International Journal of Biology and Life Sciences
ISSN: 2957-9511 | Vol. 16, No. 3, 2026

Deep Learning in Eye Image Analysis and Retinal Disease Computer Image Visual Recognition System

Shengzu Gu *

Faculty of Applied Mathematics and Control Processes, Saint Petersburg State University, 35 University pr., Petergof, St Petersburg, 198504, Russia

*** Corresponding author Email:** gusenzor@gmail.com

Abstract: Image analysis technology based on deep learning provides strong support for the early diagnosis and treatment of retinal diseases in diabetic patients. This paper focuses on a multi-feature fusion based convolutional neural network (CNN) recognition algorithm, which combines OpenCV image processing tools to realize efficient and accurate retinopathy image recognition of diabetic patients. The model extracts the deep features of retinal images by deep convolutional network and integrates the image information of different levels effectively by using multi-scale feature fusion strategy. The model further introduced the attention mechanism to enhance the attention to the lesion area, thus improving the accuracy of the lesion detection. In addition, the application of OpenCV tool greatly simplifies the retinal image preprocessing process of diabetic patients, ensuring the normalization and consistency of input data. In order to verify the effectiveness of the algorithm, a lot of simulation experiments are carried out in this paper. The experimental results show that compared with traditional methods, the proposed algorithm has higher accuracy and faster processing speed in the retinopathy image recognition task of diabetic patients. Especially in the complex background of diabetic retinopathy detection, the algorithm has shown significant advantages.

Keywords: Deep Learning; Ophthalmic Imaging; Diabetic Patients; Retinal Diseases; Image Recognition; Convolutional Neural Network.

1. Introduction

As the prevalence of diabetes continues to rise globally, Diabetic Retinopathy (DR) has become one of the leading causes of vision loss in adults. The occurrence and development of DR Is closely related to microvascular damage caused by high blood sugar levels, and early diagnosis and timely treatment are essential to prevent vision loss. The traditional ophthalmic image analysis relies on the subjective judgment of professional doctors and has some problems such as low diagnostic efficiency and poor consistency. Therefore, developing an efficient and accurate computer-aided diagnosis system is of great significance for improving the screening and diagnosis level of DR. As an open-source computer vision library, OpenCV plays an indispensable role in the technology stack of ophthalmic image processing. It provides a wealth of image processing and analysis tools, and provides strong support for the preprocessing, feature extraction and lesion detection of ophthalmic images. Traditional image recognition methods, such as feature extraction and classifier based on artificial design, can identify retinopathy to a certain extent, but because they rely on expert knowledge and manual feature design, it is difficult to capture subtle changes in lesion images, so there are shortcomings in sensitivity and specificity.

In recent years, deep learning techniques, especially Convolutional Neural Networks (CNNs), have made breakthroughs in the field of image recognition. Literature [1] proposes a fundus image classification method based on CNN, which effectively solves the limitations of traditional machine learning methods in feature extraction. In literature [2], a multi-scale convolutional network structure was adopted to improve the detection accuracy of DR Lesions by fusing features at different levels. However, these studies tend to focus on a single type of ophthalmic image analysis and fail to take full advantage of the comprehensive information from multiple image data. The literature [3] attempts to combine Optical Coherence Tomography (OCT) images with fundus photographs to improve the diagnostic performance of DR Through feature fusion technology. However, the generalization capability of this method in clinical application remains to be verified. The importance of retinopathy image recognition is self-evident, and it is the key to realize the early diagnosis and intervention of DR.

The research content of this paper mainly includes the following aspects: First, the application of existing deep learning in eye image analysis is reviewed, with special attention to the role of CNN in DR Diagnosis [4]. Secondly, a novel deep learning framework is proposed, which can simultaneously process fundus photos and OCT images, and extract more lesion information through feature fusion technology. Thirdly, a computer image visual recognition system for retinal disease in diabetic patients is constructed, which can automatically identify the different stages of DR And provide quantitative lesion assessment. Finally, the proposed system will be subjected to rigorous performance tests and compared with existing diagnostic methods to evaluate its feasibility and superiority in clinical practice.

