Fruit classification method based on deep learning

Xuemei Wang[1]

[1] Southwest University of Science and Technology
zty221@126.com

Abstract. Image recognition is a typical application in the field of pattern recognition. How to accurately and quickly carry out image recognition has always been an important topic for scholars all over the world. Fruit image recognition plays an important role in the field of intelligent agriculture and digital medicine. In the aspect of smart agriculture, the accurate cultivation of fruit trees in compound orchards and automatic fruit picking can be realized through the identification of fruits; In the field of digital medicine, fruit recognition is mainly used to assist the analysis of fruit nutrients in the later stage, so as to help patients formulate a reasonable diet.Therefore, based on the in-depth study of deep learning theory, this paper applies deep learning to fruit image recognition, so as to improve the recognition performance of fruit image.

Keywords: Image recognition, Deep Learning, Fruit classification.

1 Introduce

Fruit classification and recognition is a subject involving a wide range. At present, the research on fruit image classification and recognition mainly focuses on fruit quality classification, maturity recognition, defect detection and robot picking, and generally aims at a certain type of fruit, while there is less research on multi type fruit classification and recognition in China. The classification and recognition of many kinds of fruits also has broad application value in practice. For example, in supermarkets in developed countries, people use multi category fruit image recognition to realize self-help fruit purchase; In the production line, the classification and recognition of multiple kinds of fruits can also reduce manual errors and improve production efficiency. In addition, the classification and recognition of fruit images also has certain research significance in the fields of intelligent agriculture and digital health care. In the aspect of intelligent agriculture, the fine management of fruit trees and automatic fruit picking are realized through the identification of fruits; In the field of digital medicine, the recognition of fruits is used to analyze the types of fruits, so as to further obtain the nutrients contained therein, and help patients formulate a reasonable dietary mix in the later recovery[1].

China is a large global agricultural production country. Although the output of agricultural products is huge, it can not obtain much competitiveness only by relying on low prices and rich natural resources in the international market. At present, with the development of Internet technology and various information technologies, China's

EAI Urb-IoT 2021, December 20-21, Shenzhen, People's Republic of China

DOI 10.4108/eai.20-12-2021.2315013

agricultural model has begun to change from the original traditional agriculture to modern intelligent agriculture. Smart agriculture combines the current information technology and artificial intelligence means with agriculture to promote the more efficient and intelligent development of agriculture, so as to make China's agricultural products more competitive in the international market[2]. As the third largest industry in China after grain and vegetables, although the overall output of fruit is huge, compared with developed countries, China's fruit yield is low, the fruit quality is poor and the overall benefit of orchards is not high; There are hidden dangers in fruit quality and safety, and the awareness of pest control and comprehensive control is weak; In orchard management, most of them are manual operation, time-consuming and labor-consuming, fruit picking is time-consuming and laborious, and the efficiency is low. These problems lead to the lack of competitiveness of the fruit industry in the international market[3].

Therefore, fruit image recognition is a very key link in the digital health and medical information industry. How to realize fast and accurate fruit recognition is the premise of all work. Therefore, this paper uses the tool of deep learning to study the fruit image recognition algorithm[4].

2 Model method

Compared with voice and image, text is more complex and abstract. Human beings can have an overall understanding of the text content after reading the text according to their own understanding ability[5]. However, the semantics in natural language is difficult to be directly understood by computers. Therefore, the text content must be expressed as forms that computers can understand and process, such as 0 and 1. Text representation model is to use numerical or symbolic vectors that can be expressed by computer to represent abstract and complex natural text. In order to better represent the text, it is necessary to extract the most representative features from the text data. These features should have obvious statistical laws, which can reflect the text distribution in the feature space and minimize the computational complexity of text mapping to the feature space.

2.1 Convolutional neural network

As the most widely used deep learning model in the field of computer vision, convolution neural network (CNN) first proposed a recognition algorithm by Japanese scholar Fukushima based on the mechanism of biological vision system[6]. Later, the recognition performance was further improved through the improvement and simplification of Le Cun, simard and steinkraus. The CNN model is similar to the BP neural network model, which adopts the method of forward propagation to calculate the output value and back propagation to adjust the weight and bias; However, the biggest difference between them is that the neurons of two adjacent layers in CNN are partially connected, rather than the full connection in BP neural network. In addition, as a deep neural network, CNN will not have the same problem of non computability as the traditional deep neural network.

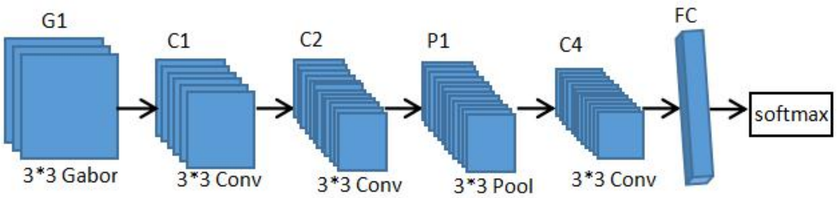

Fig. 1. The structure of CNN

CNN can directly extract the features of two-dimensional images and use BP algorithm to optimize the parameters in the network[7]. Compared with the general traditional neural network, CNN shows great advantages in image recognition: (1) directly taking two-dimensional images as input avoids the complex preprocessing process(2) Feature extraction and pattern classification are carried out simultaneously.

