International Journal of Database Theory and Application
Vol.8, No.3 (2015), pp.367-376
http://dx.doi.org/10.14257/ijdta.2015.8.3.32

The Application of Convolution Neural Networks in Handwritten Numeral Recognition

Xiaofeng Han[1] and Yan Li[2]

[1,2]*College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao, 266590, China*

[1]*hs56923@163.com* , [2] *liyan@sdkd.net.cn*

Abstract

Convolutional neural networks are a technology that combines artificial neural networks and recent deep learning methods. They have been applied to many image recognition tasks and have attracted the attention of the researchers of many countries in recent years. This paper summarizes the latest development of convolutional neural networks and expounds the relative research of image recognition technology and elaborates on the application of convolutional neural networks in handwritten numeral recognition.

Keywords: *convolutional neural networks; application; handwritten numeral recognition.*

1. Introduction

Artificial neural networks are a technology inspired from the functioning of human being. Deep learning, inspired from the recent discoveries of the biology and neurology of animal and human vision systems, has developed a deep network architecture which has hierarchical structure. Convolutional neural networks are a technology that combines artificial neural networks and recent deep learning methods. They have been applied to many image recognition tasks in recent years, such as handwritten numeral recognition, which will be dealt with in this paper.

2. The Development of Convolutional Neural Networks

Convolutional neural networks are inspired by the structure of the visual system, especially the structural model based on a cat's visual cortex put forward by Hubel & Wiesel[1].Fukushima [2] puts forward Necognition, the first network which is based on the local connection type and hierarchical organization between neurons for the transformation of image. Fukushima thinks that when a group of parameters of the same neuron are affected to the small area in different positions of the preorder layer, the features relative to data translational invariance are obtained. Then according to this viewpoint, LeCun et al [3,4] designs and adopts the algorithm based on the error gradient to train convolutional neural networks, and shows their leading performance compared with other methods then in some pattern recognition tasks. The modern physiology's understanding of the visual system is in agreement with the image processing process in convolutional neural networks[5]. By now, the pattern recognition system based on convolutional neural networks is one of the best systems of performance, especially in the handwritten numeral recognition field[3].

Convolutional neural networks are a kind of multilayer neural network specially designed for processing two dimensional data. The design of convolutional neural networks is inspired by the early Time-Delay Neural Networks[6], which were used to process speech and time sequence signals to reduce the computational complexity in the learning process. Convolutional neural networks are thought to be the first deep learning method with robustness which is really successful in using multilayer hierarchical structure networks. By the special correlation in data mining, convolutional neural networks can reduce the number of trainable parameters in the networks to improve the

ISSN: 2005-4270 IJDTA

back propagation algorithm deficiency of forward propagation networks. In convolutional neural networks, the small area which is also called the local sensing region is taken as the input data of bottom in the hierarchy. Through forward propagation, the information passes various layers in the network. Each layer consists of filters so as to obtain some significant features of observed data. Because the local sensing region can obtain some basic characteristics, such as the boundaries and corners, etc. in the image, the method can provide the relative invariant shapes of certain horizons of displacement, stretching and rotation.

The close connection and the spatial formation between levels of convolutional neural networks make them especially suitable for image processing and understanding, and can make them automatically extract the rich correlative characteristics from the images. In some experiments, the researchers make pretreatment by using the Garber filter and imitate the human's reaction process of visual stimulation through convolutional neural networks [7]. In the recent research work, convolutional neural networks have been further applied to such aspects as facial recognition [8], document analyses [9], speech detection [10] and license plate recognition [11]. Recently by using continuous frames of video data as the input data of convolutional neural networks, the researchers have introduced the time dimension data to achieve recognition of human actions in video frequency[12].

3. The Relative Research of Image Recognition Technology

Image recognition technology refers to the technology that is based on the digital image processing technology and utilizes artificial intelligence technology, especially the machine learning method, to make computers recognize the content in the image. Image recognition is one of the main fields of pattern recognition research and involves handwritten numeral recognition, facial recognition, object recognition, etc. Some relative mature technology has been applied in commerce now[13].

In image recognition tasks, handwritten numeral recognition is the field that is more researched. Handwritten numeral recognition can be applied in automatically reading the bank check information, the postal codes on envelopes, the data in some documents, etc. Handwritten numeral recognition has the following characteristics:

3.1There are fewer classification kinds of handwritten numeral recognition samples, but each classification has sufficient samples.