2. Algorithm Design

2.1. Convolutional Neural Networks

Lecun et al. introduced CNN for the first time in 1998 and established LeNet model, but due to its limited computing power and samples, it has not been widely promoted. In 2012, Alex Net, first built by Alex et al., achieved great success, triggering a new wave of deep learning technologies. In recent years, a large number of modeling methods based on convolutional neural networks have been widely used. At

present, CNN has been well used in many fields, such as voice recognition and machine vision.

2.1.1. Principle and Structure

The basic structure of CNN has five parts:

1) Input layer. When using CNN to analyze an image, it is generally a pixel array of an image. The algorithm is carried out in RGB three-way, where the length and width are expressed as the image size respectively, while the depth is 3.

2) Convolution layer. The convolutional overlay is the center of CNN. Each layer in the convolutional network has a small input value that is as large as the convolution core. The convolution layer has:

$$u_k^h = g\left(\sum_{i=M_k} u_k^{(h-1)} \cdot t_i k^h + \sigma_k^h\right) \tag{1}$$

Where h represents the h level of the neural network, M_k represents the perception region of the previous level, σ represents bias, t represents the convolutional core, and $g(\cdot)$ represents nonlinear activity. The gender-neutral activation function currently used in CNN is the ReLU function, and its specific formula is as follows:

$$g(u) = \max(0, u) \tag{2}$$

3) Lower sampling layer. The main function of the algorithm is to reduce the input characteristics of the network and reduce the operational complexity of the network [5]. There are two basic ideas of the algorithm: one is maximum pool, which makes the maximum value of the output and input features; The other is averaging pooling, which averages the input features. Sample downward in the following way:

$$u_k^h = g(\delta_k^E f(u_k^{E-1}) + \sigma_k^h) \tag{3}$$

Where $f(\cdot)$ is the pooling function, δ is the weighting, and σ is the bias.

4) Fully connected layer. In the past, convolution and down-sampling operations are essentially feature extraction and compression, so it is likely to lead to the loss of features. In this paper, complete connectivity is used to enhance these features.

5) Classifier. CNNS usually use the soft maximum as a classifier to classify features enhanced by the entire connectivity level.

2.1.2. Application in Retinal Image Analysis

In the detection of retinopathy, CNN can automatically extract features from fundus images, which are crucial for diagnosing lesions. For example, typical pathological features such as hemorrhagic spots, exudates, and microaneurysms can be identified by CNN. The automatic extraction of these features not only improves the accuracy of diagnosis, but also greatly speeds up the diagnosis speed [6]. The use of CNN in the detection of retinopathy is constantly improving. For example, some studies have proposed an automatic diagnosis system based on CNN, which can preprocess, extract features and classify fundus images to achieve early diagnosis of diabetic retinopathy, evaluation of lesion degree and prediction of disease progression. In addition, two models have been proposed for the detection and diagnosis of diabetic retinopathy through robust and interpretable convolutional neural networks to distinguish between healthy eyes and eyes with retinopathy, and between non-proliferative retinopathy and weak and severe proliferative retinopathy.

2.2. Model Design

Firstly, it can be seen from formula (2) that if the ReLU function is used as the startup function of the neural network, all the negative characteristic information helpful for recognition will be eliminated. This will inevitably lead to the loss of useful information. Leaky ReLU's nonlinear activation function is used. Its mathematical formula is:

$$p_i = \begin{cases} u_i, & u_i > 0 \\ \dfrac{u_i}{\lambda_i}, & u_i \le 0 \end{cases} \tag{4}$$

2.2.1. Multi-feature Fusion Strategy and Implementation

Secondly, the batch normalization -BN method is introduced into the construction of neural network to accelerate the learning rate and generalization effect. Its essence is to normalize the features of the upper layer, and then feed them into the neural network of this layer. Then the next step is featuring extraction. Finally, according to the VGG network architecture, the network support network convolutional network is studied. This network architecture uses two modules, type 1 and type 2. Each model consists of several smaller convolution kernels, which reduces the number of parameters in the model while ensuring the same convolution effect [7]. This method can effectively remove the negative information in the fundus image, so as to achieve the purpose of effectively using the visual information. BN is added after each model to improve the generalization ability of the model on training samples and test samples. In each module, there is a pooling layer connected to the previous convolution layer. The three levels are completely connected. The Dropout layer is added to realize effective regulation of additional parameters, thus improving the comprehensive capability of the neural network. Figure 1 shows the structure of Supplement Net (image cited in JMIR Medical Informatics, 2020, 8(7): e17580).

