The Dropout Method of Face Recognition Using a Deep Convolution Neural Network

Yi Dian
Shanghai Maritime University
china, Shanghai
1241975543@qq.com

Shi Xiaohong
Shanghai Maritime University
china, Shanghai

Xu Hao
Shanghai Maritime University
china, Shanghai

***Abstract*—With the rapid development of artificial intelligence and pattern recognition, face recognition has become a hot topic in the field of computer vision. Especially after the deep learning proposed, the performance of face recognition algorithm has been greatly improved. This paper mainly introduces the main method of face recognition using a deep convolution neural network model. First, for training network parameters, we use a fast convergence stochastic gradient algorithm (SGD). At the same time, the "dropout" method is added to each layer of the network to hide some neuron activity by a certain probability, in order to avoid the over fitting problem caused by the deep network model. Through the above process, a neural network model can recognize face images.**

Keywords—face recognition, convolution neural network, deep learning

I. INTRODUCTION

In the time of the rapid development of information technology, how to distinguish personal identity and protect the information security has become a key social problem to be solved. Traditional W ciphers, cards, documents and other identity authentication technologies are becoming more and more difficult to meet the needs of society, because they are easily forged and lost. The biometric identification technology has become the most secure and complete identity authentication in the world because of its uniqueness, concealment, anti-counterfeit, stability and universality technology.

Biometric identification is a technology that uses the human own, and uniquely identified the physiological or behavioral characteristics of its identity. Among them, the physiological characteristics are individual, including human face, fingerprint, palm print, retina, iris and so on, and behavioral characteristics are customary behavior characteristics of the individual after the day, including handwriting, gait, phonetics and keystroke action. To a certain extent, these characteristics are all common and unique, which can reflect the characteristics of different individuals. Based on these characteristics, the corresponding recognition techniques include face recognition, fingerprint recognition, palm print recognition, retinal recognition, iris recognition, speech recognition and so on. Face recognition is the most important method for human identification. Compared with other biometrics, face recognition technology has the following advantages [1]:

(1) Non mandatory. The user does not need to cooperate with the face acquisition equipment[6], and can almost get the face image in the unconscious state, so that it will not arouse the attention of people and not repugnant.

(2) Noncontact. Users do not need direct contact with the device to get face images, which is safer and more hygienic than fingerprint recognition;

(3) Intuition. When the authentication system can't determine the identity of the identified person or recognize the normal completion of the identification system, the staff generally retains the information of the identified person for the later manual check, and the human face information is easily discernable because it has better visual characteristics - the visual character of the person. Iris information, the natural person does not have the ability to identify;

(4) Simplicity. Face recognition system uses camera equipment to collect face information for recognition, and there is no special requirement for the performance of camera equipment. The common camera equipment, including mobile phones and cameras, can be used and identified without any other auxiliary equipment. In addition, the camera equipment can be placed in high places or other places which are not easy to be detected, so as to avoid human malicious damage.

Because of these advantages, face recognition technology[13] has attracted more and more attention from academia, and it has a very wide application prospect. These applications mainly include the following aspects [2]:

(1) The entrance guard system. Identify the identity of the attempted entrants through face recognition technology and prevent unreliable people from entering.

(2) Criminal investigation to break the case. The staff use network services and face recognition systems to search for fugitives all over the country through a number of ways to obtain a suspect's photo or facial features in order to quickly arrest a fugitive.

(3) Video surveillance. It can monitor people in public places such as banks, airports, stadiums, shopping malls and other public places to prevent terrorist activities.

(4) Network application[7]. Face recognition technology is used to assist the credit card in network payment, so as to prevent non credit card owners from embezzling credit cards.

(5) human-computer interaction[5]. Face recognition on personal computer is activated, face recognition is carried out on mobile phones, face recognition is used for realistic virtual game, etc.

978-1-5386-5638-9/18/$31.00 ©2018 IEEE

From the above, we can see that the research of face recognition technology[9-10] has great theoretical significance and practical application significance. Although the academic research of face recognition has a history of half a century and scholars have put forward many efficient and practical recognition algorithms, face recognition technology still faces great challenges, mainly due to the changes of attitude, expression, illumination, occlusion and so on. In recent years, it is widely used in the field of pattern recognition and image processing[14]. The Convolutional Neural Network (CNN) algorithm has a certain degree of invariance to these effects, so applying the convolution neural network algorithm to face recognition is great significance.

