

Blink Fatigue Detection Algorithm Based on Improved Lenet-5

Lei Chao

School of computer science and engineering
Xi'an Technological University
Xi'an, China
E-mail: callofduty2015@163.com

Lin Zhi

School of computer science and engineering
Xi'an Technological University
Xi'an, China
E-mail: 592880838@qq.com

Wang Changyuan

School of computer science and engineering
Xi'an Technological University
Xi'an, China
E-mail: cyw901@163.com

Huang Wenbo

School of computer science and engineering
Xi'an Technological University
Xi'an, China
E-mail: 804833079@qq.com

Abstract—Fatigue driving is the main factor in many traffic accidents. The eye behavior is the main direction in the field of fatigue driving research. It reflects the degree of fatigue of the human brain to some extent. In this paper, the publicized CEW human eye opening and closing data set is used to preprocess the human eye image, and then the preprocessed image is placed in the improved LeNet-5 convolutional neural network. In order to fully extract the human eye features, the network is added. The number of layers, at the same time, to speed up the network convergence speed in order to prevent the gradient from disappearing, the activation function is changed from Tanh to ReLU function. Experiments show that the algorithm has a blink recognition rate of 93.5% in the public CEW dataset, and an accuracy rate of 5.1% compared with the unmodified LeNet-5. This method has a good blink detection effect and has important application value in the field of fatigue driving.

Keywords-Fatigue Driving; Blinking Algorithm; Convolutional Neural Network; Network Optimization

I. INTRODUCTION

With the development of science and technology, people's material living standards have improved, and there are more and more motor vehicles on the road. However, the increase of vehicles has also caused more traffic accidents. According to reports, the main factor in traffic accidents is fatigue driving [1, 2], and currently fatigue driving equipment is relatively less used in our lives. Physiological fatigue testing is mainly analyzed by EEG signals, facial features and convolutional neural network features. Cao Ang [3] introduced the time-frequency characteristics of sEMG signal, and collected the human sEMG signal through AgCl surface patch electrode and high-precision analog front end ADS1299, and after pre-processing such as wavelet denoising, extract the time domain of sEMG signal reflecting human muscle fatigue state. Frequency domain characteristics. However, the disadvantage of this method is that the device is accessible to the human body, and the

wearing device itself has an influence on the human body. Li Weiwei [4] combined skin color detection and Adaboost method for face detection, and then processed the detected eye image by adaptive binarization and mathematical morphology to extract the blink state of the eye. However, the AdaBoost algorithm is very sensitive to the rapid opening and closing behavior of blinking. It is easily affected by the illumination and blink speed during image processing. Just the blinking behavior is a very fast process, so the recognition rate of this method is relatively low. In this paper, the improved LeNet-5 convolutional neural network is used to detect the blinking behavior. The whole process is detected in the natural state, which is not affected by other things, and has a good detection effect. Figure 1.

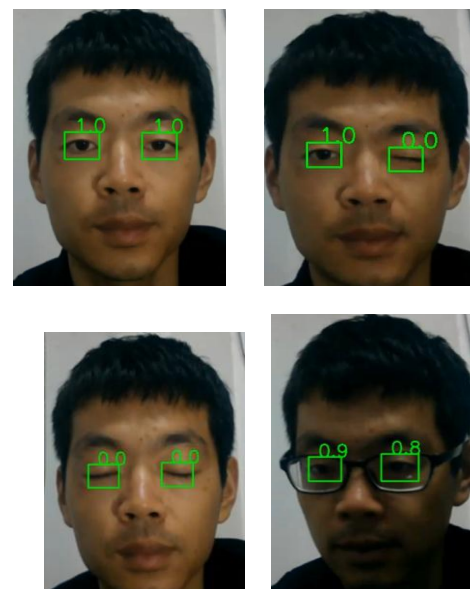


Figure 1. Recognition rate of different blink levels

II. RELATED WORK

A. Convolutional neural network principle

Deep learning technology has been widely used in many academic fields, and has achieved very good results in image classification and target recognition [5, 6]. LeCun [7] first realized the recognition of handwritten characters using LeNet-5. Subsequently, various improved LeNet-5s are widely used in many areas of life. This chapter introduces the classic LeNet-5 and the improved LeNet-5 network structure.

