

Deep Periocular Recognition Method via Multi-Angle Data Augmentation

Bo Liu

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 2502341201@qq.com

Aokui Shan

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 1569871277@qq.com

Songze Lei

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: lei_sz@xatu.edu.cn

Baihua Dong

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: dongbaihua1996@163.com

Yonggang Li

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 1984130837@qq.com

Abstract—Periocular recognition technology is a biometric recognition technology widely used in identity verification. Because of its high precision, high ease of use and high security, Periocular recognition has a broad application prospect and scientific research value. In order to solve the problem of angular rotation of eyes in practical application, this paper proposes a deep learning periocular recognition method based on multi-angle data augmentation. The method is to rotate the original data set from small angle to large angle, so that the amount of data is expanded to 7 times of the original, and the diversity of data is increased at the same time. The InceptionV3 network and MobileNetV2 lightweight network are used for experimental verification respectively, and good results are obtained from multi-angle tests, indicating that the proposed method can improve the generalization ability of the model and has good robustness.

Keywords—Periocular Recognition; Deep Learning; Convolutional Neural Network; Lightweight Model

I. INTRODUCTION

Identity identification is a basic problem in social life [1], which is not only closely related to the interests of individuals, but also affected national security and social stability. This paper studies the periocular recognition technology based on deep learning, which is to use the image of the eye area to identify the identity of people. Due to the high precision, high ease of use and high security of the eye circumference [2], it is easy to obtain eye images, analyze various information of the eye region, and integrate their respective characteristics to achieve accurate, rapid and robust identification. Periocular recognition has important application value and theoretical value.

Research on periocular recognition started late, and Park [3] initially proposed that eye peripheral area could be used for identity recognition. In the

following studies, they gradually found that the periocular recognition system performed better, but its performance was not stable, due to illumination and blurring, resulting in less texture information. Many researchers had designed different periocular recognition systems for this kind of problem. Kumar [4] had developed various biometric recognition systems using different feature sets and different classifiers to identify subjects by using some reliable segmentation methods, efficient feature extraction methods and various classifiers or classifier integration. In 2017, a hybrid method based on transformation, structure and statistics, namely deep convolutional neural networks (CNNs), was proposed by Proenca and was also used to build periocular recognition systems. At the same time, Canadian professor Hugo. Proenca [5] also proposed to use neural network to process the human eye areas outside the eye, and to take the features extracted from these areas as a mode of identity recognition and attribute recognition.

In order to solve the problem of angular rotation of eyes in practical application, this paper proposes a deep learning periocular recognition method based on multi-angle data augmentation. The method is to rotate the original data set from small angle to large angle, so that the data volume is expanded to 7 times of the original, and the diversity of data is increased. InceptionV3 network and MobileNetV2 lightweight network are used for experimental verification respectively, and good results are obtained, indicating the feasibility of the proposed method.

II. METHOD

The periocular recognition method based on deep learning in this paper is divided into deep neural network training and testing, as shown in figure 1. The training part of the network including the preparation of training data, data normalization, pre-processing operations, data enhancement, as well as the construction of the network, network parameters setting and adjustment. In the test part, the test image data can be input into the trained network to obtain the classification and recognition results.

The data enhancement method in this paper increases the number of samples, makes the data more diverse, and improves the generalization ability of the model. This method also enhances the robustness of the model and increases the practicability of the system by multi-angle and large-angle rotation of the eye image in the original database. The convolutional network adopts InceptionV3 network and MobileNetV2 lightweight network to carry out convolutional operation to extract features, train them separately, and obtain two recognition models. The recognition results of the two CNN models [6] are obtained through experiments.

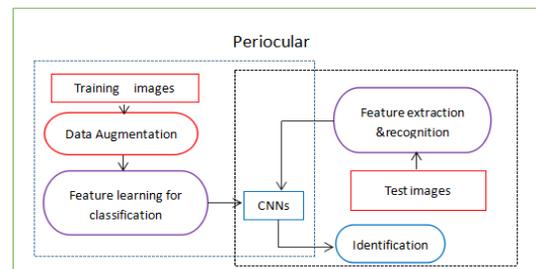


Figure 1. Data flow chart of the system

A. InceptionV3 Network Architecture

Inception module is proposed to solve the problem that the difference in image content cannot effectively extract image information. The core is to split the layered network convolution kernel into different size convolution kernels, such as 3×3 convolution kernels split into 1×1 , 3×3 convolution kernel, increases the size of the output network layer, such as 30×30 , step size is 1, use 3×3 convolution kernel, without filling, the output image size is 28×28 , If 1×1 and 3×3 convolution kernels are used to extract image information, the output depth map size is 30×30 and 28×28 , and the acquired image information is richer than a single convolution kernel.

