International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.1 (2016), pp.87-96
http://dx.doi.org/10.14257/ijmue.2016.11.1.09

Image Retrieval Based on Deep Belief Networks

Sun Ting and Qi Yingchun

School of computer science and technology, Zhoukou Normal University, Zhoukou, Henan,466001, China
sunting@zknu.edu.cn

Abstract

According to the local and global feature of image, matching the image from a lot image library, this is the image retrieval task; however, the image retrieval need to search the information in the database, we need to find a method for efficient information retrieval. Deep belief network according to the characteristic of the initiative, through the method of training a multilayer neural network to process large amounts of data, and it is very efficient, in this article, as to the characteristics of image local features and global features, it gives a deep belief network image retrieval algorithm, the experiment verify the effectiveness of the algorithm.

***Keywords**: Deep Belief Network, Image Retrieval, Local Feature, Feature Extraction*

1. Introduction

Due to the deep neural network algorithm has the performance that it can forecast and analyze big data fast, it has been widespread concern in academic circles. The deep neural network [1] is the method that automatic learning input features, and the learning characteristics can depict the essence of data and through the "layer by layer initialization" to overcome the difficulty in training, so it has been widely studied in both academia and industry, and becoming a research hotspot, so deep neural network is a method of learning complex hierarchical probabilistic model, it has been widely applied in various fields. It has been successfully applied to the field of speech recognition [2], handwriting recognition, [3], traffic sign recognition, face recognition and other image processing fields, showing the superior performance of learning.

The deep neural network includes the deep belief network [4] (DBNs), automatic coding machine, deep convolutional neural network (CNNs) structure model. In these models, the most representative structures are deep belief network and deep convolutional neural network. The deep convolutional neural network is easy to fall into local optimal solution problem of the non-convex function; at the same time, convolutional neural networks cannot deal with unlabeled data problem. The deep belief network model uses the learning process of layered, because each layer is made of many restricted Boltzmann machine RBM stack structure, so it has been used widely.

Hinton proposed RBM fast learning algorithm - the contrastive divergence (CD) algorithm [5-6] proposed Monte Carlo PT algorithm, it instead of CD algorithm, Tijmen algorithm amended the defects of CD algorithm that cannot use maximum likelihood estimation, it proposed the PCD algorithm, Cho *et al.* proposed the improvement self-adaptive learning rate and increased the gradient estimation, is used to improve the performance of the PT algorithm and CD algorithm, and then improving the learning efficiency of RBM. [7] proposed RBM recommendation algorithm based on cloud platform, the RBM process is divided into several Hadoop duty cycle, realizing the parallel computing. In 2010-2012, many researchers have proposed the Markov Monte Carlo sampling (MCMC) algorithm based on temper to improve the learning effect of

RBM. In 2014, Shusen Zhou proposed based on DBN classification, each category structure to deal with the fuzzy membership function of the deep belief network algorithm FDBN. However, a limitation of the deep neural network is which the interpretability is not strong, like a "black box", we do not know why it can achieve such good results literature [8] made high-level visual feature of deep network to visualize, it resolved some doubts from the perspective of visual, but still cannot clearly explain from the view of mathematics. On the other hand, the deep neural network obtain the very strong expression ability by increasing the hidden layer number, according to the specific model, the criteria of determining layers has not more rigorous scientific methods.

The content-based image retrieval that is mainly divided into two modules: feature extraction module and query module. The content-based image retrieval involves computer vision, pattern recognition, image understanding technology; it has four characteristics, first, using the similarity matching method. In the text-based image retrieval, as we enter the keyword or keywords, and the image tag is text, the content of the picture is irrelevant, so it can use the way that ratio of contrast, this is an exact match. In content-based image retrieval , the images of the same scene may also be due to different angles or different light intensity and different ways of expression, and the image content is rich, a strong correlation between the characteristics of the data, and is usually not a simple relationship, so commonly used similarity matching method. Second, the query using the way of direct input picture. The image retrieval of content method using direct input image based query, if the user is not familiar with the specific structure you want to query image, it can be retrieved through the browse system provides an example. Some systems also can browse through the return of the results to determine the query result is good or bad, and supervise the system to make the necessary amendments. Third, interaction is strong. A feedback function can be achieved through user feedback on retrieval methods of continuous improvement and repeatedly modified retrieval results. Fourth, it can meet the needs of multi-level retrieval requirement. Content-based image retrieval system generally includes the feature database, image database, based on knowledge base, and can meet the requirements of different retrieval.

