Research on Face Recognition Based on Deep Learning

Xiao Han
College of Software, Shenyang Normal University
Shenyang, China
941066866@qq.com

Qingdong Du
College of Software, Shenyang Normal University
Shenyang, China
421532897@qq.com

***Abstract*—With the deep learning in different areas of success,beyond the other methods,set off a new wave of neural network development.The concept of deep learning originated from the artificial neural network,in essence,refers to a class of neural networks with deep structure of the effective training methods[1].As a powerful technology to realize artificial intelligence,deep learning has been widely used in handwriting digital recognition,dimension simplification,speech recognition, image comprehension,machine translation,protein structure prediction and emotion recognition.In this paper, we focus on the research hotspots of face recognition based on depth learning in the field of biometrics,Combined with the relevant theory and methods of depth learning,face recognition technology, along the order of depth learning,based on the depth of learning face recognition,face recognition application to start research.**

Keywords—Deep Learning, Face Recognition, Biometrics, Neural Network

I. INTRODUCTION

A complete face recognition system includes two patterns of face detection and face recognition.The biological characteristics of the face have overall structural similarity and individual local differences[2].Therefore,it is necessary to extract the structural features of the face through the face detection process,and to separate the faces from the background pattern and face recognition of the separated faces. This is a process of extracting normalized face images,and then contrast and identification,the purpose is to distinguish the identity of the face from the image.

From the fifties of the last century to the eighties,face recognition has gone through the first stage,and face recognition is regarded as a general pattern recognition problem.The mainstream technology is based on the facial geometry.In the nineties of last century, experienced the second stage.In this period, the rapid development of face recognition,there have been many classic methods, such as Eigen Face,Fisher Face and elastic map match,the mainstream technology is based on face performance modeling[3].From the 1990s to the present,it is the third stage of face recognition technology.Face recognition into the in-deep development stage,the researchers began to focus on real conditions for face recognition, including the following four aspects:

First,we propose different face space models.Secondly, we deeply analyze and study the factors that affect face recognition.Then, using the new features to represent, the deep learning is one of the very important features.Finally, with new data sources to achieve face recognition: video-based face recognition,sketch-based face recognition,near-infrared face recognition.

II. DEEP LEARNING

In recent years, with the development of machine learning,deep learning as a new research direction caused widespread concern in the field of artificial intelligence[4].Deep learning is a kind of machine learning algorithm which can effectively train deep neural network, which can be used for high-level abstract modeling of data.On December 29,2012,the front page of the New York Times reports that "deep learning allows the machine to perform human activities,such as watching,listening and thinking, providing the possibility of pattern recognition and promoting the advancement of artificial intelligence."[5]In 2013,the MIT Technology Review ranked deep learning as one of the top ten technological breakthroughs in the world.

Last year,with the rapid development of deep learning,Both in industry and academia are devoted to physical and mental in the deep learning.Google,Facebook,Baidu, Alibaba and a series of large companies are claiming that artificial intelligence will be their next strategic focus.Google, Facebook,Baidu,Microsoft,Amazon and other companies have opened up their own deep learning framework.

A leading authority of the deep learning Andrew Ng joined Baidu,Yann LeCun joined Facebook,Leading figure of Computer Vision,Stanford University Professor Li Feifei joined Google,Alex Smola, the top figure of machine learning field,a professor at Carnegie Mellon University joins Amazon.May 21, 2015,Baidu led the "Deep"officially open to the public.Deep is a distributed deep of machine learning open source platform[6].

Deep learning designed a lot of machine learning technology and structure, according to these structures and technology applications can be divided into three categories:

(1)Generative model:Generating refers to the refactoring process from the hidden layer to the input data.Here mainly

ISBN: 978-1-5386-3903-0/18/$31.00 ©2018 IEEE

introduce deep belief network.Deep belief network is a generative model of deep learning, proposed by Geoffrey Hinton in 2006.Deep belief network is constructed by stacking a restricted boltzmann machine,as shown in Fig.1.Restricted boltzamann machine is a probabilistic graph model which can be interpreted as a random neural network. It is a random undirected graph model,which can be used as a basic module to construct many other deep learning models such as self-encoders and deep belief networks.The deep belief network as a model of the learning process can be divided into two stages;The first stage, the use of limited Boltzmann machine for layer-by-layer training[7].The pre-training is carried out by unsupervised greedy layer-by-layer algorithm.The second stage uses the wake-sleep algorithm to tune.In the wake phase,using the identify weights and the identify bias with the high repetition rate to estimates the generate weights and the generate bias.In the sleep phase,Repeatedly using the generate weights and the generate bais to estimates the identify weights and the identify bias.

