Advances in Computer Science Research (ACSR), volume 83
8th International Conference on Mechatronics, Computer and Education Informationization (MCEI 2018)

Comparison Optimization for Image Classification based on Deep Belief Network

Daohan Yang[1, a], Xiao Yao[2, b, *], Wensong Bai[3, c],
Bin Wang[4, d], Runyu Wang[5, e] and Zuxi Zhang[6, f]

[1, 2, 3, 4, 5, 6]The College of IoT Engineering, Hohai University, Changzhou 213000, China

[a]ydh1997831@163.com, [b]yaox@hhu.edu.cn, [c]m13365239599_1@163.com, [d]wbingtang@126.com, [e]13270710902@163.com, [f]zzx_fairy@163.com

Keywords: Image classification and identification; Intelligent robotics and control engineering; Deep learning; Image feature extraction

Abstract. In the field of intelligent robotics and control engineering, image classification technology based on deep learning is of great significance for robot image identification and has gained a wider range of applications. However, when it comes to the actual working environment of the robot, the existences of light illumination, noise and other factors will make the result of deep learning frame merely not satisfied. This paper focuses on the comparison of different feature extraction methods for optimization of DBN(Deep Belief Networks). Experimental results show that these methods can improve the accuracy of image classification.

Introduction

In the area of intelligent robotics and control engineering, the vision sensor has become one of the most important robot sensors because it can obtain large amount of information and has a wide range of application. Based on it, the intelligent robot vision servo control system is widely used [1]. The robot vision servo system refers to a system which produce closed loop feedback control of a robot using an installed vision sensor [2]. In this system, the classification and identification of images is the key component.

The image classification method, based on deep learning, aims to perform image feature extraction and classification to carry out more effective management of them [3]. In order to solve the growing problems of massive data of image, image classification technology has increasingly become a hot topic, which can extract features according to the image content and automatically divide images into the corresponding categories, so that you can not only save a lot of tedious manual classification work involved in the process, but also complete the construction of image semantic information consistent with the people's cognition [4].

In the face of today's huge data, the traditional shallow learning models have often failed to meet the expected requirements. Nowadays, deep learning has its irreplaceable advantages. It overlay multi-layer nonlinear network structure and each layer of the computational unit can represent data non-linearly well by superimposing different network layers, so the essential features of complex data are extracted [5].

The current deep learning model is basically neural network model, which can be traced back to the last century and were popular in the 1980s and 1990s. The earliest application area for deep learning was mainly in the image domain. LeCun et al. [6] use the convolutional neural network to achieve handwritten digit recognition.

In 2006, deep learning neural network has made breakthrough progress and has received extensive attention. At the time, Hinton et al. [7] used a multi-layered neural network to test large-scale databases and achieved good results. Image classification and recognition technology has therefore been greatly improved. After decades of research by scholars, the neural network model of deep learning has been greatly improved, and training methods have also made great progress. For example, in order to improve the training of deep learning networks, an unsupervised and layer-by-layer training method are proposed, which controls the back propagation process of neural networks well and makes the trained model more efficient [8].

In real life, the images obtained are often used directly for machine learning, so the result of this process is usually unsatisfied due to the influence of the external factors. The original DBN doesn't do screening to image's characteristics, so it will also learn the noise without feature extraction and this will influence the accuracy of image classification.

In order to do image classification more efficiently for intelligent robotics control engineering, this paper mainly studies on the deep learning for the neutral network, then compares various methods of feature extraction combined with deep belief networks and improve the accuracy of image classification. We will analyze how the feature extraction process affects DBN performance from the perspective of DBN intrinsic characteristics.

The paper is organized as follows: In Section 2, the principle of deep belief networks is introduced. In Section 3, we compare various methods to improve the accuracy of classification methods under the deep belief network. Experiments are performed in Section 4 to evaluate the classification performance of proposed methods. In Section 5, we discuss the result of the experiments. The conclusion is drawn in Section 6.

Deep Belief Network

In the field of deep learning, there are many models such as CNN(convolutional neural network), DBN (deep belief network) and so on. This paper selected DBN as a tool to study the parameters of image feature extraction.

In 2006, Hinton, stacked RBMs (restricted Boltzmann machines) and trained them layer by layer based on the principle of deep learning [7]. He pioneered a special multi-layer neural network structure named deep belief network. Typical deep belief models merge deep learning with feature learning and are widely used in present machine learning and application.

Its main structure is shown in Fig. 1 The network mainly includes a supervised back-propagation network and several unsupervised restricted Boltzmann machines. Deep confidence networks (DBNs) contain multiple hidden layers, each of which obtains a lot of relevant feature information from the previous layer. DBNs can be viewed as a superposition of multiple restricted Boltzmann machines. The output of each lower-level RBM is used as the input data of the next layer to train the next RBM. This group of RBMs use comparative divergence algorithm to build up a DBN.