When the database is big, due to the limited computing resources and memory, exact linear search (sequence similarity of all images and the database to retrieve) is not feasible and not necessary. There are hundreds and thousands of image data are used to describe the characteristics, so the impact is often affected by the curse of dimensionality for content based image retrieval and the performance is not high. There is an urgent need for a scalable retrieval method. The use of similarity hash indexing is a promising method, but using a single hash table can't weigh the retrieval precision rate and recall rate of the problem. Use Doha and table method can obtain very high recall rate in constant search time but the retrieval accuracy is low, a large number of non-relevant samples are returned to the user caused low efficiency. How to improve the accuracy of the Greek Doha table method and maintains a high recall rate and retrieval time becomes the key to the successful application of Dohashi.

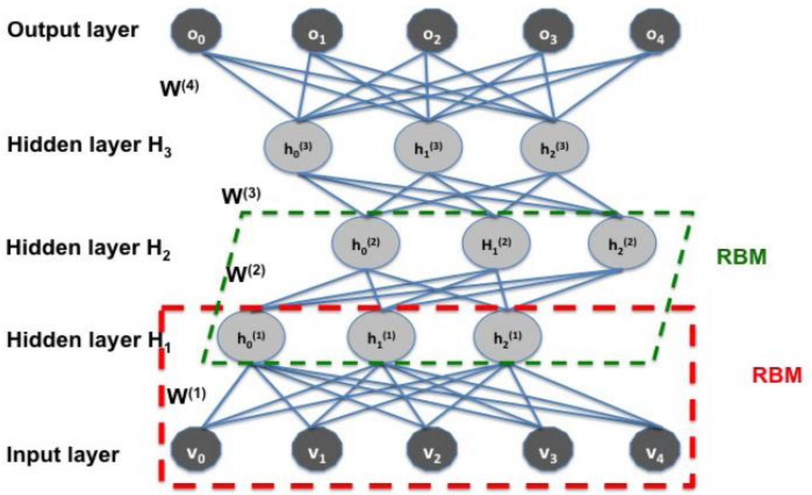


Figure 1. Deep Neural Network Structure

Image feature extraction is the key link of content-based image retrieval system, which directly affects the efficiency and the effectiveness of retrieval system. Image content features generally include the low-level visual features (color, texture and shape) and high-level semantic description of semantic and cognitive domains of human contact have strong subjective color. Compared with the high-level semantic features, not because of the consciousness change, property belonging to the image inherent, it has strong objectivity. In general, feature extraction of image refers to the image texture features, color features and shape features. The texture feature extraction methods commonly used are: gray level co-occurrence matrix, texture and Tamura wavelet transform; color feature extraction methods commonly used color histogram, color moment, color coherence vector, color correlograms and shape feature extraction method: there are two types of regional feature and contour feature, which are more representative of the direction of the border histogram, wavelet contour descriptor, Fourier descriptors, Hu invariant moments, the deep neural network technology to the application of image retrieval, the application of the ability of nonlinear mapping, self-learning ability and adaptive ability of neural network, so as the brain that oneself according to the image content to study the vision feature. Retrieval used feature is composed of two parts: one is the pre training network learn from image low-level features; the other part is the class information of the output of the model, the information of these categories are learned from labeled data, more in line with people's subjective description of image. So, in our method, the characteristics which retrieval uses include low-level features and high-level semantic features.