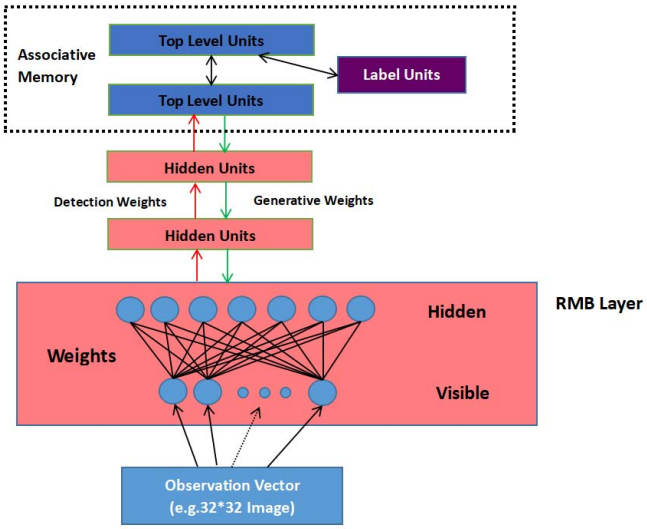


Fig.1. DBN Framework.

(2)Discriminative model:Discriminant is the protocol from the input data to the hidden layer.Discriminative models are often used to classify the intrinsic pattern of data or to describe the posterior distribution of data.We mainly introduce convolution neural networks.The standard convolution neural network is a special feedforward neural network model.The connections between adjacent neurons are non-all-connected, and the weights of the connections between certain neurons in the same layer are shared and usually have a relatively deep structure,usually composed of input layer,convolutional layer also known as detection layer,subsampling layer also named pooling layer,fully-connected layer and output layer.

The convolution operation is roughly as follows: The input image is subjected to non-linear convolution through three trainable filter group.After the convolution,the feature map is generated at each layer,and then the four pixels of each group in the feature map are summed,weighted, and offset,in which the pixels are pooled in the cell layer,Eventually reaching the output value.

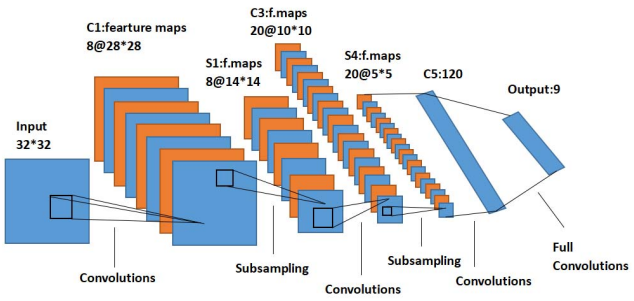


Fig.2.Example of Convolution nerual network.

(3)Mixed model:Based on the combination of standard model and variance model,through the combination of other methods,deep learning also establishes a number of mixed models[8].What is "mixed"?We should understand the meaning of "mixed" from two aspects:

□Mix the ingredients structure:Meaning the deep structure that combines the different model components.From this point of view,there are four deep mixed structure:

1)Generative+generative mixed model,such as the mixed model of deep belief network and the hidden Markov model-DBN-HMM.

2)Discriminative+discriminative mixed model,such as the mixed model of multilayer perceptron and support vector machine-MLP-SVM.

3)Discriminative+generative mixed model,such as the mixed model of convolution neural network and recurrent neural network-CNN-RNN.

4)Generative+discriminative mixed model,such as the mixed model of deep belief network and conditional random fields-DBN-CRF.

