*3, and the final network layer number reaches 20. GoogLeNet uses a series of Inception structures to replace the convolutional layer with a multi-branch and multi-scale structure to extract different image scale information. Batch normalization and Dropout are added in V3[6]. As the number of network layers deepens, the network layers close to the input layer cannot be effectively learned. He Kaiming and others proposed the residual network ResNet [7] to use a direct connection shortcut to solve the network degradation problem. When the network fitting is successful, it is equivalent to the upper and lower layers' identity mapping. Otherwise, it is to learn the difference between the upper and lower layers. Inception-ResNet[8] adds a residual structure to accelerate training. Based on the original ResNet, ResNeXt [9] introduced grouped convolution to accelerate training.

Computers' advantages are the low rate of missed diagnosis, high accuracy, high program stability, and no emotional or fatigue like people. In the case of a shortage of medical resources, the introduction of computers can reduce labor costs. Diabetic retinopathy is one of the complications of diabetes, among which 40%-45% of diabetic patients have diabetic retinopathy. A common medical diagnosis and treatment method is fluorescein fundus angiography to find microangiomas with abnormal fluorescence. It is necessary to inject fluorescein sodium during the examination, which may cause adverse reactions such as edema, nausea, vomiting, and severe cause of death. The fundus autofluorescence technique used in this article is a non-invasive and non-invasive method. Early detection through regular screening of retinal images can help reduce the number of blind people.

Kermany [10] et al. used the transfer learning method to detect OCT images to diagnose DME, CNV, and drusen, obtaining a 90% diagnostic accuracy rate. Since traditional research objects are fundus contrast images, this is the first time that deep neural networks have been applied to fundus autofluorescence images.

The first part of this article introduces the development of convolutional neural networks and autofluorescence technology. The second part introduces the research objects and the basic flow of the experiment. The third part analyzes the experimental results, and the fourth part is the summary and outlook.

II. MATERIALS AND METHOD

The research object is the fundus image photos of Ruian People's Hospital's fundus autofluorescence technology. The image size is 496*596 pixels. Among them, 36 were normal, and 59 were sick. They took photos of the left eye and right eye, totaling 95 objects and 190 pictures. The deep learning framework is TensorFlow 2.0, and the programming language is Python. The toolkits include Numpy, scikit-learn, matplotlib. The whole experiment process includes data division, data preprocessing, model building, training model, and model performance evaluation; secondly, it compares the impact of different image resolution and image enhancement on the classification results; finally explores the influence of network structure on the experimental results.

A. Data partition

The data is divided into the training set, validation set, and test set. The original data set has a total of 118 patient samples and 72 normal human samples. 10-fold cross-validation is required due to lack of data. Taking the first fold data as an example, the patient sample is divided into 94 training data, 12 verification data, and 12 test data. The normal sample is divided into 58 training data, 7 verification data, and 7 test data. For patient samples, because 12 samples per fold data will be missing 2 samples at the end, samples No. 117 and 118 will be reused in the last fold. There will be 2 samples left after 7 samples per fold for normal samples, and the numbers 71 and 72 will be discarded in the last fold.

B. Data preprocessing

Data preprocessing includes data cutting, format conversion, and data enhancement. Data enhancement is the use of basic rotation, zoom, and translation operations to increase data diversity. Part of the image below is related to the machine and not related to the condition, and this part of the redundant information needs to be removed. Cut the original 496*596 pixels into 496*496 selections. The image format conversion is to read the disk's image, divide it by 255.0 and convert it to float32 type TF data between 0 and 1. Data enhancement can generate realistic random transformation enhancement samples: generate more training data from existing training samples to ensure that the model will never see the same picture twice during training.

C. Build a model

The model is a convolutional neural network composed of three convolutional blocks, and each convolutional block has a maximum pooling layer. The penultimate layer has 512 unit fully connected layers, and the activation function uses Relu. The last layer has a 1 unit fully connected layer, and the activation function uses sigmoid. The optimizer of the model selects ADAM, and the model uses the binary cross-entropy loss function. The first layer structure of the network of different input images will change. The network structure of 256*256 input is shown in Figure 1.