2. Related Works

2.1 The Research Status of Deep Neural Network

In 2006, Professor Hinton and other scholars proposed deep neural network in [9] is called a " deep belief networks", and gave the efficient learning algorithm for neural networks of the depth, this algorithm can not only improve the network training speed, but also avoids the local train stops at the minimum value of the problem. This algorithm is the main framework of deep learning algorithms so far. In this algorithm, a neural network model is regarded as the stack of restricted Boltzmann machine, the deep neural network learning process is by training the restricted Boltzmann machine. Because restricted Boltzmann machine can be quickly trained by Contrastive algorithm of

Divergence, the frame around the high complexity of the deep neural network training directly from on the whole, the whole network training is simplified to a restricted Boltzmann machine training problem. Hinton suggested, after training which layer by layer, we can through the traditional supervised learning algorithm for the network of global adjustment, so as to make the model converges to a local optimum. The basic idea of deep learning is based on the above mechanism, the use of multi-layer neural network to simulate the human brain to the outside signal processing. The common model of a convolutional neural network, automatic coding machine, sparse coding, restricted Boltzmann machine and so on. Training a neural network with two layers, the input signal is equal to the output signal as much as possible, in order to get the characteristics of the different expressed, this is the automatic coder. The sparse coding algorithm is a kind of unsupervised feature learning methods, it by finding a set of "over complete" base vector to efficiently represent data observed. Restricted Boltzmann machine is a probabilistic generative model, the structure that a two layer (including visible and hidden layers) neural network, the connection between layers in layer, fully connected, and the connection with the symmetry of the two directions, namely one side take the same value, the connection weights can be any real number. Each node is restricted Boltzmann machines in the state of the randomly selected two values: {0-1}, each node of the state from the probability of two values is decided by all the nodes and the connected weights. The deep neural network gets by multiple auto encoder superposition or many restricted Boltzmann machines. Compared with the traditional neural network, the deep neural network has an important breakthrough lies in the fact that, to a certain extent, it overcomes the traditional act of the efficiency and effectiveness of the network training. The deep learning strategies is the global multi model before learning, the neural network is divided into some two layer neural network, layer by layer the two layer of the neural network training, and then the superposition of the trained neural network with two layers of multilayer neural network to get the initial value, then fine tuning that using the global optimization algorithm such as BP algorithm.

2.2 The Research Status of Image Retrieval

Reflect the color feature in many different ways. Swain and Ballard use the color histogram, and it is the most common way, the color feature of the image color distribution is representative of the image. Stricker and Orengo proposed the total color histogram to solve the zero which the former appeared. In addition, they also made color moment method, the matrix of order statistic of each color component. Pass *et al.* proposed the concept of color vector image, in order to solve the phase image color distribution. According to the pixel correlation distance in the image frequency statistics, Huang *et al.* put forward the color feature correlation matrix method. There are two ways of shape feature description, which are represented by the outline of the image and character of a region in the image. Fourier operator has the advantages of scale and rotation invariance, Rui and others used it to describe a closed contour image. Inspired by this idea, the closed contour of image is described by the wavelet transform operator. Detecting out image edge, and contracting image edge gradient histogram, it can be described by the gradient histogram of the object shape, no matter how objects move in the image, as long as it does not change shape, histogram of gradient value is not changed, but it has the shortcomings that if the object rotated, which its shape does not change, the value of gradient histogram will be changed. Later, Freeman *et al.* that the object is composed of some line segment which the direction and length is fixed, so it can be used to describe object, so they proposed the chain code algorithm. Hu *et al.*, the shape of the object with a moment invariants to describe. The expression of texture feature description method with common statistical method, spectral method, structure method and structure method. Haralick *et al.* proposed a gray level co-occurrence matrix, it describes each pixel in the image distance and direction, it can simply get the entropy and contrast of the whole image statistical information related to the use of the matrix. Tamura provided 6 components of the Tamura texture set, on the texture characteristics of human visual perception. Good texture representation method and Gabor transform, Fourier transform and Gabor wavelet.

