2022 4th International Academic Exchange Conference on Science and Technology Innovation (IAECST) | 979-8-3503-2000-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/IAECST57965.2022.10061910

Research and Implementation of the Algorithm of Flower Recognition Based on Deep Learning

Jie Zhang [1]
[1]Xiamen Institute of Technology,
Xiamen,Fujian,China
123871298@qq.com

***Abstract*—In the field of flower agriculture, it is often necessary to identify and process some flower varieties and monitor their growth status, but the classification and monitoring work can only be carried out smoothly under the guidance of professional technicians, which is very inefficient and greatly restricts the development of flower agriculture. Because of the similarity between different flowers and the difference of the same kind of flowers, it is difficult to solve the problem using traditional image classification methods. This paper focuses on flower recognition, which is a kind of non rigid object. By studying the new algorithm of machine learning, a deep convolution neural network model (AlexNet) algorithm for flower recognition is designed.**

Keywords-Deep Learning;AlexNet;Classification of Flowers

I. RESEARCH BACKGROUND AND SIGNIFICANCE

China is one of the countries with the largest number of flowers in the world. How to quickly identify flower species has always been an important subject in the field of botany. The earliest flower research in China can be traced back to about 3000 years ago to the late Neolithic Age. Today, it has a long history, so China is also one of the countries that started flower research very early in the world. At present, hundreds of thousands of flower species have been found in China, which is an important species resource wealth in China.

Flower classification is based on botanical classification. In general, we use early flower classification methods. First, we need to observe and understand the life habits of flowers, and observe and understand the overall morphological characteristics of flowers, including flower stamens, flower texture, flower color, flower shape, and sometimes plant information. Then we need to compare with specimens that have been collected and recorded, Finally, determine the genus of flowers. This method usually takes a long time, and needs to be carried out in the hands of researchers who have rich experience in flower classification and professional knowledge, with obvious limitations.

Deep learning is a branch of artificial intelligence and the hottest research direction of intelligence at present. With the development of computing, the structure of deep learning network with more hidden layers becomes more and more complex, and the layers become deeper and deeper. As soon as the convolutional neural network model trained by the depth learning algorithm came out, it made a significant breakthrough in many large-scale recognition tasks in the field of computer vision. After 2012, the error rate of deep learning in the field of image classification has continued to decline, and has been leading other classification methods. The essence of deep learning is to learn features from a pile of images and classify them by machines, then mark the corresponding labels of each image, and manually identify the problems of objects in the images. The most studied recognition problem at home and abroad is the recognition of rigid objects, while the recognition of non rigid objects such as flowers is still relatively small. The researchers expect to reduce the heavy work brought by early flower classification by using some methods of computer vision, and make the traditional flower classification work simple by using the fast image processing ability of the computer.

Machine learning combined with image processing recognition technology has been widely used in all aspects of life. Image recognition refers to the technology of recognizing different new targets by understanding and analyzing some characteristics of images extracted by computers. [1] With the rapid development of mobile terminal information science and technology, people have more and more ways to obtain flower photos and images. The simplest is that people can directly take pictures of various flowers through mobile devices such as mobile phones and digital cameras, but most people can not identify various flower types professionally. We still need the knowledge and guidance of botanists to classify and identify flowers. Traditional image classification methods cannot solve the problems caused by the complexity of flower images. The complex environment of flowers, the similarities of different flower species and the differences of the same flower species will all affect flower classification. In recent years, thanks to the rapid development of machine learning, using machine learning to recognize flower images can recognize and classify more deeply and more representative image features. The recognition effect is higher and more robust than traditional feature extraction methods. Among them, the depth convolution neural network model (Alex Net) has developed rapidly in the image field. Alex Net has the unique advantage of extracting deep information from images. It has a good application prospect to apply Alex Net to flower recognition to guide agricultural development. [2] This paper focuses on flower recognition, which is a kind of non rigid object. By studying the new algorithm of machine learning, a deep convolution neural network model (AlexNet) algorithm for flower recognition is designed.

