



Research and Implementation of FacialNet Based on Convolutional Neural Network

Yunxiang Liu, Chunya Wang, Jinpeng Ren and Xinxin Yuan

Abstract: Deep learning, artificial intelligence and other cutting-edge technologies are constantly being integrated into people's daily lives. Even small vending machines that can be seen everywhere in life have begun to use facial payment methods. The detection and recognition of face images is no longer unattainable, but the analysis and recognition of face information and characteristics (gender, age, race, etc.) is still not fully mature, in order to improve the accuracy of face information recognition In this paper, a face information recognition model is designed. The feature extraction part uses an eight-layer convolutional neural network, and then uses two fully connected modules as the classifiers for gender recognition and age recognition. The experimental results show that the model uses the advantages of the convolutional neural network so that the model can predict the gender and age of the face more accurately.

Keywords: Convolutional neural network, Face recognition, Gender recognition, Age recognition

I. INTRODUCTION

Face image recognition has great potential use in many fields such as scientific research and business. Therefore, the recognition of face images has attracted the attention of scholars and researchers at home and abroad. It has also developed rapidly in recent years, but it is based on human faces. The recognition of picture information (age, gender, etc.) is relatively slow. In fact, the face recognition technology has been proposed as early as the 1960s, but due to the limitation of computer processing speed at that time, it can only be studied theoretically[2]. Since 2012, deep learning has developed rapidly. It not only shines in object detection and image classification, but also has been widely used in the field of face recognition, completely changing the existing methods of face recognition. In this study, we designed a FacialNet model that classifies faces by gender and age. Compared with other similar models, this model has significantly improved gender prediction accuracy. However, the accuracy of age classification is lower than other models. Since our model is an end-to-end model, the two classifiers share a feature extraction network, so we speculate that the reason for the lower accuracy of age prediction is gender and age. The correlation

between calculations is weak. Other models use two independent networks to predict gender and age. When training the model, we train on a computer equipped with RTX 2080 GPU. The hyperparameters are set as follows: epoch is set to 100, the initial learning rate is set to 0.01, and batch_size is 25. The optimizer uses SGD, the momentum is set to 0.9, and the weight attenuation is set to 0.01.

II. Related Work

Related scholars at home and abroad have done a lot of research on this topic. In the research of Shi Xuechao [3], a face gender recognition algorithm combining adjustable supervision function and multiple feature fusion was proposed. The algorithm they proposed can be regarded as an improvement of the traditional convolutional neural network. It uses the deep convection of the complete convolution area and multi-layer feature information to combine the second and fourth convolutional layers with the fifth convolutional layer. This method enhances the image feature information from the input layer to the fully connected layer, thereby improving the accuracy of the final

Corresponding Author: *Chunya Wang: College of Computer Science and information Engineering, Shanghai Institute of Technology, 100 Haiquan Road, Haiwan Tourism District, Fengxian District, Shanghai, 201418, China, E-mail: 1973988571@qq.com*

classification; in addition, it also introduces the target monitoring function with adjustment mechanism, which can effectively guide the network to learn to make similar patterns. The distance between the classes decreases and the distance between samples of different types is larger. In the research of Zhang Zhihua [4] and other scholars, in order to make up for the error between automatic face recognition and gender estimation, they used deep convolutional neural networks for recognition and made great progress. Data set and label analysis, and used to build a simple neural network, and also have good results in gender classification. Although the use of the data set Adience is very challenging, due to the simple structure of the CNN they designed, this method is better than the previous technology.

III. Network Model Design

In order to realize the research on gender and age recognition of the face dataset Adience, this paper designs a network model with eight layers of convolution to extract facial features, and named it FacialNet. The first 4 layers are put together to extract more features, and the last 4 layers are to extract more high-level features. The 8-layer convolution can not only realize the function of feature extraction and classification, but also reduce the amount of calculation. The network structure is shown in Figure 5: The number of the first convolution kernel is set to 32, and the size of the convolution kernel is set to 3×3 . When the size of the receptive field is the same, the small size of the convolution kernel can reduce the calculation in the training process. Volume and parameters, so I choose a small-sized convolution kernel in this article. The number of the second convolution kernel is 64. The choice of the pooling layer type mainly considers two sources of error. Average pooling can reduce the error caused by the limited neighborhood size, so that more environmental information is preserved, and the maximum pooling It can reduce errors caused by the convolution process and save more texture detail information. Because the extraction of face information needs to retain more texture information, the maximum pooling is selected to extract more details. The pooling size is 3×3 . The number of third, fourth, and fifth convolution kernels is set to 128, among which the first The size of the third and fourth convolution kernel is set to 3×3 , and the size of the fifth layer convolution kernel is set to 1×1 . The number of the sixth convolution kernel is set to 256, the size of the convolution kernel is set to 3×3 , the number of the seventh and eighth convolution kernels is set to 512, of which the size of the seventh layer is set to 3×3 , and the eighth layer is volume The

size of the product core is set to 1×1 .