The InceptionV3 network architecture [7] consists of 11 Inception modules with a total of 46 layers. The package contains a total of 96 convolution layers. Table I below shows the overall structure of the InceptionV3 network model. Since Google InceptionV3 is relatively complex and the raw approach to building is resource-intensive, using the TensorFlow Slim tool can greatly reduce the amount of code needed

to design InceptionV3. So the TensorFlow Slim tool is used to help build the network.

TABLE I. OVERALL STRUCTURE OF THE INCEPTION V3 NETWORK MODEL

Type	Size of Convolution Kernel/Step Size
convolution	3×3/2
convolution	3×3/1
convolution	3×3/1
pooling	3×3/2
convolution	3×3/1
convolution	3×3/2
convolution	3×3/1
Inception modules	3 Inception Module
Inception modules	3 Inception Module
Inception modules	3 Inception Module
pooling	8×8
linear	logits
Softmax	Classification of output

B. MobileNetV2 Network Architecture

In recent years, numerous researchers have put forward various lightweight network models successively, such as SqueeZenet, MobileNetV1 [8], ShufflenetV1 [9], ShufflenetV2 [10], and MobileNetV2 [11]. In the above network architecture, the test results of MobileNetV2 and ShufflenetV2 models are relatively good. The MobileNetV2 model uses deep separable convolution, linear bottleneck and reverse residual structure to maintain a certain degree of accuracy while keeping the number of parameters and computational complexity low. MobileNetV2 draws lessons from the RESNET network in the MobileNetV1 model based design. MobileNetV1 is designed according to the conventional convolution neural network chain structure, such as VGGNet model is to pile up in the form of convolution layer to build a network model, therefore, to some extent, improved the accuracy, the innovation points of this model is that the standard separation convolution with depth, reduce the size of the model, to some extent, computational complexity has been reduced. In addition, two compression hyperparameters are used to further shrink the network. However,

because of this, too many convolutional layers will cause the problem of gradient disappearance.

ResNet networks [12] have a great fluidity of information between layers due to the use of residual units. Therefore, MobileNetV2 uses the deep separable convolution base of MobileNetV1 and ResNet's residual unit for reference. MobileNetV2 network introduces a linear bottleneck structure and combines the reverse residual structure to complete improvements in these two aspects.

Table II shows the concrete implementation structure of the core building block of MobileNetV2, which is based on the Depth Separable Convolution Block of Reverse Residual, changing the number of input characteristic channels from N to M, where S represents the step size and T represents the expansion rate. A 1×1 convolution layer was added in front of the deep separable convolution layer, and the nonlinear activation function behind the point-by-point convolution layer was deleted to change the nonlinearity into linearity. Finally, in order to realize the subsampling, the scheme adopted is to set the step size parameter in the deep convolution.

TABLE II. IMPLEMENTATION OF THE MOBILENET V2 CORE BUILDING MODULE

Input	Operator	Output
H×W×N	1×1 conv2d, ReLU6	H×W×t N
H×W×t N	3×3 dwse s=s, ReLU6	H/s ×W/s×t N
H s ×W s ×t N	linear 1×1 conv2d	H/s ×W/s ×t N

MobileNetV2 overall network structure as shown in table III, every row describes one or more than one layer of the same sequence of loop n times, all sequence of layers have the same output channel number c, the sequence of the first layer step for s, residual layer of step 1, all the size of the space convolution kernels are the specifications of the 3 x 3, expansion ratio t always used the input characteristics described in table II. The MobileNetV2 model has an initial full convolutional layer of 32 convolutional cores, and then 17 reverse residual bottleneck modules are connected. The nonlinear activation function adopts ReLU6 with good robustness in the

