Image Retrieval Method for Deep Neural Network

Sun Ting[1,2] and Geng Guohua[2]

[1] School of computer science and technology, Zhoukou Normal University, Zhoukou, Henan,466001, China
[2] Institute of Visualization Technology, Northwest University, Xi'an 710069, China

Abstract

Because of the large data in the image database, the key problem of the retrieval algorithm is to retrieve the required image in the short time. Aiming at this problem, this article given a self-learning deep belief neural network method, and through building layers, input, output, and self-learning algorithm in network architecture to get global algorithm for image retrieval. The accuracy and the convergence of the proposed retrieval method are verified by experiments.

Keywords: Deep Belief Network, Image Retrieval, Local Feature, Network Architecture

1. Introduction

Artificial intelligence is a computer to imitate human intelligence behavior, to achieve computer image recognition, natural language recognition and some specific intelligent tasks and other work. The research direction of artificial intelligence has been half a century history, is recognized as one of the modern high-tech core in the world, due to the discipline of artificial intelligence itself has characteristics of extensive, its application has been researched in-depth the various disciplines and fields and achieved significant results [1-3]. At the same time, combining and utilizing other disciplines, the research and development of artificial intelligence will bring new ideas, so as to promote the development of artificial intelligence. Bionics interdisciplinary combination of biological science, engineering and technology, through learning, imitation, copy and biological system structure, function and working principle of the reengineering, to improve the existing or even create new mechanical and technological process, modern bionics has spawned many fields [4-5]. And Research on them from the mechanism of biological evolution inspired put forward many new methods to solve the problem of artificial intelligence, such as based on the fundamental mechanisms of biological immune system that mimics the body's immune system, for the first time using artificial immune algorithm to solve the optimization problem. By drawing on the evolution law of biology, namely "survival of the fittest, survival of the fittest" the genetic mechanism was proposed for the genetic algorithm to solve the problem of searching for the large-scale and complex system. At the same time, there are methods for searching the path of food recruitment of ants and the bees forage for nectar, and the ant colony algorithm and bee colony algorithm are proposed.

In within a decade of the 21st century, mankind has been working to study how to imitate human brain work and expression of the information ability, and realize the so called artificial intelligence (AI). Human beings receive a lot of data in the perception of information, but it can accurately capture the key factors and save it for future use. Until the 1940s, with the breakthrough progress has been made in the study of neuroanatomical, neurophysiology and neuronal electrophysiological

process, people for the human brain structure and the basic working unit have more and more fully understanding and Research on personnel began to try to imitate the structure and working principle of the human brain to construct to achieve something similar to the human brain with this algorithm of recognition and memory function. The brain learning system is composed of interconnected neurons with an unusually complex network system, using the simplified signal propagation mechanism to mimic some basic functions of human brain neurons, it lays the foundation for the development of early neural computing. Based on artificial neuron model, increasing the learning mechanism proposed perceptron model to solve some of the problems in the field of character recognition, first artificial neural network theory is applied to practical problems. And it is proved that the network constitutes a two layer perceptron can be input to the linear classification, and proposed a hidden layer perceptron unit is a very important research direction. But after that due to the rapid development of computer hardware technology, and other fields, and artificial neural network in dealing with the nonlinear classification problem is not a breakthrough, the field in quiet period for a long time [6-7]. Generalizing the nonlinear data of artificial neural network of information storage and retrieval function, and through equation of dynamic equation and the learning algorithm is proposed, important formulas and parameters of network algorithm and the theoretical basis for future study and the structure of the artificial neural network are provided.