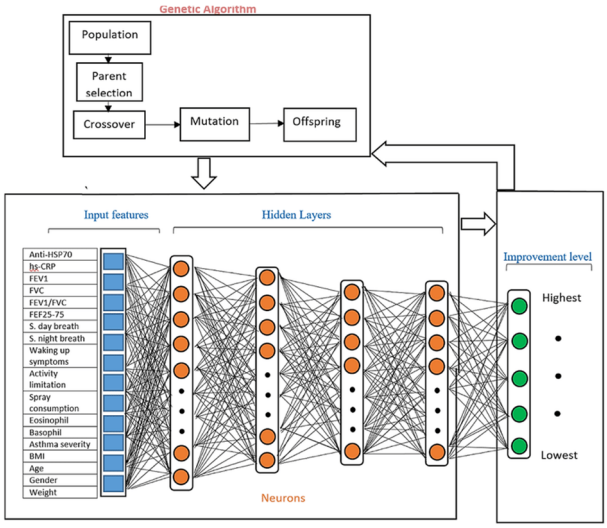


Fig 1. Supplement Net structure

2.2.2. Comparison and Advantages with Traditional Model

Convolutional neural networks (CNNS) are a special type of neural networks that are particularly suited for processing data with a grid-like structure, such as images and videos. The neurons in the convolutional layer of the CNN are only connected to a small number of their neighbors, which means that they share weights, greatly reducing the number of parameters in the model and thus reducing the risk of

overfitting. The CNN's convolution kernel slides over the image, capturing local features with each slide, which allows the CNN to capture local details of the image [8]. By stacking multiple convolution and pooling layers, CNNS are able to learn the spatial hierarchy of images, from low-level features (such as edges and textures) to high-level features (such as object shape and structure). CNN is generally superior to SVM in processing image data because CNN is able to extract useful features directly from the image, whereas SVM usually requires manual selection or design of features. In addition, CNNS generally perform better when dealing with large data sets and complex tasks.

3. Experimental Design

3.1. Experimental Environment

3.1.1. Selection and Introduction of Data Sets

The study used a publicly available fundus image dataset containing thousands of retinal images of diabetic patients from various medical institutions. These images cover a variety of retinopathy, including but not limited to microaneurysms, bleeding, hard exudation, soft exudation, and neovascularization. Each image in the dataset has been annotated by a professional ophthalmologist to ensure the accuracy and reliability of the label. In order to verify the generalization ability of the model, a part of external data sets that are not involved in training are selected for testing [9]. This article is based on the original data of major universities and competition platforms. In this study, 5 groups of samples were selected, with a total of 31845 fundus images. Of these images, 14,182 were normal and 17,663 were impaired (Table 1).

Table 1. Initial data set

Graphics	FIRE	DIARETDB1	Messidor	DR1	Kaggle-DR	Sum total
Normal image	0	28	563	716	12876	14182
Lesion image	279	65	688	341	16291	17663

For different solution sets of original samples, OpenCV technology is used to normalize and enhance them to facilitate the learning of the model. Firstly, using OpenCV's resize () technology, the resolution of the fundus image is reduced to 64x64x3. (2) Divide the data set into two main components, namely Data1 and Data Final. Using this algorithm to classify images, 24,457 eye images of different types were obtained. With a data completion of 20%, approximately 6,114 fundus images were provided for the final model trial. (3) Use the get Affine conversion function, get Perspective function and warp Perspective function in OpenCV to perform affine conversion on Data1, and use flip function to reverse up and down and other methods to extend Data1. A new Data1 containing 122,285 eye images was obtained. (4) Divide Data1 into two parts, namely Data train and Data Test. Data train accounted for 80%, and 97828 fundus images were modeled from the top down. Among them, Data Test accounted for 20%, and a total of 24,457 fundus images were used as experimental samples. Table 2 shows the set of graded test data. Figure 2 shows a dataset of retinal images of some diabetic patients.

