ИКОНИКА – НАУКА ОБ ИЗОБРАЖЕНИИ

FACE RECOGNITION: A NOVEL DEEP LEARNING APPROACH

Sh. Ch. Pang, Doctoral Student; **Zh. Zh. Yu**, Professor of Jilin University, Corresponding Author

College of Computer Science and Technology, Jilin University, Changchun 130012, China

E-mail: pangshuchao1212@sina.com, yuzz@jlu.edu.cn

We propose a novel and robust-deep learning method for face recognition, which uses an effective image representations learned automatically to handle with big data. There are two stages about the deep learning architecture in real-time application. First, on the offline training procedure, we train a stacked denoising autoencoder to learn generic image features from 80 million Tiny images dataset used as auxiliary offline training data. Second, on the supervised object recognition procedure, we construct a five layers as a feature extractor to produce an image representation and an additional classification layer, which we can use to further tune generic image features to adapt to specific objects recognition by online training of corresponding objects. Comparison with the state-of-the-art face recognition methods shows that our deep learning algorithm in face recognition is more accurate and it is a perfect processing tool in big data problem.

Keywords: *Big data, face recognition, deep learning, feature extraction, feature learning.*

OCIS codes: 100.0100, 100.3008, 100.4996

Submitted 05.10.2014

1. Introduction

In recent years, there is a large number of data including images, videos, text and so on, which have been produced by searching or uploading and downloading them on the Internet all the time. Every day, 2.5 quintillion bytes of data is to be created and 90 percent of the data coming from the world today was produced in the past two years [1]. Another example is Flickr, which is a public picture sharing site, had been receiving 1.8 million photos per day which was calculated from February to March 2012 on average [2]. Along with the above examples, the age of Big Data has arrived [3–5]. In particular, images are an important component of big data. For the research of images, image recognition is stilla popular area of research in computer vision and face recognition is one of the most successful applications of image analysis and understanding. In addition, its social and cultural implications are long-awaited and hot domain in machines and the human visual system serves in the modern world.

Because of the nature of face recognition, not only computer scientists are interested in it, but neuroscientists and psychologists also. The advanced opinions in computer vision research will provide useful insights to neuroscientists and psychologists into how human brain works, and vice versa [6] in Fig. 1. So here, we present a novel face recognition algorithm based on effective

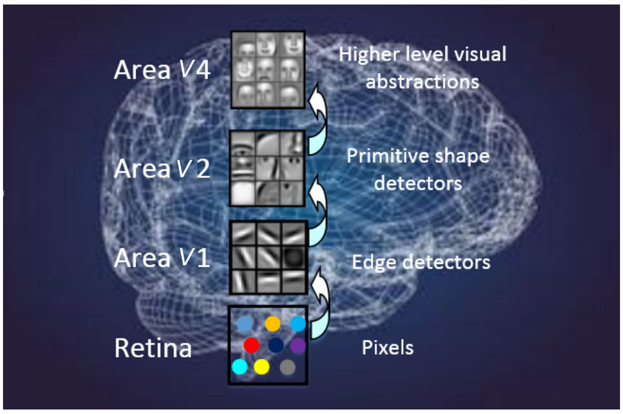


Fig. 1. The working process of the brain: the brain can extract the edge feature from the lower level area $V1$, then get the shape feature from the area $V2$, until extract the higher level feature on the process.

deep neural network architecture, which includes five layers to produce the extremely compact face feature representation. Further, proposed face recognition system is perfect in this research field in that it uses the deep learning framework to learn face features automatically instead of the conventional and fine engineered features. The advantage of deep learning is that it can handle with large training sets and it has achieved much success in many diverse domains, like vision, speech and language modeling [7]. And it is known that learning new representations [8], and especially hierarchical representations is the promising approach in computer vision [9]. Specifically in face domain, our deep learning framework can get the further representation from raw images using a simple knowledge transfer method in our system. Our deep learning method is also an unsupervised training and learning process, so it is possible to learn the feature from the raw pixel values and produce a feature representation without combining any engineered descriptors. At last, we can use the feature representation learned from our deep learning framework to classify face or object and certify who it is. In our system, it is the autonomous learning process and there are no many parameters to tune unlike the traditional methods. Once the deep learning framework has been trained on a very large dataset, we can use this network to face recognition and object recognition and so on. In the offline training process of our deep learning network, we randomly sample 1 million subset from Tiny Images dataset [10] as auxiliary data. However, the reason why we have chosen the deep learning framework is that it is intelligent and effective to control the network and implement face recognition and the training time and recognition rate is shorter and faster than before methods. And the most important characteristic is that the recognition ratio is higher than traditional face recognition methods, like LPP [11], PCA [12], Sparse [13], NPE [14] and so on.

In this paper, we propose a novel deep learning face recognition algorithm. In our system, we use a stacked denoising autoencoder (SDAE) [15] to learn generic image features from the large Tiny images dataset [10] as auxiliary offline training data and then construct a five layers as a feature extractor to produce the image representation and an additional classification layer, which we can use to further tune the generic image features to adapt to the specific objects recognition by online training the corresponding objects. Moreover, through the process of our deep learning framework, the image representations obtained are more expressive and intelligent than those traditional methods based on PCA [12] which must be chosen several principal components. After several experiment and comparison with other methods, we can get to conclusion that the proposed deep learning algorithm is significantly more efficient and suitable for the real-time application of face recognition based on big data.

