Research on Image Classification Method Based on DCNN

Chao Ma[1]

Shanghai Academy of Agricultural Sciences, Shanghai, Shanghai, China
machao@saas.sh.cn

Xianyong Yi[2]

King Abdullah University of Science & Technology, Thuwal, Jeddah, KSA
xianyong.yi@kaust.edu.sa

Chenglong Yu[3]

China Industrial and Commercial Press, Beijing, Beijing, China
yuchenglong@byaic.cn

Shuo Xu[4]

University of Edinburgh, Edinburgh, Scotland, UK
sososoxu@outlook.com

Linyi Li[5*]

Shanghai Academy of Agricultural Sciences, Shanghai, Shanghai, China
*Corresponding author's e-mail: lly@saas.sh.cn

Abstract—**Image classification is a kind of image processing technology, which can recognize different things by the feature information given by pictures. With the rapid development of science and technology and people's higher and higher demand for quality of life, image automatic classification technology has been applied to various fields of development. When we classify the image, the traditional image classification method can not accurately grasp the internal relationship between the recognition objects, and the traditional method also has the limitation of the recognition object's feature expression because of the too high characteristic dimension of the data, so the experimental results are not ideal. In view of the above content, this paper proposes an image detection method based on convolutional neural network. The experimental algorithm mainly refers to deep learning and convolutional neural network. Different from the traditional image classification methods, the deep convolution neural network model can be used for feature learning and image classification at the same time. By improving the structure of each part of the experiment and optimizing the convolution neural network model, the over fitting phenomenon can be prevented, and then the accuracy of image detection can be improved. The experiment on cifar-10 database shows that the improved deep learning model of this method has achieved effective results in image detection.**

Keywords- feature; classification; dimension; convolution

I. INTRODUCTION

With the rapid development of the Internet, the progress we have made is obvious to all, and we are living in an era of information explosion. At home and abroad, instagram, WhatsApp, wechat, QQ, Sina Microblog and other chat software generate hundreds of millions of pictures every day [1]. In our daily life, we will produce a large number of images through different ways of communication. Through these image data, we can receive all kinds of information transmitted through it. In order to get more valuable and effective information, we can rely on convolutional neural network model. Traditional deep learning methods, such as SVM, Bayesian network, time-frequency domain analysis and other deep learning methods [2-3]. In recent years, with the continuous progress of science and technology, image recognition has been widely used in agriculture, medicine, aerospace and other fields; in addition, it also involves in our daily life, such as face recognition, license plate recognition, fingerprint recognition and other technologies. From this point of view, image recognition technology plays an important role in our life, so it is of great significance to be able to effectively identify the useful information in the image. The traditional method of image recognition is divided into image data acquisition and preprocessing, as well as the extraction of image features to describe the effective information that the image wants to express. The most important step in the deep learning is feature extraction.

II. CONVOLUTIONAL NEURAL NETWORK

Convolution neural network is one of the key points of machine learning and deep learning. Convolution is also a special formula in mathematical operation. Convolution neural network is a special single layer or multi-layer neural network. Its structure is shown in Figure 1.
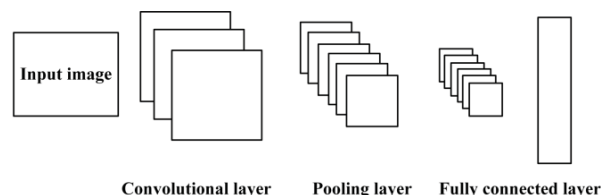

Figure 1. Convolutional neural network structure

Convolution neural network (CNN) is used for feature extraction. It has developed from scale invariant feature transform (SIFT), Harr, histo-program of original gradients (HOG) to now [4-6]. In 1998, Lecun [7] put forward LeNet-5, which combines convolution with neural network. This is a epoch-making and far-reaching convolution neural network structure. Two new concepts of convolution and pooling are proposed. With the development of deep learning theory, Alex Krizhevshy [8] proposed Alexnet in 2012, which is an 8-layer deep convolutional neural network. This model won the title of ImageNet competition in 2012. Since aAlexnet, researchers have proposed various convolutional neural net work models with better and better performance from different network structures, among which the more famous ones are visual geometry group (VGGNet), GoogleNet, deep residual network (ResNet) [9-10].