Table 2. Experimental data set

Data set	Data Train	Data Test	Data Final
Image quantity	101904	25476	6369

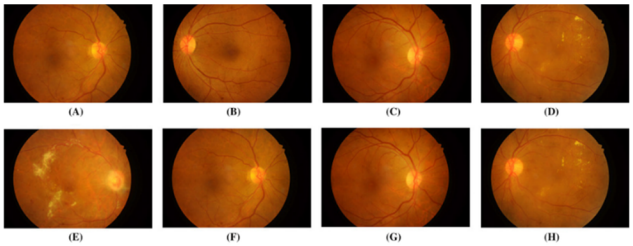


Fig 2. Retinal image dataset of some diabetic patients

3.1.2. Software and Hardware Configuration

The establishment of experimental environment is the key to ensure the model training and testing. The experimental platform configuration of this paper is as follows:

Hardware configuration: The experiment uses a high-performance GPU cluster equipped with NVIDIA Tesla V100 graphics card with ample video memory and powerful floating point computing power to support large-scale parallel computing. At the same time, the server is equipped with Intel Xeon Gold processor and high-speed SSD storage, ensuring the efficiency of data reading and processing.

Software configuration: The experiment runs on the Ubuntu operating system and uses Python as the main programming language. The deep learning framework has chosen TensorFlow, a widely used open-source platform that offers a rich API and efficient computational graph optimization. In addition, this paper uses Keras as a high-level neural network API to facilitate rapid prototyping and debugging of the model.

In the experiment process, this paper adopts distributed training technology, through the way of data parallel and model parallel, make full use of GPU cluster computing resources, accelerate the training process of the model [10]. At the same time, in order to monitor the training state and performance of the model, this paper integrated the Tensor Board tool to record and visualize key indicators such as loss function and accuracy in the training process in real time.

3.2. Experimental Results

3.2.1. Determination of Evaluation Indicators

This project intends to establish a training sample library based on F1-score. In this paper, the method of correct rate and response rate is adopted to improve the stability and reliability of evaluation. The relevant calculation formula is:

$$precision = \frac{TP}{TP + FP} \tag{5}$$

$$recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 - score = \frac{2 \cdot precision \cdot recall}{presion + recall} \tag{7}$$

In the formula, TP represents the correct number of categories for active sampling. FP represents the number of negative samples mislabeled as positive. FN represents the number of positive samples that were mislabeled as negative.

The range of F1-score is [0,1]. When the range is larger, the recognition rate is higher.

3.2.2. Comparison with Experimental Results of Other Methods

The model was confirmed in data testing and final data, and the results are shown in Table 3 below. Table 3 shows that the identification accuracy of the model in Data Test and Data Final are 94.96% and 98.34%, respectively. The main reason is that in the training process, the training model adopted 5 different enhancement methods, including a large number of eye image features, so the obtained model has better generalization performance [11]. In addition, the features in different convolutional networks are fused effectively, which makes the algorithm better reflect the eye details, and thus achieves good recognition effect and robustness on data sets with a small number of features.

Table 3. Recognition of the model on the test data set

Data set	Correctly identify the number of images	Image volume	Recognition accuracy rate /%
Data Test	23224	25476	94.96
Data Final	6013	6369	98.34

Compared with the commonly used Compact-Net methods of Alex-Net, Google-Net, ResNet-101 and other literatures [4], the effectiveness of the proposed test method was tested. The values of F1-score in Data Test and Data Final and the average accuracy of these two data sets are listed in Table 4.

Table 4. Comparison of F1-score values of different network models (%)

Algorithm	Data Test_F1-score	DataFinal_F1-score	Average accuracy
Alex-Net	86.70	87.61	87.16
Google-Net	89.94	90.13	90.03
Compact-Net	90.99	92.80	91.90
ResNet - 101.	99.29	96.98	98.14
Textual algorithm	96.19	100.13	98.16

The method achieved 92.80% F1-score in the data test, compared with Alex-Net, Google-Net and Compact Net, The F1-score of this method is 8.25,5.24 percentage points higher than that of Alex-Net, Google-Net and Compact-Net. In Data Final, compared to Alex-Net, Google-Net, Compact Net, and ResNet-101, Compared with the F1-score of Alex-Net, Google-Net, Compact-Net and ResNet-101, the F1-score proposed in this paper has an average increase of 11.25,10.2,7.42 and 4.21 percentage points, and its performance index is the best. It shows that the method completely satisfies the design index in practice. Compared with the other four methods, the average accuracy of this method is 94.23%, which is 11.24,8.21,6.25 and 0.03 percentage points higher than that of the other four methods, which proves that this method can be used well for fundus image classification.