3.2. The image borders of handwritten numeral recognition samples are relatively clear and single.

3.3. Due to the diversity of image recognition problems, the concrete methods for image recognition generally aim at the concrete recognition problems, for example, the best method of handwritten numeral recognition cannot be well applied to facial recognition and other image recognition problems. So a large part of recognition systems needs a lot of work and algorithm research to get the breakthrough of performance (such as improving the recognition rate and accelerating the training speed) in the specific recognition problems. Therefore it is highly necessary to find a generally used method which can obtain better recognition effects in different recognition problems.

Convolutional neural networks represented by LeNet and put forward by LeCun [4] have obtained good results in different image recognition tasks, which are recognized as one representative of the general image recognition system so far. The convolutional neural networks of permutation encoding technology put forward by Krussul et al also get good results in such recognition tasks as handwritten numeral recognition, facial recognition and the recognition of small objects.

4. The Application of Convolutional Neural Networks in Handwritten Numeral Recognition.

4.1.The MNIST Database of Handwritten Digits

The MNIST Database of Handwritten Digits[14] consisted of MNIST Database 3 and Special Database 1 when they were first established. SD-3 was used as the training set and SD-1 as the test set. In order to make the training result and the test result more independent of his other adopted database sets, American scientist LeCun combined the characteristics of the 2 sets to establish the MNIST Database of Handwritten Digits. The training set of the MNIST Database of Handwritten Digits equally chose 30,000 samples from SD-3 and SD-1. The chosen 60,000 samples were from around 250 different individual handwritten data. Similarly, the test set chose respectively 5,000 samples from SD-3 and SD-1.

4.2.The Working Mode of Convolutional Neural Network Model

The classification process of trained convolutional neural networks is similar to that of a multilayer feed forward process. The image is used as input data to transmit layer by layer until the output layer when the classification result is output. As shown in Figure 1, after receiving the input data of image(In Figure1, the image is the number‘4’) ,having passed the 6 filters of the first convolutional layer, the networks form 6 characteristic graphs i.e. Layer C1. These characteristic graphs contain all the characteristics that the image acquires after passing each filter. Then through a 2×2 to 1 down sampling operation, Layer C1 obtains Layer S2. Compared with Layer C1, Layer S2 reduces the size of the characteristic graphs and enhances the network robustness with respect to noise and minor perturbations. Just as shown in the characteristic graph at the bottom in Layer C1 and Layer S2, the notch of the lower right corner of the number ‘4’in Layer C1 becomes less evident in Layer S2 . The convolutional neural network repeats this process until Layer C5 is obtained which contains many more 1×1 characteristic patterns.

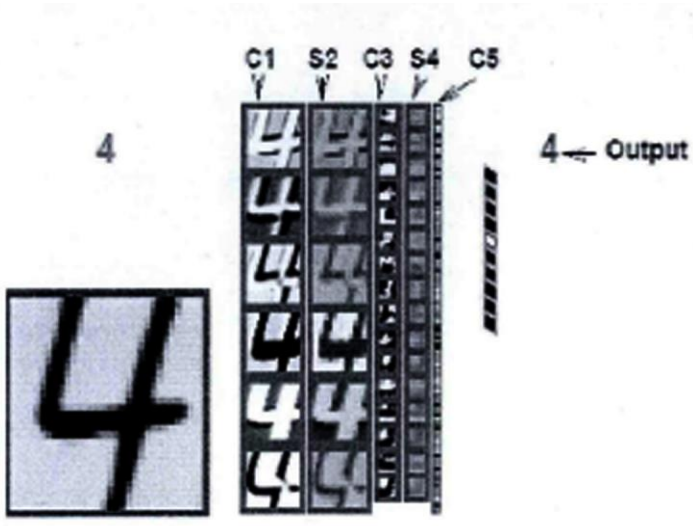

Figure 1. The Diagram of Each Layer’s Characteristic Graphs in the Process of Convolutional Neural Network Classification

