The Classification of Stored Grain Pests based on Convolutional Neural Network

Dexian Zhang[1], Wenjun Zhao*, [1]

[1]School of Information Science and Engineering, Henan University of Technology, Zhengzhou, 450001,China

Keywords: Deep Learning; Convolutional Neural network; Stored-grain Pests Recognition; Feature Extration

Abstract. With the advent of the era of big data, convolutional neural network (CNN) of deep learning has been widely used in the field of image recognition. The CNN has higher recognition rate and faster extraction speed of characteristic than traditional machines in learning methods. The CNNs theory is introduced for the recognition of stored grain pests in the granary environment. Firstly, stored grain images with pests are normalized and the implicit characteristics are extracted using a trained convolution kernel. Then, the maximum pool method is used to reduce the dimensionality of the extracted features. Finally, the Softmax classifier is used to classify the image of the test sample. The results show that CNN has a good ability to identify reserves and generalization ability.

Introduction

China is a big country of grain, and the security of grain storage is related to the country's economic lifeline. The party's the fifth Plenary Session of the 18th CPC Central Committee 2020 on China's agricultural modernization of the relevant requirements, pointed out the direction for the grain storage industry modernization. To implement the State Council leadership of Comrade Li Keqiang on the "wide grain, grain product is good, the spirit of the important speech of good grain" the State Grain Bureau has officially launched the "grain supply security projects. Grain during storage of harmful stored grain pests is very serious. According to relevant reports, the world every year at least 5% of the food is spoiled by pests, if human, material and technology can not keep up with 20%~30%. In many developing countries, the loss of grain postpartum is about 10%~15%. In China, every year Treasury grain loss is about 0.2%, therefore, the classification and recognition of stored grain pests has become an imminent problem. Only accurate classification, can do to control.

Fortunately, there's deep learning how to solve the automatic learning "quality characteristics" of the problem [1-2].Through the analysis of the mechanism of human brain analysis, it introduces the hierarchical information processing process to the feature representation. The feature representation of the sample in the original space is transformed into a new feature space by the method of layer by layer feature transformation. As a common model of deep learning, the depth convolution neural network has become one of the hot topics in the field of scientific research. Compared with the traditional algorithm, it does not use any artificial input features, avoids the complicated manual feature extraction process, can realize the automatic. CNN is widely used in the field of image recognition in the field of image recognition, because it has the same advantage in large scale image recognition[3].

This paper presents a method of classification and recognition of stored grain pests based on convolutional neural network. Firstly, the pest images are normalized and the implicit features are extracted by using the trained convolution kernel. Then the maximum pool method is used to reduce the dimensionality of the extracted features. Finally, the Softmax classifier is used to identify the pests in the sample image.

Deep Learning and Convolutional Neural Network

On 2006, Dean of the field of machine learning, University of Toronto professor Hinton [4-5] first proposed the deep learning theory, published an article using the deep structure of the neural network model is implemented in Science data reduction papers [4]. This paper expresses two main points: (1) a lot of hidden layer artificial neural network has excellent feature learning ability, learning characteristics of the data has a more essential characterization, which is conducive to visualization or classification; (2) the difficulty of the deep structure of the neural network in training can be used on the "layer initialization" to overcome. Researchers from Stanford University, University of Montreal, New York University and other institutions have published the research results of the deep structure model. The United States Defense Advanced Research Agency(DARPA) was established in 2009, the depth of the learning project group (Project). Google, Microsoft, Baidu and other well-known high-tech companies with big data want to invest resources to capture the depth of learning technology commanding heights. In 2012, Google researchers in X's lab with 1000 computers (16 thousand processors) to construct the world's largest artificial neural network -- "Google brain", from the YouTube video on the site to extract about 10 million still images as training samples. The training of the “Google brain” can learn to recognize the face of the Internet, the human body, and even cats and other categories of images from the Internet video, showing the potential for unsupervised learning. Microsoft Corp also successfully applied the depth learning method to the speech recognition system, which reduced the word error rate by about 30% compared with the previous optimal method. In addition, deep learning has achieved good results in both target and behavior recognition [6-7].

Deep learning is essentially a general term for a class of models that have a deep structure. The depth structure is relative to the shallow structure. The shallow structure model usually contains no more than nonlinear feature transform one or two layers, such as the Gaussian Mixture Model (GMM), Support Vector Machine (SVM) and Multi-layer Perceptron (MLP). Related studies have shown that the shallow structure relative to the internal structure is complex, the constraint is not strong data with good results, but when dealing with complex internal structure of data in the real world (such as voice, sounds of nature, natural images, video, etc.), these models will appear to characterize the ability shortage. However, the deep structure model is characterized by hierarchical representation, which is more powerful than the shallow structure model in the study of the highly nonlinear relationship and complex function representation in the large data.