The prediction accuracy of this method is 94.23%, and the prediction accuracy of ResNet-101 is 95.32%, which is very similar. This is mainly because ResNet-101 uses a 101-level network model, and generally a deeper network model can better identify images. The experiment was built on a VGG-1616 class network architecture, which resulted in a 3.12% improvement in the F1-score score obtained by ResNet-101 under data testing conditions compared to existing methods. The results show that better learning results can be obtained by using 16-level network for feature fusion.

3.2.3. Effect Analysis of Complex Lesion Recognition

Figure 3 shows the image feature extraction and feature fusion methods based on convolutional neural networks. It can be seen from the figure that most of the features in the original figure are extracted from Cov1_1, Cov_2, Cov1_3, and there are only boundaries in Cov2_1, Cov2_2, Cov2_3. Further study shows that with the increase of network depth, Cov3_1, Cov3_2, Cov3_3, Cov4_1, Cov4_2, Cov4_3 can only extract higher order image features. From the above convolution operation, it can be seen that different convolution functions can extract different features, and as the depth of the neural network increases, the extracted features become more and more abstract. The features in each convolutional network are fused to form an input layer of a complete connected layer FC1. By analyzing the working principle of convolutional neural network, it can help to optimize the model in the future.

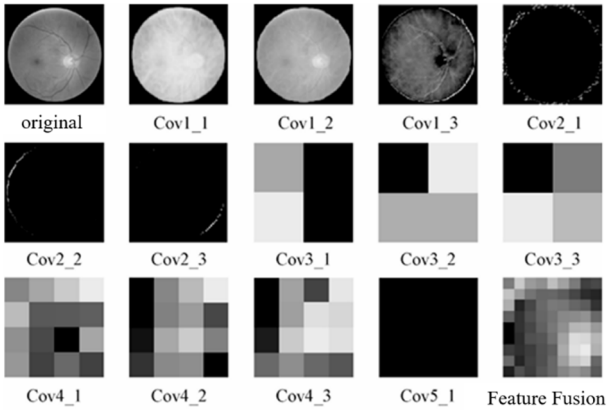


Fig 3. Visualization of image features in part of the convolution process

3.2.4. Evaluation of Model Accuracy, Efficiency and Robustness

During the trial, each node of the data task was set to 30, and a maximum of 3,000 iterations could be performed. The initial value of the weight is set as a truncated normal distribution, that is, an average of 0.01 and an average of 0.1, to prevent overweighting or overweighting caused by the use of normal distribution. In the loss function, the weighted fading coefficient is 0.1. In order to accelerate the convergence of the algorithm, the initial value of 0.07 is used to accelerate the convergence of the algorithm. In this paper, the soft classification algorithm is used to study binocular vision. When the repetition time increases, the relationship between Loss and Accuracy is shown in Figure 4.

After approaching 1000 turns, the value of Loss decreases from 1.0 to 0.18, and the accuracy increases from 0 to 0.82, because it looks for the fastest drop by the value of the directional derivative of the lost function. From 1000 rounds to 1750 rounds, the values of Loss and Accuracy changed slightly, and the variation range was between [0.06,0.27] and [0.73,0.94], which was caused by the random gradient reduction method being easy to fall into local extreme values. After 1750 training sessions, the value of Loss function loss tends to be consistent at about 0.06, and the precision value also becomes stable at 0.94. The above research proves that the established experimental model achieves good results in image recognition.