Although the query is a relatively simple process, but for large image database, image database is big, the efficiency and speed of retrieval is the key to the problem, although the image feature database for the original image data, the data volume is much smaller, but it still has the characteristics of these feature vectors the high dimension, such as color feature vector is 256 dimensional, the traditional search methods, its computational complexity is still huge, the retrieval efficiency and the speed may be unbearable and the data structure of the traditional database used by three angle inequality index cannot be well organized these high-dimensional feature vector, so the high dimensional vector dimensionality reduction and clustering, and on this basis to establish a database index efficiently to achieve fast and efficient retrieval is a very meaningful thing with database index method, it has positive theoretical significance and practical value.

The present research mainly focuses on the low-level visual features, semantic expression and understanding of the user is higher than that of low-level features, in order to improve intelligence of the search, we should research the semantic image high-level description based on the CBIR technology, therefore the expression of image semantic features and semantic content-based image retrieval will become one of the hot spot in the future. The international standards organization MPEG MPEG-7 standards, the goal is to achieve the content based image retrieval in the high-level semantic features and low-level visual features in one of the. However, there are many difficulties to realize image retrieval based on semantic content, it is difficult to transform the trend and "based on the development of content based image retrieval from low-level features to high-level semantic features is the traditional problem of computer vision research, but there is no breakthrough" how to achieve this transformation is still a the key problem in CBIR.

Image retrieval technology based region is a similar to retrieval technique which the human intelligence understand, it is the main object segmentation technique to extract image through the image, and then for each region using local feature to describe the general characteristics, each region can be described characteristics of the image, finally, using similar measurement standard suitable to image retrieval, image segmentation is a development is not mature and very difficult technology, it still cannot make the object region and image segmentation in the perfect match, so this kind of method of retrieval accuracy rate is not too high. Content based retrieval is a kind of approximate matching in the retrieval process, it uses the similarity matching of images in the image library, so as to obtain the results of the query, the retrieval of single feature, there are two possible image similarity measure is close to the difference in semantics, and the use of multiple features a comprehensive search for users, more flexible, more effectively express the query request; but the similarity measure has relation to people's subjective feelings, currently, similarity calculation basically is based on mathematics, it also has a certain gap between the characteristics of human visual perception.

3. The Proposed Scheme

In this paper, according to the structure of deep neural network algorithm, based on content, designing input data, the determination of the network layer ,node and adaptive learning algorithm for deep neural network, finally getting the image with executable retrieval scheme, according to the plan, we the effectiveness of the proposed scheme.

3.1 The Input of Deep Belief Network

In the deep belief network input, the input of the past, using direct input mode, however, in deep belief network, the training is difficult, so it needs to be improved. According to the structure of deep belief networks, from the perspective of information theory to the image features related to the calculation of information entropy, such as

$$H = -\int_{-\infty}^{\infty} f(x) \ln f(x) \tag{1}$$

Where, the information entropy can be color, stripe, gray and other global content, also can be partial content, according to the comparison of a number of global and local information, the information entropy feature selection as the starting characteristics, according to the starting characteristics, computing conditional probability:

$$H(x_1 | x_2) = -\int_{-\infty}^{\infty} f(x_1 | x_2) \ln f(x_1) \tag{2}$$

Where, the algorithm is suitable for the numerical data type, and for the conditional probability, the biggest conditional entropy should be correlation between two features is very small, which represents the information needed to learn is large. So, we can get more optimized input information. But it needs to determine the number of input features, we can get the number of input features according to (3)

$$\max \sum_{i,j=1}^{K} H(x_i | x_j) \tag{3}$$

3.2 The Determination of Learning Network Layer

According to the analysis of information theory, the proposed network layer which the information after learning obey the normal distribution is optimal, the optimal network layers to meet the learning data， and followed normal distribution, so it is necessary to determine the network layer based on the principle of. The learning model of the deep belief network is shown in Figure2, according to the normal distribution, the weighted distance as the fitting line. As (4)

$$y = \Phi(x) = ax + b \tag{4}$$

Where, $\Phi(x)$ represents the function of the cumulative error in the normal distribution.