A typical depth learning model has a Deep Belief Networks (DBN)[5], the Stacked Auto-encoder (SAE) [8], convolutional neural network (CNN) [9-10], etc. DBN consists of a number of structural units stacked, the structure of the unit is usually Restricted Boltzmann Machine (RBM). RBM is the Boltzmann machine is a special form of graph model variables between the connection form is limited, only the connection weights between visible nodes and hidden nodes. There is no connection between the visible and invisible nodes and hidden nodes. The number of visible neurons in each RBM cell in the stack is equal to the number of hidden neurons in the previous RBM unit. DBN automatically learns the level of abstraction by bottom-up, and finally gets the nonlinear description of the feature, presents an automatic feature extraction process which does not depend on manual selection. DBN has been successfully applied in many fields such as handwritten numeral recognition. However, DBN ignores the spatial structure information of the image and the local structure of adjacent pixels, and it is difficult to learn the local features of the image; at the same time, the learning process of DBN is slow, and the improper parameter selection leads to the convergence of learning to the local optimal solution. The structure of the stack type self encoder (SAE) is similar to that of DBN, which is composed of a number of structural units. The difference is that the structure of the unit is self encoder and not RBM. Convolutional neural network (CNN) can be directly used as the input to the pixel value of the image, autonomous learning through the training sample image data, implicitly obtained image features more abstract expression of the image, zoom, rotation review, or other forms of transformation is robust, more suitable for the identification of two-dimensional images.

CNN is a kind of deep neural network which contains the volume of layers. The model is inspired by the research of the brain neuroscience, and imitates the process of the simple cells in the visual cortex and the processing of the visual information by the complex cells. Simple cells respond to the edge information from different directions, and the complex cells accumulate the results of similar simple cells, called Hubel-Wiesel structures [11]. CNN contains a multi-stage Hubel-Wiesel structure. Each stage usually consists of a basic simulation of the convolution operation of simple cells and the pooling of complex cells. In CNN, sub block in the image (local experience area) as the lowest level of the input information, and then transmitted to different layers, each layer by a digital filter to obtain the most obvious characteristics of observed data. This method can obtain the significant features of translation, zoom and rotation invariant observation data, because the characteristics of the local area of the image or feel the abyss of processing unit allowed access to the most basic, such as directional edge or corner.

In recent years, CNN in the ImageNet [12] Large-Scale Visual Recognition Challenge(ILSVRC) has been continuously updated image classification and target location recognition rate record. 2012, Krizhevsky and other [13] will be the first application of CNN in the ILSVRC, the depth of the training of convolutional neural networks in the ILSVRC-2012 challenge, made the image classification and target positioning of the first task. Among them, the image classification task, the first 5 options error rate of 15.3%, well below the error rate of second of the 26.2%; in the target positioning task, the first 5 options error rate of 34%, well below the second 50%. In the ILSVRC-2014 challenge, almost all teams have adopted a convolution neural network machine deformation method. The GoogleNet team uses a combination of multi-scale Hebbian model theory proposed by convolution neural network, first obtain the graph classification “to specify the data” group to the classification error rate of 6.7%; CASIAWS group by using the method of weakly supervised neural network combined with positioning and convolution, made the first graph classification" additional data group, its classification the error rate is 11%. The CNN is first applied to the ILSVRC challenge and achieved outstanding results, to challenge 2014 almost all teams are using CNN learning method based on depth, and the classification error rate is reduced to 6.7%, can see the manual extraction feature of deep learning compared with the traditional method has great advantage in the field of image recognition.

CNN Structure Design

CNN is a feed-forward neural network, which can extract features from a two-dimensional image, and optimize the parameters of the network by back propagation algorithm. In this paper, we design a CNN structure for the identification of stored grain pests, as shown in Figure 1. Not including the input layer, the network consists of 7 layers, which include the 3 layer volume (C1, C2 and C3 layer), 2 layer pool layer (S1 and S2), the 1 layer fully connected layer and the 1 layer Softmax layer. The input layer is the original pixel matrix of 96*96, and the convolution layer and the pool layer (Pooling) layer are connected in some way. Convolution layer C1, C2, C3, respectively, using 32,64,128 convolution kernel convolution operation, each convolution layer using the convolution kernel size are 5*5; pool layer S1, S2 using the sampling window size of 2*2; the fully connected layer contains 300 neurons, which are fully connected with the S2; there is a layer of softmax.