B. LeNet-5 Convolutional Neural Network

LeNet-5 [7], proposed by Yann LeCun, is a very efficient convolutional neural network for handwritten character recognition. Although this network structure is small, it contains the basic modules of deep learning: convolutional layer, pooling layer, and full-link layer. LeNet-5 is the basis of other deep learning models. There are 8 layers from input layer to output layer. Each layer contains trainable parameters. Each layer has multiple Feature Maps. Each Feature Map is extracted by a convolution filter. Enter the characteristics. LeNet-5 network structure is shown in Figure 2.

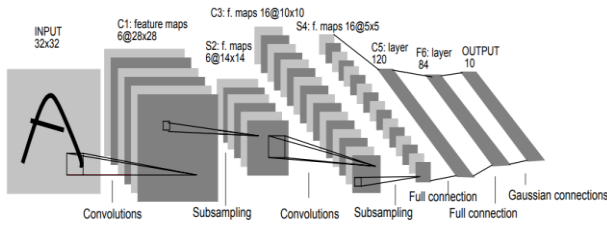


Figure 2. LeNet-5 network structure diagram

1) INPUT-input layer

The first is the data input layer. The size of the input image is uniformly normalized to 32*32. The layers below the layer C represent the convolutional neural network layer, the S layer represents the pooled layer, and the F layer represents the fully connected layer.

2) C1 layer - convolutional layer

The first convolution operation is performed on the input image, and six convolution kernels of size 5*5 are used to obtain six 28*28 feature maps.

3) S2 layer - pooling layer

The first convolution is followed by the pooling operation, using the 2*2 core for pooling, and the S2 layer is obtained: six 14*14 feature maps. The pooling layer of S2 is obtained by summing the pixels in the 2*2 region of C1 by a weight coefficient plus an offset.

4) C3 layer - convolutional layer

After the first pooling, it is the second convolution. The output of the second convolution is C3, 16 10x10 feature

maps, and the convolution kernel size is 5*5. Because S2 has 6 14*14 feature maps. C3 is a 16 feature map calculated by special combination of the feature maps of S2. The specific calculation is shown in Figure 3:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X		X	X	X	X	X	X	X	X
1	X	X			X		X	X		X	X	X	X	X	X	X
2	X	X	X				X	X	X		X		X	X	X	X
3		X	X	X			X	X	X	X		X	X	X	X	X
4			X	X	X		X	X	X	X		X	X	X	X	X
5				X	X	X		X	X	X	X		X	X	X	X

Figure 3. Connection method of S2 and C3 feature maps

From 0 to 15 are the numbers of 16 convolution kernels of the C3 convolutional layer, From 0 to 5 are the numbers of the six feature maps of the S2 pooling layer, and the intersecting marks represent the corresponding number positions of the C3 layer and the S2 layer.

5) S4 layer - pooling layer

S4 is the pooling layer, the window size is still 2*2, a total of 16 feature maps, and 16 10x10 graphs of the C3 layer are respectively pooled in 2x2 to obtain 16 5x5 feature maps.

6) C5 layer - convolutional layer

The C5 layer is a convolutional layer. Since the size of the 16 maps of the S4 layer is 5x5, which is the same as the size of the convolution kernel, the size of the map formed after convolution is 1x1. This produces 120 convolution results.

7) F6 layer - full connection layer

The F6 layer is a fully connected layer. The F6 layer has 84 nodes.

8) Output layer - full connection layer

The Output layer is also a fully connected layer with a total of 10 neurons, resulting in a row matrix of length 10 to determine the recognized characters.