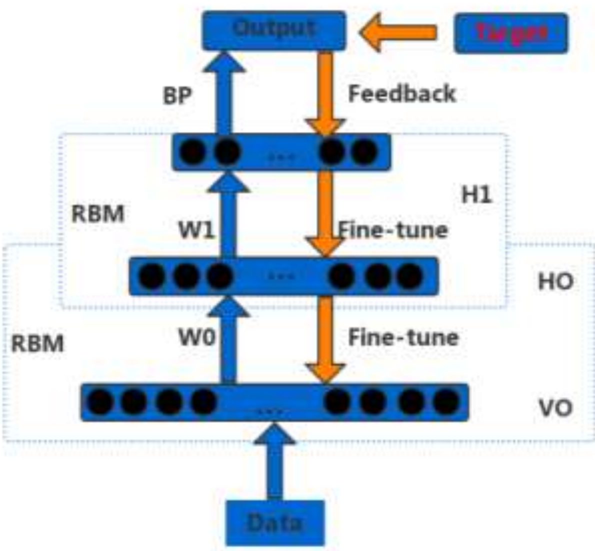

Figure 1. The structure of DBN

After several trainings, the hidden layer of RBM can not only accurately represent the characteristics of the visible layer, but also restore the visible layer. If several RBMs are "serialized", they can be stacked into one DBN structure. Among them, the hidden layer of the last RBM is the visible layer of the next RBM. The output of the previous RBM is the input of the next RBM. Each layer of RBM is fully trained from top to bottom until the last layer, and training of the DBN network is basically completed. Learning under unsupervised conditions can also achieve good training results. Therefore, DBN is used as a structural framework for unsupervised networks, and it has good results in the area of intelligent robotics and control engineering for image classification and identification.

Feature Extraction Methods

This paper proposes several methods combined with the DBN to improve the performance of image classification.

Canny Operator and Prewitt Operator. The edge of the image reflects the appearance of the object and is an important feature of image analysis and pattern recognition. And the edge of the gray image is where the gray value of the pixel changes intensely. These changes are usually roof changes or step changes. The size of the image roof or step change generally uses the size of the first or second derivative of the gray image to describe [9]. We mainly use first-order differential image edge detection operators as the edge detection methods for gray images.

First-order differential image edge detection operators typically threshold the amplitude of the image gradient to extract edges. The gradient of the gray image reflects the magnitude and direction of the gray value's change in the image. The gradient is defined as [10]:

$$\nabla I(x,y)=\begin{bmatrix} G_x \\ G_y \end{bmatrix}=\begin{bmatrix} \dfrac{\partial l}{\partial x} \\ \dfrac{\partial l}{\partial y} \end{bmatrix} \tag{1}$$

The gradient magnitude and direction are defined as [10]:

$$\left|\nabla I(x,y)\right|=\sqrt{G_x^{\ 2}+G_y^{\ 2}} \tag{2}$$

$$\alpha=\arctan\left(\frac{G_y}{G_x}\right) \tag{3}$$

Threshold the amplitude of the gradient, and then you can get the edge of the image. According to the difference of image's horizontal gradient component Gx and the vertical gradient component Gy when we calculate, there are different edge detection operators. We take one of them. It is the prewitt operator:

$$S_x=\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}, S_y=\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \tag{4}$$

In 1986, Canny proposed an optimal edge detection operator, which is considered to be one of the most successful and widely used edge detection methods [11]. The canny operator first uses Gaussian smoothing filter to suppress the noise in the image, and calculates the gradient magnitude and direction of the image through the first-order difference operator. Then, the gradient amplitude is non-maximum suppressed, and finally the edge is extracted by the double-threshold method.

The canny operator is:

$$S_x = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, S_y = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \tag{5}$$

Low-frequency Filtering and High-frequency Filtering Made by Wavelet Analysis. Signal analysis generally aims to obtain the relationship between time and frequency domains. The Fourier transform provides information about the frequency domain, but the localized information in time is essentially lost. Different from the Fourier transform, the wavelet transform can obtain the time information of the signal by shifting the mother wavelet, and the frequency characteristics of the signal can be obtained by scaling the width (or the scale) of the wavelet. The scaling and shifting operations on the mother wavelet are used to calculate the coefficients of the wavelet, which represent the relationship between the wavelet and the local signal.

Discrete wavelet transforms can be used to analyze or decompose signals. This process is called decomposition or analysis. The process of restoring the decomposed coefficients to the original signal is called wavelet reconstruction or synthesis, and mathematically it is called inverse discrete wavelet transform (IDWT).