2 The Sigmoid function has the defect of saturation, so the activation function in this model is the ReLU function. This paper uses the learning rate adjustment function LearningRateScheduler to monitor the changes in the value of loss to adjust the learning rate, and tests the impact of different initial values of the learning rate on the final result on a small-scale data set. The test range is from 0.1 to 0.001. One person is divided into 3 test numbers for experimentation. It can be obtained that when the learning rate is 0.002, the correct rate is the highest, and the loss is also very low, so the initial value of the learning rate in this article is set to $2e-3$. The detailed structure of the model is shown in Table 1, Table 2 and Table 3.

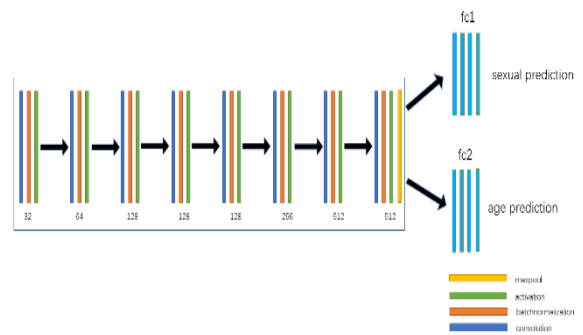


Fig.1 FacialNet Architecture

Table.1 Feature extraction network

Type	Channel	Filter size	Stride	Padding
Conv1	32	3×3	2	1
Conv2	64	3×3	1	1
Conv3	128	3×3	2	1
Conv4	128	3×3	1	1
Conv5	128	1×1	1	0
Conv6	256	3×3	2	0
Conv7	512	3×3	2	1
Conv8	512	1×1	1	0
maxpool	--	3×3	2	0

Table.2 Gender classifier

Type	Input Neurons	Output Neurons	Dropout	Activation Function
Fc1	2048	512	0.5	relu
Fc2	512	64	0.3	relu
Fc3	64	2	--	softmax

Table.3 Age classifier

Type	Input Neurons	Output Neurons	Dropout	Activation Function
Fc1	2048	512	0.4	Relu
Fc2	512	64	0.20	Relu
Fc3	64	8	--	softmax

This experiment uses SGD as the network training optimizer, and uses the cross-entropy loss function to calculate the loss for the predicted category.

(1) Two categories

In the case of dichotomy, there are only two cases where the model needs to predict the final result. For each category, the predicted probabilities are p and 1-p.

The expression at this time is:

$$L = \frac{1}{N} \sum_i \frac{1}{N} \sum_i - [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \quad (1)$$

y_i —Sample label

p_i —Indicates the probability that the sample prediction is positive

(2) Multi-category

The multi-category situation is actually an extension of the two-category: among them:

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i - \sum_{c=1}^M y_{ic} \log(p_{ic}) \quad (2)$$

M—number of categories

y_{ic} —Indicator variable

p_{ic} —For the predicted probability of the observed sample belonging to the category

IV. Model training process and experimental results

4.1 Data preprocessing

Keras, an advanced neural network framework, can use CNTK, TensorFlow or Theano as the computing

backend. Because it is based on Python, users can directly describe the model in Python language, so it has strong scalability and ease of use. At the same time, it is designed with people-oriented and emphasizes rapid modeling. Users can customize the selection of modules. Constructing a network model, and being able to simply and quickly transform your ideas into an achievable model, can be a good help for experimental research. Support GPU and CPU operation, suitable for various hardware environments.

This experiment uses the Adience data set as the experimental data, which can reflect the real camera environment to a large extent. Remove duplicate data and undetected face images, and finally get preprocessed experimental face images, and use image data enhancement methods to increase the amount of data during training and verification. The data distribution is shown in Figure.2 and Figure.3.