The recent discovery of neuroscience provides some clues to the basic rules of the brain's expression of the mammalian brain. A key finding is a lot of perception related cerebral neocortex neocortical and no clear pre process sensory signals, but let these signals through a complex hierarchy module communicate, and with the passage of time, this module can learn based on observed signals showed some regular characteristics to describe the observed signals. And this feature in the primate visual system performance is more obvious. The process can be divided into some successive processing stages: edge detection, the shape of the then gradually increased to more complex visual shape. The findings lead to the emergence of the depth in the field of machine learning, the commitment to research with the neocortex that can show some similar characteristics of the presentation of information capacity calculation model. 50 years ago, the pioneer Bellman in the field of dynamic programming theory and optimization control pointed out that the high dimensions in the data were the fundamental obstacle to many scientific and engineering applications. In pattern recognition, a lot of difficulties are learning, the complexity of the algorithm is relative to the dimension of the data to present the index level of growth, Richard bellman, this phenomenon is known as the curse of dimensionality. To avoid the curse of dimensionality of the mainstream approach is by preprocessing the data to achieve the data dimension reduction to can be some of the existing methods to effectively deal with the degree, this reduction process dimension is generally referred to as the feature extraction. Therefore, the intelligent process of many pattern recognition systems is transformed to the high difficulty of manual design and the feature extraction process based on the application of specific applications. However, if the feature extraction process is errors or imperfect, the performance of the classifier will inevitably be limited. In addition to the real life of the data in addition to the space dimension, the time dimension is also very important. A group of observed continuous data is capable of conveying a number of specific information, or the actual meaning of the event or observed data is generally inferred from the similar data of the time. Thus, the modeling of temporal components in the observation data is especially important in the information expression. Therefore, the consistency of the observation data to capture the dependence of the data in the time and space, is considered to be the fundamental

goal of the depth learning system. If can get a learning system with the depth of the robustness, can through this hierarchy system training on a large data set, then the extraction system of information as input to follow a relatively simple classification system, came to an equally robust pattern recognition system. Here the robustness of the said classification results for data in the transformed and distorted can keep invariant features, this kind of transformation and distortions include noise, scaling, rotation, illumination changes, displacement and so on.

Deep neural network algorithm is widely concerned in academic field, which is based on fast data analysis and forecasting. The depth of the neural network is an automatic learning sample characteristics of the input method, and to study the characteristics of data is more essential characterizations and through layer by layer initialization "to overcome the difficulty in training. Therefore, it has been a wide range of academic and industrial circles, and became as research fields. Therefore, the depth of the neural network as a learning model of complex hierarchical probabilistic method in various fields has been widely used. At present, it has been applied to the field of speech recognition, recognition of handwriting font, traffic signs, face recognition and other image processing fields, showing superior performance of learning.

Deep neural network includes deep belief network (DBNs), automatic coding machine and deep convolutional neural network (CNNs) structure model. In these models, the most representative structures are two kinds of deep belief networks and deep convolutional neural networks. The depth convolution neural network is easy to get into the local optimal solution of the non-convex function, while the depth convolution network also has the problem that cannot deal with the problem of label free data. In the deep belief network model, the layered learning process, because each layer is used many restricted Boltzmann machine RBM stack constitute, and widely used.

Content based image retrieval is divided into two main modules: feature extraction module and query module. Content based image retrieval involves computer vision, pattern recognition, image understanding and other technologies, which has four features, first, the use of similar matching. In the text-based image retrieval based on, as we enter the keyword or keywords, and image labels text, pictures and irrelevant, so we can use ratio compared to the way, which is an exact match. In content-based image retrieval based on, because the images of the same scene may also be because of the different angles or light and shade different and different ways of expression and content of an image is generally more abundant, between the strong correlation, and characteristics of the data is usually not the simple relationship. It is generally used similarity matching. Second, the use of direct input image query. Based on image content retrieval generally uses the method of directly enter the picture query, if the user is not familiar with the specific structure you want to query image can browse through select the system to provide the sample picture retrieval. Some systems can also check the results of the results of the query to determine the good or bad, and then urge the system to make the necessary amendments. Third, interactive strong. The feedback function of the system can be improved by user feedback and the retrieval results are corrected repeatedly. Fourth, can meet the multi-level retrieval requirements. The content-based image retrieval system generally includes feature base, image database and knowledge base, which can meet the different retrieval requirements.

