Advances in Computer Science Research, volume 80
International Conference on Computer Science, Electronics and Communication Engineering (CSECE 2018)

Human Face Age Recognition Based on Convolutional Neural Networks

Zijiang Zhu*, Yi Hu, Dong Liu, Xiaoguang Deng and Junshan Li
South China Business College of Guangdong University of Foreign Studies , Guangzhou 510545, China
*Corresponding author

***Abstract*—In the field of image recognition, the issue of face age recognition has attracted the attention of many scholars, and a lot of outstanding algorithms have been proposed, but the correctness rate of age recognition is not high. To improve the AGE identification accuracy, this paper proposes a face AGE recognition based on convolutional neural network (CNN) -- the AGE model, the IMDB - WIKI database and Caffe framework for training and testing, and the AGE recognition has the highest accuracy of 52%. Through experiments, this model is proved to be scientific and provides new ideas and methods for the study of face age recognition.**

Keywords—human face age recognition; convolutional neural networks; CNN-AGE; caffe

I. INTRODUCTION

In the field of machine vision and image processing, age recognition based on face image is an important research subject, which has attracted the attention of scientific researchers at home and abroad in recent years [1]. The issue of age recognition has broad application prospects in academic researches and business applications. For example, it has important application value in security monitoring, personalized human-computer interaction, image retrieval and criminal investigation, etc. [2] But it is relatively difficult to identify age from face image, mainly because of the following three reasons [3][4]: 1) The positioning of age recognition is fuzzy because it can be either classification issue or regression issue; 2) It is difficult to gather data of human face image ages; 3) Facial image features of the same age are significantly different.

The key of extracting age feature through face image is facial feature model. At present, there are corresponding studies on the features of the local appearance and the global appearance. J. Ylioinas et al. [5] used the improved Local Binary Patterns (LBP) to extract features and carried out the test in Gallagher database, getting an accuracy rate of 51.7%. Yan S et al. [6] divided the age into 0-2 categories, namely, 3-7, 8-12, 13-19, 20-36, 37-65, >66, 7, and the improved LBP feature extraction method was used. Finally, the accuracy rate in Gallagher database was 50.3%. Eidinger et al. [7] used FPLBP feature extraction method to extract the age characteristics of face images, and achieved a 66.6% accuracy rate on Gallagher database, and achieved 45.1% accuracy on Adience data base. F. Gao et al. [8] proved that the method of extracting the age feature of the face image by Gabor feature extraction is better than the Local Binary Patterns (LBP). Ueki et al. [9] used the Expectation Maximization (EM) algorithm to build 11 kinds of Gauss models in the low dimensional 2DLDA+LDA feature space and achieved good results. In the research of Ueki et al, the accuracy rate reached 50% when the age classification interval was 5 years old, and the accuracy rate reached 82% when the age interval was 10 years old. However, the face image test set they used were all males or females.

It can be seen that it is very difficult to carry out the research of face age recognition. The accuracy of age recognition is affected by objective factors such as the time span of face image (age classification interval), sex and individual senescence speed. Based on convolution neural network, a neural network model of face age recognition is constructed in this paper in a bid to contribute to the research of face age recognition.

II. CONVOLUTIONAL NEURAL NETWORKS (CNN)

Convolutional Neural Networks (CNN) is an improved algorithm based on Back Propagation (BP). The common point between CNN and BP neural network is that during the training process, the output value of the network is calculated in the forward propagation phase of the signal, and the weights and biases between the neurons are adjusted in the reverse propagation stage of the error. The main difference between CNN and BP neural networks is that the neuron nodes between the adjacent layers of the BP neural network are all connected, while the CNN is partially connected [10]. As shown in Figure 1, CNN mainly includes three parts: the convolution layer, the pool layer and the full connection layer.

In CNN, there are three basic concepts: local receptive fields, shared weights and pooling.