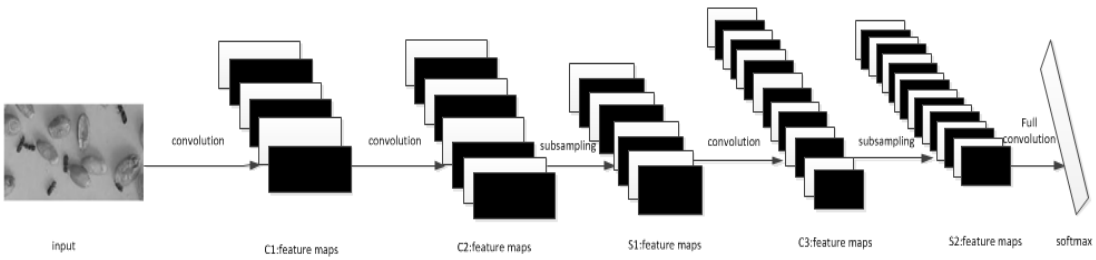

Fig.1. CNN structure of stored grain insect pests identification

Convolutional layer

The statistical characteristics of different sub blocks in natural images usually have consistency, which means that it can be used as the detector from the block to learn a feature image, then traverse the whole image of all sub blocks, to get the other sub blocks the same feature activation value. Convolution layer CNN is the use of the inherent characteristics of the image, with the training of different convolution kernel respectively with a layer of all feature maps of the convolution summation, and biased, then the result is output through the activation function of the formation of the current layer neurons, which constitute the different layer feature map sign. Generally, the expression of the convolution layer is

$$y_j^l = \theta\left(\sum_{i=1}^{N_j^{l-1}} W_{i,j} \otimes x_i^{l-1} + b_j^l\right), j = 1,2,\cdots,M \quad (1)$$

The l layer is the current layer and the l-1 layer is the previous layer; y_j^l represents the current layer of the j feature map; $w_{i,\ j}$ represents the convolution of the j and i features of the previous layer; x_i^{l-1} represents the previous layer of the i feature map; b_j^l represents the bias of the j feature map of the current layer, experiment make $b_j^l = 0$,the network can be trained quickly, while reducing the learning parameters; N_j^{l-1} represents the number of all the features of the previous layer that is connected to the j feature map of the current layer; M represents the number of feature maps in the current layer; $\theta(\cdot)$ represents activation function. In the experiment, we use Rectified Linear Units (ReLU) function [14], rather than the usual sigmoid or hyperbolic tangent function (tanh), because ReLU is more sparse. The expression of the ReLU function is

$$\theta(x) = \max(0,\ x) \quad (2)$$

It has been proved that the network trained by the ReLU activation function has the appropriate sparsity. At the same time, it can well solve the problem that the traditional activation function may disappear in the process of adjusting the parameters of the back-propagation, and accelerate the convergence of the network.

The convolution layer is the feature extraction layer, and the input of each neuron is connected with the local receptive field (i.e. the image sub block) of the previous layer . The convolution layer C1 uses 5*5 convolution to check the input image of 96*96 pixels, which means that each neuron specifies a 5*5 local receptive field, So after convolution operation to get the feature map size (96-5+1) * (96-5+1) =92*92. Through the convolution operation of 32 different convolution kernels, 32 feature maps are obtained, which are extracted from the 32 different feature images. The same characteristic graph of each neuron (the value of sharing right with the same convolution kernel), but they receive from different local receptive fields. The input layer C2 by convolution check 5*5 convolution pool S1 output feature map layer convolution operation, 64 feature maps, each feature map size (92-5+1) * (92-5+1) =88*88. The convolution layer C3 uses 5*5 convolution to check the feature map of the S1 output of the pool, and then obtains the 128 feature maps, each of which is (44-5+1) * (44-5+1) =40*40.

Pool layer (down sampling layer)

The number of feature map increases with the increasing of convolution layers, the characteristic dimension of the learned will increase rapidly, if directly used to learn all the features to train the Softmax classifier, will inevitably lead to the dimension disaster problem. In order to avoid this problem, a feature pool is usually used to reduce the feature dimension. The role of the pool is to use the maximum pool (Max-pooling) operation for the down sampling, so the pool layer is also called the down sampling layer. The down sampling does not change the number of feature maps, but the feature map will be smaller, so that the output of the feature map to translation, scaling, rotation and other sensitivity reduction. If the size of the sampling window is n*n, after the down sampling, the size of the feature map becomes 1/n*1/n. The general expression of pool

$$y_j^l = \theta\left(\beta_j^l down\left(y_j^{l-1}\right) + b_j^l\right) \quad (3)$$

y_j^l and y_j^{l-1}, respectively, the current layer and the previous layer of the j feature map, $down(\cdot)$ represents a down sampling function; the β_j^l and b_j^l represent the multiplicative bias and additive bias of the j characteristic graphs of the current layer, respectively, experiment makes $\beta_j^l = 1, b_j^l = 0; \theta(\cdot)$ represents activation function, the identity function is used in the experiment.

