Natural Image Classification Method based on Deep Learning

Mingwen-Chi
Baotou Light Industry Vocational Technical College， Baotou,014030，China
chimingwen@163.com

ABSTRACT
In this paper, natural image classification based on deep learning is studied. Image classification is an important research direction in the field of computer vision. With the rapid development of Internet and mobile terminals in recent years, the number of pictures in various social networking sites is growing geometrically, However, the diversity and disorder of these images make it difficult to obtain the effective information completely. Convolutional neural network is an important application of deep learning in image processing. Compared with other machine learning algorithms such as SVM, it can convolute image pixels directly and extract features. It can also use massive image data to train network parameters to achieve better classification effect.
Keywords: *Deep learning; Image classification; Deep convolutional neural network*

1 INTRODUCTION

At present, with the improvement of computer computation and running speed, it is possible for image classification algorithm based on neural network learning. Therefore, in order to find a more accurate, effective and fast image classification technology, a large number of researchers are actively engaged in this research. As a popular machine learning algorithm, the preprocessor of deep learning is neural network. At that time, neural network did not give a strict definition. Its basic feature is to imitate the biological processing between brain neurons for its own learning. The brain stores things and concepts not in a single nerve cell, but in the whole nerve cell. Deep learning has the advantage of distributed representation of image information[1]. Therefore, learning image features and realizing accurate image classification based on deep learning algorithm has become a research hotspot of machine learning.

Image classification technology is a wide-ranging subject research field, which covers many fields such as UAV detection, intelligent machine and object dynamic monitoring. Its research purpose is to extract useful information efficiently and quickly from massive data, and replace manual selection with machine recognition, which will save a lot of workload for human beings[2]. The research on high accuracy natural image classification can not only help human get image information quickly from computer, but also provide basic guarantee research for advanced machine behavior such as object detection and recognition, human-computer interaction and so on. At the same time, image classification technology has great research significance in the fields of intelligent security system, UAV and so on.

The pattern of feature extraction and classifier continued until 2012, and great changes have taken place in the whole field of computer vision. Deep learning has made unprecedented breakthroughs in the fields of image classification, image retrieval, target detection, target tracking, face recognition, text recognition, semantic segmentation and so on. The main reason is that the work of neural network completes the work of feature extraction and classifier at the same time. Through the end-to-end operation mode, these two parts of work can cooperate with each other, so as to make the extracted features more discriminative[3]. Therefore, deep learning can surpass the previous artificial feature extraction algorithms in various fields. The research on high accuracy natural image classification can not only help human get image information quickly from computer, but also provide basic guarantee research for advanced machine behavior such as object detection and recognition, human-computer interaction and so on. At the same time, image classification technology has great research significance in the fields of intelligent security system, UAV and so on.

2 RELATED WORK

2.1 Research status of image classification

Image classification technology simulates the process of human image recognition through the quantitative analysis of images by computer. At first, it is a text-based image classification method, which mainly manually labels the images, and then classifies them by text keyword matching. This method consumes a lot of human and material resources, and the classification effect depends on the subjective consciousness of the labeling person. The content-based image classification technology was proposed in the 1990s. It mainly uses the similarity matching method to classify the visual features of the image, such as texture, color, shape and spatial information. However, it is difficult to understand the deep meaning of the image in practical application and can not meet the new requirements of image classification.

In 2000, the bag of words (bow) model and its improved algorithm were the most used image classification methods. The whole image was described mainly by manually extracting local features and feature coding composition. The model took the extracted image features as a group of vocabulary and summarized the occurrence frequency of each word in the image, After that, the frequency histogram is used as the feature representation of the image[4]. Common local feature extraction algorithms include: in 1999, David Lowe and others proposed scale invariant feature transform (SIFT), which extracts the scale, position and direction information by constructing the scale space and finding the extreme points, so as to obtain more stable local features. In 2005, DALAL et al. Proposed the histogram of oriented gradient (HOG), which is characterized by summarizing the gradient direction histogram of local areas in the image and applying it to pedestrian detection. In 2006, Herbert Bay and others proposed surf (speed up robust features) feature after improving the SIFT feature. The calculation speed is greatly improved. At the same time, its matching effect is better when the brightness changes. In the same year, s Lazebnik and others proposed spatial pyramid matching (SPM), which divides the image into regional blocks of various sizes according to space, extracts the features respectively, and adds spatial information, which greatly improves the feature description ability of the word bag model.

Most of the image classification technology is the research of machine learning. It mainly uses the training data to construct the statistical model, which makes the computer have the ability to predict and analyze the untrained data, and there are many effective classification algorithms. Common classification algorithms include linear regression, decision tree, neural network, support vector machine, Bayesian classifier, ensemble learning and clustering. These classification algorithms are good for simple images, but they are not suitable for complex image classification. In recent years, with its powerful modeling and data representation ability, deep learning has rapidly become a research hotspot of computer vision, pointing out a new direction for us in the field of image classification.