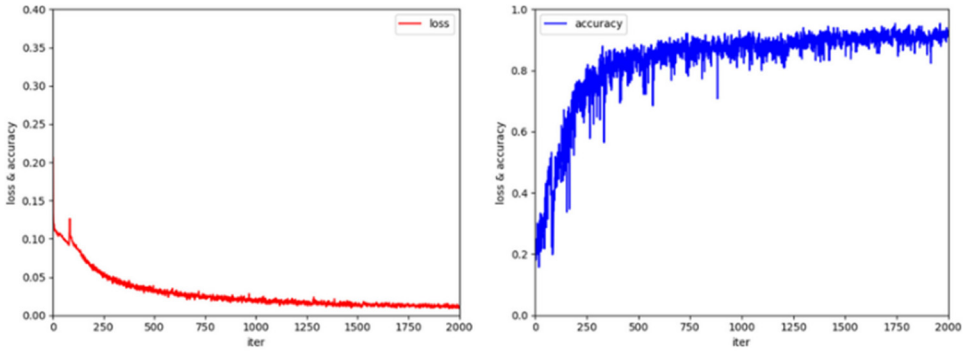


Fig 4. Change curves of Loss and Accuracy

Then the method is tested and the results show that the method has good robustness. In order to ensure that there is only one variable in the data test library, randomly select samples with the same degree of data completion, which is called data test *. The F1-score difference of the algorithm in data test * and data final is shown in Table 5. Difference represents the difference between the method and the training sample identification results, and its scale reflects the robustness of the method in a sense. The greater the difference, the better the robustness. The results show that the proposed method has good robustness, while ResNet-101 has the weakest robustness.

Table 5. Difference (%) of F1-score in Data Test* and DataFinal

F1-score	ResNet-101	Google-Net	Alex-Net	Compact-Net	Textual algorithm
Data Test*_F1-score	96.93	89.98	86.82	91.08	98.03
DataFinal-F1-score	96.98	90.13	87.61	92.80	99.65
Difference value	0.05	0.15	0.79	1.72	2.09

4. Conclusion

This study explores the application of deep learning technology in the field of ophthalmic image analysis, especially the role of computer image visual recognition system in diabetic retinopathy (DR). By constructing and training the convolutional neural network (CNN) model, the automatic detection and classification of DR Lesions in fundus images are successfully realized. The results show that the deep learning approach has demonstrated excellent performance in identifying various clinical manifestations of DR, such as microaneurysms, bleeds, exudates, etc. Through the effective fusion of CNN features at different levels, the model in this paper can capture more abundant and diversified image information, so as to extract more discriminative features from complex fundus images. This strategy significantly improves the generalization ability of the model, allowing it to maintain a stable recognition effect on previously unseen data sets.

References

[1] Liu, T. L., Zhang, Z. K., & Wu, J. G. (2022). System model and access control for medical image collaborative analysis. Computer Science and Exploration, 16, 1779–1791.

[2] Lian, X. F., Liu, Z. Y., Zhang, L., et al. (2021). An image recognition method for retinopathy based on deep learning. Computer Application and Software, 38, 179–185.

[3] Gu, T. F., Hao, P. Y., Bai, C., et al. (2021). Classification of diabetic retinopathy combined with multichannel attention. Journal of Image and Graphics, 26, 110–114.

[4] Wei, Z., Zhao, G., Wang, N. C., et al. (2020). Research on retinal classification of diabetes based on deep learning. Chinese Medical Equipment, 17, 52–55.

[5] Yu, X. Y., Liu, J. X., Xue, W. C., et al. (2021). Study on image classification model of diabetic retinopathy based on improved CNN. Modern Electronic Technology, 44, 168–172.

[6] Lu, Y. Q., Chen, M. H., Qin, K. B., et al. (2023). OCT image super-resolution reconstruction based on Pyramid Transformer. Chinese Journal of Lasers, 50, 61–72.

[7] Yang, A. P., Lu, X., & Zhao, Y. W. (2021). Application of artificial intelligence in retinal diseases. International Journal of Ophthalmology, 21, 43–48.

[8] Chen, Y., Zhou, Y. J., & Ding, H. (2020). An XNet-CNN retinal image classification method for diabetes mellitus. Journal of Harbin University of Science and Technology, 25, 74–78.

[9] Zhang, X. X., Zhang, M. Y., Zhao, W. X., et al. (2024). Research progress of psychological distress in diabetic retinopathy patients. Chinese Journal of Modern Nursing, 30, 2101–2106.

[10] Lian, X. F., Liu, Z. Y., Zhang, L., et al. (2021). An image recognition method for retinopathy based on deep learning. Computer Application and Software, 38, 72–77.

[11] Fan, W., Wang, X. L., Ma, X., et al. (2022). Multimodal deep learning model for staging diabetic retinopathy based on ultra-wide-angle fluorescein fundus angiography images. Chinese Journal of fundus Diseases, 38, 7–14.