C. Convolution and pooling

The convolution layer convolves the feature map of the previous layer and the convolution kernel of the extracted feature, and activates the function operation to form a feature map of this layer. The mathematical expression of the convolution function:

$$H_i = f(H_{i-1} \otimes W_i + b_i)(i \geq 1) \quad (1)$$

H_{i-1} represents the input image, W_i represents the convolution kernel weight matrix of the i layer, b_i represents the offset value corresponding to the convolution kernel, \otimes represents the convolution operation, and $f(\cdot)$ function represents the activation function.

The pooling layer mainly reduces the size of the feature map output from the convolutional layer and is a nonlinear downsampling method. On the one hand, reducing the size of the picture is a feature that becomes smaller and reduces the amount of network calculation. On the other hand, extract important features and discard secondary features. More abstract features can be obtained through the pooling layer. Generally take the maximum pool and average pool as shown in Figure 4, we take the maximum pooling as an example:

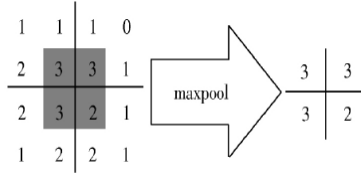


Figure 4. Connection method of S2 and C3 feature maps

During the maximum pooling process, the largest element in the pooled area is used as the output value of the pooled area. The formula is as follows:

$$\text{maxpool} = \max(x_i) \quad (2)$$

D. Improved LeNet-5 Convolutional Neural Network

This article has been improved on the basis of LeNet-5, the improved network is shown in Figure 5.

1) The size of the input layer image is changed from 32*32 pixels to the size of the blink image 24*24 size.

2) The activation function is a mixture of the Relu function and the sigmoid function.

3) In order to fully extract the eye features, a node with a full connection layer setting of 32 is added.

4) To increase the Dropout function in order to prevent overfitting

a. Relu function

The Relu function is more capable of generating feature sparsity than the tanh function. There is no gradient saturation when inputting positive numbers. In addition, the Relu function has only a linear relationship and the calculation speed is very fast. The calculation formula is:

$$f(x) = \max(0, x) \quad (3)$$

Figure 6 below shows the function of Relu:

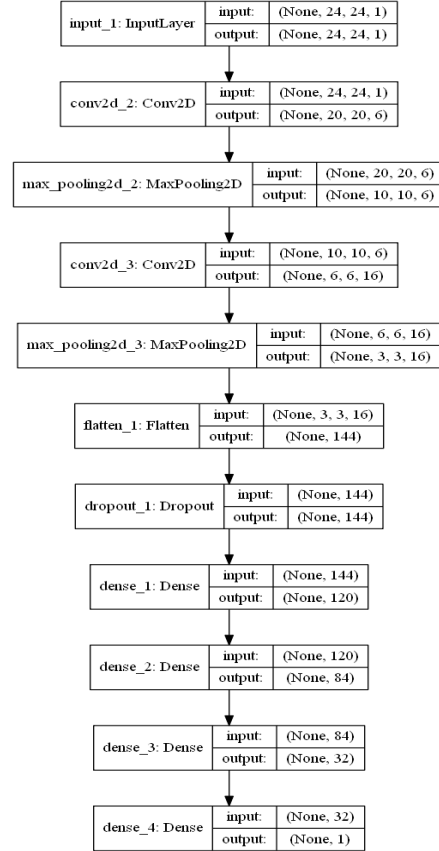


Figure 5. Improved LeNet-5 network structure

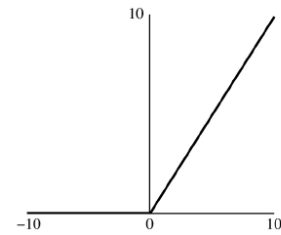


Figure 6. Relu function image

b. Dropout function

Dropout is a regularization method for neural network models proposed by Srivastava [8]. Dropout can randomly ignore some neurons during training, weaken the joint adaptability between neuron nodes, enhance the ability of pan-China, and effectively Prevent over-fitting during training.