When the database is large when, due to the restriction of computational resources and memory, the exact linear search (sequence comparison between the query image and database of all image similarity to retrieve) is not feasible and not necessary. It is used to describe the features of image data, so the content-based image retrieval is often affected by the curse of dimensionality and the performance is not high.. This

urgently needs a highly scalable retrieval method. It is a promising method to use the similarity hash to establish the index, but the use of a single hash table can not measure the accuracy of the retrieval and the recall rate.. The use of the Doha Greek method can get a high recall rate in constant retrieval time, but the retrieval accuracy is low, and a large number of uncorrelated samples are returned to the user resulting in low efficiency. How to improve the accuracy of the Doha Greek method, and maintain the high recall rate and the small retrieval time become the key of Dohashi's successful application.
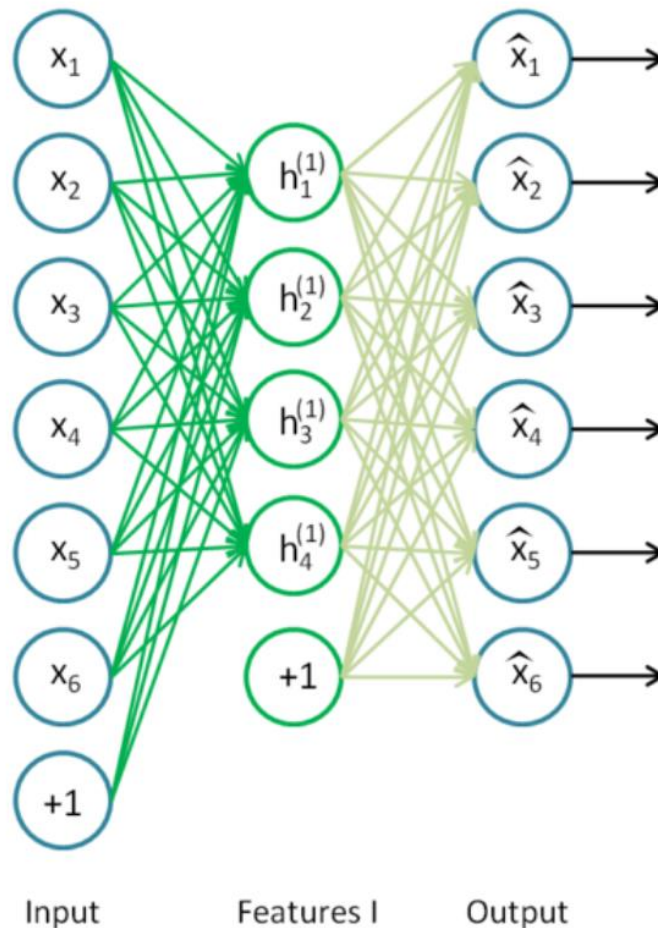

Figure 1. Deep Neural Network Structure

2. Related Works

Image recognition is the technology of digital image processing, the use of artificial intelligence technology, especially machine learning method, so that the computer can identify the contents of the image. Image recognition is a major area of research of pattern recognition, relates to handwritten character recognition, face recognition, object recognition and so on, and some of them relatively mature technology has been commercial applications. In image recognition tasks, handwritten numeral recognition and face recognition are studied more and more fields. Handwritten numeral recognition can be used to automatically read the bank check information, the postal code on the envelope and documentation of data and so on, and face recognition can used to log on to the computer system, video security monitoring system and so on. Due to the diversity of image recognition, currently be

used for the image recognition method is generally based on specific recognition problems, such as the best handwritten numeral recognition method is often in face recognition and other image recognition problem and not very good performance. So a large part of the identification system needs a lot of work and algorithm research to get a breakthrough in the performance of specific identification issues. Therefore, it is necessary to search for a relatively general machine learning method that can obtain good recognition results in different recognition problems.