1.1. The main face recognition methods

Dimensionality reduction is a main idea for face recognition in many classical recognition algorithms and its goal is to discover the hidden structure from raw data. Here, we simply describe several famous and classical face recognition algorithms. Principal Component Analysis (PCA) [12] is choosing the principal components of raw images to reduce their dimensionality and preserve the global Euclidean structure. Besides, Neighborhood Preserving Embedding (NPE) [14], Linear Discriminant Analysis (LDA) [16] also. These approaches are classical and easy to implement and exploit popularly. Unfortunately, they fail to discover the underlying nonlinear structure using this traditional linear method. Later, more and more nonlinear techniques have been proposed, such as Locally Linear Embedding (LLE) [17], Laplacian Eigenmap (LE) [18]. These nonlinear methods successfully preserve local structures in small neighborhoods among training data, but they cannot show the explicit maps on new testing data points for recognition problems. To overcome the drawback from nonlinear methods, a landmark method based manifold learning, named Locality Preserving Projection (LPP), have been proposed by He et al. [11]. LPP can approximate eigenfunctions of the Laplace-Beltrami operator on the manifold and the new testing points can be mapped to the learned subspace without trouble. However, LPP can accurately derive local structures among the neighborhood images and be incapable of extracting the intrinsic feature structure of the raw image. With the development of sparse representation, John Wright et al. [13] proposed a new theory from sparse signal representation to address face recognition problem. And this method show that if sparsity in the recognition problem is properly harnessed, the choice of feature is no longer critical.

So the key point of this approach is that the sparse representation is correctly computed and it needs to be artificially set, not automatically. In the past, neural networks are a good method to reduce dimensionality and automatically extract features from raw images, see [19]. However, there are many parameters to tune on training neural network and it maybe lead to produce error due to overfitting problem. Besides, these neural networks cannot effectively train the large of raw images, which causes decline of the accuracy of feature extraction along with the increase of the training data. And the usual neural networks are not high-performance enough to effectively handle the long run time from the training the network process.

In particular, these methods possess three failings in this visual angle: (I) these features produced by them are shallow in the sense because they cannot extract more depth features from the raw images; (II) for good performance these methods must incorporate hand-engineered feature, which maybe bring unexpected human errors in this procedure; (III) they cannot automatically extract useful features from images without any human help and they will have difficulty in dealing with large data.

1.2. The innovation of our algorithm

For Big Data applications, the most fundamental challenge is to explore large number of data and extract useful information for the future actions [20]. We firstly apply deep learning method to face recognition and here are some key differences which are worth noting, comparing to classical and popular face recognition methods.

(1) We can learn and extract the generic image features from a more general and larger dataset rather than a smaller dataset constructed by only some chosen image classes. Also, this alleviates the issue that much of the data are no label in the actual application. So for the big data problem, our method is best choice for research and analysis.

(2) The most highlight of our method is its autonomy. In other words, we can learn image features and extract useful information from the original images automatically instead of using hand-engineered features.

(3) After offline training process of our method, further feature learning is allowed in order to tune and adapt better to the specific face database being recognized during the online face recognition process.

2. Stacked Denoising Autoencoder

In 2008, Bengio [21] proposed Denoising Autoencoder (DAE) method that can add random noise to the data input layer in the network to get more robust features. As we all know, Denoising Autoencoder is the basis of the Stacked Denoising Autoencoder. DAE is a one-layer neural network and is a more recent variant of the conventional AE. The most characteristic of DAE is that it can recover a data sample from its corrupted version and get robust features since this neural network has a hidden layer with fewer units than the input units. The random noise is Gaussian noise or salt-and-pepper noise. As we can see from Fig. 2a, DAE can encode the data X' which is the corrupted vision of X to the hidden units Y and then decode the hidden layer to the approximative original data sample Z. To make the same between the data sample Z and the original data sample X, DAE can come true by solving the following optimization formula:

$$\min_{W,W',b,b'} \sum_{(i=1)}^{k} \| X - Z \|_2^2 + \lambda\left(\| W \|_F^2 + \| W' \|_F^2\right), \quad (1)$$

where

$$Y = f(WX + b), \quad (2)$$

$$Z = f(W'Y + b'), \quad (3)$$

Here, there are k training data samples, and W, W' respectively denote the weights for encoder and decoder. Besides, $f(\cdot)$ is a typically nonlinear logistic sigmoid function and b, b' denote the bias terms. Furthermore, the parameter λ is used to balance the reconstruction loss and weight penalty terms, and the symbol $\|\cdot\|_F$ refer to the Frobenius norm. In a word, we can learn from the network architecture of DAE that the hidden layer Y can represent the original data sample X and retain a significant amount of information about it, and Y can be called the robust feature of the original data sample by the autonomic learning way. At the same time, DAE is like the vision system of human beings, which can recognize this object even if a part of the object is occluded. Further, the network architecture of DAE can better handle with object occlusion and illumination problems than AE.

As we can see from the DAE′s features, they are trained locally to denoise corrupted versions of their inputs and get higher level feature representation. Based on stacking layers of DAE, we explore and use an original strategy for building deep layered networks. This resulting algorithm has a better performance and gets more perfect image features than stacking of ordinary autoencoders. Furthermore, using the denoising criterion and stacking layers can guide the learning of useful higher level representations and get more abstract and compact image feature representations learnt in this purely unsupervised fashion. The reason why the stacked denoising autoencoders (SDAE) [15] work much better is that it can initially use a local unsupervised criterion to pretrain each layer in turn and learn to produce a useful higher level representation from the lower level representation output by the previous layer. After this pretraining phase, the SDAE can be regarded as a feedforward neural network and the whole network is fine-tuned using the classical backpropagation algorithm. Besides, conjugate gradient decent method is applied into it to increase the convergence rate. The complete procedure for learning the generic feature of stacked denoising autoencoder is shown in Fig. 2b.