2.2 Deep learning

There is no strict formal definition of deep learning. It is not only the general name of a series of multi-layer network structure models, but also a part of machine learning. Its basic feature is to imitate the transmission and processing of information between brain neurons. A calculation model to be divided into neural networks, as shown in Figure 1, usually requires a large number of interconnected nodes with the following two characteristics. First, each node needs to solve the weighted input values from other adjacent nodes through a specific activation function; Second, the so-called weighted value is used to define the intensity of information transmission between nodes, and the algorithm adjusts the weighted value through continuous self-learning.

By simulating the multi-level abstraction of brain and learning things with multi-layer neural network, the abstract expression of certain things or data is realized. In recent years, scholars at home and abroad have proposed a variety of deep learning models. The so-called depth and shallow learning model refers to the number of levels of network learning model structure[5]. In terms of data expression ability, the multi-level structure model has stronger expression ability. When it is applied to the field of image classification, it means that the depth model can train more data than the shallow model, And the classification effect is more accurate.

(1) the essence of deep learning

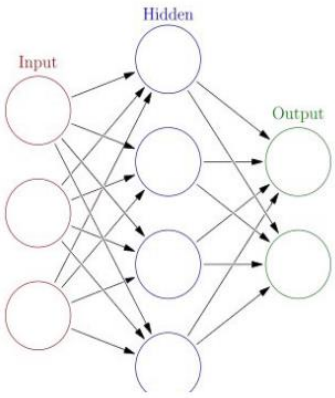

Fig. 1. Neural network model

The essence of deep learning is to express abstract data features by using multi-layer nonlinear units, or to express complex functions by using multi-layer nonlinear network structure. In image processing, it is necessary to learn image features from a large number of samples[6].

(2) convolutional neural network

Compared with traditional methods, convolutional neural network has the advantages of strong generalization ability, simultaneous feature extraction and classification, and strong applicability. It has become one of the research hotspots in the field of deep learning. In this chapter, we will analyze convolutional neural network from the following aspects

(3) neural network

Firstly, a neural network composed of only one "neuron" is introduced, as shown in Figure 2

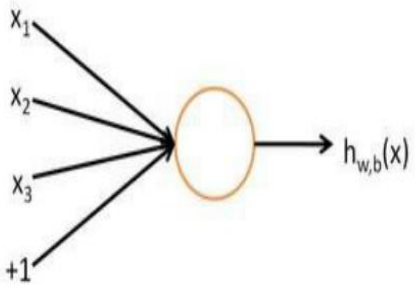

Fig. 2. Simple neural network

The corresponding formula is as follows:

$$h_{w,b}(x) = f\left(w^T x\right) = f\left(\sum_{i=1}^{3} w_i x_i + b\right) \quad (1)$$

Where, X_1, X_2, X_3 and intercept + 1 are input values, and the output is $h_{w,}b(x)$, f (-) is "activation function". Generally, sigmoid function and tanh are selected as activation functions, sigmoid function and tanh], and their corresponding formulas are as follows:

$$f(z) = \frac{1}{1+\exp(-z)} \quad (2)$$

$$f(z) = \tanh(z)\frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (3)$$

The input-output mapping of a single neuron is essentially a logical regression. The neural network model is formed by connecting several single "neurons" in the form of hierarchical structure.

In image processing, we usually represent the image as a vector of pixels, such as a 1000 × 1000 image, can be expressed as a 1000000 vector. In the neural network, if the number of hidden layers is the same as that of input layer, that is, 1000000, then the parameter from input layer to hidden layer is 10 ^ 12, so it is difficult to train the neural network. So we must reduce the number of parameters to speed up the training of neural network.

Convolutional neural network can reduce the number of parameters in two different ways. The first way is local sensing. In the spatial connection of images, the local image pixels are closely related, while the remote ones are weakly related. Generally speaking, people's perception of the outside world is also from local to global. Therefore, each neuron only needs to perceive the local image instead of the global image[7]. Finally, the global information can be obtained by synthesizing the local information at a higher level. As shown in Figure 3 below

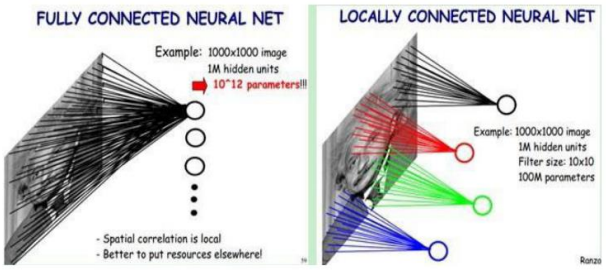

Fig. 3. Full link and local link neural networks

3 DATA ANALYSIS

3.1 Image feature extraction method

The features extracted from images are usually divided into two categories: low-level visual features and high-level semantic features. The underlying visual features include simple features such as shape, color and texture, as well as some complex local invariant features such as surf, sift, hog and so on. High level semantic features are the recognition and interpretation of image content, which need the help of human understanding and reasoning. They are more abstract, such as human behavior features, emotional features, face analysis, etc. they are also extracted and learned from the underlying features. This section will introduce three low-level visual feature extraction methods.