□Mixed learning structure:Combined with the deep structure of different learning strategies.From this point of view,there are six deep mixed structure:

1)Unsupervised + supervised learning mixed structure: Firstly, use the way of unsupervised to pre-training deep neural network, and then use a supervised way to tune,such as learning in auto encoder.

2)Deep+Bayesian learning mixed structure:Combined with Bayesian learning methods to train deep neural networks, such as learning in a composite layered deep model.

3)Deep+kernel learning mixed structure:Combining nuclear learning methods to train deep neural networks,such as learning in a multi-layered nuclear machine.

4)Deep+transfer learning mixed structure:Combined with the migration learning method to train deep neural networks,such as high-level abstract features,in the same deep structure to share different types of data.

5)Deep+reinforcement learning mixed structure:Combining the enhanced learning method to train the deep neural network.In this hybrid structure,a subject of intensive learning must find a way to interact with a dynamic,initially unknown environment,with the aim of maximizing its expected cumulative reward signal.

ISBN: 978-1-5386-3903-0/18/$31.00 ©2018 IEEE

6)Deep+other learning mixed structure:Combine alternative learning methods to train deep neural networks,such as learning in a deep coding network.

III. FACE RECOGNITION BASED ON DEEP LEARNING

A. The main method of face recognition

The process of face recognition is the process of feature extraction and contrast recognition of normalized face images. The purpose is to distinguish the identity of the face from the image.A traditional face recognition system is shown in the following figure.

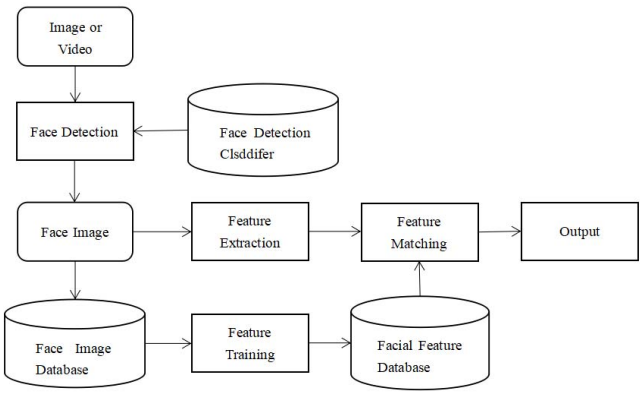


Fig.3.The Basic Components of Face Recognition System.

The main methods of face recognition are:subspace method,geometric structure method, local feature method and deep learning method.

(1) Subspace method:The basic idea is to transform the high-dimensional image features into a low-dimensional subspace by spatial compression.So that the original sample features in this low-dimensional sub-space easier to classify.The representative algorithm has linear discriminant analysis LDA(Latent Dirichlet Allocation), principal component analysis PCA(Principal Components Analysis),independent component analysis ICA(Independent Component Analysis) and so on[9].

(2) Geometric structure method: By using a set of geometric feature vector representation of the face topology.Which converts the recognition process into the matching between the feature vectors and uses the Euclidean distance to make a similar measure.

(3) Local feature method: the basic idea is to split the face image into a number of local features,so that interference factors only affect a part of the characteristics.The feature of the face is described by local features that are not affected by the interference factor.Methods based on local features:local descriptor Gabor features,local binary pattern(LBP)feature, scale invariant feature transform (SIFT) feature and so on.

(4) Deep learning method: As the facial features will change with external or internal conditions,so the study of high complexity.This inspires us to let the machine imitate the human brain structure and the way of thinking,use of neural networks to complete this goal.

B. Face Recognition Based on Deep Learning

The first artificial neural network used to identify the face is a single-layer adaptive network.But its robustness to interference factors such as emotions, gestures,backgrounds and shelters,can not get better output.Deep learning solves this problem,not only to ensure accuracy,but also to ensure the robustness of the algorithm.