In a word, we are looking for unsupervised and autonomous learning principles likely to lead to the learning of feature detectors that detect important structure in the input images. What a coincidence, the SDAE algorithm is able to learn Gabor-like edge detectors from natural image patches and larger stroke detectors from digit images. At last, in each individual layer of the SDAE we use the DAE algorithm to detect the key image structure and extract the useful higher level feature representation. Using this deep layered networks of the SDAE, we can achieve robust features and superior performance in extracting the image feature problem.

3. Deep learning framework

3.1. Deep learning

Hand-designed features such as PCA [12], LPP [11], SIFT [22] and HOG [23] are applied to many successful object recognition approaches. However, these only capture low-level feature information and it has proven to be difficult to effectively design and capture the mid-level feature and high-level feature information for images [24]. But deep learning method, known as a new area of machine learning research, have presented how hierarchies of features can be learned in an unsupervised manner directly from original data. Furthermore, the traditional recognition performance depends on how we choose the hand-designed features and the way of hand-designed feature is a kind of difficult and heuristic methods

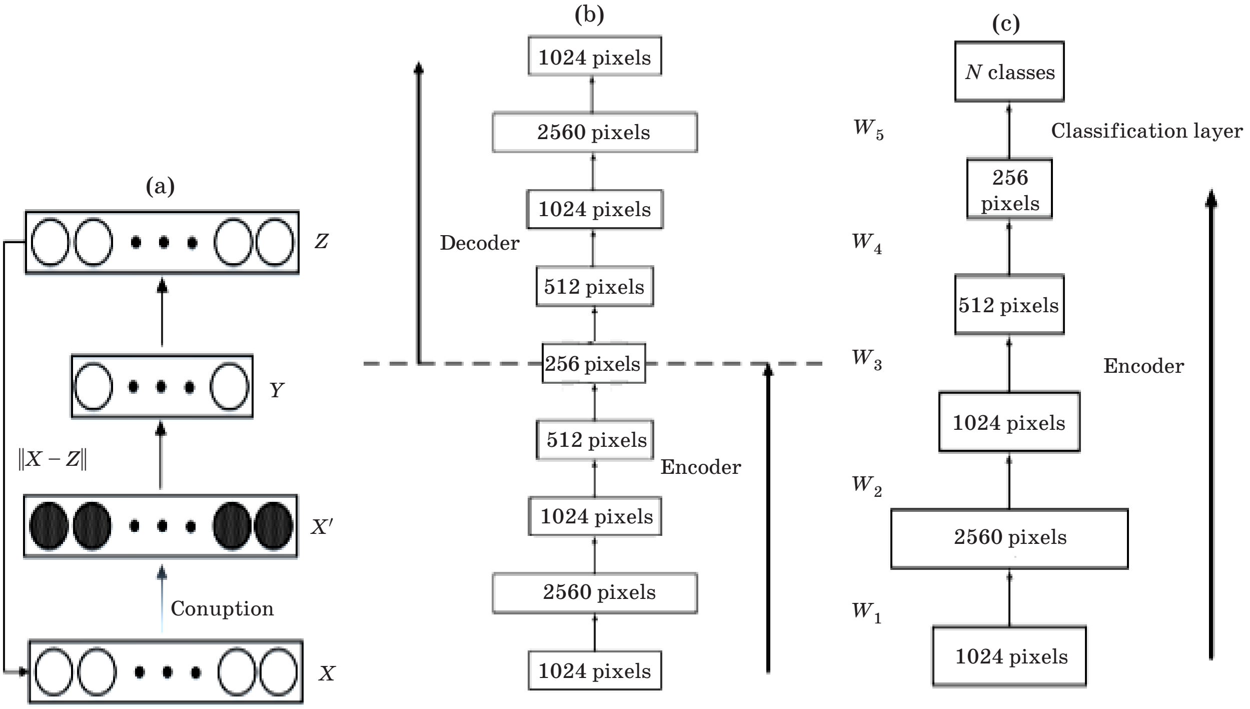


Fig. 2. The architectures of several networks: (a) denoising autoencoder, (b) stacked denoising autoencoder, (c) network for supervised face recognition with N that is the number of classes considered in this recognition task.

by experience and luck from human, which cannot be defined by the way of automatically learning the features from original images. However, deep learning method can extract the image features automatically and model high-level abstractions in data by using architecture composed of multiple non-linear transformation [25]. Now we will explain what the deep learning is.

We assume that there is a network S, which consists of n layers ($S1$, $S2$, ..., Sn), and its first layer receives input images I and its last layer produces the output O. So the total system can be picturesquely represented as I => $S1$ => $S2$ => ... => Sn => O. Through this system, if the output O is almost equal to input I, we can get the conclusion that the representation Si from each layer can be regarded as other representations of the input images I. By tuning the parameters from the system S to make the equality between input I and output O, we can automatically get a range of layered feature representations of the input data I, which is respectively $S1$, $S2$, ..., Sn. For the deep learning, its idea is stacking several layers in the network, whose each level learns a compressed representation of the observations that is fed to the next level. By this way, hierarchical representations of the input data can come true and further extract the mid-level or high-level feature information from the original images.

Deep learning algorithms are based on distributed representations, a concept used in machine learning [26]. Different concepts are learned from other concepts, with the more abstract, higher level concepts being learned from the lower level ones. To model this idea, many architectures are often constructed with a greedy layer-by-layer method. All in all, deep learning can help to disentangle these abstractions and pick out which features are useful for learning.

3.2. The implementation process

Now we present our deep learning face recognition framework. First, we train the SDAE by using the large Tiny images dataset [10] as auxiliary offline training dataset to learn the generic natural image features, which is called the unsupervised feature learning process in our deep learning framework. After the offline training stage, we can get the network weight W for the encoder. Then with the labeled face dataset, we can continue to training the feature extractor which is constructed from the encoder part of the trained SDAE by the layer-by-layer way and fine-tuning the network weight W and the whole deep learning framework. During this stage, we add an additional classification layer to the encoder part of the trained SDAE and train this layer to classify and distinguish the object using the testing data from the rest of the labeled face dataset. More details are obtained from the other parts of this section.