II. DEEP CONVOLUTION NEURAL NETWORK

A. Composition of Deep Convolutional and Subsample layer

The number of layers involves the work of Deep Convolution Neural Network (D-CNN) is scaled down by the convolutional layer and a subsample layer in a single layer. This concept[18] was popularized by Simard, which was after known by Mamalet and Garcia.In this task, we change back to back sub-sample layers and convolutional the single convolutional layer using two strides. A pattern on an image can be extracted by the following expression:

$$p_i^t(q,p) = F\left(\sum_{i=0}^{B}\sum_{u=0}^{R_q^t}\sum_{v=0}^{R_p^t} p_i^{t-e}\left(s_q^t q + u, s_q^t p + v\right)m_{ef}^t(u,v) + \theta_j^t\right) \quad (1)$$

where p_i^{t-e} and p_i^t are the input and output pattern map respectively, F is function i.e known by activation function which we have used in our work, m_{ef}^t is the convolutional kernel weight θ_j^trepresents bias denotes total number of input feature mapping, sq(t)q represents horizontal convolution step size, sp(t)p represents vertical convolution step size, and Rq(t) and Rp(t) are width and height of convolutional kernels, respectively. where M(t-e)and A(t-e) and height and width of input feature mapping.

$$A(t) = (A(t\text{-}e) - Rp(t))/sp(t) + 1 \quad (2)$$

$$M(t) = (M(t\text{-}e) - Rq(t))/sq(t) + 1 \quad (3)$$

B. A CNN for gender estimation

The convolution neural network[3] in Figure 1 is a multi-layer perceptron inspired by biological vision and the simplest preprocessing operation. It is essentially a forward feedback neural network. The biggest difference between the convolution neural network and the multilayer perceptron is that the front layers of the network are made up of the convolution layer and the pool layer, which are used to simulate the visual cortex. In the high level feature extraction, simple cells and complex cells alternate cascade structure.

It uses a feature based method to identify the face, which is different from the traditional artificial feature extraction and the feature based high performance classifier design. Its advantage is to extract the features by layer by layer convolution, and then through multi-layer nonlinear mapping, so that the network can be from the training samples without special processing. Dynamic learning forms a feature extractor and classifier that should identify tasks. This method reduces the requirements for training samples, and the number of layers of the network is more, and the characteristics of the learning are more global.

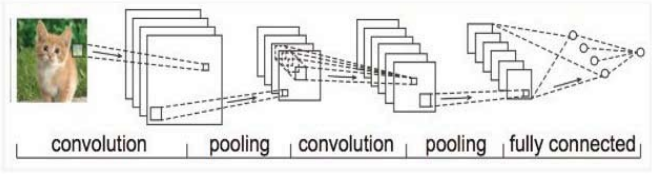

Fig.1 convolution neural network

The basic structure consists of two layers, one is the convolution layer that is used to simulate a simple cell with local receptive field by means of local connection and weight sharing, and the process of extracting some primary visual features is obtained. The local connection refers to each neuron on the convolution layer to connect with the neurons in the fixed area in the previous feature map; the weight sharing means that the neurons in the same feature graph are connected to the previous layer with a set of same connection strengths, which can reduce the network training parameters.

The second is the pooling layer. The pool layer simulation complex cell is the process of screening and synthesizing the more advanced and abstract visual features of the primary visual features. In the network, the number of the output feature maps is changed after sampling by the pooling, but the scale of the feature map will be smaller, and the computational complexity is reduced in order to resistance to changes in micro displacement.

The next layer is the softmax regression layer, which is the essence of the RBF classifier, using the numerals in the ASCII code table as the error correcting code or the equal probability of selecting -1 and 1 as the weight value of RBF. Each RBF input unit calculates the Euclidean distance between the input features and the parameter vectors, but the face features are more complex, and there is no uniform template for face classification, and softmax regression with nonlinear classification ability can be used as a classifier to deal with these problems better.