S1 is a layer of the convolution layer of the C2 output characteristics of the image using the 2*2 window to get the next sampling, so the characteristics of the size of the 44*44, the sampling does not change the number of feature maps, so the number of feature maps is 64 . Similarly, the use of the S2 layer of the 2*2 layer of the convolution layer of the output characteristics of the C3 sampling operation, get 128 feature maps, each feature size of 20*20.

Fully connected layer

The connecting layer input should be a one-dimensional array, each feature map of a layer of the pool layer S2 output is two-dimensional array, so the two-dimensional array of each feature map corresponding to the transformation of the one-dimensional array, and then the 128 one-dimensional array connected in series 51200(20*20*128= 51200) dimensional feature vector as a whole, the connection layer of each neuron input. The output of each neuron is

$$h_{w,b}(x) = \theta(w^T x + b) \tag{4}$$

Which, $h_{w,b}(x)$ represents the output value of neurons; x represents the input feature vector of the neuron; w represents the weight vector; b represents the bias, experiment makes b=0; $\theta(\cdot)$ represents activation function, ReLU function was used in the experiment.

The fully connected layer is fully connected with the S2, and the number of neurons will affect the training speed and fitting ability of the network. The results show that when the number of neurons is 300, the effect is the best.

Softmax layer

The last layer of CNN uses Softmax classifier. Softmax classifier is a multi output competitive classifier. When a given sample is given, each neuron outputs a value between 0 and 1, which represents the probability that the input sample belongs to the class. Therefore, select the output value of the corresponding categories of neurons as the classification results.

CNN Parameter Training

CNN is essentially a mapping between input and output, it can learn a lot of mapping relationship between input and output, do not need any precise mathematical expression between input and output. As long as the training of the CNN with a known pattern, the network has the ability of mapping between input and output. CNN is a supervised training, before the start of training, with a small number of different random numbers of the ownership value of the network initialization.

The training of convolutional neural network is divided into 2 stages:

(1) Forward propagation stage. A sample x is extracted from the training samples, and the corresponding class label is y. The x is input to the CNN network, and the output of the upper layer is the input of the current layer. Then through the activation function, the output of the current layer is calculated. Finally, the output of the Softmax layer is obtained $\tilde{y}$. $\tilde{y}$ is a 7 dimensionlal vector whose elements represent the probability that x is assigned to each class.

(2) Back propagation stage, also known as error propagation. The error category label vector y to calculate the output $\tilde{y}$ and a sample of the Softmax level (y is a 7 dimensional vector, only with the category label y corresponding to that element 1, the other elements are 0), method of adjusting the weights and parameters used to minimize the mean squared error cost function.

Analysis of Experimental Results

The experiment is based on the Python language depth learning library Theano, the hardware platform for DELL: Intel (R) Core (TM) i3-3240 CPU, frequency: 3.4GHz, memory: 6GB. Experimental procedure ①The use of digital cameras filming the pest images, color images, each training image are adjusted to 128*128. to avoid overfitting in all image image cropping and normalized as shown in Figure 2 the 96*96 gray training process as the input of CNN. ②Use the Caffe framework to install and configure the Caffe environment. ③ImageNet training to get their own model.

Fig.2. The use of stored grain pests image sample

In order to improve the reliability of the results, 500 cross validation methods were used to divide the image into 5 parts, each of which was divided into 100 parts by the method of 5 cross validation. The numbers are in the beginning of the 3,4,5,6,7, each of which is a class, I choose 20 from each class as a test, and the remaining 80 as a training. In each experiment, 4 of them were used as the training samples, and the remaining 1 were used as the test samples. The recognition experiments were repeated for 5 times, and the average recognition rate was taken as the final recognition performance for the last 5 times. In order to verify the effectiveness of CNN in the identification of insect pests, the experimental results are compared with those of MLP and CNN, and the results are shown in Table 1.