(1) Color features

Color is the most intuitive visual perception of an image, which is related to the specific targets and scenes in the image. Different kinds of targets may have different color features, and color images need to express their color features in a certain color space, such as RGB and HSV. The stability of color features is good, and it remains unchanged for rotation, scale change and translation. It has strong robustness and simple calculation. Its description methods include color set, color histogram and color moment.

(2) Shape feature

Shape is a relatively fixed image feature and will not change with other image features. It is divided into two extraction algorithms: contour based and region based. The former describes the contour of the target area in the image, including Fourier shape description, boundary moment, chain code, etc. the latter describes all the areas of the target. The common representation methods include dispersion, geometric invariant moment and eccentricity.

(3) Texture feature

Texture features belong to global features, which can describe the contour properties of corresponding targets in the image, are not affected by color and brightness, have no rotation deformation and strong resistance to noise. Texture feature extraction methods are divided into statistical method, structural method, model method and signal processing method, such as gradient histogram method (HOG) based on statistical method and local binary mode method (LBP) based on structural method, the calculation method is as follows:

$$E(t)\dot{x}_{d+1}(t) - E(t)\dot{x}_{k+1}(t) = E(t)\Delta\dot{x}_{k+1}(t) = f\big(t, x_d(t)\big) + B(t)u_d(t) - f\big(t, x_k(t)\big) - B(t)u_k(t) = f\big(t, x_d(t)\big) - f\big(t, x_{k+1}(t)\big) + B(t)\Delta u_{k+1}(t) \quad (4)$$

3.2 *Classification of fine image based on convolution neural network*

With the image data becoming larger and more complex, the need for computer to accurately classify it has become more and more complex. Through research, it is found that shallow convolution neural network is far from enough for this kind of complex image data, and deeper convolution neural network is needed to meet its needs. Because the deeper the network, the more comprehensive the characteristics of learning.

Fine grained image classification is one of the research focuses of this paper. It is between semantic level image classification and instance level image classification. It is used to distinguish the sub categories of objects without subdividing them into individuals. For example, in Cub database, the subject of all images in the whole database is birds, The goal of classification is to distinguish 200 categories of birds according to their biological names. Because of this, the research in this field often draws on the research results of the first two fields[8].

Figure 4 shows the early classification framework based on response graph proposed by Li Fei Fei team in 2012. In this framework, the author abandoned the previous ideas based on codebook and annotation, and adopted the method of randomly selecting local bounding box instead of the bounding box in annotation, On this basis, the response maps of each image for different templates are drawn, and then the feature expression of the image is obtained through these response maps. Finally, the extracted features are classified by SVM. At the same time, many algorithms also adopt similar processing methods[9]. The advantage of this method is that it can reduce the additional database cost caused by manual annotation, so as to reduce the training cost of fine-grained image classification network, So that fine-grained image classification can be applied to people's daily life faster.

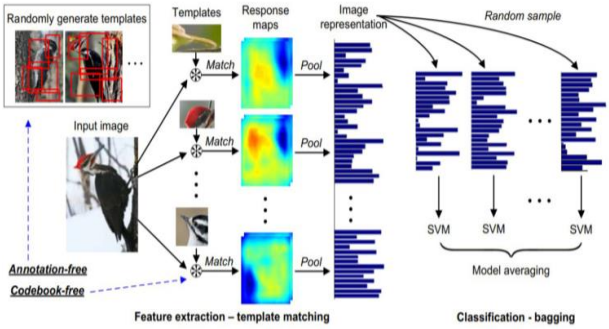

Fig. 4. Fine grained image classification algorithm

It can be seen that the introduction of deep learning technology has brought a significant change in feature extraction for fine-grained image classification. The original method based on artificial features has been rapidly eliminated, and replaced by the feature map extraction technology based on neural network. However, through the previous description, we can still find that many ideas of innovation in this field can still be used. The reason is that neural network is only a tool, not a final solution. In order to achieve better results, we should conform to the characteristics of the database itself, and refer to the methods used by researchers in similar databases to build a classification framework, instead of blindly seeking the expansion of the database and building a more complex network structure.