III. MATERIALS AND METHODS

A. Benchmark dataset

YouTube Faces Database, a database of face videos designed for studying the problem of unconstrained face recognition in videos. The data set contains 3,425 videos of 1,595 different people. All the videos were downloaded from YouTube. An average of 2.15 videos are available for each subject. The shortest clip duration is 48 frames, the longest clip is 6,070 frames, and the average length of a video clip is 181.3 frames.

This section will use the YouTube face dataset. The face image in the dataset (5.4GB) is intercepted from the YouTube video site, with a total of more than 300 people, each with at least 2 pictures in the scene, and there are about 40 pictures in each scene. There shows a part of the YouTube face image Figure .2

Fig.2 YouTube face image

B. SGD algorithm

With the increase of the number of neural networks[15], the number of parameters increases multiplicative. When training network[4], it is a very important problem to update parameters. For updating the parameters of supervised learning convolution neural network, the stochastic gradient descent algorithm with fast convergence rate is usually adopted. Its updating speed is higher than gradient descent algorithm, and it also ensures that the parameters converge in the direction of the optimal solution.

$$f(x) = \sum_i^n f_i(w_i, x_i, y_i) \tag{4}$$

$$w_{i+1} = w_i - \eta_{i+1} \sum_i^n \nabla f_i(w_i, x_i, y_i) \tag{5}$$

$$w_{i+1} = w_i - \eta_{i+1} \nabla f_i(w_i, x_i, y_i) \tag{6}$$

In Eq(1),where f is the activation function, w is the weights, and x_i, y_i are the input and outpit of layer i for each other. Parameter weight updating of gradient descent method is Eq(2). The idea of stochastic gradient descent is to randomly select one ∇f_i in each medium instead of the $\sum_i^n \nabla f_i(w_i, x_i, y_i)$ above, taking this randomly chosen direction as the direction of descent, as is seen in Eq(3).

C. dropout

With the limitation of training data and the increase of training parameters, almost all large convolutional neural networks are facing the problem of overfitting. All over fitting is nothing more than the lack of training samples and the increase of training parameters. In general, a lot of training parameters are needed to get a better model, which is one of the reasons why the CNN network is getting deeper and deeper. And if the training sample lacks diversity, the more training parameters are meaningless, because this creates a fitting, and the generalization ability of the training model will be poor.

Dropout can be regarded as an average model[19], so called model average. As the name implies, it means that the estimation or prediction from different models can be averaged through a certain weight. In some literature, it is also called model combination, which generally includes combination estimation and combination prediction.

Where the "different model" in Dropout is, the secret is that we randomly choose to ignore the hidden layer nodes. In each batch training process, the hidden layer nodes are different at random, so that each training network is different, and each training practice can be a "new" model. In addition, the hidden nodes appear randomly in a certain probability, so that every 2 hidden nodes can't be guaranteed at the same time each time, so the update of the weight is no longer dependent on the joint action of the fixed relation hidden nodes, which prevents some of the features only having the effect under other specific features[16].

During the training phase:

1. Dropout[20] is on the structure of the standard BP network, which makes the activation value of the hidden layer of the BP network changed to 0 with a certain proportion of v, that is, a portion of the hidden layer nodes are randomly invalid according to a certain proportion of v. In the rear benchmark experiment test, some experiments make the hidden layer nodes Invalidation on the basis of a certain proportion (20%) is part of the input data failure (this is a bit like denoising auto-encoder), so that a better result.

2. Removing the weight penalty item, replace the limiting the scope of the weight and seting an upper limit for each weight value. If the weight exceeds the upper limit during the training and the new process, the weight value is set to the value of the upper limit

No matter how large the update is, the weights will not be too large. In addition, the algorithm can also make the algorithm use a larger learning rate to speed up the learning speed, so that the algorithm can search for better weights in a wider weight space without worrying that the weight is too large.

Test phase:

The output values of the hidden layer nodes in the front of the network to the output layer should be reduced to (1-v) times; for example, the normal hidden layer output is α, and it needs to be reduced to α(1-v) at this time.

In this way, the dropout process is a very effective neural network model averaging method, which can predict the average probability by training a large number of different networks. Different models are trained on different training sets (the training data for each batch are random selected), and finally "fusion" is used in each model with the same weight.