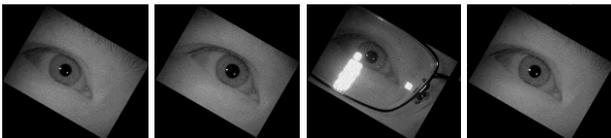
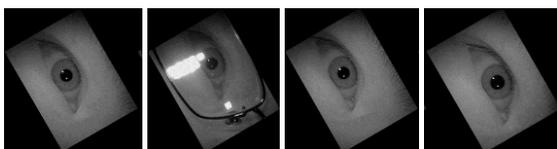
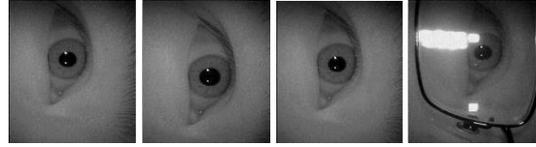
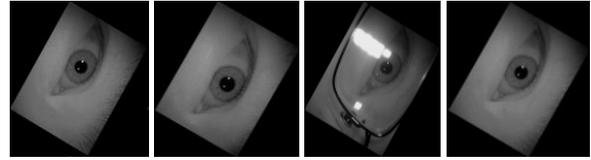
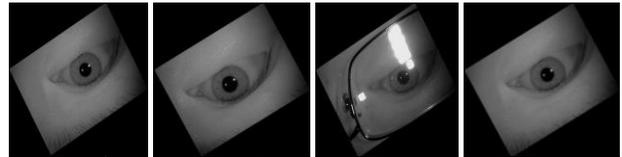
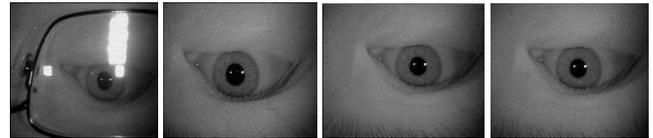
calculation with low precision, and the size selection of convolutional cores is 3×3 [14].

TABLE III. OVERALL ARCHITECTURE OF MOBILENET V2

Input	Operator	t	c	n	s
$224 \times 224 \times 3$	Conv2d	-	32	1	2
$112 \times 112 \times 32$	Bottleneck	1	16	1	1
$112 \times 112 \times 16$	Bottleneck	6	24	2	2
$56 \times 56 \times 24$	Bottleneck	6	32	3	2
$28 \times 28 \times 32$	Bottleneck	6	64	4	2
$14 \times 14 \times 64$	Bottleneck	6	96	3	1
$14 \times 14 \times 96$	Bottleneck	6	160	3	2
$7 \times 7 \times 160$	Bottleneck	6	320	1	1
$7 \times 7 \times 320$	Conv2d 1×1	-	1280	1	1
$7 \times 7 \times 1280$	Avg pool 7×7	-	-	1	-
$7 \times 7 \times 1280$	Conv2d 1×1	-	1000	-	-

C. Data Augmentation

In the process of neural network training, the stability of the model is proportional to the robustness, the number, and diversity of the training set data. However, the quality of the actual eye image acquisition may be poor. Therefore, the eye image in the original database is rotated from multiple angles to make the original database more diverse [15]. At the same time, after the multi-angle rotation expansion, the problem of an insufficient number of eye images in the original data set is also solved, so as to improve the stability of the neural network training process. The schematic diagram of a partial processing of the eye peripheral sample is shown in Figure 2 to 7. The eye peripheral sample is rotated at 30° , 60° , 90° , 120° , 150° , and 180° respectively. Without the above method, in real life, if the user is slightly tilted or the angle of the device changes, the model will not be able to correctly recognize and judge, resulting in wrong recognition results. This will greatly affect the robustness of the model.

Figure 2. Treatment diagram of the sample rotated by 30° around the eyeFigure 3. Treatment diagram of the sample rotated by 60° around the eyeFigure 4. Treatment diagram of the sample rotated by 90° around the eyeFigure 5. Treatment diagram of the sample rotated by 120° around the eyeFigure 6. Treatment diagram of the sample rotated by 150° around the eyeFigure 7. Treatment diagram of the sample rotated by 180° around the eye

After rotation of 30° , 60° , 90° , 120° , 150° and 180° , the data volume of each group was amplified 6 times. Before the training set, each group had 6 pieces, and after amplification, each group had 42 pieces. Before amplification, the verification set and test set were 2 pieces in each group, and after amplification, 14 pieces in each group. The total number of data sets is 42,000.