4 EXAMPLE ANALYSIS

The quality of extracted features will directly affect the accuracy of subsequent image classification, so this is the focus of digital image processing research. In traditional image processing, feature extraction algorithm is based on very complex mathematical formula, so researchers have to spend a lot of time to get good features, such as hog, sift and so on. Participants need to verify and debug the extracted features repeatedly, which will increase the workload. The current deep convolution neural network can make the process of image feature extraction easy and simple, but most of us don't know what the deep convolution neural network learned before massive data. The

content learned from massive data is like a black box. Although it can record useful information, we can't get it. In order to enable the network to automatically extract features that are beneficial to classification tasks, convolution neural networks propagate forward the features extracted from the convolution of the image and propagate the difference between the network output value and the data label back-propagation, and then adjust the network parameters. It has been mentioned that convolutional neural network itself has a black box property, which will increase the difficulty of optimizing network parameters. However, if the features learned by convolutional neural network in each network layer can be displayed in the form of images, it will help the participants to optimize the network parameters more conveniently and quickly[10].

After the input image is processed by convolution neural network convolution layer conv1, the output feature map is obtained. Through the visualization image after the convolution layer conv1, we can clearly see that a lot of information learned by the convolution layer conv1 is reflected in the edge contour information of the input graph. From different visual angles. As shown in the figure, the first and second columns are more likely to observe the input image from the left, the third and fourth columns are more likely to observe from the front, and the fifth and sixth columns are more likely to observe from the right. Through these conclusions, it can be predicted that the edge contour information of the cat shown in the figure will be learned by the convolution kernel of the convolution layer conv1 after a lot of training. It further shows that the theory of image edge extraction is suitable for the learning method of the underlying convolution layer. That is to say, based on the principle of extracting the contour information of the input image, the contour information of different directions can be extracted by convolution operators of different angles. The contour information in the vertical direction can be extracted by the vertical gradient operator, and so can the horizontal direction. Since it is known that the convolution kernel of the convolution layer conv1 will mainly learn the edge information of the image object, we should pay attention to the following when designing the convolution layer conv1:

(1) enrich the number of convolution kernels. Because there is only slight difference in the viewing angle of the adjacent input images, and the viewing angle difference is increased when the distance is longer. The number of convolution kernels represents to look at objects from different angles. Furthermore, the more convolution kernels, the more feature information will be extracted, which is very helpful to the classification results.

(2) The more convolution kernels, the better. If the number exceeds the upper limit, redundancy will be generated, and different convolution kernels may extract feature information from the same angle. Moreover, there is a contradiction between the number of convolution kernels and the training speed of depth network, which needs to be designed reasonably.

5 CONCLUSION

With the increasing amount of image data, it can quickly extract image features from the mass data and identify them with the highest accuracy, and then image classification is bound to become one of the important research topics to obtain image information. It uses computer to replace human visual reading, which lays the foundation for the subsequent intelligent object detection and tracking. Therefore, the research of natural image classification technology is very necessary. In this paper, the key technology of image classification is studied, and the research status of image classification technology at home and abroad is expounded. The convolution neural network in deep learning is analyzed emphatically. The limitation of shallow learning is analyzed, and the advantages of deep convolution neural network in image classification are found. Its advantages mainly lie in that it can extract global features from massive image data, and can fully train the parameters of each network layer, so that this paper can optimize the network parameters of each layer, and finally achieve better image classification accuracy.

REFERENCES

[1] C.Szegedy,W. Liu，Y. Jia.Going deeper with convolutions[RJ. In ar Xiv:1409-4842, Sept.17,2014.

[2] O. Shamir，T.Zhang.Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes[C]. In ICML,2013:530-597.

[3] J. Schmidhuber. Multi-column deep neural networks for image classification[C]. CVPR,2012:656-732.

[4] M. Lin, Q.Chen, S. Yan. Network in network[C].ICLR, 2014:207-312.

[5] C. Corinna and V. Vladimir. Support-vector nerworks[J]. Machine Learing, 1995,20(3):273-297.

[6] I S. P. Daryle Niedermayer. An introduction to Bayesian networks and their contemporary appications[C]. Innovations in Bayesian Networks, 2008:452-524.

[7] K. Andrej, S. Sanketh, T. George, S. Rahul, L. Thomas and F. Li. Large-scale video classification with convolutional neural networks[C]. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2014:322-404.
[8] S. Lazebnik, C. Schmid and J. Ponce. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories[C]. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. 2006:2169-2178.
[9] T. Zhang, B. Ghanem, S. Liu, C. Xu and N. Ahuja. Low-rank sparse coding for image classification[C]. In Proceedings of the IEEE International Conference on Computer Vision, 2013:281-288.
[10] K. P. Bennctt and E. J. Brodenstcincr. Duality and geometry in SVM clasifiers[C]. In Proceedings of the 17th International Conference on Machine L earning, 2000:57-64.