II. DEEP LEARNING MODEL

The deep learning model was proposed by Geoffrey Hinton,the leader of computer and artificial intelligence, and his students in 2006. They made the following two aspects of research by analyzing the shortcomings of traditional neural networks, including that traditional neural networks need good enough features as input and the training difficulty of neural networks with multiple hidden layers: [3]

The neural network under the condition of multiple hidden layers can independently learn more essential features from the input image. These high-dimensional features have a more essential description of the original image than the manually designed low-dimensional features, and can better perform image classification.

Neural networks with multiple hidden layers are prone to fall into local minima during training, which can be solved by layer by layer initialization.

Deep learning is trained through massive data and more non-linear combined computing units. Such a multi-layer model can reduce the function complexity of the traditional model, and gradually increase the difficulty through hierarchical methods, so that the model can analyze more complex classification patterns. [4] Compared with the traditional shallow model, the advantage of the deep learning model is that it can change the image of the data layer by layer, making the feature space constantly changing, and the data features can be more easily recognized.

In this paper, a deep learning model is used to study flower recognition of non rigid objects. A depth learning system includes three distributions, namely, the input layer, hidden layer and output layer. The mark of whether a model is a depth model is that the model has multiple hidden layers as the middle layer. Extracting various image features with essential description significance through the hierarchical relationship between multiple hidden layers is the biggest advantage of the depth learning model. Using the depth learning model no longer requires manual design and extraction of complex image features, The whole model will automatically extract reasonable features for image classification according to the parameters. In addition, neural networks with multiple hidden layers are more consistent with the structure of human brain neurons, and such system structure is more suitable for classification and recognition of complex things.

III. AUTOMATIC FLOWER CLASSIFICATION SAMPLE

Now people have more and more means to obtain flower images, and it is easier to collect flower images. Due to the complexity of the environment background where the natural flower images are located, the inherent attributes of flowers include the similarity of different kinds of flowers and the difference of the same kind of flowers. This inherent attribute makes it difficult for flower images to be studied in image recognition, among which the image segmentation effect has a greater impact. The quality of the image segmentation effect will affect the extraction of image features, thus affecting the final classification results; In addition, in flower classification, image feature extraction is particularly critical. The traditional method is to manually design and select single or multiple features or multi feature fusion. The accuracy of feature selection will also affect the final result. The predecessors of flower image classification research have also made great achievements through continuous efforts and methods improvement.

The flower data used in this paper is the most popular image dataset, ImageNet, which started in 2009 and is organized according to the hierarchical structure of WordNet. In ImageNet, the goal is to show that each synset provides an average of 1000 images. Each concept image is quality controlled and human annotated. At present, there are 14197122 images in ImageNet, which are divided into 21841 categories. From this dataset, 80 flower images of common categories are extracted as the training set of depth learning model. There are 350 images of this category available for use in each type of image, 300 of which are used as the training set, and the remaining 50 images are used as the test set. Figure 1 shows some flower image samples.

Figure 1. Some flower image samples

IV. FLORAL RECOGNITION MODEL BASED ON ALEXNET DEEP LEARNING

A complete convolutional neural network structure includes input layer (INPUT layer), convolution layer (CONV layer), activation layer (RELU layer), pooling layer (POOL layer), and full connection layer (FC layer).

1) Input layer

The input layer is the bottom layer of each layer of the convolutional neural network. Its function is to process the input data, such as removing the average value. De averaging means that the dimension of the input data is concentrated to zero, reducing the excessive deviation in the data, and improving the training effect.

2) Convolute layer

In convolutional neural network, convolution layer is the core layer of the whole network. The main function is to sum linear products.

3) Activate Layer

Relu activation function is generally used in convolutional neural network. The main reason for using the Relu activation function is that Relu converges faster than Sigmaid.