4.3 The Structure of the Convolutional Neural Network Models Adopted in This Paper

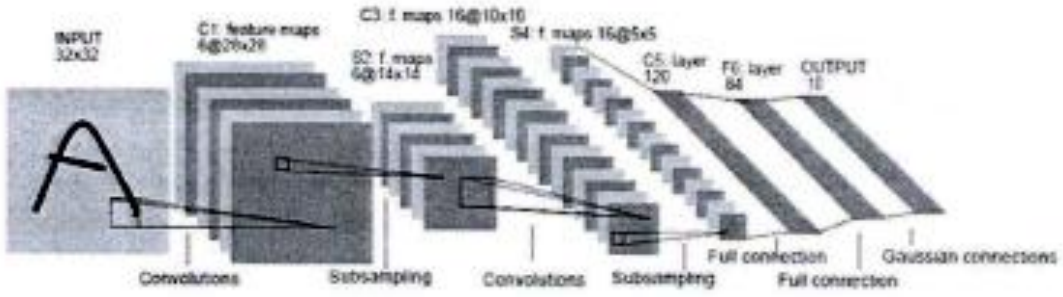

Figure 2. The Diagram of LeNet-5 Network Structure

In recognizing handwritten numerals, we use the network structure similar to that of LeNet-5, the connection mode of which is shown in Figure 2. The input data are the matrixes which are composed of 32×32 pixels. The first feature layer (C1) contains 6 characteristic graphs, which use the 5×5 window for the convolutional operation on the input image. The size of each acquired characteristic graph is 28×28.Then the first down sampling layer (S2) makes down sampling operation on C1 to obtain the same 6 characteristic graphs, but the size of each graph is reduced to 14×14. Layer C3 is a convolutional layer and the convolutional kernel size is 5×5. S4 makes down sampling on the basis of C3. The C5 layer makes convolutional operation on the S4 layer, using the full connection mode i.e. each convolutional kernel in the C5 layer performs convolutional operation in all the 16 characteristic graphs of Layer S4. The C5 layer contains 120 characteristic graphs and the size of each graph is 1×1. The C5 layer finishes the process of feature extraction. On the basis of the C5 layer, the output result of 1×10 is finally obtained through a fully connected network.

4.3.1 Network Model 1 (CNN-1)

CNN-1 uses a structure which is basically similar to LeNet-5, but makes the following modifications: (1)All the activation functions use the sigmoid function i.e. the output values of all layers in the network, including the results of the output layers ,are all in the (0,1) range, while LeNet-5 uses the tanh function. (2)The output layer is connected with the C5 layer, that is, omitting the F6 layer and adopting the full connection mode instead of LeNet5's radial basis function network structure (RBF). (3) In the training process, the learning rate is fixed at 0.002, while LeNet-5 adopts a special learning rate sequence. (4)The original input data size of 28×28 can be expanded to 32×32 by the method of filling frame with 0 (0 frame filling method).

4.3.2 Network Model 2 (CNN-2)

CNN-2 makes the following modifications on the basis of CNN-1. The C1 layer is reduced to four characteristic graphs. Also the corresponding S2 is reduced to four characteristic graphs, too. The corresponding C3 and S4 are reduced to 11 characteristic graphs. C5 is reduced to 80 characteristic graphs. The rest parts remain the same.

Compared with CNN-1, CNN-2 reduces a certain number of filters and the relative characteristic graphs in the C1 layer, the C3 layer and the C5 layer so as to reduce the parameters of the network that can be trained and at the same time restrains the number of filters which the network can learn and the corresponding number of possible extracted features.

4.3.3 Network Model 3 (CNN-3)

In contrast with CNN-2, CNN-3 increases the number of filters in the network layers on the basis of CNN-1. Among them the C1 layer increases to 8, the corresponding C3 layer increases to 24 and the C5 layer increases to 180. The connection between the C3 layer and the S2 layer is chosen according to the same principle as CNN-1 to ensure the combination of the existing main characteristic graphs.

Compared with CNN-1, CNN-3 increases more filters and the corresponding characteristic graphs. The number of characteristic graphs is more than twice that of CNN-2, therefore the related parameters that can be trained in the network greatly increase.

4.4. The Experiment Result

Because there are limited experimental resources, we have adopted a subset of the MNIST Database of Handwritten Digits and from its training set and test set respectively chose 400 data as the training set and 400 data as the test set . It can be seen that compared with the 60,000 training samples contained in the MNIST Database of Handwritten Digits, the 400 samples adopted by us only account for only a small part of it. Besides, compared with the around 4,000-5,000 weights needed to be trained in the network, there are fewer sample data, which make the result finally trained in the network fail to obtain such effect of 1.15% misclassification rate which LeNet-5 has got. But we focus on the influence which the number of filters and other network structures in the network layers of convolutional neural networks have on the recognition performance of the system, especially the recognition and training performances of convolutional neural networks under the condition that there are fewer data which can be provided for training. In the course of real experiment, we adopt the convolutional neural network structure described in this paper and such data scale as 400 training examples and 400 test examples. The training process converges after 15 to 18 iterations and takes 8 to 10 hours.