Image feature extraction is the key link of content-based image retrieval system, which directly affects the operating efficiency and the retrieval effect of the system. The content features of the image generally include the underlying visual features (texture, color and shape) and high-level semantics to describe the high-level semantics and human cognitive domain, a relatively strong subjective color. Compared with the high-level semantics, the underlying features are not changed by human consciousness, which belongs to the inherent attributes of the image, and has a strong objectivity. Generally, the feature of the image is texture, color and shape[8]. Extracted texture features of commonly used methods: gray level co-occurrence matrix, Tamura texture and wavelet transform *etc.*, color features commonly used method for extracting color histogram and color moment and color polymerization vector, color and other related plans to extract shape features of two kinds of methods: regional feature and contour feature, more representative is boundary direction histogram, wavelet contour descriptors, Fourier descriptors, Hu invariant moments. The depth of the neural network technology is applied to image retrieval and application of the nonlinear mapping ability, self-learning ability and adaptive ability of self, let neural network as a human brain that oneself according to the image content to learn the low-level visual features. Image feature extraction are generally classified into two parts, a part is pre trained that network learned the low-level features from image, another part is the fme-tuming model output of the categories of information, the information of these categories is learned from manually annotated data, more in line with the people the image of subjective description. Thus, in our method, the features of the retrieval are both low-level features and high-level semantic features.

Deep learning was proposed by Hinton *et al.* in 2006, which is a new area of machine learning. Deep learning has made great progress in speech recognition. It depends not only on the ability of the parallel processing of the large data, but also on the algorithm, and this algorithm is the depth of learning. June 2012, "New York Times" disclosed the Google brain project, the project by Stanford University Professor Andrew Ng and computer systems for the top expert Jeff Dean co dominant, 16000 CPU core parallel computing platform training a known for the depth of the neural network of the machine learning model. The human visual system is classified by the processing of information.. From low-level edge features extracted to shape (or target), to the higher target, target behavior, namely low - level features into the high-level features, characterized by a low to high said more and more abstract. That depth study is how to learn from this process, the process is the process of modeling. For a pair of images, the pixel level feature has little value. The smaller graphics can be made of basic hook, so the complicated concept graphics need a higher level feature, that is, the high-level representation is composed by the bottom layer.

Based on the region of the image retrieval technology is a close to understand human intelligence to retrieval technology, mainly through the image segmentation technique to extract image object, then for each region using local features to describe, integrated the characteristics of each region can be to the image feature description, finally, using appropriate similar measurement criteria to retrieve images, and image segmentation is a development is not mature and with

considerable difficulty, and now it is impossible to make objects of segmented regions in the image correspond perfectly, so this kind of method of the retrieval accuracy is not too high. Content based retrieval is a kind of approximate matching in the retrieval process, which uses a similarity measure to match the images in the image database, to obtain the query results, single feature retrieval may appear similar to measure proximity of the two images in semantic difference is bigger than that based on the multi feature integrated retrieval, convenient for the user to the more flexible, more effective to express queries require; but the similarity measure and the subjective feelings about, the similarity calculation basically is based on mathematics, and human visual perception is also has a certain gap.

3. The Proposed Scheme

According to the structure of the deep neural network algorithm, the content, on the deep neural network input data, the number of network and node design and adaptive learning algorithm design, finally obtaining image retrieval scheme which is executable, according to the plan, the validity of scheme can be verified.

3.1. System Model

DBN visual layer node number for p, the first layer number of hidden layer nodes is q, the input of the training sample images for X, pixels for $m \times p$, visual layer node offset $b = (b_1, b_2, \cdots, b_q)$, the offset of the first hidden layer nodes. When visual input layer node training samples $x_i = (x_1, x_2, \cdots, x_p)$, the first to corresponding to the output of the hidden layer nodes

$$h_j = \sum_{i=1}^{p} x_i \cdot w_{ij} + b_j \tag{1}$$

Each value of visible nodes and hidden nodes between the values of all have a energy. Energy function is defined as

$$E(x,h) = -\sum_{i=1}^{p}\sum_{j=1}^{q} x_i w_{ij} h_j - \sum_{i=1}^{p} a_i x_i - \sum_{j=1}^{q} b_j h_j \tag{2}$$