4) Pooled layer

Pooling means that the average area is the largest. Its principle is shown in Figure 2. Take the maximum value of the 2*2 matrix of the upper left corner, upper right corner, lower left corner, and lower right corner to obtain the results of 6834.

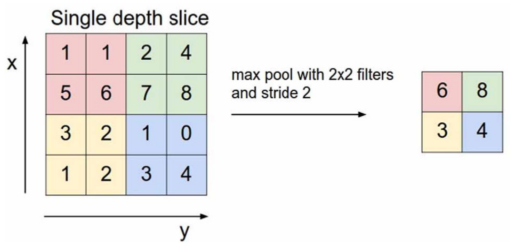

Figure 2. Pooling layer schematic

5) Full connection layer

The full connection layer is the layer in the neural network that can be classified, which can be simply understood as a classifier. If the operations of the convolution layer, pooling layer and active functional layer map the original data to the hidden layer feature space, the full connection layer is used to map the learned "distributed feature representation" to the sample label space.

Convolution operation is the most important operation in convolutional neural network. This operation can recombine the previously extracted features through the rules of convolution kernel to obtain more complex features. As we all know, the complex features in an image can be combined by using simple features in some way. In the convolution neural network, the convolution operation can change a single image feature into a more complex image feature. The more complex the image feature, the more it can be used for classification. Digital images are stored by two-dimensional discrete data. If the original image is f (x, y) and the convolution kernel is g (x, y), the convolution can be defined as:

$$f_{M-1N-1}(x,y)*g(x,y)= \sum_{m=0}\sum_{n=0} f(m,n)g(x-m,y-m)x=0,1,2...M-1,\ y=0,1,2...N-1 \tag{1}$$

The above formula can be used to calculate the convolution of two-dimensional images, which can be mapped to the corresponding convolution obtained by continuously sliding the convolution window.Figure 3 shows the convolution operation between the image and the convolution kernel. In this paper, the convolution operation of the first convolution layer uses 256 different convolution cores, and 256 different convolution results will be generated in the second convolution layer for subsequent convolutions.

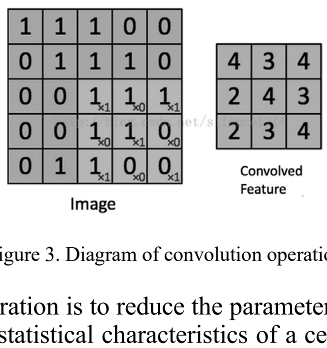

Figure 3. Diagram of convolution operation

Pooling operation is to reduce the parameters in all areas by calculating the statistical characteristics of a certain area as the representative of the characteristics of the area. Because static images generally have aggregation attributes, features in one area will also have certain effects in other areas. Therefore, for more neural network parameters, we can aggregate multiple parameters to get fewer parameters, but this parameter can still represent the specific characteristics of the region.

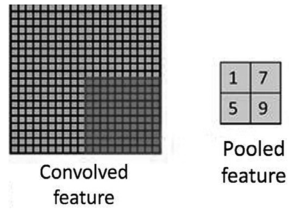

Figure.4 Pool operation

This paper adopts mean pooling. After the convolution results of the first, second and fifth layers, a pooling layer is added to reduce the training parameters from 1 million to about 600000. The experimental results show that the pooled neural network has less over fitting and is more suitable for image classification and recognition.

This paper selects the deep convolution neural network model (AlexNet), and adopts the neural network architecture of 5 convolution hidden layers and 3 full connection layers. Input 3 channel image 224*224*3D data, extract 4096 dimensional features as output through neural network, and finally get 80 recognition results, corresponding to 50 flower image categories. Figure 5 shows the convolutional neural network architecture used in this paper.