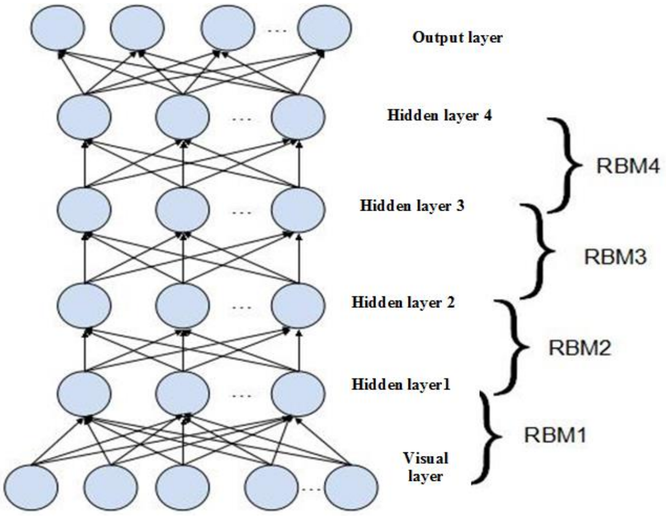


Figure 2. Deep Learning Model Contain Four Hidden Layers

According to this method, we can obtain:

$$gaurate = \frac{N}{N_{total}} \times 100\% \tag{5}$$

3.3 Setting Hidden Nodes

In the setting of hidden nodes, the Boltzmann machine is limited; based on the energy model, the model method can be expressed as:

$$E(\mathbf{x}, \mathbf{h}; \theta) = -\sum w_{ij} x_i h_j - \sum c_j x_i \tag{6}$$

Where, θ is the model parameter, the joint distribution is:

$$P(\mathbf{x}, \mathbf{h}; \theta) = \frac{1}{z} \coprod_{i,j}^{k} e^{w_{ij} x_i h_j} \tag{7}$$

By adjusting the parameters of network in the Boltzmann machine in low layer is equal to the input data of energy can be improved significantly in the condition of restricted Boltzmann machine layer is equal to the probability that the input data of the state. If it has the training data, the form of the derivatives of the log likelihood function of network parameters is simple.

3.4 Adaptive Learning Method

The learning rate selection is important, if it is small may result in long training time, slow convergence, and if it is large may lead to system instability. Under normal circumstances, in order to keep the system stable, we tend to choose smaller learning rate. Error curve decreased fast indicates learning rate is more appropriate, if the larger shocks the learning rate is too large. So, it should according to the different network to select a suitable learning rate. The restricted Boltzmann machine adopts automatic adjustment method of learning rate, in order to make the network to adjust. We will give an initial learning rate is 0.01 in training, in the training, if the in the reconstruction error has declined, then the learning rate unchanged, if the reconstruction error unchanged or increased, it will be divided by 2. When the learning rate is less than 0.0001, stop training.

According to the setting and optimization of the above, using self-learning process, the specific steps are as follows:

(1) The input set, setting the initial value and the hidden layer nodes and the maximum period of training;

(2) The output settings, the weight matrix W, and visual layer and the hidden layer nodes;

(3) The training stage;

Initializing nodes and weights;

Using cycle; solving conditional distribution and maximum joint information entropy;

end

4. Experiment Results and Analysis

In this article, compared with experimental results of BP neural network and the deep belief networks (in comparison to the results of the experiments here before), according to the above, the threshold is 0.33, Figure3 can be obtained. Experimental analysis: as can be seen from Figure3, in the 12 categories we randomly selected, the F score of DBN except the sixth and tenth, others bigger than F of the BP neural network. This means that the DBN in accuracy and recall rate was better than the BP neural network method. Thus it can be seen, the pre training process of DBN achieved relatively good results, and we can learn the useful characteristics from the pictures, so as to improve the retrieval precision and recall. In the pre training on the deep network, we found that the size of the data set, the network learning characteristics of different effects, the better the network dataset to learn, the better the retrieval results, otherwise, it is relatively poor. If the data set is small, the network may not learning useful feature, but it will cause interference to retrieve. Therefore, selecting the appropriate data sets for pre training on the network is more important.