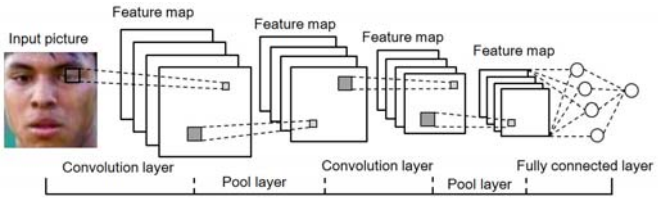

FIGURE I. BASIC STRUCTURE OF CONVOLUTIONAL NEURAL NETWORKS

- 1) Local Receptive Fields.

The local sensing domain of CNN is not connected to all neuron nodes while connecting the neurons at the input layer to the hidden layer. It only connects some neuron nodes, that is, the local sense. Taking 8 * 8 gray level image as an example,

the number of neurons in the input layer is 64, and the window of the local receptive field is set to 4 * 4. Each neuron node in the hidden layer connects only part of the neuron nodes in the input layer, that is, the local 4 * 4=16 connected to the input layer, as shown in Figure 2. Through mobile local receptive field window (from top to bottom, left to right), all input layer neurons are covered, and each neuron node with local receptive field is connected to different neuron nodes of hidden layer, respectively. This is the learning process of the convolution neural network from the local feature of the input layer to the global feature.

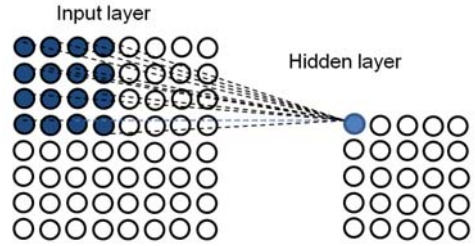

FIGURE II. LOCAL EXPERIENCE PROCESS

Assume that the receptive field is only connected to the 1 neuron nodes of the hidden layer, and the mobile step length is 1. From Figure 2, it can be seen that if the field moves down 4 times to the right direction, it can reach the edge of the image, and the hidden layer has 5 * 5=25 neuron nodes. For 8 * 8 images, the receptive field is 4 * 4=16, the first hidden layer has 25 neurons, and the number of neurons is 16, and the total weight is 25 * 16=400. Compared with the BP neural network, this method greatly reduces the number of parameters of the weights, that is, the number of parameters of the training weight is reduced.

- 2) Shared Weight.

It is known from the above 1) that the 8 * 8 image requires 400 weight parameters. If each neuron corresponds to the 16-weight parameter is equal, the total weight parameter is only 16, 25 neurons share is 16 weights, which reduces the number of weight parameters, i.e., the concept of shared weight. Suppose that the neurons are located in the hidden layer (i, j), its $o_{i,j}$ output can be expressed as follows:

$$o_{i,j} = f(\sum_{k=0}^{3} \sum_{m=0}^{3} w_{k,m} a_{i+k,j+m} + b) \quad (1)$$

In equation (1), f is a neuron activation function, $w_{k,m}$ is 4 * 4 weight sharing matrix, $a_{i,j}$ is the input value of the neuron in the hidden layer (i, j), and the B is a shared bias parameter. It can be seen that the feature extraction method of CNN is convolution operation, and the neuron extracts the characteristics of the same statistical characteristics at different positions in the image. The feature map is the input and output of the hidden layer. The weights and biases on the feature map are shared weights and biases. All the shared weights and biases are called convolution kernels, also known as filters. In practical applications, there are often multiple convolution cores in CNN, indicating that there are multiple feature graphs. The more the feature maps, the more the content of the image is expressed.

- 3) Pooling.

CNN classifies sample instances through feature graphs, but because of too many feature maps, the classifier has a large amount of computation. When a large number of feature graphs are input, the training of the classifier is difficult, and it will lead to the over fitting (over-fitting) phenomenon of CNN. To solve this problem, the feature graph is put into the pool layer for pooling and aggregate the statistics of different regions, calculate the maximum or average value of a region's characteristics, reduce the dimension of statistical features and prevent the over fitting phenomenon.