IV. EXPERIMENT

A We use the VGG-16 model for testing[8], and VGG-16 has 13 coiling layers, 5 maximum pool layers, 3 full connection (FC) and 1 SoftMax layers, which are configured as shown in Table.1

Table.1 The configuration of each layer of VGG-16

layer	Layer name	configuration
0	Input	RGB images of 224x224 size
1	conv1_1	Kernel_size=3，stride=1，pad=1，num_output=64
2	conv1_2	Kernel_size=3，stride=1，pad=1，num_output=64
3	pool1	Kernel size=2，stride=2，pad=0
4	conv2_1	Kernel_size=3，stride=1，pad=1，num_output=128
5	conv2_2	kernel_size=3，stride=1，pad=1

		，num_output=64
6	pool2	kernel_size=2，stride=2，pad=0
7	conv3_1	kernel_size=3，stride=1，pad=1，num_output=256
8	conv3_2	kernel_size=3，stride=1，pad=1，num_output=256
9	conv3_3	kernel_size=1，stride=1，pad=1，num_output=256
10	pool3	kernel_size=2，stride=2，pad=0
11	conv4_1	kernel_size=3，stride=1，pad=1，num_output=512
12	conv4_2	kernel_size=3，stride=1，pad=1，num_output=512
13	conv4_3	kernel_size=1，stride=1，pad=1，num_output=512
14	pool4	kernel_size=2，stride=2，pad=0
15	conv5_1	kernel_size=3，stride=1，pad=1，num_output=512
16	conv5_2	kernel_size=3，stride=1，pad=1，num_output=512
17	conv5_3	kernel_size=1，stride=1，pad=1，num_output=512
18	pool5	kernel_size=2，stride=2，pad=0
19	fc6	Output of 4096 dimensions vector
20	fc7	Output of 4096 dimensions vector
21	fc8	Output of 4096 dimensions vector
22	SoftMax	Output classification results

Because of the lack of depth models[11-12] in face recognition, most of them are focused on face verification. Therefore, the VGG model is used as feature extraction in this section, and the feature vectors extracted from the fc7 layer are sent to a nearest classifier (Nearest Neighbor, NN) for classification. We randomly selected 50% YouTube database data for training, the other as a test. Its performance is like Table. 2:

Table.2 example result

Model	accuracy
VGG-16	97.35%

As can be seen from Table.2, the accuracy of the network model and VGG16 designed in this paper is 97.35%. At present, the model of deep learning is gradually developing in a deeper and more complex direction. The deeper, more complex model is more variable and learning ability will be stronger because of the increase of layer series, and how many levels and how to configure the network model without detailed theoretical support, depending on the type of task and experimental results. To evaluate, the experimental results show that the deep convolution neural network designed in this chapter can meet the system requirements.

The convergence of deep learning methods is still unsolved and beyond our topic. Here we simply depict the curves of objective function value and classification rate to show the convergence of DCNN. Fig. 3 shows the objective function loss versus different number of iterations on dataset. It is easy to see our DCNN fast converges a local minimum after a few iterations. Fig. 4 shows the classification rate versus different number of iterations on dataset. We can see that our DCNN converges in 30 iterations and the classification comes to a stable value after 30 iterations.

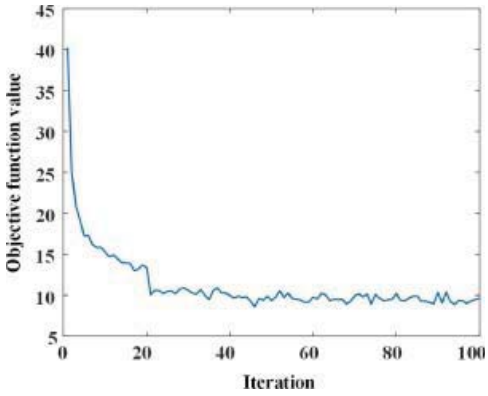

Fig. 3 Convergence curve. Here one iteration means the training data input into DCNN once.

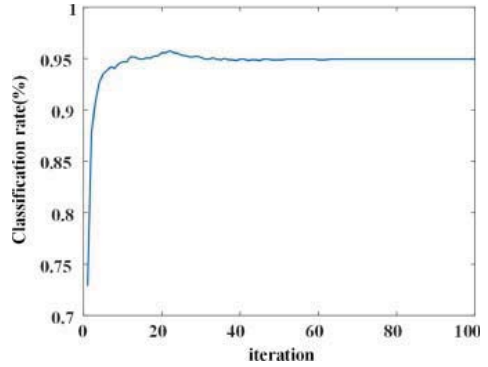

Fig. 4 Classification rate versus different number of iterations of DCNN .