G.Hinton and others use the greedy layer-by-layer algorithm to train DBNs to ensure that the lower bound of the logarithmic likelihood function of the data is improved,and this method also applies to face recognition.Marc Aurelio Ranzato uses the gated Markov random field(MRF)as the front end of DBNs to learn the deep generation model of face images,which can learn facial expressions and have better robustness.Osadchy.M uses convolution neural networks for face detection and attitude assessment.Huang.G.B uses convolution deep belief network(CDBN) to learn multi-level features.Sun.Y uses a three-layer convolution neural network to construct a hierarchical regression structure[10],and proposes a network model of mixed convolution neural network (ConvNet) and restricted Boltzmann machine(RBM). The algorithm uses deep convolution neural network to extract the face image Local related features.

Lin M proposes to employ a deep belief network to deal with non-linear problems caused by different postures.Brain Cheung proposed by training convolution neural networks to make it possible to distinguish between natural face images and computer-generated face images[11].Chen, in order to solve the problem of image size,divide the image into non-overlapping blocks,and then use the deep neural network(DNN)to train each block,and finally connect the local features of the training to identify.

In 2012,the Rennes Miller team took the lead in learning the face recognition for the LFW (labeled face in the wild) database with unsupervised feature learning methods that achieved 87% recognition.The recognition rate of the classical face recognition algorithm Eigenface in LFW is only 60%, and the recognition rate of the latest deep learning algorithm is 99.47%,even more than the recognition rate of the human eye,the recognition rate of human eye is 99.25%[12].In recent years,many of the international projects will be the deep learning to successfully applied to face recognition,such as: DeepFace, DeepID, FaceNet and so on.

The above algorithm is based on massive training data,so that the deep learning algorithms from the massive data to learn for the light,expression, angle and other characteristics of the same.In those algorithms, the highest accuracy is FaceNet, in the LFW database has reached 99.63%, the accuracy rate, more than the human eye recognition results.The LFW database was completed by the computer vision laboratory at the University of Massachusetts University in Amst,and was used to study the face recognition problem in unrestricted cases,which has become a standard reference for academic evaluation.The LFW project is the driving force for the face recognition technology from the initial stage to the prototype system stage[13].The data of the LFW test has become the reference standard of the present paper.The database has

ISBN: 978-1-5386-3903-0/18/$31.00 ©2018 IEEE

greatly promoted the progress of face recognition technology in deep learning.Table 1 and 2 respectively list the recognition rates of deep learning and other methods in the LFW face database.

TABLE □. RECOGNITION RATE OF DEEP LEARNING IN THE LFW FACE DATABASE

Method	Registration keywords	Training sample	Recognition rate/%
ConvNet	3	unsupervised	87
FaceNet	5	87620	92.52
DeepFace	73	7000000	97.35
DeepFace	73	500000000	98.40
DeepID	5	202599	97.45
DeepID2	18	202599	99.15
DeepID2+	18	450000	99.47

TABLE □. RECOGNITION RATE OF EACH METHOD IN THE LWF FACE DATABASE

Method	Recognition rate/%
COST-S1[29]	3.0
COST-S1＋S2	56.7
DeepFace	66.5
DeepFace	64.9
DeepID2	91.1
DeepID2＋	95.0

Compared with other methods,the use of deep learning method to extract facial features,so that the complex extraction process becomes simple,Recognition accuracy,robustness, efficiency have a significant advantage.We will mainly introduce the deep learning based on the face recognition method.

IV. The Application of Face Recognition Based on Deep Learning

The emergence of deep learning makes face recognition technically breakthrough[14].The latest research results of face recognition show that the facial features expressed by deep learning have important characteristics which are not possessed by manual feature expression.

A. Face Recognition Based on Deep Self-coding Network

Algorithm of face recognition based on Deep self-coding Network.Deep self-encoder network structure as shown in Figure 4,it has an input layer containing number n of neurons, 2r-1 hidden layer and an output layer.

Firstly,the image is preprocessed, the face image is normalized to [0,1] and the size of the image is adjusted to 32 * 32.And then use the network model for training,face feature extraction, the network model as shown in Figure 5,and use the Softmax classifier to identify the face.