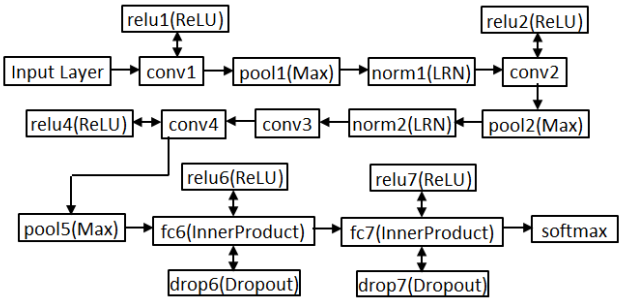

FIGURE III. CNN-AGE MODEL STRUCTURE

In Figure 3, Input Layer is the input layer. Conv1, conv2, conv3 and conv4 are 4 volumes. Pool1, pool2 and pool5 are 3 pools. fc6 and fc7 are 2 full link layers, and pool5 is classifier. In CNN, because the RelU function has a fast convergence property [11], the RelU function is used as the activation function of the convolution layer and the fully connected layer. In the pool layer, in order to achieve the purpose of local normalization, the LRN (Local Response Normalization) function is proposed by Krizhevsky et al. In order to prevent training over fitting, the Dropout processing is added to the entire connection layer, and droupout sets the output neuron to zero [12, 13] according to a certain probability. After convolution layer conv3 is convoluted, considering that the feature graph is very small, then pooling is meaningless, but it increases resource consumption. Therefore, the pool layer is canceled, and the next volume layer is directly connected to extract the deep features of the image. Softmax is a classifier that classifies and outputs the age category.

III. TRAINING PROCESS

A. *Experiment Environment*

The experiment in this paper is carried out using the depth learning framework (Caffe) of BVLC. The training data set is a IMDB-WIKI database, because the data is too big, just select 62 thousand pictures from Wikipedia. Face detection is done for each picture. Only face images are selected for training and testing. The face images are pre-processed (aligned and clipped) before the experiment. The experimental environment is shown in Table 1 as follows.

TABLE I. EXPERIMENT ENVIRONMENT OF HUMAN FACE AGE RECOGNITION

System	Ubuntu 16.04 Beta
Memory	DDR4 16G
CPU	NVIDIA Ge Force GT 4G
CPU	inter(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz

B. Data Set Division and Age Classification

The pictures provided in the IMDB-WIKI database contain detailed information files, and two Matlab codes written by the author can be used to realize data partition and age classification. Because the data in different age groups of IMDB-WIKI dataset are not balanced, it is possible to get less than 5 years old or more than 85 years old. There are relatively few sample photos. So we use 5-84 year old age photos, and configuration information is like Figure 4. The age division of 10 years is divided into 8 categories, and the span of each class is 10 years. The log information of the age classification is shown in Figure 5.

```
ratio = 8;
sel = 1;
class = 10;
min = 5;
max = 84;
maxsel = 84;
minsel = 5;
```

FIGURE IV. DATA SET DIVISION

```
train.list and val.list created by age_div_train_test.m
dataset: 35699
sel: 1.00000, ratio: 8:1
selected: 35699 (about 31732 train(88.89%), 3967 test(11.11%))
class capacity: 10, class num: 8 (about 4462 per class)
guaranteed num of every class:
    235 4462 4462 4462 4462 3078 1786 867
min: 5, max: 84
minsel: 5, maxsel: 84
```