The basic idea of a wavelet transform is to represent a function or signal, such as an image signal, using a set of wavelet functions or basis functions.

The steps of wavelet transform are as follows:

(1) Decomposition: Select a wavelet to perform wavelet decomposition on the signal;

(2) Extraction of different frequency signals: Extract the low frequent coefficients(which represent approximate component of the signal) and high frequent coefficients (which represent detailed component of the signal) of the signal using the functions provided;

(3) Reconstruction: Reconstruct the low-frequency or high-frequency information of an image using the extracted low-frequency or high-frequency coefficients. And this is the low-frequency filter and high-frequency filter made by wavelet analysis. After this step,we can obtain the low-frequency and high-frequency part of the signal.

Comparison of Feature Extraction

After introducing the tools used to compare the feature extraction methods, the process of comparison is described below. The flow chart is shown as Fig. 2 .

(1) The MNIST database is used as deep learning database. MNIST is a classic entry-level demo for deep learning which is composed of 60,000 training pictures and 10,000 test samples. Each picture is 28*28 in size. These image are different numbers from 0 to 9 handwritten.

(2) Read into the MNIST database, and then use the discussed four feature extraction methods to process them separately.

(3) Input the processed data into DBN for training. We train a 100-100 hidden unit DBN and use its weights to initialize a NN, this DBN has two RBMs and its structure is shown as figure 2.2.

(4) Save the trained DBN and read the test samples.

(5) Compare networks' reconstruction error.

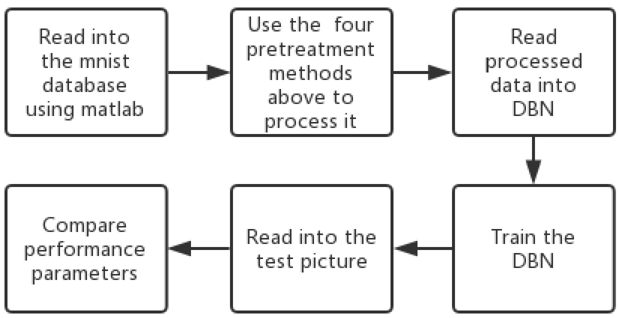

Figure 2. Flow chart of comparison

Result and Discussion

The reconstruction error. The reconstruction error means the percentage of samples that are different from the original label after the DBN's reconstruction. It represents the performance of each feature extraction methods. So, smaller reconstruction error explains the better classification performance is achieved.

In Fig. 3 , the horizontal axis represents training times, and the vertical axis represents the reconstruction error. Experimental results show that the feature extraction method with canny operator get the best performance, and prewitt operator, canny operator and Low frequency filter using wavelet analysis are all perform better than the original DBN without any feature extraction methods.

In addition, it should be noted that with the increase of the number of training, the error rate of the reconstructed DBN network will be significantly reduced, and the error rate of the original DBN without feature extraction will not drop significantly. Since methods with the feature extaction converges quickly, we can conclude that feature extraction has an important significance for improving the recognition rate of DBN network. Especially in the case of high recognition rate is needed and sufficient time and sufficient training are provided, using the feature extraction methods (the canny operator performs best) will obviously improve the experimental results.

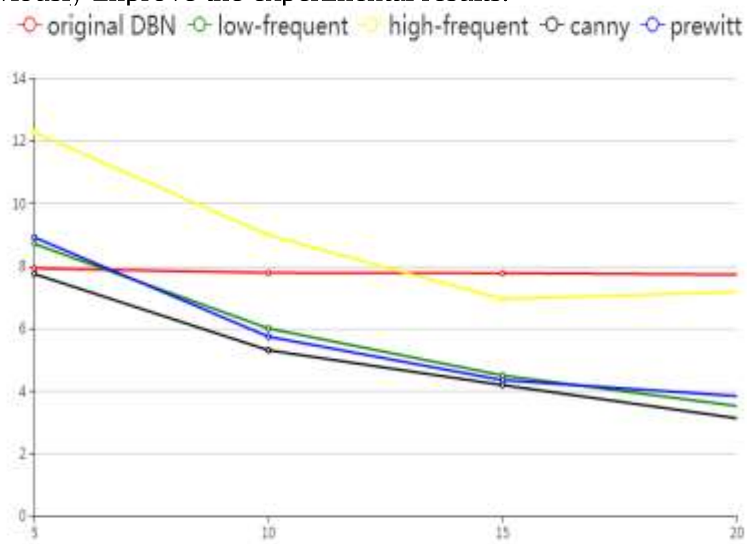

Figure 3. The reconstruction error of feature extraction methods

Discussion. In order to explore the advantage of the canny operator, we used these four feature extraction methods to process the same image in the MNIST database. The result is shown as Fig. 4 :