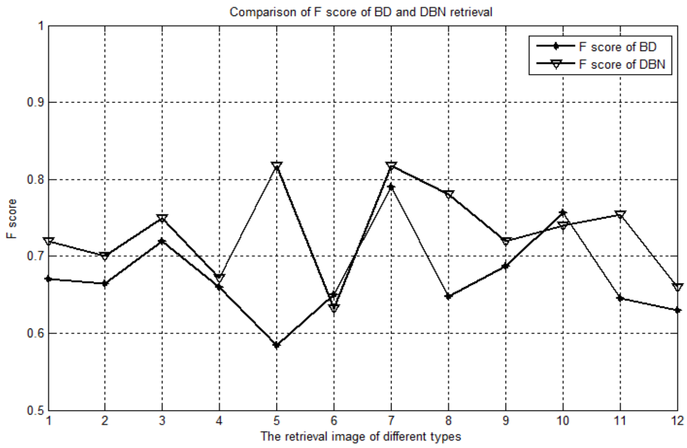


Figure 3. Comparison of Neural Network and the Deep Neural Network

After the clustering process and the derivation of prototype, using the test to evaluate performance of retrieval, it is used for the standard of MPEG7 data set in the literatures. This is a left column test, one of the 40 most similar shapes were determined for each query shape (*i.e.* two times of 40 the shape, the shape of which belong to the same category of the query shape). The final score is the sum of the number of retrieved for each query shape, the highest number can be related to retrieval shape (in the case of 20*1400) ratio. In Figure4 gives three examples of results. The query is a very different ratio under the conditions of the shape from three. For each query (the left line), the first 10 results are shown. We can see, in b) 5 unrelated shape case (belonging to different classes) are retrieved, but they are similar to the query at visually. For case c) the effect is poor, but they are proved made between the retrieved shapes visual similarity.

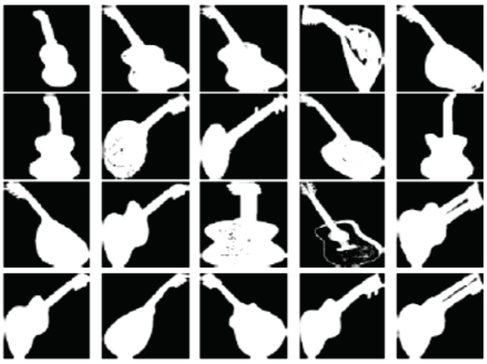
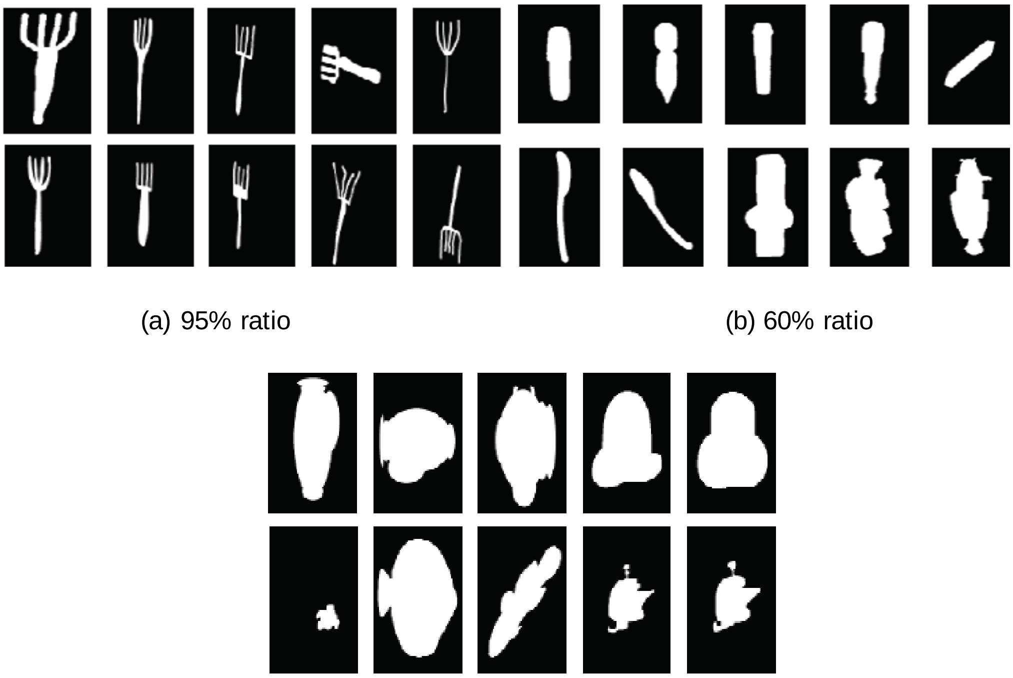