FIGURE V. AGE CLASSIFICATION LOG

C. Training Results

Through repeated experiments and repeated adjustment of the relevant parameters, the best accuracy rate of face age recognition is 9000 iterations, and the accuracy rate is 52% as shown in Figure 6. The experiment was carried out with random initialization of Gauss distribution function, the volume weight average layer is set to 0, the random drop rate is set to 0.5 fully connected layer Dropout, the number of training samples set Batchsize to 256, on a gradient weight momentum is set to 0.9, the initial learning rate of base_lr is set to 0.0001.

```
d_solver.cpp:106] Iteration 9009, lr = 0.0001
lver.cpp:337] Iteration 9010, Testing net (#0)
lver.cpp:404]     Test net output #0: accuracy_test_top01 = 0.519
lver.cpp:404]     Test net output #1: accuracy_test_top02 = 0.635
lver.cpp:404]     Test net output #2: loss = 1.35133 (* 1 = 1.35133 loss)
lver.cpp:228] Iteration 9010, loss = 1.31241
lver.cpp:244]     Test net output #0: loss = 1.31241 (* 1 = 1.31241 loss)
d_solver.cpp:106] Iteration 9010, lr = 0.0001
lver.cpp:228] Iteration 9011, loss = 1.21332
```

FIGURE VI. TRAINING RESULTS

IV. ANALYSIS OF EXPERIMENT RESULTS

A. Scientificity of CNN-AGE Model

According to the above training results, select two pictures from 8 age groups at random for human face age identification experiment, and the experiment results are as shown in the following Table 2.

From the experimental results of the last table 2, it can be seen that the accuracy of the test at the age of 15-44 is higher and the maximum is 52%. The accuracy of the tests at the age of 5-14 and 55-64 is generally close to 30%, and the accuracy of the test at the age of 65-84 is low, and the maximum is only 24%. The accuracy of the tests at all ages is shown in Figure 7.

As can be seen from Figure 7, with the change of age, the CNN-AGE model has a larger age recognition rate for facial features. Figure 7 shows that the changes in the facial features at the age of 15-44 are relatively slow, and the changes in other age groups are relatively large. This is consistent with the natural law of human being from maturity to maturity, from maturity to aging. It also proves the scientificity of the face recognition model based on convolutional neural network.

TABLE II. HUMAN FACE RECOGNITION EXPERIMENT RESULTS OF ALL AGES

No.	**Physical age**	**Experiment results of all ages**							
		5-14	***15-24***	***25-34***	***35-44***	***45-54***	***55-64***	***65-74***	***75-84***
1	***11***	***0.273***	0.255	0.194	0.067	0.053	0.076	0.059	0.023
2	***8***	***0.297***	0.306	0.254	0.073	0.021	0.022	0.027	0
3	***19***	0.172	***0.369***	0.31	0.094	0.043	0.008	0.004	0
4	***21***	0.091	***0.471***	0.325	0.111	0.001	0.001	0	0
5	***26***	0.006	0.232	***0.512***	0.200	0.043	0.007	0	0
6	***34***	0.011	0.148	***0.479***	0.321	0.019	0.019	0.003	0
7	***38***	0.006	0.093	0.247	***0.465***	0.173	0.015	0.001	0
8	***39***	0.003	0.122	0.191	***0.517***	0.161	0.007	0	0
9	***49***	0.004	0.003	0.122	0.277	***0.296***	0.205	0.089	0.004
10	***53***	0.004	0.018	0.016	0.013	***0.310***	0.323	0.193	0.123
11	***55***	0.003	0.011	0.009	0.026	0.293	***0.307***	0.257	0.094
12	***61***	0	0	0	0.013	0.301	***0.298***	0.255	0.133
13	***73***	0	0.001	0.001	0.021	0.243	0.305	***0.233***	0.196
14	***70***	0	0.001	0.003	0.016	0.221	0.363	***0.240***	0.156
15	***81***	0	0	0.002	0.038	0.090	0.265	0.379	***0.226***
16	***79***	0	0	0.009	0.032	0.230	0.239	0.286	***0.204***

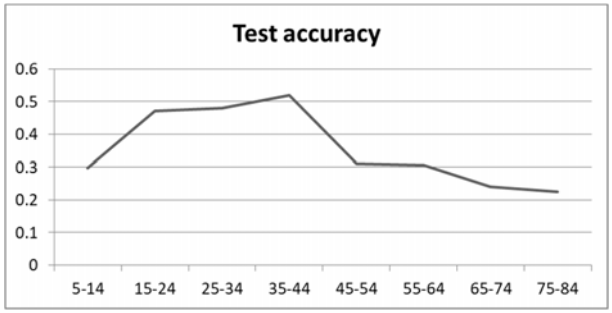

FIGURE VII. CORRECT RATE OF TESTING FOR ALL AGES

B. Summary of the Experiment

In the process of using Caffe to realize the CNN-AGE model, there are many problems, and the main problems are summarized as follows:

- 1) When the training set is converted into LevelDB or LMDB, shuffle treatment is necessary, otherwise train loss will quickly return to 0, leading to the failure of network training.
- 2) Corresponding pre-treatment must be carried out on the data set, particularly the human face data. After the human face testing, aligning and cutting, the training speed and accuracy will be greatly improved.
- 3) The mean value of the image data must be deduced, or the accuracy may reduce.
- 4) Learning rate and momentum cannot be set too large, otherwise, it may lead to the rapid and infinite trend of the weight of some point. Normalization of scale may improve this problem.
- 5) A deep network will consume a lot of computing resources and memory resources, and it takes a very long training time if the situation allows for a smaller network. Too simple networks may be difficult to extract the appropriate classification features, and the appropriate network should be built or selected according to the situation.
- 6) The number of samples in the data set must be enough, otherwise it is easy to fall into the case of fitting.

V. CONCLUSION

Many scholars have carried out research on the age prediction of face information, and many more successful cases have been put forward. According to different methods, the data set, age classification, prediction accuracy rate of high and low, but were not used in the actual production. The main reason is that the accuracy of age identification is not high, and it cannot meet the needs of practical application. But this does not mean that the research is of no value, but it means a better way has not been found to solve the problem. The face age recognition model based on the convolution neural network, which is proposed in this paper with the highest accuracy of 52%. Although it has not reached the standard of practical application, it provides new ideas and methods for the research of face age recognition, and provides an important reference for future research.

ACKNOWLEDGEMENTS

This article has received the support of the Characteristics innovation project of colleges and universities of Guangdong Province (Natural Science), 2016, No.2016KTSCX182; and also received the support of the Youth Innovation Talent Project of colleges and universities of Guangdong Province,2016, No. 2016KQNCX230.

REFERENCES

[1] Savvides, Marios, Seshadri, Keshav, Luu, Khoa. Face age-estimation and methods, systems, and software therefor[J]. 2016.

[2] Wang Xianmei, Liang Lingyan, Wang Zhiliang, et al. On the Technology to Estimate Ages from the Image of Human Face [J]. Journal of Image and Graphics, 2012,17(6):603-618.

[3] Fu Y, Guo G, Huang T S. Age synthesis and estimation via faces: A survey[J]. PatternAnalysis and Machine Intelligence, IEEE Transactions on, 2010, 32(11): 1955-1976.

[4] Zhou Feng, Li Deqiang, Huang Danyi, et al. Design of a Face-age Decision Method [J]. Software, 2015(7):46-52.

[5] Ylioinas J,Hadid A, Pietikainen M. Age Classification in Unconstrained Conditions Using LBP Variants[C]// Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012:1257-1260.

[6] Shan C. Learning local features for age estimation on real-life faces[C]// Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis. ACM,2010:23-28.

[7] Eidinger E, Enbar R, Hassner T. Age and gender estimation of unfiltered faces[J]. Information Forensics and Security, IEEE Transactions on, 2014, 9(12): 2170-2179.

[8] [8] Gao F, Ai H. Face age classification on consumer images with gabor feature and fuzzy lda method[M]. Advances in biometrics. Springer Berlin Heidelberg, 2009: 132-141.

[9] Ueki K, Hayashida T, Kobayashi T. Subspace-based Age-group Classification Using Facial Images under Various Lighting Conditions[C]// International Conference on Automatic Face and Gesture Recognition. 2006:13-18.

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

[11] Bartlett M S, Littlewort G, Frank M, et al. Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior[C]// 2005:568-573.

[12] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014,15(1):1929-1958.

[13] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012, 3(4): 212-223.