III. EXPERIMENT

The experimental data used the published blink data set CEW [9]. In real-world scenarios, detecting a blink of an eye is a challenging task due to individual differences and environmental changes. The data set contained 2,423 subjects, of which 1192 were closed. Subjects were collected directly from the Internet, and 1231 subjects with open eyes were selected from the labeled wild face LFW [10] database. Eye spots are automatically acquired based on the

rough face area and eye position, and are estimated by the face detector and eye position, respectively. First, adjust the size of the roughened surface to 100*100 pixels, and then extract 24*24 eye spots centered on the eye position. Figure 7 shows the face image in the data set.

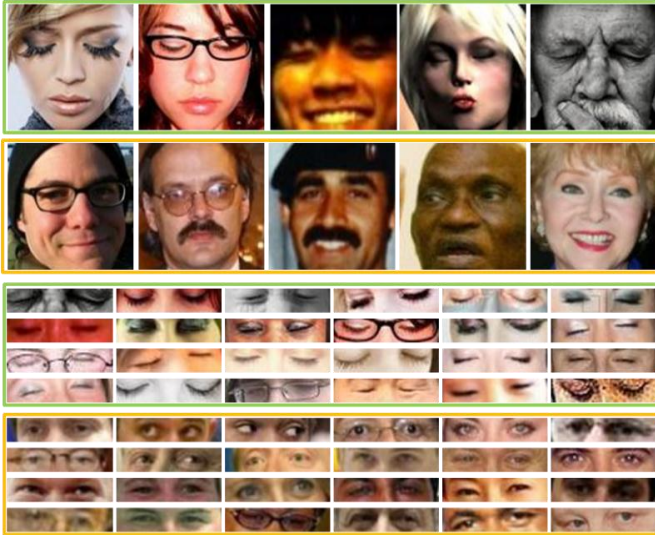


Figure 7. Face and human eye image

A. LeNet-5 and improved LeNet-5 results

Tested on public datasets, the experimental results show that the improved LeNet-5 is 5.1% better than the traditional LeNet-5. The following are the training data of the accuracy and model loss function of Lenet-5 and the improved Lenet-5 model. Figure 8 .a and b