3.2. Extract the Image Spatial Information

To solve the matter of the goal is to all visual node value and the value of hidden nodes energy accumulation, and minimize the energy, the solving process is complicated, now more commonly used method is the introduction of probability. The following analysis combined with entropy said. Free energy is defined as
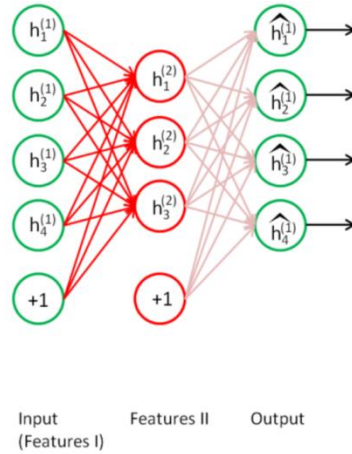
$$FreeEnergy(x) = -\ln \sum_{h} e^{-E(x,h)} \tag{3}$$

Figure 2. Deep Learning Model Contain Four Hidden Layers

$p(x)$ can be written as

$$p(x) = \frac{e^{-FreeEnergy(x)}}{Z} \tag{4}$$

3.3. The Output Image Information

The formula (4) to logarithm on both sides

$$\ln p(x) = -FreeEnergy(x) - \ln Z \tag{5}$$

Or

$$H_x(p) = \sum_x p(x)FreeEnergy(x) + \ln Z \tag{6}$$

On the left side of the equation for the visual input image layer of information entropy, use $H_x(p)$, the equation $\ln Z$ is a constant in the second on the right, o probability is still, after an average entropy function is associated with the probability distribution of x only, has nothing to do with the space position x, which has nothing to do with the space information system after convergence results.

After the sample translation image rotation compared with the original sample image, although the space position has changed, but the general image content is constant, the proportion of each pixel is certain, sample image information entropy is certain, free energy of the system is also did not change, the objective function. Therefore, the neural network based on the depth of space information of images, rotation training sample images after the translation with the original sample image has the same objective function value, still can effectively identify, robust.

3.4. Learning Process Based on Markoff

The selection of learning rate is very important, which may lead to the training time is too long, the convergence is slow, and big may cause the system become instability. Generally, in order to maintain the stability of the system, it tends to select the smaller learning rate. The error curve fell fast that the learning rate is appropriate, if the shock is too large, the study rate is too large. So, it can accord to different network to select a suitable learning rate. In this experiment, restricted Boltzmann machine adopts automatic adjustment learning rate method, making the network to adjust according to the situation. We have given an initial learning rate of

0.04, when training is given, if the error of refactoring is decreased, the learning rate is a constant, if the error is a constant or beginning to rise, so it is divided by 2. When the learning rate is less than 0.0004, stop training.

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

(1) Input settings, Initializing the input variables,

(2) Output settings, loop learning the sample value and visual layer ,hidden layer nodes,

(3) The training and learning stage, using Markov transfer probability maximization method,

Determining the information transition probability;

Solving conditional distribution and maximum joint information entropy;

end

4. Experiment Results and Analysis

In this paper, compared the neural network algorithm with the deep belief network, according to the results of image retrieval matching, the threshold is 0.33, and then we can get Figure 3. It can be seen from the figure, with the different kinds of pictures, the corresponding matching degree F value also appeared to fluctuate up and down, however, for the overall effect, based on the deep neural network has accurate matching degree and recall better than the genetic algorithm. The deep neural network training and self-learning, if learning more features, the characteristics will be more complex, so the better retrieval results can be got. However, the training set is small, the advantages of the deep neural network is not too obvious, however in different trials, the picture features that required to be retrieved is more complex, the picture database gradually become complex, so the deep neural network is suitable for the actual situation.
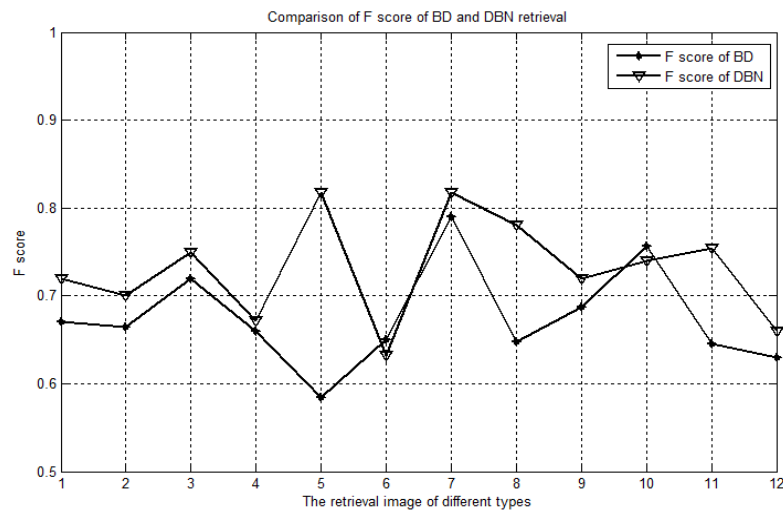