3.2.1. Dataset and preprocessing

We choose the 80 Million Tiny Images database [10] as auxiliary data to offline train our deep learning framework. We use Wordnet [27] to extract 75,062 nonabstract nouns, then collect images for these nouns which have dense coverage of all visual forms. This large images are freely available online and collected from the Websites that are seven independent image search engines: Altavista, Ask, Flickr, Cydral, Google, Picsearch, and Webshots. At last, downloading all the images provided by each engine for all these nonabstract nouns, we use these images to offline train our deep learning framework. Here, for suitable our unsupervised feature learning procedure, we uniform all the images and reshape the size of every image to 32×32 pixels. We can download the large Tiny images dataset from http://people.csail.mit.edu/torralba/tinyimages.

Consequently, whether this is an unlabeled dataset or other labeled datasets in our implementation process, we train our deep learning face recognition system with a vector 1024 dimensions corresponding to 1024 pixels and each dimension is linearly scaled to the range [0, 1] to further facilitate the computation of our framework.

3.2.2. The unsupervised feature learning procedure

The offline training stage aims at learning generic image features with a stacked denoising autoencoder by the unsupervised learning way. As we all know, a DAE is the basic building block of SDAE, and has a very important characteristic that it can recover a data sample from its corrupted version and get robust features. By reconstructing the input data sample from its corrupted version, the DAE is more effective than the conventional autoencoder in discovering more robust features. Further, the SDAE can handle with object occlusion and illumination problems in the face recognition research process. In this case, we train the network architecture of SDAE by using large Tiny images dataset as auxiliary offline training dataset.

The whole structure of the SDAE is depicted in Fig. 2b, as we construct the SDAE architecture whose encoder and decoder parts consist five layers respectively. To capture the image structure better, we further speed up pre-training in the first layer by dividing each 32×32 tiny image into five 16×16 patches, which is respective upper left, upper right, lower left, lower right, and the center one which overlaps with the other four. Then we respectively train five DAEs each of which has 512 hidden units. After that, we use the trained weights of the five small DAEs to initialize a large DAE and then train the large DAE normally [28]. So, our SDAE model has 2560 hidden units of the second layer totally. The next step is that the number of units is reduced by half whenever a new layer is added until there are only 256 hidden units, serving as the bottleneck of the autoencoder. Here, we should point out that robust features are learned since the neural network contains a “bottleneck” which is a hidden layer with fewer units than the input units.

Using unsupervised and autonomous learning way, the SDAE algorithm can learn feature detectors from natural image patches and detect some important and useful structure. This is in line with the neurophysiological mechanism in the area $V1$ visual cortex from the Fig. 1. After this unsupervised feature learning procedure, we can get the feature detectors from natural image patches in the Fig. 3a. After trained on the 80 Million Tiny Images database, the SDAE can learn the feature detectors from the first hidden layer of it and the feature detectors from this layer are straightforward to visualize. Each hidden neuron $a\{i\}$ has an associated vector of weights $W\{i\}$ which has the same dimensionality as the input so that it uses to compute a dot product with an input example. And these $W\{i\}$ vectors, called the filters, can be displayed as little images showing what aspects of the input each hidden neuron is sensitive to.

3.2.3. The supervised face recognition procedure

By using auxiliary natural images on the offline training stage, we train a stacked denoising autoencoder model to learn generic image features that are more robust against variations. This is then followed by knowledge transfer from offline training to the online object recognizing process. Specially, face recognition is an important part of object recognition in visual research area. On the online and supervised face recognition procedure, our deep learning recognition networks involve a feature extractor which is constructed from the encoder part of the SDAE and an additional classification layer that is added on the feature extractor part. Here, we choose the logistic sigmoid function [29] to classify the result. Both the feature extractor and the classifier can

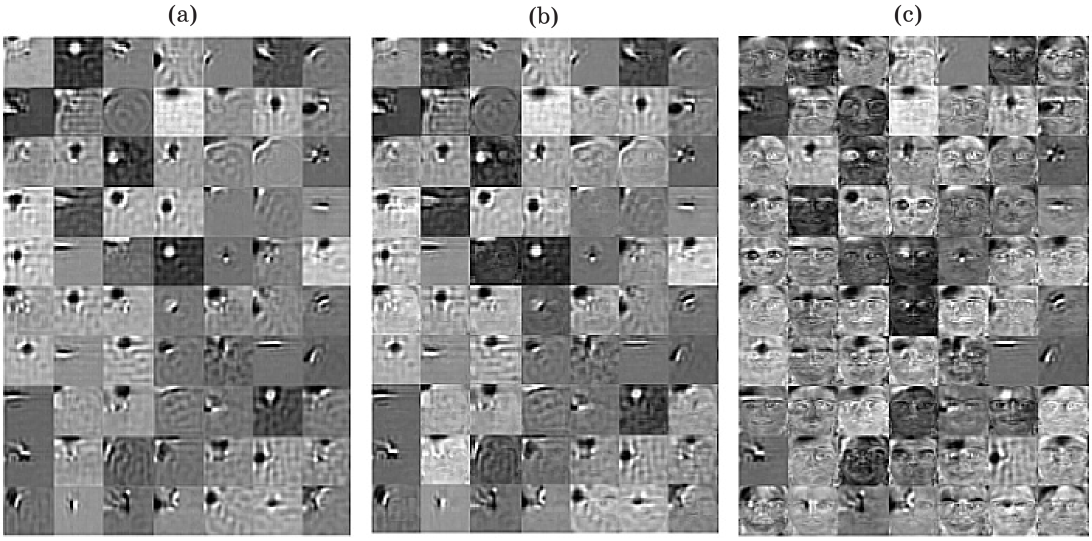


Fig. 3. Feature detectors. (a) is the general feature detectors that correspond to the first hidden layer of the SDAE trained on the 80 Million Tiny Images database. (b) shows the feature detectors after 80 times training on the AR database. (c) is the face feature detectors which looks like face structure from the AR database after 500 times training on it.

be further tuned to adapt to the changes of the recognized object.