Table 1 Recognition performance comparison of different algorithms on ImageNet

Algorithm	Recognition rate/%
MLP(200-600-300)	86.9
MLP(600-1200-600-300)	91.82
CNN	98.41

In Table 1, MLP (200-600-300) represents a 3 layer MLP network, each containing a total of 200,600,300 neurons; MLP (600-1200-600-300) represents 4 layer MLP layer network, each layer contains a total of 600,1200,600,300 neurons. As can be seen from the experimental results in Table 1, CNN achieved the highest average recognition rate, compared with MLP (600-1200-600-300) increased by nearly 7%.

Conclusions

Deep learning is a very hot research direction, using convolution layer convolution neural network, the basic structure of pool layer and connection layer, can make the network structure of their own learning and extracting relevant features, and make use of them. This feature provides a lot of convenience for many studies, can be omitted in the past very complex modeling process. In addition, deep learning is now in the image classification, image segmentation and other aspects have been very great achievements and progress. On the one hand, the depth of learning application is very wide and versatility, can continue to work hard to extend it to other applications. On the other hand, there are still many potential for deep learning, it is worth exploring and discovering. As far as the future is concerned, although much of the previous discussion is supervised learning (for example, the last layer of the network will be calculated according to the actual value of a loss value, and then adjust the parameters), and supervised learning has achieved great success. Deep learning in unsupervised learning application is likely to be the trend in the future. After all, most people do not know what a name is, either a person or an animal. In the field of computer vision in the future, it is expected that the convolutional neural network based on deep learning will become a very popular network model, and it will make a better breakthrough and progress in more application

research.

Convolutional neural network is a multilayer perceptron in order to identify the two-dimensional image and design, with local perceptions, hierarchical, feature extraction and classification process combining global features of the training, has the following advantages in recognition of stored grain pests: (1) can be input directly to the pixels of the image as autonomous learning through the training sample image data, implicitly obtained image features more abstract expression, avoiding the explicit feature extraction process of traditional recognition of stored grain pests. (2) the connections between neurons is not fully connected, the connection weights between neurons in the same layer and the feature map is shared, the non fully connected network structure and weight sharing to reduce the complexity of network model, reduce the number of weights (training parameters), but also conducive to parallel learning. (3) The down sampling operation of the pool layer enhances the robustness of the convolutional neural network and can tolerate a certain degree of distortion of the image. In this paper, the convolution neural network is used to identify the stored grain pests. The experimental results show that the method has high recognition rate and good generalization ability. The next step is to further optimize the convolutional neural network structure, which will be used to identify the stored grain pests.

Acknowledgement

This work was supported by the National high technology research and development program (863 program)(2012AA101008)corpus, the National Nature Science Foundation of China (61173054, 61428207), the “Twelfth five-year” national science and technology support plan corpus-grain circulation monitoring sensor technology research and equipment development(2013BAD17B04).

References

[1] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks[C]//Acoustics, speech and signal processing (icassp), 2013 ieee international conference on. IEEE, 2013: 6645-6649.

[2] Liu Jian-Wei,Liu Yuan,Luo Xiong-Lin. Research and development on deep learning. Application Research of Computers,2014, 31(7): 1921-1930.

[3] Najafabadi M M, Villanustre F, Khoshgoftaar T M, et al. Deep learning applications and challenges in big data analytics[J]. Journal of Big Data, 2015, 2(1): 1.

[4] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. science, 2006, 313(5786): 504-507.

[5] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.

[6] Choi S, Kim E, Oh S. Human behavior prediction for smart homes using deep learning[C]//RO-MAN, 2013 IEEE. IEEE, 2013: 173-179.

[7] Yin Z, Quanqi C, Yujin Z. Deep learning and its new progress in object and behavior recognition [J][J]. Journal of Image and Graphics, 2014, 19(2): 175-184.

[8] Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising autoencoders[C]//Proceedings of the 25th international conference on Machine learning. ACM, 2008: 1096-1103.

[9] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[10] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code

recognition[J]. Neural computation, 1989, 1(4): 541-551.

[11] Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. The Journal of physiology, 1962, 160(1): 106-154.

[12] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 248-255.

[13] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[14] Lucey P, Cohn J F, Kanade T, et al. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression[C]//Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. IEEE, 2010: 94-101.

[15] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 248-255.

[16] Fei-Fei L, Fergus R, Perona P. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories[J]. Computer vision and Image understanding, 2007, 106(1): 59-70.

[17] Goodfellow I J, Mirza M, Xiao D, et al. An empirical investigation of catastrophic forgetting in gradient-based neural networks[J]. arXiv preprint arXiv:1312.6211, 2013.

[18] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[19] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

[20] WANG X, TANG J, WANG N. Gait recognition based on double-layer convolutional neural networks[J]. Journal of Anhui University (Natural Science Edition), 2015, 1: 006.