Figure 4. 20 Images of Guitar



(a) 95% ratio

(b) 60% ratio

(c) 25% rate

Figure 5. Retrieval Example under Different Ratio

5. Conclusion

In this article, according to the image retrieval based on local features and global features, according to the characteristics of local features and global features, based on the structure of deep neural network, the input, initialization, network, the number of network layer, nodes and self-learning algorithm, which are based on information theory, it can be seen from the experimental results, the proposed algorithm has a high recognition probability.

Acknowledgements

This work was supported in research projects of Henan province under Grant 122400450356 and 142400411133

References

[1] Lee H., Grosse R. and Ranganath R., "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations", Proceedings of the 26th Annual International Conference on Machine Learning. ACM, **(2009)**, pp. 609-616.

[2] Mohamed A., Dahl G. E. and Hinton G., "Acoustic modeling using deep belief networks", Audio, Speech, and Language Processing, IEEE Transactions on, vol. 20, no. 1, **(2012)**, pp. 14-22.

[3] Mohamed A., Sainath T. N. and Dahl G., "Deep belief networks using discriminative features for phone recognition", Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on. IEEE, **(2011)**, pp. 5060-5063.

[4] Lee H., Grosse R. and Ranganath R., "Unsupervised learning of hierarchical representations with convolutional deep belief networks", Communications of the ACM, vol. 54, no. 10, **(2011)**, pp. 95-103.

[5] Raudies F., Zilli E. A. and Hasselmo M. E., "Deep Belief Networks Learn Context Dependent Behavior", PloS one, vol. 9, no. 3, **(2014)**, pp. e93250.

[6] Sainath T. N., Kingsbury B. and Ramabhadran B., "Making deep belief networks effective for large vocabulary continuous speech recognition", Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop on. IEEE, **(2011)**, pp. 30-35.

[7] Rioux-Maldague L. and Giguere P., “Sign Language Fingerspelling Classification from Depth and Color Images Using a Deep Belief Network”, Computer and Robot Vision (CRV), 2014 Canadian Conference on. IEEE, **(2014)**, pp. 92-97.

[8] Deng L., Hinton G. and Kingsbury B., “New types of deep neural network learning for speech recognition and related applications: An overview”, Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, **(2013)**, pp. 8599-8603.

[9] Deng L., Hinton G. and Kingsbury B., “New types of deep neural network learning for speech recognition and related applications: An overview”, Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, **(2013)**, pp. 8599-8603.

Authors

Sun Ting, He was born in HeNan, China, on June 20, 1972. He received the M.S. degree in Computer Software and Theory from Zheng Zhou University and The PhD in Computer Software and Theory from Northwestern University in 2002 and 2011 respectively. His research interests include digital image processing, Cloud computing and Internet of Things.

Qi Yingchun, He received B. Eng. Degree in computational mathematics from HeNan Educational College and the M. Eng. Degree in computer application technology from University of Electronic Science and Technology in 1997 and 2007 respectively,He is currently researching on the analysis and design of Intelligent algorithm.