In this way, the process of the supervised face recognition is described below. Firstly, we choose the database which is used to recognize the object. Here we choose the several face databases in our experiments. Secondly, we use the network weight W learnt from the generic feature detectors to extract the useful features of the face images. After supervised training on the face database, we can further fine-tune the deep learning network weight W. Here, we can get the feature detectors from the identified face database in the Fig. 3b, c. Thirdly, like the structure from the Fig. 2c, the lower level features make up of the higher level feature and the feature is more and more abstract from lower level to higher level. When the level is up to the top layer, the feature is apparent to recognize the faces like our brain in recognizing the object process in Fig. 1. Fourthly, we use the identified network weight W and the additional classification on the testing data from face database to define and recognize the testing faces.

Based on this supervised face recognition process, our deep learning face recognition algorithm can be more easy and accurate in classifying and distinguishing the testing face. All in all, the key to success using our deep learning architecture is to learn fully richer invariant features in each hidden layer via multiple nonlinear transformations. At last, the entire procedure of our method is organized as in Algorithm 1.

Algorithm 1 Deep Learning Face Recognition

Offline pretraining: Using 80 Million Tiny Images to offline train a five-layer stacked denoising autoencoder (SDAE) in order to unsupervised learn generic image features $W\{i\}$.

Online recognition:

(1) Select a recognition database for the experiment.

(2) Randomly choose a part of the database as the training data, and the rest data as the testing data.

(3) Build the five-layer back-propagation (BP) structure and use the training data and non-linear sigmoidal transformation function sigm() to calculate the hidden layer feature representation $a\{i\}$ and fine-tune the weights $W\{i\}$. Specific calculation formula is as follows:

$$a\{1\}=X,\ X \text{ is the input data (the training data)}, \tag{4}$$

$$\mathrm{a}\{i\}=\mathrm{sigm}\left(a\{i-1\}W\{i-1\}'\right),\ i=2,3,4,5,6, \tag{5}$$

$$e=y-a\{n\},\ d\{n\}=-e.*\left(a\{n\}.*\left(1-a\{n\}\right)\right),\ y \text{ is the label value of } X, \tag{6}$$

$$d\{i\}=\left(d\{i+1\}*W\{i\}\right).*\left(a\{n\}.*\left(1-a\{n\}\right)\right),\ i=5,4,3,2;\ n=6, \tag{7}$$

$$\Delta W\{i\}=d\{i+1\}'*a\{i\},\ W\{i\}=W\{i\}-\Delta W\{i\},\ i=1,2,3,4,5. \tag{8}$$

(4) Use the testing data and updated weights $W\{i\}$ to test our trained deep learning network and utilize the sigmoid classification layer to realize the face recognition.

```
For $i = 1 : 6$
    If $i == 1$: $a\{i\} = X$, $X$ is the original testing data;
    Else: $a\{i\} = \mathrm{sigm}(a\{i-1\} * W\{i-1\}')$;
    End If
End For
$T = \mathrm{Label}[\max(a\{n\})]$, $n = 6$, $t$ is the predicted class of testing data $X$;
If $y == t$: the recognition is right;
Else: the recognition is wrong;
End If
```

(5) Count the number of right recognition of all the testing data and calculate the recognition ratio of this face database.

Ratio = Num/N

Variable Ratio is the recognition ratio,
Num denotes the number of right recognition,
N is the number of testing data.

4. Experiments

In this section, as the most important part of the paper, we will examine the current performance obtained by our proposed deep learning approach on a range of standard vision experiments compared with other face recognition methods. By these qualitative experiments comparing to the famous face recognition methods on publicly available databases, we will demonstrate the efficiency of our algorithm and validate the illustrations of the previous sections. Then, we will first simply introduce the comparing algorithms and several face databases and why we choose them to compare and experiment. Next, we will demonstrate the robustness of our proposed method to expressions, occlusion, illumination and pose, comparing to the popular face recognition algorithms on different common databases. In particular, we emphasize the characteristic of our deep learning architecture on the autonomic feature learning than other methods with manually looking for best recognition ratio on every dimension.

4.1. Several popular face recognition methods as contrast tests

As the most popular subspace method, the PCA approach [12] is also known as eigenface method and as the representative of traditional linear methods. Recently, LPP [11] has been given more attention in face recognition, the reason why it is so famous is that it can find a best embedding and subspace by preserving local information and detecting the essential face manifold structure. So LPP is usually called a landmark method based manifold learning. NPE [14] is less sensitive to outliers and defined using in everywhere not only on the training data by preserving the local neighborhood structure method. At present, face recognition via Sparse representation [13] is a hot research point in object recognition and also a classical classification algorithm for image recognition, which shows that if sparsity during the recognition process is properly harnessed, it is no longer critical of the choice of features.