III. EXPERIMENT AND RESULTS

The experiment in this paper is carried out under Linux Ubuntu 16.04 system, using NVIDIA GeForce RTX2080Ti $\times 4$ graphics card, Intel Xeon Gold 6254 $\times 2$ CPU, 128G memory, 960G SSD hard disk +8TB mechanical hard disk. GPU acceleration is used to increase computing efficiency and speed, and CUDA10.0 is configured. The Python environment is version 3.6.5, with TensorFlow as a neural network framework [16].

A. The Data Set

The eye peripheral data set used in this experiment is the CASIA-IRIS-Thousand periocular image in the IRIS database CASIA-

IRIS 4. 0 of Chinese Academy of Sciences. In this group, there are a total of about 1000 people's eyes and a total of 20, 000 eyes. The eyes of about 500 people were selected. Table IV shows the sample distribution of the experimental eye peripheral data set. As shown in Figure 8, the partial eye peripheral sample of 000L category in the eye peripheral data set is shown.



Figure 8. Diagram of 000L partial eye sample

TABLE IV. SAMPLE EYE PERIPHERAL DATASET

Eye Peripheral Data Set	Original Training Set	Raw Verification Set	Raw Test Set	Total Original Sample	Original Sample Type
CASIA-Iris-Thousand	6000	2000	2000	10000	1000

After the above data augmentation, the sample of the data set changed to the eye peripheral data set amplified by rotation, as shown in Table V. 30°, 60°, 90°, 120°, 150° and 180° image data are added in the training set, verification set and test set. Training, validation and testing experiments are performed on the amplified eye peripheral data set.

TABLE V. AMPLIFIED SAMPLES OF THE DATA SET AROUND THE EYE

Eye Peripheral Data Set	Training Set	Verification Set	Test Set	Total sample	Sample Type
CASIA-Iris-Thousand	42000	14000	14000	70000	1000

B. InceptionV3 test results

After the eye peripheral data set is prepared, specific experimental parameters are set, as shown in Table VI: The maximum number of steps is set to 20000 rounds, the batch size is set to 24 pieces, the learning rate decay type indicates whether the learning rate drops automatically, which is set to fixed, the learning rate is set to 0.001, RMSProp is selected by the optimizer and weight decay of all parameters in the model is set to 0.00004. As shown in figure 9, in the training process of the periocular recognition model based on the InceptionV3 network on the eye peripheral data training set, its loss function decreased rapidly

from the beginning of training to 5000 rounds, after which the loss function value tended to be flat.

TABLE VI. PARAMETER SETTINGS

Parameter Types	Parameter Settings
Max number of steps	20000
Batch size	24
Learning rate	0.001
Learning rate decay type	fixed
optimizer	RMSProp
Weight decay	0.00004

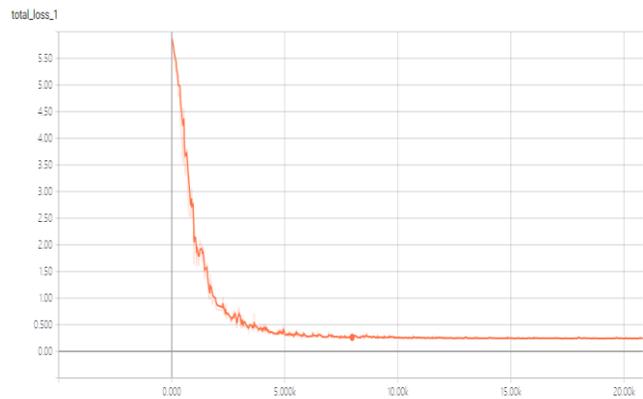


Figure 9. Loss function diagram of InceptionV3 network model

After model training, validation set is used to verify the model. The verified accuracy rate is 98%. Finally, all the samples around the eyes of the test set are tested and the accuracy rate is 98%. It can be seen that the rotating samples with small and large angles, such as 30°, 60°, 90°, 120°, 150° and 180°, can be well recognized.