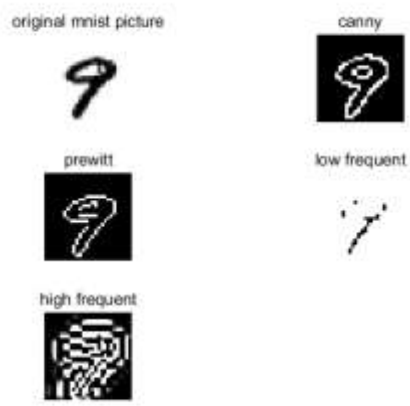

Figure 4. The performance of the four methods on a particular picture

As the figure shown, the canny operator performs best in the area of the extraction of image edges. Compared to the others, canny's edge is clearer and thicker. So, it can extract the boundaries of the image well. It is important to handwritten digital recognition since the edge information is its main feature.

After exploring why canny operator performs best, we believe that the low-frequency filtering and high-frequency filtering made by wavelet analysis methods are insufficient in characterizing the image since they only extract a part of the frequency's effective information.

Moreover,We can learn from paper [12] that the canny operator uses two thresholds to detect strong and weak edges. And if they are connected to the edge, then the output contains only weak edges. In comparison, prewitt has a good detection effect on gray gradient low-noise images, but the effect of processing is not ideal for images with multiple complex noises. Therefore, the canny can detect the edge more clearly.

People haven't do fearture extraction to DBN before since they believe it has the ability to generalize feature. However, when we perform feature extraction, the reconstruction error rate is significantly reduced under the same training times. We believe the lower reconstruction error results from the limited number of hidden layer units in the first layer of the DBN, so the data read in should contain effective information as much as possible. And this explain why do feature extraction is meaningful to DBN. Among them, canny operator perform best to DBN when do feature extraction. We conclude that the DBN network is more sensitive to sharp edges compared to the frequency domain.

Moreover, the convergence rate of DBN is faster and this is of significance to deep learning. So, the training is also more efficient, and the computing resources and time are saved. And the feature extraction method is significant when applied to robotics' actual working environment with many interference factors.

Conclusions

In order to improve the accuracy of image classification in the condition of intelligent robotics and control engineering, this paper studies and compares four feature extraction methods based on DBN, and obtained the best effect of deep learning using canny operator to do feature extraction. This paper prove that do feature extraction to image before deep learning is meaningful since this will significantly improve the accuracy rate (especially in the case of actual working environment of the robotics).

Acknowledgements

This work was supported by the National Nature Science Foundation of China under grants (61501170, 61401148, 61471157, and 61671202), The Fundamental Research Funds for the Central Universities 2015B25214 , Key Development Program of Jiangsu Province of China under grants BE2017071 and BE2017647, and the Projects of International Cooperation and Exchanges of Changzhou of China under grants CZ20170018.

References

[1] Z.J. Yan: *Research on intelligent robot visual servo system*, (Xi' an University of Technology, China, 2004).

[2]T. H. Park and B. H. Lee: An approach to robot motion analysis and planning for conveyor tracking, *Proc. IEEE Int. Conf. Robotics and Automation*, 1991, p. 70-75.

[3] Z. K. Huang: *A Study of feature extraction technique based on color image & plant classification*, (University of Science and Technology of China, China, 2006).

[4] S. Majeed, Z. ul Qayyum and S. Sarwar. SIREA: Image Retrieval Using Ontology of Qualitative Semantic Image Descriptions, *International Conference on Information Science and Applications* (ICISA), 2013, p. 1-6.

[5] B. Y. Liu: *Research on Feature Representation and Image Classification Based on Deep Learning*, (University of Chinese Academy of Sciences, China, 2015).
[6] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner: Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 1998, 86(11), p. 2278-2324.
[7] G. E. Hinton: Learning multiple layers of representation, Trends in Cognitive Sciences, 2007, 11(10), p. 428-434.
[8] G. E. Hinton, S.Osindero and Y. W. Teh: A fast learning algorithm for deep belief nets, Neural Computation, 2006, 18(7), p. 1527-1554.
[9] B. He and T.Y. Ma: *Visual C++ Digital Image Processing*, second ed. (Posts and Telecom Press, Beijing, 2002, pp. 394-398).
[10] R. C. Gonzalez and R. E. Woods: *Digital Image Processing*, second ed. (Publishing House of Electronics Industry, Beijing, 2003).
[11] J.canny: A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, 8(6), p. 679-698.
[12] G. M. Shi, Q. T. Wei and F. S. Meng: Image Edge Detection Algorithm Based on Canny Operator, Modern electronic technology, 2015, (12), p. 92-93, 97.