A. Convolution layer

First, we introduce the convolution layer of the convolution neural network in detail. The main purpose of downsampling layer is to extract the features of massive data image effectively. Convolution layer is the most basic operation in convolution neural network. In the convolution layer operation, the designated neurons in the convolution neural network are connected with the corresponding neuron nodes in the previous layer, and each connection between neurons and neurons corresponds to a weight W (weight), that is, in the convolution neural network, the connection between neurons and neurons is not fully connected. The weights and offsets of the connections between some neurons in the same layer are the same, which can greatly reduce the number of training parameters [11]. The image is transformed into a digital image to get a two-dimensional matrix. The two-dimensional matrix and its corresponding convolution kernel are de sampled. Through different convolution kernels (also known as convolution parameters, convolution kernel or convolution filter), we can get different image characteristic images. The size, number and step size of the designed convolution kernel determine the size of the image characteristic image. In the convolution operation, if the step size of the convolution kernel is large, it will affect the loss of some information of the image. At this time, we need to pad the studied image. The more convolution kernels in the convolution neural network model are gotten, the more effective information will be extracted from the image data. At the same time, it will increase the complexity of the convolution neural network model and prone to over fitting phenomenon.

The output size f of the effective characteristic graph of the downsampling operation and the subsampling operation can be expressed as:

$$f = \frac{n + 2p - w}{s} + 1 \tag{1}$$

The image size studied in the formula is $n \times n$, and the convolution size in the convolution neural network model is $w \times w$, in which the step size is s, and p is the number of pixels for image boundary completion.

B. Activation function

The commonly used activation functions are sigmoid function, than function and rule function [11]. After the first step of convolution operation, the activation function will be used for a nonlinear mapping to increase the expression ability of the whole convolution neural network. Through different research objects, the activation function we used is also different. It has a certain impact on the final convergence speed of convolution neural network experiment. The activation function imitates the characteristics of biological neurons, and accurately represents and outputs the received effective signals. The formula of ReLU function is:

$$rectifier(x) = \max\{0, x\} = \begin{cases} x & if \ x \geq 0 \\ 0 & if \ x \leq 0 \end{cases} \tag{2}$$

In conclusion, we can get a gradient of 1 when x≥0 and 0 when x < 0.

C. Pool layer and full connection layer

Pooling operation is a special convolution operation, which has three properties: pooling operation pays more attention to the invariable characteristics than its specific location. Under the effect of pool operation, the calculation amount of the whole experiment can be reduced as much as possible. At the same time, it can effectively reduce the occurrence of over fitting. The training process of the improved convolutional neural network is a supervised learning method, which is roughly divided into two stages [11-12]: forward propagation for transmitting image message characteristics and back propagation for optimizing the learning model. In the convolution neural network model, the role of the full connection layer is "classification", which is to map the feature information learned from the model to the marker space of the studied object, and then use the output of softmax classifier to predict. The cost function expression of softmax regression algorithm as:

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k}1\left\{y^{(i)} = j\right\}\log\frac{e_j^T x^{(i)}}{\sum_{l=1}^{k}e_j^T x^{(i)}}\right] \tag{3}$$

In the formula, j= 1, 2, 3, ..., k are type labels. When the equation y(i)=j holds, the final value is 1; otherwise, it is 0.

D. Forward and back propagation

Train a convolution with a convolution stride of S, a convolution kernel size of K, and data as a multi-channel image V, defined as C (K, V, S), which minimizes the loss function J(V, K). In the forward propagation process, we can get the

loss function J. To implement this function, we can use this function:

$$g(G,V,s)_{i,j,k,l} = \frac{\partial}{\partial K_{i,j,k,l}} J(V,K) = \sum_{m,n} G_{i,m,n} V_{j,(m-1)\times s,(n-1)\times s+l} \quad (4)$$