The experimental content is divided into 4 parts: LeNet-5 , CNN-1, CNN-2 and CNN-3. The training is made on the same data set. Compared with LeNet-5, we simplify the activation function used in the network and adopt the simple full connection as the connection mode of output layer. The depths of CNN-1, CNN-2 and CNN-3 networks are the same. The difference is the number of characteristic graphs in each layer. With the C1 layer as an example,CNN-1, CNN-2 and CNN-3 respectively include 6, 4 and 8 characteristic graphs. By comparing the network's convergence and the misclassification rate, etc through each network's line graphs of misclassification rate in the experimental training process, we can compare such performances as the network's training speed, the recognition rate, etc.

4.4.1 The Experimental Result of LeNet-5 in the MNIST Database of Handwritten Digits

The experimental result shows that the test MCR of LeNet-5 reaches the lowest point: 9% after the 15th iteration. Though the training MCR slightly decreases in the subsequent training process, the test MCR increases after the 18th iteration and afterwards keeps basically stable. So the network is thought to achieve the best effect of training at present upon completion of the 15th iteration, i.e. the network training converges, its training MCR is 1.15% and its test MCR is 9%.The operating results of the common classification methods in MNIST Database of Handwritten Digits are shown in Table 1.

Table 1. The Classification Results of Several Conventional Methods in the MNIST Database of Handwritten Digits

Classifier	Pretreatment	Misclassification rate (%)
The linear classifier of single layer perception	None	15.0[15]
K proximity algorithm	None	7.0[15]
The SVM using the Gauss nuclear	Resistance to distortion	1.25[15]
The three-layer feed forward network containing 500+150 hidden units	None	3.05[15]
LeNet-5	None	1.15[15]
Multi-column deep network	None	0.25[16]

The classification result of LeNet-5 in the MNIST Database of Handwritten Digits is 1.15%, which has a comparatively bigger performance advantage than the other methods. The multi-column deep network in Table 1 is also one of convolutional neural networks. It uses a large number of filters and a competition mechanism between filters to construct networks, which is one of the best methods of experimental results in the general recognition tasks so far.

The misclassification rate of 9% in LeNet-5 in this experiment has a bigger difference with 1.15% in Table 1. The reason is mainly that the networks in Table 1 are under the condition that the training sets in the whole NMIST Database of Handwritten Digits are used as the training samples, and such enormous data as 60,000 training samples and 10,000 test samples ensure the training effect of LeNet-5. But in many classification problems, there are no such sufficient data, which makes the methods of these larger scale network structures fail to obtain good training. In our experiments, we only chose the 400 training samples from them at random, which are mainly used to compare the learning capacity and the effect of the networks under the condition that there are fewer sample data.

4.4.2 The Experimental Result of CNN-1 in the MNIST Database of Handwritten Digits

The network structure of CNN 1-1 is basically the same as that of leNet-5. The main difference is that it doesn't adopt some empirical parameters in LeNet-5 and the final classifier portion is implemented by the simple and fully connected network. The misclassification rate of CNN-1 changes in the training process. The test MCR of CNN-1 after convergence is slightly higher than the test MCR of LeNet-5 and the test MCR and training MCR in the training process are slightly higher than those of LeNet-5.

But at the same time it can be seen that the changing processes of the test MCR and training MCR of CNN-1in the training process are relatively stable. After the test MCR after the 13th iteration reaches the lowest point of 10.5%, the test MCR has no significant fluctuation in the subsequent training process. It can be thought that the network training process at this time becomes convergent and achieves a stable state. The experimental result also shows that the CNN-1 network formed after reducing some empirical parameters in LeNet-5 and simplifying the structure of the output layer, has the same classification performance as LeNet-5. In the training process, its training MCR and test MCR are relatively stable.

4.4.3 The Experimental Result of CNN-2 in the MNIST Database of Handwritten Digits

According to our experiment, the misclassification rate of CNN-2 changes in the training process. The test MCR of CNN-2 reaches the lowest point of 13% after the 11th iteration and the corresponding training MCR is 6.75%. The test MCR of CNN-2 is

slightly higher than that of CNN-1, but the convergence speed of CNN-2 in the training process is faster than that of CNN-1 and it achieves the best network state after the 11th iteration. Compared with it, CNN-1 needs 13 iterations, which owes mostly to the reduced number of filters in each layer in CNN-2 network to make the number of parameters which are needed to learn in the network greatly reduced.