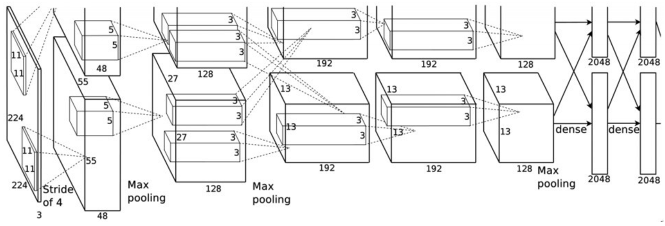

Figure 5. The architecture of convolutional neural network

LRN and maximum down sampling layer are added after partial convolution layer. The input of the network is a 224 × 224 × 3. The feature map needs to integrate the results of the upper and lower GPU. Dropout technology sets the output of the hidden layer neuron to zero with a certain probability, as if the neuron itself has been deleted, and the neuron will no longer participate in the forward and back propagation process of the network. Dropout weakens the dependency between neurons, reduces over fitting, and enhances the robustness of network learning. The Dropout diagram is shown in Figure 6.

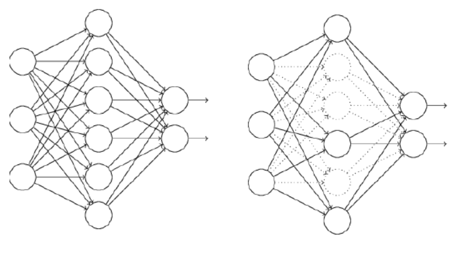

Figure 6. Full connection (left) and Dropout (right)

Data enhancement refers to a method of adding a database by randomly clipping, horizontal flipping and color lighting transformation of the original image. AlexNet replaces the traditional sigmoid and tanh functions with ReLU activation functions, speeding up the convergence of the network.

$$f(x) = \max(0, x) \tag{2}$$

When the input signal is<0, the output is 0. When the input signal is>0, the output is equal to the digital input. The derivative less than zero is abandoned, which speeds up the BP speed. Local response normalization (LRN) is added after ReLU activation function, which can increase the generalization performance of the network。

$$b_{x,y}^{i} = a_{x,y}^{i} I(k + \alpha_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (\alpha_{x,y}^{j})^2)^{\beta} \tag{3}$$

$a_{x,y}^{i}$ is the characteristic graph of convolution kernel i after passing ReLU in (x, y) coordinates. Formula 1 represents a value and its N values before and after it are standardized. Super parameter k=2, n=5, $\alpha = 10^{-4}$, $\beta = 0.75$。 N is the total number of convolution kernels, and max and min represent the standardization of edge layers.

The model convolution and shared weights greatly reduce the weights in the hidden layer of the neural network, making the training faster, and a small number of parameters easier to adjust and maintain, which will also improve the recognition rate for the classification results. Convolution neural network is developed from neural network. Traditional neural network includes an input layer, a hidden layer and an output layer. The pooling operation is used after the first, second and fifth convolution layers to reduce the dimension of the current convolution result and further reduce the size of the parameters. Input three channels of image 224 * 224 * 3D data, extract 4096 dimensional features through neural network as output, and finally use Convert imageset function to test training samples and test samples, corresponding to 80 flower image categories. The recognition results of the depth convolution neural network model (AlexNet) are used.

V. MODEL TRAINING

The number of flower database samples used in this paper is limited, so the ImageNet database is used to train the network. There are 1000 categories of ImageNet, and each category has about 1000 samples. To ensure the effectiveness of the model training, the learning rate decreases by 0.1 every 10000 iterations in the training process.

There are 350 flower database samples used this time, and ImageNet training network is used. There are more than 1000 categories of ImageNet, and more than 1000 samples for each category. To ensure the effectiveness of model training, the learning rate decreases by 0.1 every 1000 iterations in the training process. This model uses Stochastic Gradient Descent (SGD) for network back propagation.