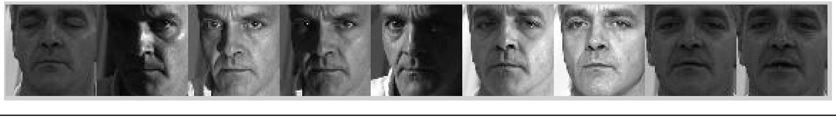
4.2. Several face databases used in our experiments

In our experiments, we choose Yale, AR and PIE databases to test our proposed deep learning face recognition algorithm with the several comparing methods together. The reason why we have chosen them on experiments is that these common face databases contain the crucial cases: expression, occlusion, illumination and corruption. First, Yale database contains 165 images, coming from 15 individuals and 11 images of each person, whose size of each image is 100×100. Second, there are 100 people (50 males and 50 females) in AR database, and each person has 26 face images with the size of 120×165. Here, because there are many similar images of each person, so we choose 8 images (1th, 2th, 3th, 5th, 8th, 9th, 11th, 12th) of each person, totally 800 images used in our experiments on AR database. Third, PIE database, named by Pose, Illumination, and Expression, consists of 68 individuals. For each individual, there are 49 images produced by various disturbances and its size is 64×64. In our experiments, we carefully choose a subset of PIE database, which is that 68 individuals and 9 images (1th, 4th, 9th, 12th, 19th, 26th, 43th, 48th, 49th) of each individual.

One thing we should declare is that the original AR or PIE databases have so many images of every individual so that it's easy to get good performance and recognition ratio for most face recognition methods. In order to increase the difficulty for recognition and feature extraction, we select several varied images of every individual which are used to train and test these famous face recognition methods. Moreover, these selected images of every person are interferential by the varying expression and occlusion, as well as illumination and corruption. Besides, in order to improve the reliability and accuracy of our proposed face recognition system, the training set we choose should not contain several similar images of every individual.

For example, we can see all the faces of one person from every database in Table.1, which we choose to perform these experiments.

Table 1. Faces of one person from every database.

Num	Database	All the faces of one person
a	Yale	
b	AR	
c	PIE	

3.3. Experiments on these face databases

We perform several experiments on Yale, AR and PIE databases, respectively. In one database, there are also some groups of experimental and each group is repeated 20 times to get mean of the values. And the important point is that we randomly select the training data each time and the rest data as the testing data. At the same time, there is no overlap between training data and testing data of every individual each time. Here, we use *Gm*/*Pn* to denote that there are m training data and n testing data of each individual. For the compared experiments with these popular face recognition methods, we use the original algorithm code provided by their authors and also perform 20 times for each group *Gm*/*Pn*.

3.3.1. On Yale database

From Table 1(a), we can see that there are various interferences to the face images under the conditions of lighting angles, facial expressions and wearing glasses. Among them, the various lighting angles contain left-light, center-light and right-light; whether wearing glasses or not maybe seen as an interference; in particular, the most primary disturbance is the facial expressions, including normal, happy, sleepy, surprised, sad and wink.

After that, aiming at these interferences from Yale database, we use our deep learning method to test the effect with these popular face recognition methods, Table 2.

From Table 2, qualitative experiments show that,

(1) Our proposed method and Sparse representation method obviously outperform other classical methods on Yale database. Besides, the first and second methods are robust to various face expressions and are easy to distinguish which facial expressions are coming from the same person.

(2) Look from whole, with the increase of the training data, the mean recognition ratio of every method has improved in some extents. However, it should be noted for PCA, NPE and LPP methods, there is such a phenomenon that when the training data is increased to a certain degree, the recognition ratio decreases. The strange phenomenon means that, after training many images from the one individual, these methods will be wrong to think that a new testing image from the same individual is another category and it appears the issue of overtraining. Besides, it also shows that these classical methods cannot handle big data problem.

(3) From our proposed method and sparse method, we can find that when the training data is less, the result of sparse is slightly better than ours. However, with the increase of training data, our method is getting better and better. In particular, at the age of big data, the research of big data is more and more hot and it is difficult to handle big data, but our deep learning method can just right explore and analyze the big data and extract the useful information or feature from the big data.

3.3.2. On AR database

This is a difficult experiment for these methods to recognize the face who it is. Here, we use 100 individuals and choose 8 faces per person. We all know that if the number of categories is far less than the number of samples of each class, the recognition task becomes very easy. But we don′t talk about this problem, we will emphasize the key point on the issue that the class number is far greater than the number of samples of each class. Besides, this new AR database we choose is seriously affected by the illumination and occlusion, which we can see from Table 2(b). In a word, this experiment is very valuable to research and summarize.

Table 2. Comparison of 5 face recognition methods on Yale database. The value in the table represents the mean recognition ratio (%) and the red color denotes the best recognition result of each group, the blue color shows the second best

Yale	*G*2/*P*9	*G*3/*P*8	*G*4/*P*7	*G*5/*P*6	*G*6/*P*5	*G*7/*P*4	*G*8/*P*3	*G*9/*P*2	*G*10/*P*1
Ours	75.19	81.75	85.81	88.89	90.80	92.33	93.78	95.67	96.67
Sparse	76.96	84.50	85.05	88.61	89.60	90.67	91.33	91.83	95.00
PCA	65.37	72.00	74.29	75.72	76.53	75.75	77.67	75.83	80.00
NPE	63.48	71.50	74.76	75.83	77.53	77.75	78.78	76.00	79.00
LPP	58.70	70.92	74.24	76.56	78.20	78.50	79.44	78.00	80.00

Fig. 4 shows the recognition ratio of these popular methods as the increase of the training data. More details we can get are that: First, when the number of categories is far greater than the number of samples of each class, our method is the best one among them, also, Sparse and NPE methods are the second one and third one. Second, for the experiment data by the intense illumination and occlusion effect, our method can better handle these disturbances than others. At last, our method is unsupervised learning algorithm and other methods are mostly supervised and need to waste time finding the best performance from every dimensionality which original images are mapped into. However, our method does not need to traverse each dimensionality and then reach the accurate result, which shows that our method has a lower computation cost. In addition, the most highlight of our method comparing with others is that it can learn the image features from the original images automatically instead of using the hand-engineered features. Besides, we supply the code interface of GPU to users and so it can provide the computing time more quickly.