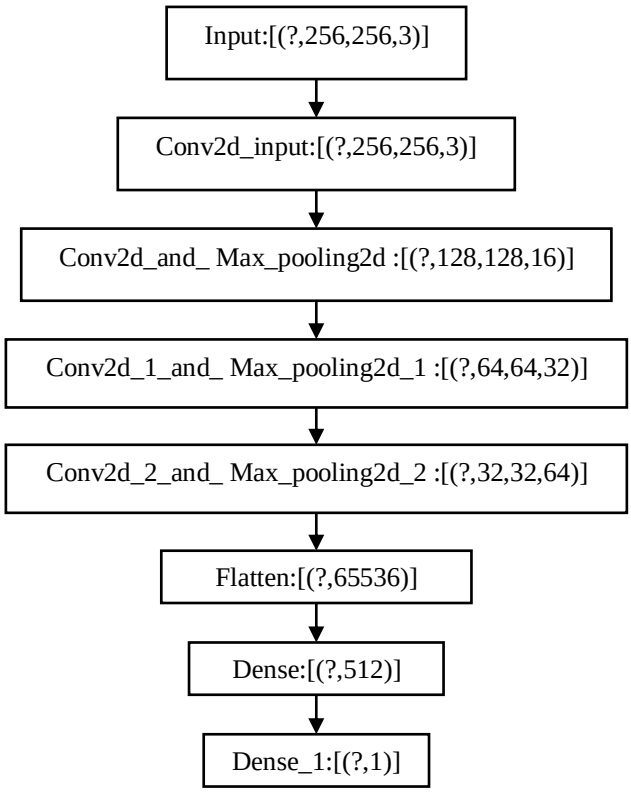

Fig. 1 Input model structure of 256*256 image

III. RESULTS AND DISCUSSION

This paper uses the 10-fold cross-validation accuracy rate as the evaluation index. The sample is divided into 10 sample subsets of equal size. Traverse these 10 subsets, in turn, each time using the current subset as the test set and the remaining part as the training set. Use the training set to train the model and the test set to evaluate the model to calculate the accuracy. Finally, calculate the average of the accuracy rates obtained 10 times to obtain the final evaluation index. The image enhancement operations that can be taken include rotation_range, width_shift, height_shift, horizontal_flip, and zoom_range. The values set here are 45, 0.15, True, and 0.5, respectively. So finally, there are 16 image enhancement operations to choose from. Make the following mark: if the four operations are executed, take 1 in the corresponding position; otherwise, take 0. Use 0001 for this. This mark indicates that only the image is scaled by 0.5. After confirming the operation,

compare the effects of 16*16, 32*32, 50*50, 64*64, 100*100, 128*128, 256*256 resolution on the experimental results.

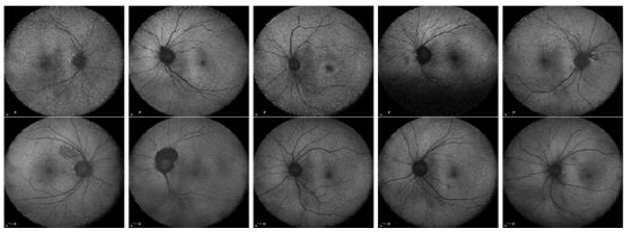

Fig. 2 Top: Fundus image with 128*128 resolution Bottom: Fundus image with 256*256 resolution

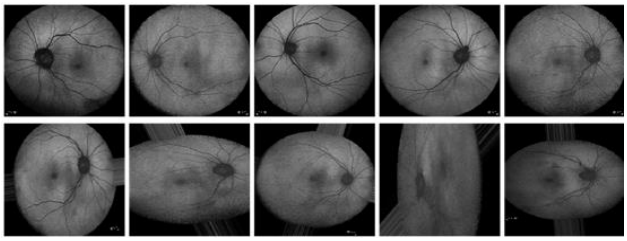

Fig. 3 Top: Horizontally flipped image. Bottom: Horizontally flipped and randomly rotated 45 degrees and zoomed 50% of the image

Obtain a training sample and apply the enhancement to the same image five times. Figure 2 shows fundus images with resolutions of 128 * 128 and 256 * 256. Figure 3 shows the results of various operations for the horizontally flipped image, the horizontally flipped and randomly rotated 45 degrees and zoomed 50% of the image. The corresponding identification codes are 0010 and 1011, respectively.

Table 1 shows the test accuracy of 7 image resolution types and 16 types of image processing methods. Each column of Table 1 indicates that the different resolutions of the input image are 16 * 16, 32 * 32, 50 * 50, 64 * 64, 100 * 100, 128 * 128, 256 * 256. Each row represents different image enhancement operations. The operations corresponding to the position are rotation_range, width_shift and height_shift, horizontal_flip, zoom_range.

Modify the network structure of Table 1 to view the impact on the test data, training data, and all data classification results, as shown in Table 2.Each column of Table 2 represents the basic network, adding 128 layers to the original 512, adding 64 layers to the original 512, adding two layers of 64 and 16 to the original 512, adding 256 layers of 512 to the original 512, Replace 512 with 1024, and replace the original 512 with 256. Each row represents the classification accuracy of test data, training data, and all data.