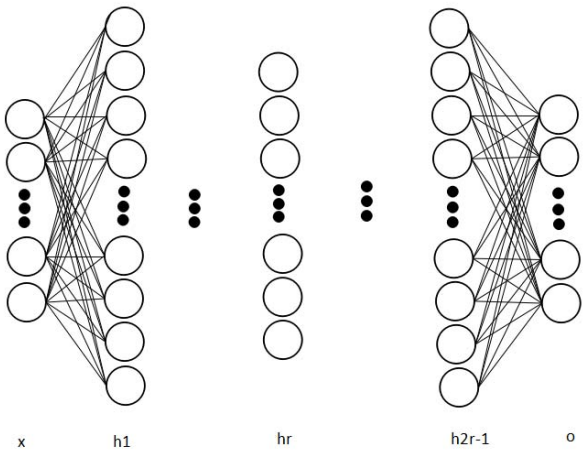


Fig.4. Deep Auto Encoder Network Structure.

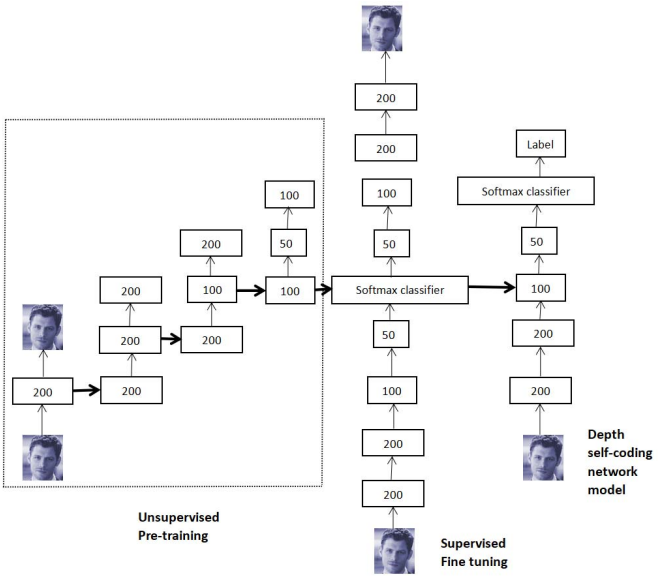


Fig.5.The Model of Deep Auto Encoder Network.

Since the encoder was first proposed by Rumelhart in 1986,it can be used to dimension the high dimensional data[15].In 2006, Hinton improved the self-encoder learning method,formally proposed a deep self-encoder.

The model used in this paper is a deep self-coding network with four hidden layers,which is used to extract the features of the face.The softmax classifier is used to classify the extracted features.Softmax regression is a kind of supervised learning.

It is used in conjunction with unsupervised learning-self-coded neural networks,with hidden nodes of 200-200-100-50 for each hidden layer.These nodes are randomly selected and integer nodes selected for the convenience of the experiment.The whole learning process is divided into pretreatment and tuning.During the preprocessing process, from the input layer of the encoder to the middle layer,the adjacent two layers are regarded as a restricted Boltzmann machine,from the bottom of the limited Boltzmann machine began to use unsupervised learning algorithm layer by layer

ISBN: 978-1-5386-3903-0/18/$31.00 ©2018 IEEE

training,this method is called greed by layer unsupervised pre-training.The process of tuning,the network regard as a whole,and then the overall tuning, so as to get the optimal network structure.

Self-coded neural network is an unsupervised learning algorithm.Through the analysis of the brain,it is found that adding the sparseness restriction in the self-coding neural network is that the self-coding network also can learn the inherent correlation between the input data and the hidden neurons.

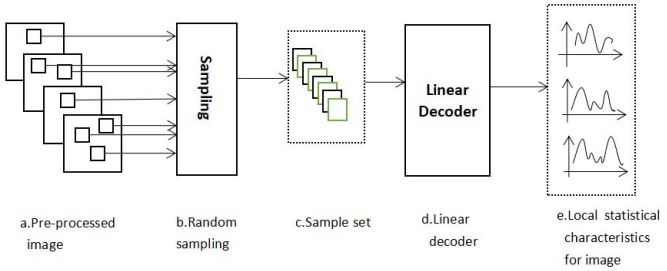
B. Face Recognition Based on Covolution Neural Network

Convolution neural network (CNN) is a kind of deep under the supervision of learning under the machine learning model[16],which can extract the local features of the data, extract the global training characteristics and classification, and its weight sharing structure network makes it more similar to the biological neural network and has been successfully applied in various fields of pattern recognition.