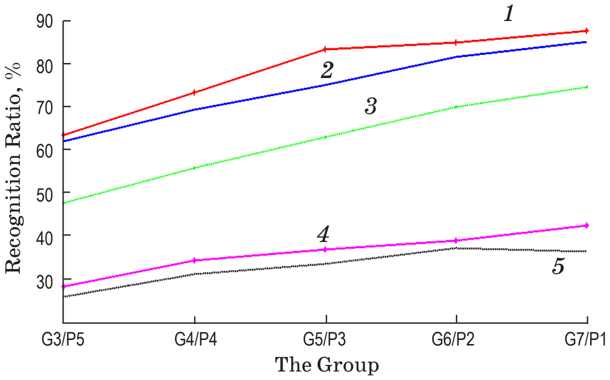


Fig. 4. The recognition ratio of every method on AR database. *1* – Ours, *2* – Sparse, *3* – NPE, *4* – LPP, *5* – PCA.

4.3.3. On PIE database

From Table 1(c), one could argue that the PIE database is produced by the interference coming from the severe illumination, expression and pose. Furthermore, we can find that the left half of 2th face is dark and no light, like the left half is shaded. And the right half of 5th face is occlusion by dark. Other faces of one person are also apparent in light intensity and pose and slight expression.

In Fig. 5, we show the recognition ratio of every method on PIE database corresponding to the six intervals of the number of training data and testing data *Gm*/*Pn*. It's obvious that as the increasement of the ratio of training data, the recognition results are more and more better using every method. For the treatment of serious illumination problem, our proposed method and Sparse method are more effective than NPE, LPP and PCA methods. However, our method is inherently based on extracting the most useful feature for data representation and recognition, and Sparse is not. On the whole, our method is more robust than Sparse method. Besides, like the PCA and LPP methods as the traditional and classical methods, when the database is suffered from serious illumination and occlusion and etc, the effect of recognition is very poor, comparing the Table 2, Fig. 4 and Fig. 5.

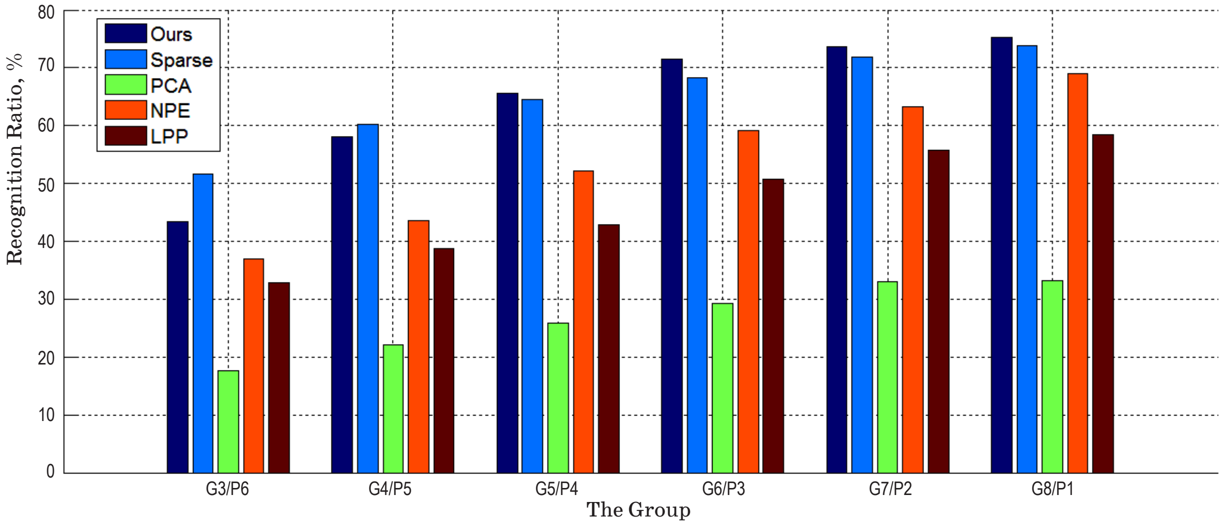


Fig. 5. A summary of the performance of these methods for which we have results using the PIE database.

5. Conclusion

In this paper, we have proposed a novel deep learning algorithm as a predictor using in face recognition and achieved superior performance than other popular face recognition methods at present. First of all, by analyzing and imitating how human brain works, we propose our deep learning thought to design this deep learning face recognition architecture. Next, based on traditional denoising autoencoder (DAE), we first construct the multilayered stacked denoising autoencoder (SDAE) as ourselves offline unsupervised training model, which aims at learning generic image features using the larger auxiliary natural images. After offline training stage, by adding the logistic regression layer to the encoder part of the proposed SDAE model, we then train a classification neural network to further extract the specific feature during the online face recognition process. This is then followed by knowledge transfer from offline training to the online object recognizing process. At last, Comparison with the state-of-the-art face recognition methods on several common face datasets, these experiments show that our deep learning algorithm in face recognition is more accurate and low computational cost and is a perfect processing tool in big data problem. Moreover, our method is more robust in handling with the external interference, including occlusion, illumination, facial expressions, pose variance and so on. Besides, we supply the code interface of GPU to users and so it can provide the computing time more quickly.

As the basic visual work on applying deep learning to the face recognition, there are so many opportunities for further research. For example, the classification layer in our proposed method is just a linear and simple classifier and if extending it to more excellent classifiers, the performance of deep learning in face recognition will become better in the future. Moreover, this approach of using the deep learning architecture as a face recognition system also could be applicable to other problems, such as building recognition, traffic sign recognition or speech recognition. On the other hand, implementing the multi-layered features in a camera APP on a smartphone is another possible application.