The experimental result shows that in the case of 400 training examples, after reducing the characteristic graphs, the network can better capture the characteristic information of the input data in order to achieve better classification results. At the same time the smaller size of the network can shorten the time required to train the network in the practical application so as to improve the applicability of network.

4.4.4 The Experimental Result of CNN-3 in the MNIST Database of Handwritten Digits

When CNN-3 is compared with CNN-1, the number of filters in each layer of network increases by half, which makes the parameters needed to learn in the network increase at the same rate. Though the number of features that can be learned in the network and the network capacity are increased, more parameters mean needing more sample data for network training. The experiment shows that the training MCR and the test MCR change unstably and often fluctuate in the training process. Though the test MCR touches the buttom after the 14th iteration, then it soon rises and the curve is not stable. In the 400 sample cases of CNN-3 network, the training process cannot converge. The misclassification rate of 21.5% (after the 14th iteration) is much higher than 10.5% of CNN-1 and 13% of CNN-2.

The chief reason why such training process cannot converge and can't get a good classification cognition rate is that there are too many fillers in the layers of CNN-3 models, correspondingly making the learning weights too many which are required for the training process. Compared with too many weights needed to train in the network, the size of the data set that can be supplied in the training process is too small, which cannot meet the learning needs. From the point of the filters and the characteristic graphs, too many filters and insufficient training samples make the network fail to get the stable and effective filter combination through learning and make the training process fail to converge and fail to obtain better classification effects.

The experimental result also shows that under the condition of the limited number of samples, increasing the number of characteristic graphs in the network can increase the number of parameters needed to learn in the network, which possibly makes the network training process fail to reach a relatively stable state and fail to get a better classification recognition rate.

4.4.5 The Comparison of Experimental Results

Table 2 synthesizes some data in the above sections for comparison. CNN-1 is a simplified LeNet-5 network structure and the experimental result shows that this simplified network can also obtain a better classification recognition rate. Based on CNN-1 and formed by appropriately reducing the number of filters in each layer in the network, CNN-2 model can also get a better classification recognition rate, improve the learning speed to some extent and reduce the iterations needed in the network training process. CNN-3 is also based on CNN-1. The experimental result shows that in the case of 400 training examples, by appropriately increasing the number of filters in each layer in the network, CNN-3 can't get very good convergence in the training process and its classification recognition rate is worse than those of the other network models.

Table 2. The Comparison of Several Learning Performance of Four Kinds of Network Structures in the MNIST Database of Handwritten Digits

	Training MCR(%)	Test MCR(%)	The Number of iterations
LeNet-5	1.15	9	15
CNN-1	2.25	10.5	13
CNN-2	6.75	13	11
CNN-3	2.75	21.5	That can't be converged

It needs to be pointed that the training samples of the data set adopted in our experiments are only 400, which are relatively few. In the case of adequate samples and according to the characteristics that convolutional neural networks are layered and extract features from local sensing regions, if the quantity of each layer perception is appropriately increased, the number of features extracted by each layer of the network can be increased. In this way, network recognition performance can be enhanced and the network has better robustness with respect to noise, translation and disturbance. The characteristics of being unable to converge and having the higher error recognition rate shown by CNN-3 mode in the experiments are to a great extent limited by the relatively small number of training samples.

By comparing the above the experimental results of network training, we can see that under the condition of the 400 training samples, i.e. there are not sufficient training samples, CNN-1 can obtain recognition performance similar to LeNet-5. By appropriately reducing the number of filters in each layer of the network to make the network maintain a certain recognition rate, CNN-2 can accelerate the training speed of the network. On the basis of CNN-1, CNN-3 increases a certain number of filters in each layer of the network. Its training process is not well converged and the classification recognition rate is relatively poor.

5. Conclusion

The deep web characterized by convolutional neural networks brings a new research focus to artificial neural networks. In this paper, the authors summarize the latest development of convolutional neural networks, and based on the famous LeNet-5 convolutional neural network, construct some convolutional neural models and apply them in handwritten numeral recognition. We can draw the following conclusions by comparing the experimental data:

1. Convolutional neural networks obtain a relatively good result in recognizing handwritten numerals. In this task, too much additional adjustment work is not needed. In our research, we have only made 2 aspects of adjustment: the transformation of input image sizes and the matching of output unit numbers. The facts show convolutional neural networks can also be used in more different recognition tasks.