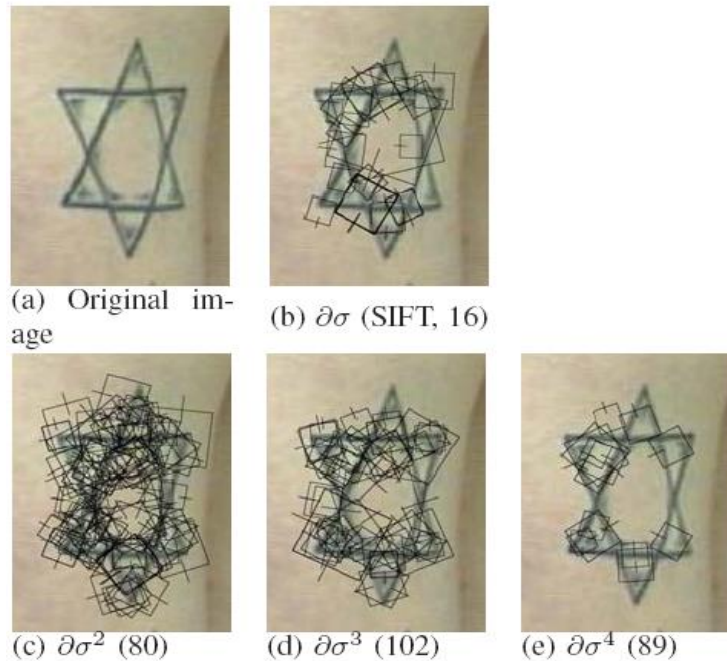**Figure 3. The Comparison of Neural Network and the Deep Neural Network**

Figure 4. The Number of Different Scale Space Derivative Order Point Extraction

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

References

[1] A. Krizhevsky, I. Sutskever and G E Hinton, "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, **(2012)**, pp. 1097-1105.

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

[3] J. Wang, S. Kumar and S. F. Chang, "Semi-supervised hashing for scalable image retrieval", Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, **(2010)**, pp. 3424-3431.

[4] G. E. Hinton, "Learning multiple layers of representation", Trends in cognitive sciences, vol. 11, no. 10, **(2007)**, pp. 428-434.

[5] S. Osindero and G. E. Hinton, "Modeling image patches with a directed hierarchy of Markov random fields", Advances in neural information processing systems, **(2008)**, pp. 1121-1128.

[6] K. Simonyan, A. Vedaldi and A. Zisserman, "Deep fisher networks for large-scale image classification", Advances in neural information processing systems, **(2013)**, pp. 163-171.

[7] G. E. Dahl, D. Yu and L. Deng, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition", Audio, Speech, and Language Processing, IEEE Transactions on, vol. 20, no. 1, **(2012)**, pp. 30-42.

[8] L. Bo, K. Lai and X. Ren, "Object recognition with hierarchical kernel descriptors", Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, **(2011)**, pp 1729-1736.

Authors

Sun Ting was born in HeNan, China, on June 20, 1972. He received the M.S. degree in Computer Software and Theory from ZhengZhou 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.

Geng Guohua was born in 1955. Ph.D., professor, Ph.D. supervisor, Her research interests include Intelligent information processing, Database and knowledge base.