The face recognition based on deep convolution neural network is divided into the following four parts:1. Using ZCA(Zero-phase Component Analysis) whitening to pre-train the data.2.Using the linear decoder to extract the local features of the face image as the convolution kernel of the convolution neural network.3.Using the convolution kernel convert the face image,extract the convolution characteristics of the image and pool it.4.Using the Sofemax classifier to classify the convolution characteristics and complete the face recognition. And the deep convolution neural network has good robustness to scaling, tilting and translating[17].

The algorithmic network model is shown in the following figure,and the algorithm needs to train the linear decoder and Softmax classifier.The training steps are as follows:

1. Randomly sample the processed face image to obtain a sample set.
2, The sample set for the linear decoder input, and then training linear decoder.
3, Using the trained linear decoder as a convolution kernel, extract the convolution characteristics.
4. Use the average pooling strategy and then pool the convoluted image.
5. Put the convolution characteristics after the pooling together and then normalized to get face image features,which are used as input to the Softmax classifier to train the Softmax classifier.



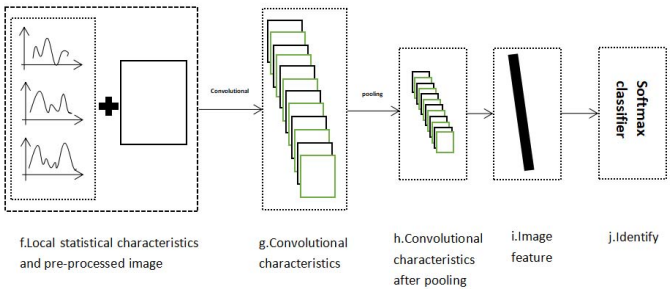


Fig.6.The Model of Deep Learning Convolutional Neural Network.

The normalized image data is pretreated with ZCA whitening,which is modeled by the method of treating the image with the retina.With two hidden layer of linear decoder, the input of the sample block is 9 * 9,two hidden layers of hidden nodes to 50, the network structure as shown below, he linear decoder training steps are as follow,

First,pre-trained the linear decoder,and then the linear decoder is tuned to obtain the optimal parameters.The convolution characteristics are pooled by averaging pooling,The reason is: First,theoretically,you can use the Softmax classifier to directly train the convolution feature,but the computation is too large.Second,the pooling has good robustness to the image feature translation.So the average pool can not only reduce the complexity of network computing, but also can improve network performance.

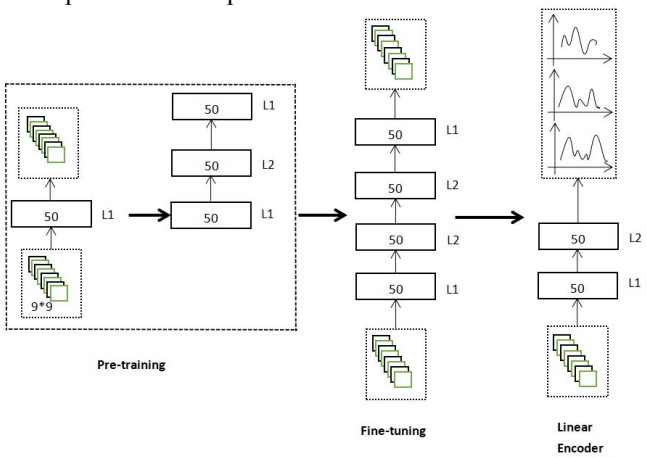


Fig.7.The Model of Linear Encoder Network.

V. CONCLUSION

With the progress of science and technology,artificial intelligence more and more frequently appear in our field of vision,and face recognition has also become an important means of changing people's lives[18].Deep learning has a key advantage over machine learning for other face recognition techniques.First,low-level features can be learned from raw data that is almost no processing.Second,complex interactions can be detected from the features.Therefore,the deep learning can not only learn to get more useful data,but also to build a more accurate model.But the current deep learning have problems need to be resolved:

ISBN: 978-1-5386-3903-0/18/$31.00 ©2018 IEEE

(1) What degree of deep learning can really achieve artificial intelligence?
(2) How to evaluate and judge the input characteristics after learning is the optimal solution?
(3) Can we combine deep learning with other face recognition algorithms and upgrade existing face recognition models based on deep learning into 3D models?
(4) How to further enrich the database resources?