2. The number of filters in each layer of the convolutional neural network structure has a relatively great influence on the training speed and the final training effect. Selecting a suitable number of filters in each layer can appropriately shorten the training time and keep a certain recognition rate. But under the condition of insufficient samples, the excessive number of filters can make the network training fail to converge and fail to achieve effective recognition.

Convolutional neural networks have developed so well that they have formed a relatively mature technology and method in many aspects. If we can further accelerate the training speed of the network and improve the performance and the versatility of the models, we can greatly improve the technical level of convolutional neural networks in image recognition and they can be applied in more different fields.

Acknowledgements

This work is supported by the National Natural Science Foundations of China(61402265 and 61170054)

References

[1] D. H. Hubel and T. N. Wiesel, "Receptive Fields, Binocular Interactive and Functional Architecture in the Cat's Visual Cortex", Journal of Physiology, vol.160, **(1962)**, pp.106-154.
[2] K. Fukushima, "Neocognition: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position", Biological Cybernetics, vol.36, **(1980)** pp.193-202.
[3] Y. LeCun, L. Bottou, Y. Bengio and P. Heffner, "Gradient-based Learning Applied to Document Recognition", Proceedings of the IEEE, **(1998)**.
[4] Y. LeCun, B. Boser, J.S. Denker, D.Henderson, R. E. Howard, W. Huberd and L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition.Neural Computation", vol.1, **(1989)**, pp.541-551.
[5] T. Serre, G. Keriman, M. Kouch, C. Cadieu, U. Knoblich and T. Poggio, "A Quantitative Theory of Immediate Visual Recognition, Progress in Brain Research", Computational Neuroscience. Theoretical Insights into Brain Function, vol.165, **(2007)**, pp.33-56.
[6] A. Waibel, T. Hanazawa, G. Hinton, K. Shiano and K. Lang, "Phoneme Recognition Using Time-delay Neural Networks", IEEE Transactions on Acoustics, Speech and Signal Processing, vol.37, **(2005)**, pp.551-556.
[7] B. Kwolek, "Face Detection Using Convolutional Neural Networks and Gabor Filters, Artificial Neural Networks", Biological Inspirations, vol.3696, **(2005)**, pp.551-556.
[8] F. H. C. Tivive and A. Bouzerdoum, "A New Class of Convolutional Neural Networks and Their Application of Face Detection", Proceedings of the International Joint Conference on Neural Networks, **(2003)**.
[9] P. Y. Simard, D. Steinkrans and J .C. Platt, "Best Practice for Convolutional Neural Network Applied to Visual Document Analysis", The 7th International Conference on Document Analysis and Recognition, **(2003)**.
[10] S. Skittanon, A. C. Surendran, J. C. Platt and C. J. C. Burles, "Convolutional Neural Networks for Speech Detection", Inter-speech, **(2004)**, pp.1077-1080.
[11] Y. Chen, C. Han, C. Wang, B. Jeng and K. Fen, "The Application of a Convolutional Neural Network on Face and license Plate Detection", The 18th International Conference on Patten Recognition, **(2006)**.
[12] S. Ji, N. Xu, M. Tang and K. Yu, "3D Convolutional Neural Networks for Human Action Recognition", IEEE Transaction on Pattern Analysis and Machine Intelligence, **(2012)**.
[13] E. M. Kussul, T. N. Baidyk, D. C. Vunsch, O. Makeyev and A. Martin, "Permutation Coding Technique for Image Recognition Systems", IEEE Transaction on Neural Networks, vol.17, no.6, **(2006)**, pp. 1556-1579.
[14] Y. LeCun and C. Corinna, "The MNIST Database of Handwritten Digits. Available".
[15] Y. LeCun, L. Battou, Y. Bengio and P. Haffner, "Gradient-based Learning Applied to Document Recognition. Proceedings of the IEEE, **(1998)**.
[16] D. Ciresan, U. Meier and J. Schmidhuber, "Multi-column Deep Neural Networks for Image Classification", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **(2012)**.

Authors

Xiaofeng Han, he is a lecturer in the College of Mathematics and Systems Science, Shandong University of Science and Technology in China. He has published 7 books and 22 articles. He is now working on his doctorate at Shandong University of Science and Technology in China.