Experiment 0000 indicates that no operation is performed on the image, and the effect of different resolutions on the experimental results is compared. The result shows that the higher classification accuracy rates are 0.88947 and 0.87368 in 32*32 and 64*64. Therefore, low-size images can get better classification performance without image enhancement. By comparing experiment 1 with other image enhancement experiments, it can be seen that by flipping the image horizontally and using a resolution of 64*64, the optimal classification result is 0.92105. Explore the network structure by flipping horizontally and using a resolution of 64*64. Modify the network structure in Table 1. There are a total of 6 modification results: adding a layer of 128 to the original 512, adding a layer of 64 to the original 512, adding two layers of 64 and 16 to the original 512, and adding a layer of 256 to the original 512. Replace the original 512 with 1024, and replace the original 512 with 256. Table 2 shows the impact of different network structures on the test data, training data, and all data classification performance during the experiment. Whether it is to increase the network layer or replace the network layer, the classification performance has been reduced. Therefore, there is no need to adjust the fully connected layer further.

TABLE 1THE TEST ACCURACY OF 7 TYPES OF IMAGE RESOLUTION AND 16 TYPES OF IMAGE MANIPULATION METHODS

Image enhancement	Different image resolution						
	16	32	50	64	100	128	256
0000	0.74737	0.88947	0.85263	0.87368	0.8	0.73158	0.71579
1000	0.62632	0.62105	0.62105	0.66842	0.62632	0.66842	0.68947
1100	0.64737	0.61053	0.62632	0.65263	0.64211	0.71579	0.65263
1110	0.59474	0.60526	0.63684	0.62632	0.64737	0.66316	0.73684
1111	0.63684	0.64737	0.62105	0.59474	0.61053	0.64211	0.66842
1010	0.61579	0.64211	0.59474	0.63684	0.60526	0.65789	0.7
1001	0.64211	0.63684	0.63684	0.64737	0.6	0.68947	0.58947
1101	0.63158	0.59474	0.62105	0.63158	0.63158	0.68421	0.58421
1011	0.63158	0.63684	0.62632	0.63158	0.63684	0.67368	0.61053
0100	0.63158	0.63684	0.62632	0.67368	0.74211	0.72105	0.65789
0110	0.63158	0.64211	0.64737	0.63158	0.62632	0.74211	0.85263
0101	0.63158	0.62632	0.62632	0.64737	0.63158	0.70526	0.58947
0111	0.64737	0.63684	0.62632	0.64211	0.61053	0.66316	0.64211
0010	0.65789	0.86842	0.73684	0.92105	0.81053	0.76316	0.72105
0011	0.64211	0.61053	0.63684	0.61579	0.62632	0.70526	0.63158
0001	0.60526	0.64211	0.64211	0.69474	0.62105	0.74737	0.60526

TABLE 2 THE IMPACT OF DIFFERENT NETWORK STRUCTURES ON CLASSIFICATION PERFORMANCE

Data	Different network						
	Base	**Add128**	**Add64**	**Add64|16**	**Add256**	**R1024**	**R256**
test	0.92105	0.92105	0.91053	0.81579	0.88421	0.87368	0.65789
train	0.94094	0.91462	0.91345	0.81871	0.86608	0.92456	0.69649
all	0.93895	0.91526	0.91316	0.81842	0.86789	0.91947	0.69263

IV. CONCLUSIONS

This paper studies the application of a convolutional neural network algorithm on autofluorescence images of the diabetic retina. Compare the effects of different image resolutions and image enhancement operations on the results. The optimal image resolution is 64*64, the image enhancement operation is a horizontal image flip, and the optimal test accuracy rate is 0.92105. In the future, more networks can be considered for experimentation, such as modifying the convolutional layer to update the network; use the capnets [11] network proposed by Hinton. The network is inherently stable to transformations such as rotation and translation while using fewer data sets, which may be more conducive to processing medical images; Prognet [12] adjusts the network structure according to different images; The shufflenet [13] with separate convolution is used to remove useless parameters in the model. After visualizing the network, doctors can formulate more detailed treatment plans based on the patient's condition.

REFERENCES

[1] Shannon C E. A mathematical theory of communication[J]. The Bell system technical journal, 27(3) (1948) 379-423.

[2] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series[J]. The handbook of brain theory and neural networks, 3361(10) 1995.

[3] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. science, , 313(5786) (2006) 504-507.

[4] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[5] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[6] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 2818-2826.

[7] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 770-778.

[8] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet, and the impact of residual connections on learning[J]. arXiv preprint arXiv:1602.07261, 2016.

[9] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. (2017) 1492-1500.

[10] Kermany D, Zhang K, Goldbaum M. Large dataset of labeled optical coherence tomography (oct) and chest x-ray images[J]. Mendeley Data, v3 http://dx. doi. org/10.17632/rscbjbr9sj, 2018, 3.

[11] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[C]//Advances in neural information processing systems. (2017) 3856-3866.

[12] Zhang Z, Ning G, Cen Y, et al. Progressive neural networks for image classification[J]. arXiv preprint arXiv:1804.09803, 2018.

[13] Zhang X, Zhou X, Lin M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. (2018) 6848-6856.

[14] Narendra Mohan, Recognition of Skin Diseases using Deep Neural Network Optimized by Group Teaching Algorithm, International Journal of Engineering Trends and Technology 68(9)(2020):109-120.