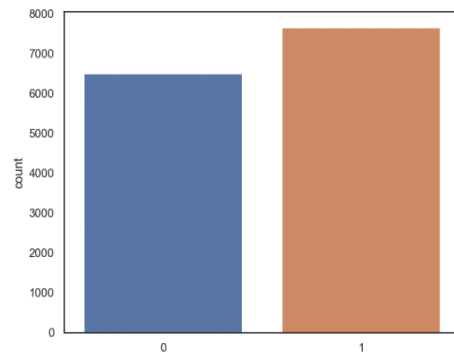


Fig.2 Distribution of gender data

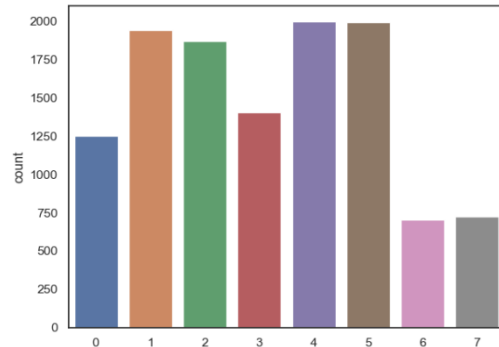


Fig. 3 Distribution of age data

This experiment uses a 75*75 picture size for face detection. First, the face part of the face picture is segmented, and then the face is aligned, and the picture is converted into a grayscale picture. The face information is recognized and Color does not matter much, but will affect the training and recognition to

learn irrelevant factors, and then the face image training and verification data will be enhanced and then passed to the model for learning. The display of some pictures in the data set after data preprocessing is shown in Figure 4.

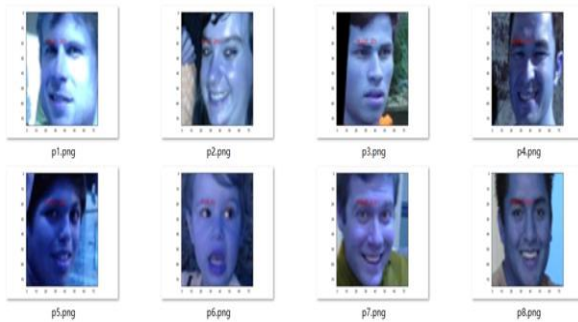


Fig.4 Partial data display of Adience data set after preprocessing

4.2 Image data enhancement

In the era of deep learning, the larger the scale and the higher the quality of the data, the better the generalization ability of the model, and the data directly determines the upper limit of model learning. However, in actual engineering, it is difficult for the collected data to cover all the scenes, such as the lighting conditions of the image. The pictures taken in the same scene may be very different due to different lighting, so you need to add lighting when training the model. Aspects of data enhancement. On the other hand, even if you have a large amount of data, you should also perform data enhancement, which helps to increase the amount of data in the relevant data set, prevent the model from learning unwanted models, and avoid overfitting. There are two commonly used data enhancements, supervised and unsupervised. This article uses supervised data enhancement. Supervised data enhancement is based on the existing data set, by analyzing the completeness of the data, and using certain rules to expand the existing data. Supervised data enhancement can be subdivided into single-sample data enhancement and multi-sample data enhancement. In actual engineering applications, single-sample data enhancement is used more. There are some open source data enhancement projects with better performance on git, which have more complete functions and The processing speed is also very fast, developers can directly call, such as imgaug and albumations. In pytorch, many data enhancement function packages can be integrated through torchvision's transforms module. Specific operations include cropping, zooming, and

shifting [17-21].

This article uses the method of image data set enhancement to increase the amount of image data. This method not only allows the limited image data to generate more equivalent image data, but also prevents overfitting during model training. The effect of the final image data enhancement is shown in Figure 5.