V. CONCLUSION

In this paper, a convolution neural network is used to construct the model, mainly because the convolution neural network is invariant to geometric transformation, deformation, illumination and so on, and can take less time to deal with high dimensional data because of the sharing of convolution kernel. The trained convolution neural network can scan the whole image to be detected at a lower computational cost. The main contents are as follows:

(1) Analyzing the current situation of convolution neural network in face recognition, and explaining the key issues of its research. Discuss the view of research.

(2) Constructing a face recognition convolution neural network model which is a deep neural network model which consists of the convolution layer and the pool layer alternately to form the front half part model, and the latter part of the model is composed of multiple full connection layers and the last softmax layer. For the training of parameters in the network, the stochastic gradient descent algorithm is adopted. At the same time, in order to prevent the overfitting problem caused by the stochastic gradient descent algorithm, we add the "dropout" method to each layer of the network.

REFERENCES

[1] Paul Viola, Michael J. Jones. Robust Real-Time Face Detection[J]. International Journal of Computer Vision, Volume 57, Issue 2, 137-154, 2004.

[2] P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model[C]. In Proceedings of CVPR, 2008.

[3] Sachin, Mohammad, and Li-Jia Li. Multi-view Face Detection Using Deep Convolutional Neural Networks[C]. International Conference on Multimedia Retrieval, 2015

[4] Y. LeCun, L. Bottou, G. B. Orr, and K.-R.M¨uller. Efficient backprop. In Neural Networks: Tricks of the Trade, pages 9–50. Springer, 1998.

[5] Brunelli, R., Poggio, T. Face Recognition through Geometrical Features[C]. European Conference on Computer Vision (ECCV) , S. 792–800, 1992.

[6] Turk and Pentland. Eigenfaces for recognition[J]. Journal of Cognitive Neuroscience 3, 71–86 ,1991.

[7] M. Lin, Q. Chen, and S. Yan. Network in network. arXiv:1312.4400, 2013.

[8] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015. [T.Kohonen. The self-organizing map[J]. Proceeding of the IEEE, 78:1464-1480, 1990.

[9] Y. LeCun, K. Kavukcuoglu, and C. Farabet. Convolutional networks and applications in vision. In Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on, pages 253–256. IEEE, 2010.

[10] Steve Lawrence, Lee Giles, Ah Chung Tsoi, and Andrew D.Back. Face Recognition: A Convolution Neural Networks Approach[J]. IEEE

[11] C. Szegedy,W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In CVPR, 2015.

[12] K.Huang and S.Aviyente. Sparse Representation for Signal Classification[J]. Neural Information Processing System, 2006.

[13] Jhon Wright, Yi Ma, and et al. Robust Face Recognition via Sparse Representatioin[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.31, NO.2, February, 2009.

[14] Zhenyao Zhu, Ping Luo, and Xiaogang Wang. Deep Learning-Preserving Face Space[C]. ICCV, 2013,

[15] O. Barkan, J. Weill, L. Wolf, and H. Aronowitz. Fast high dimensional vector multiplication face recognition[J]. ICCV, 2013.

[16] [T.-H. Chan, Y. Ma, et al. PCANet: A simple deep learning baseline for image classification?[J]. IEEE Transactions on Image Processing, 2015 .

[17] Cao, Z., Yin, Q., Tang, X. Face recognition with learning-based descriptor[J]. In Computer Vision and Pattern Recognition (CVPR), 2707–2714, 2010.

[18] Ranzato, M., Susskind, J., Mnih, V., Hinton, G. On deep generative models with applications to recognition[J]. In Computer Vision and Pattern Recognition (CVPR), 2857–2864, 2011.

[19] Lone, M.A., Zakariya, S.M., and Ali, R. Automatic Face Recognition System by Combining Four Individual Algorithms[C]. Computational Intelligence and Communication Networks (CICN), 2011.

[20] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving neural networks by preventing coadaptation of feature detectors. arXiv:1207.0580, 2012