C. MobileNet V2 test results

MobileNetV2 is used as the recognition method of Convolutional Neural Network and InceptionV3 network. The eye peripheral data sets are all expanded eye peripheral images. A total of 1000 groups of eye peripheral samples, 42 samples for each group, 42,000 training sets, 14,000 verification sets and 14,000 test sets.

The experimental parameters of the MobileNetV2 lightweight network model are set as Table VII. The maximum number of steps is set to 100000 rounds, the batch size is set to 32, the learning rate decay type indicates whether the learning rate automatically drops, which is set to

fixed, and the learning rate is set to 0.001. RMSProp is selected by the optimizer and weight decay is set to 0.00004 for all parameters in the model. As shown in figure 10, in the periocular recognition model based on the MobileNetV2 lightweight network, in the training process of the training set using the eye peripheral data, the change of its loss function drops sharply at the beginning, but with the progress of training, when the iteration reaches 20000 rounds, the loss function becomes flat.

TABLE VII. NETWORK MODEL PARAMETER SETTING

Parameter Types	Parameter Settings
Max number of steps	100000
Batch size	32
Learning rate	0.001
Learning rate decay type	Fixed
optimizer	RMSProp
Weight decay	0.00004

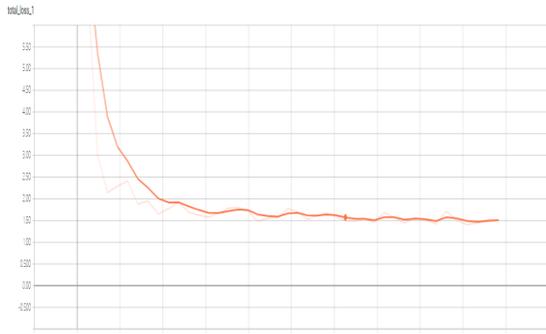


Figure 10. Change diagram of loss function of the MobileNet V2 lightweight network model

The verification accuracy of the MobileNetV2 lightweight network model is 98.21%. The validation accuracy of the InceptionV3 network model is 98.55%. After the training, the model is generated. The file size generated by the InceptionV3 network model is 93MB, and the MobileNetV2 lightweight network model is 24MB. The MobileNetV2 lightweight network model is 3.8 times smaller than the InceptionV3 network model when the verification accuracy is only 0.3% lower. Therefore, the periocular recognition based on lightweight convolutional neural network model is feasible.

The same method is used to test the test set. The test set samples are rotated from multiple angles to test the robustness of the model. After the rotation of the test set, the model is still able to be recognized normally. The accuracy of the InceptionV3 network model test is 98.5%, and the MobileNetV2 lightweight model test is 98.4%.

TABLE VIII. COMPARISON OF INCEPTION V3 AND MOBILENET V2 METHODS

Methods	Verification Accuracy	Test Accuracy	Model Size
InceptionV3	98.55%	98.5%	93MB
MobileNetV2	98.21%	98.4%	24MB

The eye images of the test set are tested on the two models respectively. As shown in figure 11, when testing the samples belonging to 302R, models with or without glasses can be correctly identified and the category of the tested samples can be correctly output.

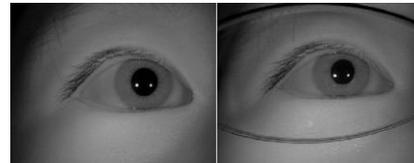


Figure 11. 302R test sample diagram

It can be found from the eye samples that failed in the test, as shown in figure 12, that the feature information of the eye can no longer be accurately extracted due to a large amount of specular reflection interference. In fact, such samples do not meet the conditions of sample collection, which will have a great impact on the recognition accuracy of the model. However, the accuracy of the InceptionV3 network model test is 98.5%, and the MobileNetV2 lightweight model test is 98.4%, even when large specular interference samples are included. Therefore, it can be concluded that the model in this paper still has a good recognition effect under the interference of such factors.