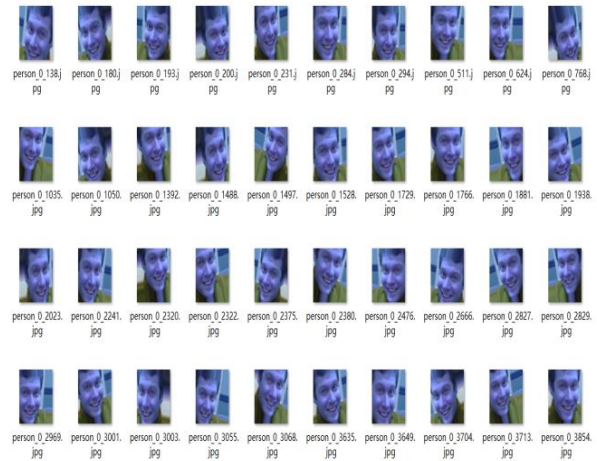


Fig.5 Image data enhancement

4.3 training configuration

This model is implemented using pytorch and trained on RTX 2080. The TensorBoard tool is used to visualize the parameters in the training process. And use the data set Adience for face information recognition, and build a unique CNN network model to solve the problem of gender and age recognition. Below we will elaborate on the model framework and experimental process. The convolutional neural network models Gender_CNN and Age_CNN proposed in this paper are implemented using the advanced deep learning framework Keras, using TensorFlow as the computing background and using GPU for training. The hardware environment of this experiment is shown in Table 5-1. The visualization of each parameter in the training process of this article uses the TensorBoard tool. Enter the picture again on the trained model, pass the input image through each layer of convolutional layer, and extract the feature map through convolution. The result is shown in Figure 5.1. It can be seen that as the depth of the convolutional layer increases, The extracted image features are becoming more and more abstract.

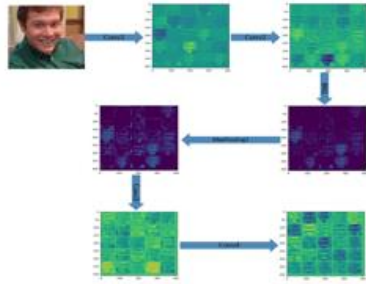


Fig.6 Visualize the training process

Table 4 Experimental hardware environment table

operating system	Windows 10 Pro 1809
RAM	DDR3L 16G
hard disk	Samsung SSD 860 EVO 500G
CPU	Intel Core i7-4720HQ CPU @ 2.60GHz
GPU	NVIDIA GeForce GTX 860M 2G

4.4 training result

The first method of this gender recognition experiment is also to use the original data set to train the model, and the correct rate on the Adience test set is 88.6%. The second method is to perform data enhancement on the original data set. After training, the correct rate of the verification set is 90.3%; the correct rate of the first method of the age recognition experiment is 68.1%, and the correct rate of the verification set of the second method is 72.8%. The comparison of the effects is shown in Table 5.

Table 5 Comparison of gender and age recognition effect on Adience

Methods	Gender accuracy	Age accuracy
The model in literature [3] works best	---	54.1%

The model in literature [4] works best	88.2%	----
The CNN+RFC model in literature [6] works best	86.5%	60.2%
The Modified CNN model in literature [7] works best	91.8%	87.0%
Our Method without data enhancement	88.6%	68.1%
Our Method after data enhancement	94.2%	72.8%

The trend curve of gender recognition accuracy rate during training is shown in Fig. 7:

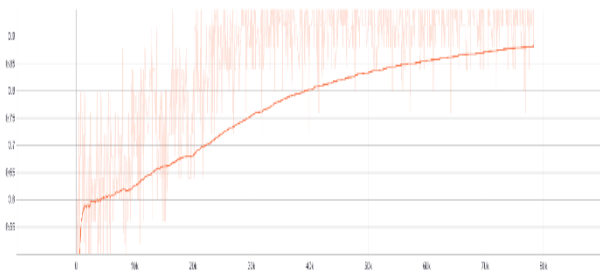


Fig..7 Gender_CNN verification accuracy trend curve

The trend curve of age recognition accuracy rate during training is shown in Figure 8:

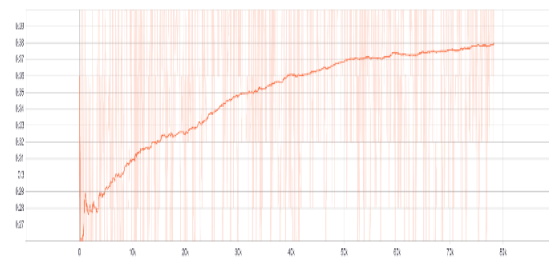


Fig.8 Age_CNN verification accuracy trend curve

The trend curve of loss function during training is shown in Figure 9:

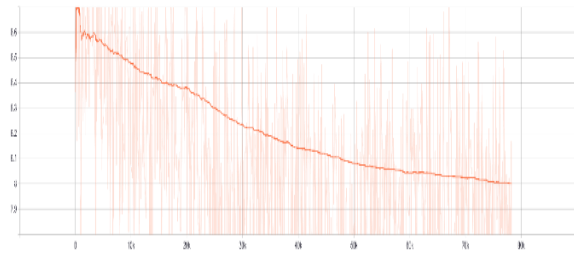


Fig.9 Loss function trend curve

V. CONCLUSION

This article describes the theory of convolutional neural network and its application background in the direction of face information recognition. A convolutional neural network model Gender_CNN is designed for gender classification, and a convolutional neural network model Age_CNN is used for age classification. From the above comparison results, it can be seen that the accuracy of this model in gender prediction has a small increase compared with other models, but there is still a gap compared with the excellent age model. Integrating various methods together, finally got good results. In the next work, we will use more complex and large face data sets to build more efficient and accurate models for further research in improving the application scenarios and accuracy of face recognition.

REFERENCES

- [1] Li Huai. Research on Face Age Recognition Based on Integrated Convolutional Neural Network [D]. South China University of Technology, 2016.
- [2] Wang Jimin. Research on face detection and gender recognition based on convolutional neural network [D]. Nanjing: Nanjing University of Science and Technology, 2015.
- [3] Zhang Zhihua, Zhang Qi, Zhang Weixin. Fast gender detection based on convolutional neural network[J]. Industrial Control Computer, 2018, 1: 036.
- [4] Yang Ziwen. Research on face recognition based on deep convolutional neural network [D]. Guangxi Normal University, 2017.
- [5] Zhang Juntao. Face detection and face age and gender recognition method[D]. University of Science and Technology of China, 2017.
- [6] Ai Xin, Luo Qun. Research on age and gender recognition based on improved convolutional neural network [EB/OL]. Beijing: China Science Paper Online [2018-12-03].
- [7] Zhou Jie, Lu Chunyu, Zhang Changshui, et al. Overview of automatic face recognition methods [J]. Chinese Journal of Electronics, 2000, 28(4): 102-106.
- [8] Li Hui, Shi Bo. Face recognition algorithm based on convolutional neural network[J]. Software Guide, 2017, 16(3): 26-29.
- [9] Sun Yanfeng, Qi Guanglei, Hu Yongli, et al. Deep Convolutional Neural Network Recognition Algorithm Based on Improved Fisher Criterion[J]. Journal of Beijing University of Technology, 2015, 6: 835-841.
- [10] Cao Pin, Mo Yating, Huang Chen, et al. Face age analysis algorithm and implementation based on deep convolutional neural network [J]. Software Engineering, 2016, 19(8): 14-18.
- [11] Ma Wenjuan, Dong Hongbin. Research on Face Age Classification Algorithm Based on Integrated Convolutional Neural Network[J]. Computer Science, 2018, 45(1): 152-156.
- [12] Qiu Shuang, Nie Rencan, Zhou Dongming, et al. Face recognition method based on self-normalized convolutional neural network[J]. Journal of Yunnan University (Natural Science Edition), 2018, 40(4): 659-664.
- [13] Yin X, Liu X. Multi-task convolutional neural network for pose-invariant face recognition[J]. IEEE Transactions on Image Processing, 2018, 27(2): 964-975.
- [14] Li H, Lin Z, Shen X, et al. A convolutional neural network cascade for face detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 5325-5334.
- [15] Parkhi O M, Vedaldi A, Zisserman A. Deep face recognition[C]//BMVC. 2015, 1(3):
- [16] Levi G, Hassner T. Age and gender classification using convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2015: 34-42.
- [17] Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond Empirical Risk Minimization[J]. 2017.
- [18] Ye Yun. Deep Learning and Computer Vision: Algorithm Principle, Framework Application and Code Implementation [M]. Beijing: Mechanical Industry Press, 2018: 194-195.
- [19] [1]Terhörst Philipp, Kolf Jan Niklas, Damer Naser, Kirchbuchner Florian, Kuijper Arjan. Post-

comparison mitigation of demographic bias in face recognition using fair score normalization[J]. *Pattern Recognition Letters*,2020,140.

- [20] [2]Zhou Li,Liu Mingfan,Ye Baojuan,Wang Xinqiang,Liu Qiaosheng. Sad expressions during encoding enhance facial identity recognition in visual working memory in depression: Behavioural and electrophysiological evidence[J]. *Journal of Affective Disorders*,2021,279.
- [21] [3]Massoli Fabio Valerio,Falchi Fabrizio,Amato Giuseppe. Cross-resolution face recognition adversarial attacks[J]. *Pattern Recognition Letters*,2020,140(prepublish).