Acknowledgments

This work has been supported by Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No.20120061110045) and supported by the Project of Science and Technology Development Plan of Jilin Province, China (Grant No. 20150204007GX).

* * * * *

ЛИТЕРАТУРА

1. IBM What Is Big Data: Bring Big Data to the Enterprise, http://www-01.ibm.com/software/in/data/bigdata/. IBM. 2012.
2. *Michel F.* How Many Photos Are Uploaded to Flickr Every Day and Month? http://www. flickr.com/photos/franckmichel/6855169886/. 2012.
3. *Wu X., Zhu X., Wu G. Q., Ding W.* Data mining with big data // Knowledge and Data Engineering. IEEE Transactions on. 2014. V. 26. № 1. P. 97–107.
4. *Mervis J.* U.S. Science Policy: Agencies Rally to Tackle Big Data // Science. 2012. V. 336. № 6077. P. 22.
5. *Labrinidis A., Jagadish H.* Challenges and Opportunities with Big Data // Proc. VLDB Endowment. 2012. V. 5. № 12. P. 2032–2033.
6. *Benayed S., Eltaher M., Lee J.* Developing Kinect-like Motion Detection System using Canny Edge Detector // American Journal of Computing Research Repository. 2014. V. 2. № 2. P. 28–32.
7. *Hinton G., Deng L., Yu D., Dahl G., Mohamed A., Jaitly N., Senior A., Vanhoucke V., Nguyen P., Sainath T., Kingsbury B.* Deep neural networks for acoustic modeling in speech recognition // IEEE Signal Processing Magazine. 2012. V. 29. № 6. P. 82–97.
8. *Potapov A.S., Malyshev I.A., Puysha A.E., Averkin A.N.* New paradigm of learnable computer vision algorithms based on the representational MDL principle // Proc. SPIE. 2010. V. 7696. P. 769606.
9. *Potapov A.S.* Theoretico-informational approach to the introduction of feedback into multilevel machine-vision systems // JOT. 2007. V. 74. № 10. P. 694–699.

10. *Torralba A., Fergus R., Freeman W.T.* 80 Million Tiny Images: a Large Data Set for Nonparametric Object and Scene Recognition // PAMI. IEEE Transactions on. 2008. V. 30. № 11. P. 1958–1970.

11. *Niyogi X.* Locality Preserving Projections // Neural Inform. Proc. Systems. 2004. V. 16. P. 153.

12. *Averkin A., Potapov A.* Learning Representative Features for Facial Images Based on a Modified Principal Component Analysis // Proc. AIP Conf. 2013. V. 1537. P. 76–84.

13. *Wright J., Yang A. Y., Ganesh A., Sastry S. S., Ma Y.* Robust Face Recognition via Sparse Representation // PAMI. IEEE Transactions on. 2009. V. 31. № 2. P. 210–227.

14. *He X., Cai D., Yan S., Zhang H.J.* Neighborhood Preserving Embedding // Computer Vision. 2005. ICCV 2005. Tenth IEEE Intern. Conf. on. IEEE. 2005. V. 2. P. 1208–1213.

15. *Vincent P., Larochelle H., Lajoie I., Bengio Y., Manzagol P.A.* Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion // The Journal of Machine Learning Research. 2010. V. 11. P. 3371–3408.

16. *Belhumeur P., Hepanha J., Kriegman D.* Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection // IEEE Trans. PAMI. 1997. V. 19. № 7. P. 711 720.

17. *Saul L.K., Roweis S.T.* Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds // The Journal of Machine Learning Research. 2003. V. 4. P. 119–155.

18. *Belkin M., Niyogi P.* Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering // NIPS. 2001. V. 14. P. 585–591.

19. *Bouattour H., Fogelman-Soulie F., Viennet E.* Solving the Human Face Recognition Task Using Neural Nets // Artificial Neural Networks. 1992. V. 2. P. 1595–1598.

20. *Rajaraman A., Ullman J.* Mining of Massive Data Sets. Cambridge: Univ. Press, 2011.

21. *Vincent P., Larochelle H., Bengio Y., Manzago P.A.* Extracting and Composing Robust Features with Denoising Autoencoders // Proc. of the 25th Intern. Conf. on Machine Learning. ACM. 2008. P. 1096–1103.

22. *Lowe D.G.* Distinctive Image Features from Scale-Invariant Keypoints // International Journal of Computer Vision. 2004. V. 60. № 2. P. 91–110.

23. *Dalal N., Triggs B.* Histograms of Oriented Gradients for Human Detection // Computer Vision and Pattern Recognition.2005. CVPR 2005. IEEE Computer Society Conference on. IEEE. 2005. V. 1. P. 886–893.

24. Deep Learning Methods for Vision, http://cs.nyu.edu/~fergus/tutorials/deep_learning_cvpr12/. CVPR. 2012.

25. *Bengio Y., Courville A., Vincent P.* Representation Learning: a Review and New Perspectives // IEEE Trans. PAMI. Special Issue Learning Deep Architectures. 2013.

26. *Hinton G. E.* Learning multiple layers of representation // Trends in Cognitive Sciences. 2007. V. 11. № 10. P. 428–434.

27. *Miller G., Fellb*aum C. Wordnet: an Electronic Lexical Database, http://www.cogsci. princeton. edu/wn. 1998.

28. *Wang N., Yeung D.Y.* Learning a Deep Compact Image Representation for Visual Tracking // Advances in Neural Information Processing Systems. 2013. P. 809–817.

29. *Zadeh M.R., Amin S., Khalili D., Singh V.P.* Daily Outflow Prediction by Multi Layer Perceptron with Logistic Sigmoid and Tangent Sigmoid Activation Functions // Water Resources Management. 2010. V. 24. № 11. P. 2673–2688.