Figure 12. Specular interference around the eye sample

Follow-up, the eye images with specular reflection that has affected the image feature information in the test set are removed. The remaining samples are being tested. The accuracy of the InceptionV3 network model test is 99.8%, and the MobileNetV2 test is 99.4%. It is concluded that the accuracy of both the traditional model and the lightweight model are improved when the image samples with great influence of mirror radiation are removed.

IV. CONCLUSION

This paper mainly studies the periocular recognition technology based on deep learning, and uses the convolutional neural network InceptionV3 model to take the eye circle image as the input of the model, greatly shortening the steps of the entire recognition process. In fact, there are few angular rotation of eyes, so the original data can be more diversified by enhancing the eye image in the original database and carrying out multi-angle rotation. Then, according to the demand of mobile terminal, the application of lightweight model MobileNetV2 is proposed, which is used as the feature extraction classification architecture, and the periocular recognition technology based on lightweight convolutional neural network model is obtained. Experimental results show that both models can achieve satisfactory results after multi-angle data augmentation.

ACKNOWLEDGMENT

This work was supported by the National Joint Engineering Laboratory of the New Network and Testing Foundation (grant number: GSYSJ2016008). The CASIA-IRIS collected by the Institute of Automation, Chinese Academy of Sciences (CASIA) was used in part of this study.

REFERENCES

- [1] Miao Di. Multi-modal biometric feature fusion based on feature layer and fractional layer [D]. University of Science and Technology of China, 2017:1-2.
- [2] Zhang Xiaobo. Fusion of eye circumference and iris information for identification [D]. Chinese Academy of Sciences, 2012:1-2.
- [3] U. Park, A. Ross, A. K. Jain. Periocular biometrics in the Visible spectrum: A feasibility study[C]. In Proceedings of IEEE International Conference on Biometrics: Theory Applications and Systems, 2009:1-6.
- [4] Saiyed Umer, Alamgir Sardar, Bibhas Chandra Dhara, Ranjeet Kumar Rout, Hari Mohan Pandey. Person identification using fusion of iris and periocular deep features [J]. Neural Networks, 2020, 122.
- [5] He Yong. Research on the preprocessing and recognition method of human eye biometric image [D]. Hunan Industrial Science, 2020:6-7.
- [6] Wu Xue, Song Xiaoru, Gao Song, Chen Chaobo. A Survey of Target Detection Algorithms Based on Deep Learning [J]. Sensors and Microsystems, 2021, 40(02):4-7.
- [7] Ketan Joshi, Vikas Tripathi, Chitransh Bose, Chaitanya Bhardwaj. Robust Sports Image Classification Using InceptionV3 and Neural Networks [J]. Procedia Computer Science, 2020, 167.
- [8] Sun Yan, Ding Xuewen, Lei Yuting, Chen Jing, Kong Xiangxin. Image Recognition of Dogs and Cats Based on SSD_Mobilenet_v1 Network [J]. Journal of Tianjin Polytechnic Normal University, 2020, 30(01):38-44.
- [9] Zhang Xiangyu, Zhou Xinyu, Lin Mengxiao ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices [J]. 2017:14-15.
- [10] Ma Ningning, Zhang Xiangyu, Zheng Haitao. ShuffleNetV2: practical guidelines for efficient CNN architecture design [J]. 2018:33-34.
- [11] Lan Tianxiang, Xiang Ziyu, Liu Mingguo. Fast Semantic Segmentation Network Combining U-Net and MobileNet-V2 [J]. Computer Engineering and Applications, 2015:1-8.
- [12] He Kaiming, Zhang Xiangyu, Ren Shaoqing. Deep Residual Learning for Image Recognition [J]. 2015:13-15.
- [13] Han D, Kim J, Kim J. Deep pyramidal residual networks [J]. 2016:18-19.
- [14] Bi Pengcheng, Luo Jianxin, Chen Weiwei. Research on Lightweight Convolutional Neural Network Technology [J]. Computer Engineering and Applications, 2019(16):25-35.
- [15] Chen Chen. Research on Hyperspectral Image Classification Based on Convolutional Neural Network Based on Data Enhancement [D]. Beijing University of Chemical Technology, 2020, 45.
- [16] Wang Hengtao. Image Recognition Integrated System Based on TensorFlow, Keras and OpenCV [J]. Electronic test, 2020(24):53-54.