During backpropagation, we can get a tensor:

$$G_{i,j,k} = \frac{\partial}{\partial K_{i,j,k,l}} J(v,k) \quad (5)$$

Which needs to find the gradient of V, we can use this formula:

$$h(K,G,s)_{i,j,k} = \frac{\partial}{\partial V_{i,j,k}} J(V,K) = \sum_{\substack{l,m \\ s.i.}} \sum_{\substack{n,p \\ s.i.}} \sum_q K_{q,i,m,p} G_{q,l,n}$$

$$(l-1)\times s + m = j(n-1)\times s + p = k \quad (6)$$

III. MODEL BUILDING

A. Convolution neural network model

In this paper, AlexNet model is mainly used to construct convolutional neural network. At present, this deep learning model is widely used in the field of image processing. As shown in Figure 2, we design the overall framework of the model.
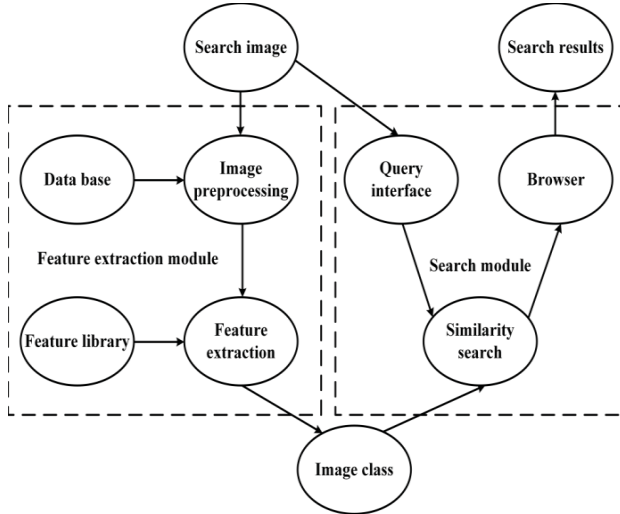

Figure 2.   Model framework

B. Image classification model

The process of image classification experiment includes data preprocessing, network training and parameter adjustment. Preprocessing is to pre train convolution kernel, which is divided into three parts: gray processing of image data, adding noise of certain probability distribution to input data, dimension reduction of PCA data and region distortion of image. The popular stochastic gradient descent (SGD) is used to initialize the weights in the convolution kernel, and the back propagation algorithm is used to fine tune the network parameters. In this way, the convolution kernel can be characterized more essentially and the convergence of the training results is easier. Through repeated forward and backward propagation, the parameters of convolution layer and full connection layer are adjusted continuously, and finally the network is close to the optimal solution. The overall classification model is as follows:
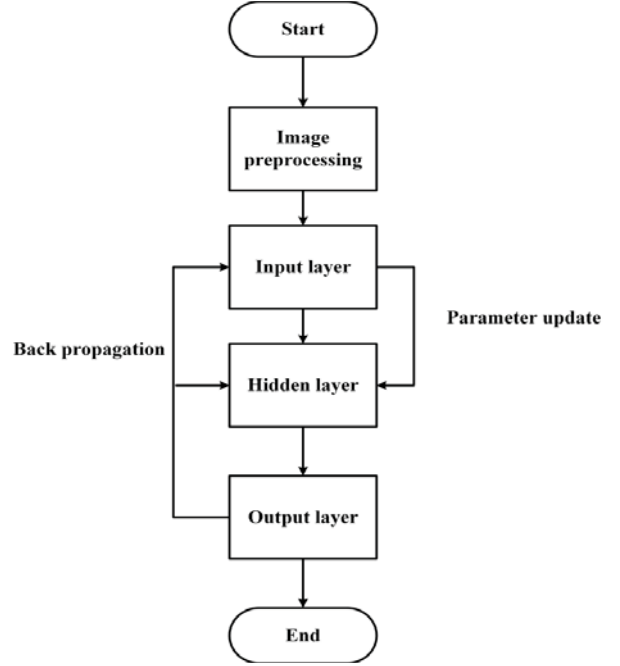