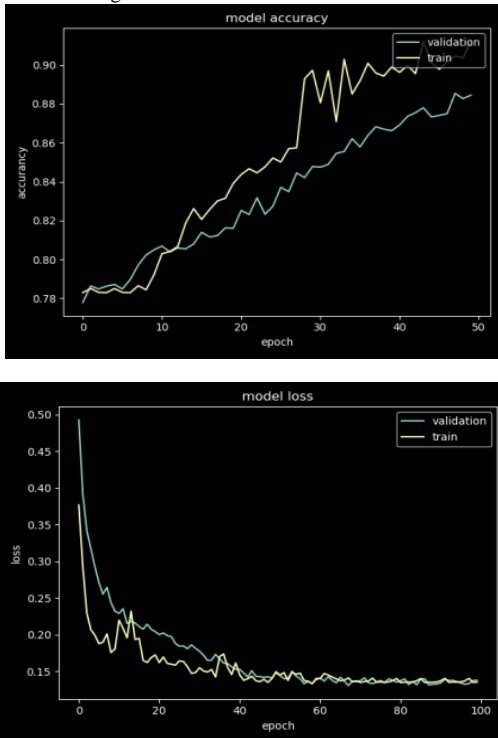


Figure 8. (a) Accuracy and loss function graph for LeNet-5 data set

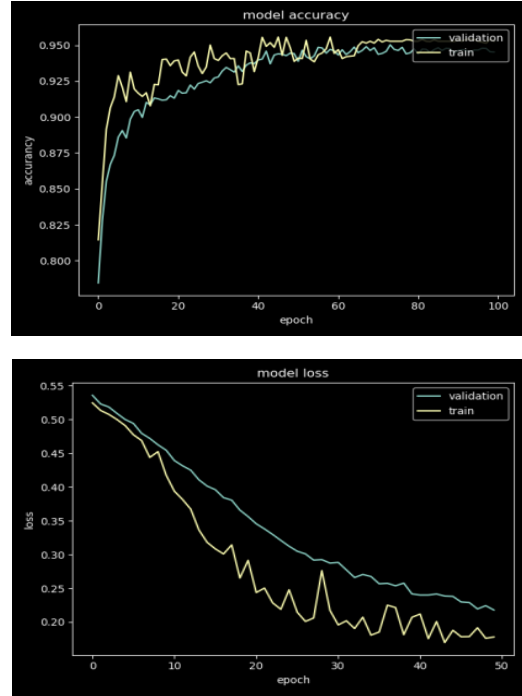


Figure 8. (b) Accuracy and loss function of the improved LeNet-5 data set

B. Compare with other literature

TABLE I. COMPARED WITH OTHER LITERATURE

literature	Accuracy
LeNet-5	88.4%
11	91.5%
12	92.95%
Method of this paper	93.5%

IV. CONCLUSION

In this paper, the LeNet-5 convolutional neural network is improved. Experiments show that in the blink detection process, the improved convolution network has an accuracy of 5.1% higher than that of the unimproved convolution network. Compared with other literatures, Very good effect, this algorithm can effectively detect the fatigue state of blinking, and has good practical value in the field of fatigue driving. What to improve in the later stage: Because the experimental samples are mostly foreign eye features, the sample data set will be further expanded to enhance the generalization performance of the algorithm, so that the convolutional network can learn more accurate eye features and can be widely used in the field of life.

REFERENCES

[1] Li Rui, Chen Yingjie V, Zhang Linghao. Effect of Music Tempo on Long-Distance Driving: Which Tempo Is the Most Effective at Reducing Fatigue?[J]. i-Perception, 2019, 10(4).
 [2] Huang Zhiqiang, Zhang Fangfang, Gu Zhuoyu, Chen Tao, Gu Su, Wang Lingyan. Design and development of fatigue detection system

- for high-speed rail driver before driving [J]. *Railway Computer Application*, 2019, 28 (08): 14-17.
- [3] Cao Ang, Zhang Yujia, Liu Rui, Zou Lian, Fan Ci-en. Muscle fatigue state classification system based on surface EMG signal[J]. *Journal of Computer Applications*, 2018, 38(06): 1801-1808.
- [4] LI Weiwei, HU Guiming, HE Longling, LI Ming.Face Fatigue Detection Based on Adaboost Algorithm[J].*Automation Technology and Application*,2014,33(02):46-48+53.
- [5] Zheng Yuanpan, Li Guangyang, Li Wei. A review of the application of deep learning in image recognition[J]. *Computer Engineering and Applications*, 2019, 55(12): 20-36.
- [6] ZHANG Qi,ZHANG Rongmei,CHEN Bin.Review of Image Recognition Technology Based on Deep Learning[J].*Journal of The Hebei Academy of Sciences*,2019,36(03):28-36.
- [7] Lecun Y , Bottou L , Bengio Y , et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11):2278-2324.
- [8] .Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov; Dropout: A Simple Way to Prevent Neural Networks from Overfitting 15(Jun):1929–1958, 2014.
- [9] F.Song, X.Tan, X.Liu and S.Chen, Eyes Closeness Detection from Still Images with Multi-scale Histograms of Principal Oriented Gradients, *Pattern Recognition*, 2014.
- [10]G. B. Huang, M. Ramesh, T. Berg and E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, Tech. Rep. 07-49, University of Massachusetts, Amherst (October 2007).
- [11]Ren Anhu, Liu Bei. Face Recognition Blink Detection Based on Adaboost[J]. *Computer and Digital Engineering*, 2016, 44(03): 521-524.
- [12]Lu Xue, Wei Bing, Wu Xiaopei. A new model of human-computer interaction based on continuous blinking[J]. *Computer Engineering and Applications*, 2018, 54(20): 263-270.