On the current development trend,based on the deep learning face recognition will be a development trend,there are still many problems waiting for us to discover and improve[19],For example,in the case of multi-hidden layer based on the deep of self-coding face recognition,parameter settings are still lack of further research;in the convolution neural network based on face recognition,how to extract a higher order image features,how Improve the image recognition rate and robustness,is still a problem to be solved.

Therefore,we must combine the knowledge to solve unresolved problems,,improve the performance to be improved.

REFERENCES

[1] Lee H,Grosse R,Ranganath R,et al. Unsupervised learning of hierarchical representations with convolutional deep belief networks[J]. Communications of the ACM,2011,54(10): 95–103.

[2] Sun Y,Wang X,Tang X. Deep learning face representation from predicting 10000 classes[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE,2014: 1891–1898.

[3] Huang G B,Lee H,Miller E L. Learning hierarchical representations for face verification with convolution deep belief networks[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE,2012: 223–226.

[4] Ouyang W,Wang X. Joint deep learning for pedestrian detection[C]//Proceedings of International Conference on Computer Vision. Piscataway: IEEE,2013: 2056–2063.

[5] Liu M Y,Shan S G,Wang R P,et al. Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE,2014: 1749–1756.

[6] Sun Y,Wang X G,Tang X. Deeply learned face representations are sparse selective ,and robust[EB/OL]. (2014–12–03)[2016–02–23]. http//arxiv.org/abs/1412. 1265.

[7] Hu Y,Liao S C,Lei Z,et al. Exploring structural information and fusing multiple features for person reidentification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE,2013: 794–799.

[8] Schroff F,Kalenichenko D,Philbin J. FaceNet: A unified embedding for face recognition and clustering [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE,2015: 815–823.

[9] Sun Y,Wang X D,Tang X. Deep learning face representation by joint identification-verification[J]. Advances in Neural Information Processing Systems 2014,27: 1988–1996.

[10] Lu C C,Tang X. Surpassing human-level face verification performance on LFW with GaussianFace [EB/OL]. (2014–04–15)[2016–02–23]. http: //arxiv.org/abs/1404. 3840.

[11] Ranzato M,Susskind J,MnihV,et al. On deep generative models with applications to recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE,2011: 2857–2864.

[12] Susskind J,Hinton GE,Memisevic R,et al. Modeling the joint density of two images under a variety of transformations[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway : IEEE ,2011: 2793–2800.

[13] Mnih V,Kavukcuoglu K,Silver D,et al. Human-level control through deep reinforcement learning[J]. Nature,2015,518(7540): 529–533.

[14] Wang Y Q,Xie Z G,Xu K,et al. An efficient and effective convolutional auto-encoder extreme learning machine network for 3d feature learning[J]. Neurocomputing,2016,174: 988–998.

[15] Tissera M D,McDonnell M D. Deep extreme learning machines: Supervised autoencoding architecture for classification[J]. Neurocomputing,2016,174: 42–49.

[16] Ji S,Xu W,Yang M,et al. 3D convolutional neural networks for human action recognition[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence,2013,35(1): 221–231.

[17] Kumar N,Berga A C,Belhumeur P N,et al. Attribute and simile classifiers for face verification[C]//Proceedings of the 12th International Conference on Computer Vision,Piscataway: IEEE,2009: 365–372.

[18] Cui Z,Chang H,Shan S,et al. Deep network cascade for image super-resolution[M]//Computer Vision-ECCV 2014. Springer International Publishing,2014: 49–64.

[19] Liu M Y,Li S X,Shan S G,et al. Deeply learning feformable facial action parts model for dynamic expression analysis[M]//Computer Vision-ACCV 2014. Springer International Publishing,2014: 143–157.