Figure 3.   Overall model of image classification

IV. EXPERIMENTAL ANALYSIS

The learning model uses the CIFAR-10 database for image classification. The database contains 10 types of pictures that can be classified. The dataset has 60,000 color pictures with a size of 32 * 32. Among them, there are 50,000, and 10,000 for experimental testing. The experiment was performed on a Windows computer system, by using deep learning structures and Python for coding. In actual operation, due to the high requirements for the collection standards of pictures and the number of extended data sets, some data omissions will be caused. Therefore, the correct neural network model is selected for training to promote the full development of the work, which can get more guarantees. It will be a problem that the deep convolutional neural network needs to solve later. As shown in Figure 4, after training the convolutional neural network model, the experimental results show that in the convolutional neural network model, the training time and the final classification effect of the obtained effective features through the convolution operation and the pooling operation have the important impact, as the learning time gone that test accuracy gradually improves, it eventually converges to 80.54%.
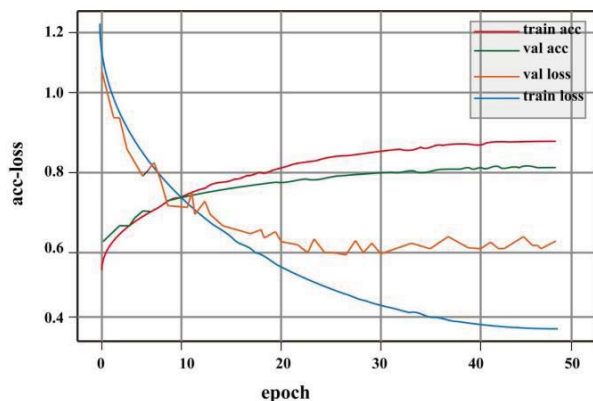
Figure 4.   Experimental results

V. CONCLUSION

Aiming at the disadvantages of the traditional image classification method because the characteristic dimension of the data is too high, the feature expression of the recognition object is limited, and at the same time it relies too much on past judgments, and it is difficult to train. This paper proposes a construction strategy based on convolutional neural networks. The learning model is used for image classification, and the network structure of each part of the experiment is controllable and optimized. At the same time, the pertinence of the model algorithm is strengthened, and the difficulty of the learning model is greatly reduced, and the desired effect is achieved by fewer training times. The experimental results show that the proposed convolutional neural network model is effective. The experimental results of the convolutional neural network model proposed in this paper are better than traditional methods. Of course, this article only classifies things with fewer categories. The training model still has many shortcomings. In future research and learning, we hope to improve the accuracy of classifying things on the premise of shortening the running time.

REFERENCES

[1] Samson, C., Blanc-Féraud, L., Aubert,G., et al. (2000) A Level Set Model for Image Classification. International Journal of Computer Vision, 40:187-197.

[2] Alham, N.K., Li, M., Liu, Y.,et al. (2013) A MapReduce-based distributed SVM ensemble for scalable image classification and annotation[J]. Computers & Mathematics with Applications, 66:1920-1934.

[3] Liu, Y. Guo, J., Lee, J. (2011) Halftone Image Classification Using LMS Algorithm and Naive Bayes[J]. IEEE Trans Image Process, 20:2837-2847.

[4] Lowe, D.G. (2004) Distinctive image features from scale——invariant keypoints. International Journal of Computer Vision. , 60 : 91-110.

[5] Moeskops, P., Viergever, M.A., Mendrik, A.M., et al. (2016) Automatic Segmentation of MR Brain Images With a Convolutional Neural Network. IEEE Transactions on Medical Imaging, 35:1252-1261.

[6] Zhang, S.B., Wang, X.J. (2013) Human detection and object tracking based on Histograms of Oriented Gradients. In: Natural Computation. Shenyang. DOI:10.1109/ICNC.2013.6818189

[7] Lecun, Y., Bottou, L., Bengio, Y., et al. (1998) Gradient-base learning applied to document recognition. Proceedings of the IEEE. 86: 2278-2324

[8] Nguyen, K., Sridha,C., Sridharan, S. (2015) Improving deep convolutional neural networks with unsupervised feature learning. In: Image Processing. Quebec City. DOI: 10.1109 / ICIP.2015.7351206.

[9] Szegedy, C., Liu, W., Jia, Y.Q. (2015) Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston. DOI: 10.1109 / CVPR.2015.7298594.

[10] Lu, J.W., Liong, V.E, Wang, G., et al. (2015) Joint Feature Learning for Face Recognition. IEEE Transactions on Information Forensics and Security., 10: 1317–1383.

[11] Gao, S.H., Zhang, Y.T., Jia, K., et al. (2015) Single Sample Face Recognition via Learning Deep Supervised Autoencoders. IEEE Transactions on Information Forensics and Security., 10: 2108–2118.

[12] Lu, J.W., Wang, G,, Zhou, J. (2017) Simultaneous Feature and Dictionary Learning for Image Set Based Face Recognition. IEEE Transactions on Image Processing., 26: 4042–4054.