The SGD function can be written as:

$$J(\omega) = \frac{1}{m} \Sigma \frac{1}{2} (y^i - h_\theta(x^i))^2 = \frac{1}{m} \Sigma \cos t(\omega, (x^i, y^i)) \tag{4}$$

$$\cos t(\omega, (x^i, y^i)) = \frac{1}{2} (y^i - h_\theta(x^i))^2 \tag{5}$$

Use the function shown in each sample to derive W to obtain the corresponding gradient, which is used to update W;

$$\omega_j^i = \omega + (y^i - h_\theta(x^i)) x_j^i \tag{6}$$

Where, h (x) represents the function to be fitted, J (w) is the loss function, W is the value to be solved iteratively, m is the number of records, and J is the number of parameters.

VI. CONCLUSION

The test results are given according to the identification results of five categories of flowers, namely azalea, carnation, jasmine, peony and hyacinth. [5] Table 1 shows the accuracy of different algorithms and algorithms proposed in this paper for the five categories of flowers. The comparison of recognition rate between the algorithm in this paper and the traditional algorithm is shown in the following table 1:

Table 1. Recognition accuracy of common flowers in several different algorithms

Algoritm	**Fernando** [6]	**Gehler** [7]	**Xie** [8]	**Chai,Bics** [9]	**Khan** [10]	**This paper**
Rhododendron	91	83.5	92	91.1	88	92.1
Carnation	91.2	82.5	91	91.3	89	91.1
Peony	90	82.5	91	91.2	87	91.2
Jasmine Flower	92	84.5	92	92.0	88	92.1
Hyacinth	91.3	85.5	9.1	91.0	90	91.3

It can be seen from the results in the above table that theaccuracy of the depth convolution neural network model (AlexNet) proposed in this paper has obvious advantages over the previous algorithm, and has improved a lot in recognition accuracy. It shows that thedepth convolution neural network has a good advantage in feature extraction. The main reason is that the depth convolutionneural network features are self-learning from the data set, more consistent with the data, and the number of convolution layers is relatively deep, which can extract the abstract, shape approaching nonlinear features in the data.

Although flower classification is a part of image classification, it is often difficult to identify flowers because of the similarity of different types of flowers, the difference of the same type of flowers, the complexity of the environment in which flowers are located, and the external factors to obtain flower images. In addition, because the deep convolution neural network is also a kind of neural network, it needs a lot of time to train the neural network model, and the requirement for training data is also greater than that of traditional machine learning methods. How to reduce the training time is also the direction we will study in the future.

REFERENCES

[1] Ke Dao, Chen Xiaofen, Li Shaozi. Flower image retrieval based on multi feature fusion [J]. Computer Science, 2010, 37 (11): 282-286

[2] Sun Yanfeng, Qi Guanglei, Hu Yongli Depth convolution neural network recognition algorithm based on improved Fisher criterion [J] Journal of Beijing University of Technology, 2015, 6:835-841

[3] Yin Baocai, Wang Wentong, Wang Lichun, et al Review of Research on Deep Learning [J] Journal of Beijing University of Technology, 2015, 1:48-59

[4] Tian Haobing, Zhu Jiagang, Lu Xiao Rough one class support vector machine based on feature contribution weighted Gaussian kernel function [J] Computer Science, 2015, 42 (6): 239-242R.

[5] Zhang Fan,Set sail Image classification based on dense SIFT features and its pooling model [D] Central South University, 2014

[6] Fernando B, Fromont E, Muselet D, etc Discriminant feature fusion for image classification [C]. Computer Vision and Pattern Recognition (CVPR), IEEE Conference 2012, IEEE, 2021:343441.

[7] Gehler P, Nowozin S. On feature combination of multi class object classification [C]. Computer vision. IEEE 12 International Conference 2009. IEEE， 2009:221-228.

[8] Xie Xiaodong, Research on Fine Image Classification for Flower Images,Xiamen University, 2014.

[9] Chai Y, Rahtu E, Lemdisky V, etc. Tricos: a three-level classification discrimination co segmentation method for image classification [M]. Computer 2012, Springer, 2012:794-807.

[10] Khan F S, Van De Weijer J, Vanrell M. Top color for object recognition [C]. IEEE 12 International Conference on Computer Failure, 2009. 2009:979-986.