Feature extraction and pattern classification are combined to obtain the required parameters through continuous optimization, and the results are given in the output layer (3) By means of sparse connection and weight sharing, the number of parameters to be trained in the network is greatly reduced and the generalization performance is further enhanced..

The structure of CNN is generally composed of five parts: input layer, convolution layer, down sampling layer, full connection layer and classifier. Where C represents the convolution layer. Each convolution layer is composed of multiple feature maps. Each feature map contains many independent neurons, and each neuron is locally connected with the neurons in the input (feature map) of the upper layer. Due to the certain relationship between different local features, when a local feature is extracted, Its position relative to other features is approximately determined without determining its exact position. Different feature maps in the convolution layer are obtained by convoluting the input image (feature map) of the previous layer with different filters. Each feature map represents a feature, and the weights of all neurons in the feature map are equal. Layer s represents the lower sampling layer, which mainly performs simple scaling mapping on multiple feature maps in the convolution layer, reduces the resolution of the feature map and reduces the number of training weights in the network. After convolution and down sampling, a full connection layer is usually connected before the output layer of the network to combine the features, and finally identify and classify them.

2.2 Transfer learning

When using convolutional neural network, there must be a large amount of data support to obtain a more accurate model Good generalization. However, under complex practical conditions, it is sometimes difficult to collect so much information and data. Even if there is so much data, it will take a lot of time to label the data; And usually, a complex deep learning model takes at least a few days or weeks of training time; This will not only consume manpower and time, but also increase the project cost. Therefore, in order to solve these practical problems, we can use transfer learning to complete. In human understanding, transfer learning is the ability to draw inferences from one instance. For example, if you learn to ride a bicycle, it is generally easy to learn to ride an electric car. In machine learning, that is, in order to solve a learning task in the source domain, the model trained through a large amount of data can be applied to the learning task in the target domain after simple adjustment.

The data sets of the source domain and the target domain are different. Transfer learning can save users a lot of time and cost under the condition of ensuring the robustness of the model, but its effect is still worse than the model trained from scratch with a large amount of data. In general, the learning task in the source domain and the learning task in the target domain have certain commonalities, so the transfer learning effect will be good. For example, the model for cat face detection can be transferred to face detection, but if it is transferred to the task for predicting the age range of characters, the effect will be very poor. There are two uses of transfer learning. One is to extract features, that is, remove the last fully connected layer in the original convolutional neural network, ensure that other layers and their parameters remain unchanged, extract features from new data sets, and then put these features into the classifier for training. In most cases, softmax classifier can be used; This method is generally applied when there is less data in the target domain. The other is fine tuning, which can be used when the target domain data is large. Starting from the top layer of the neural network, freeze part of the network and its parameters, and then retrain the model to update the parameter values of the remaining network. If the data is large enough, the parameters can be retrained from the top. In this study, due to the small amount of data, transfer learning is adopted.

2.3 Activation function

CNN selects the features extracted from the network by means of nonlinear mapping with the help of activation function to avoid the problem of insufficient expression ability of linear operation. In practice, the common activation function can be divided into saturated nonlinear function (such as sigmoid function) and unsaturated non-linear function.

Sigmoid function was widely used in early neural networks and was once regarded as the core of neural networks because it well explained whether neurons were activated and passed back after stimulation. Mathematically, although the sigmoid function has a large signal gain for the central region and a small signal gain for the two side regions, it has a good effect on the feature space mapping of the signal. However, it has a major defect that it is easy to produce gradient dispersion effect during training. The main reason for this effect is that sigmoid function, as an S-type function, has very small gradients on both sides. In a deep neural network model, many neurons are on both sides of this function, This makes the accumulated value of the gradient smaller and smaller in the process of back propagation. It is confirmed in some literatures that the gradient will indeed be smaller and smaller, resulting in the fact that the previous layers cannot be trained effectively. This disadvantage makes the neural network unable to learn in many scenarios.

ReLu is the abbreviation of the modified linear unit. It forces the data less than 0 to be compressed into 0, while the data greater than 0 remains unchanged, and when greater than 0, the derivative is a constant value. The derivative of sigmoid function is not a constant value, s, but a curve shape similar to Gaussian function. When both sides are close to the target, its derivative decreases. Too small derivative will increase the BP back propagation error during training, resulting in slower network convergence, which can be well avoided by re Lu function. Re Lu only needs a threshold to get the activation value, saving a series of complex operations. However, re Lu function also has some disadvantages. Re Lu function is very fragile during training. For example, a very large gradient flows through a re Lu neuron. After updating the parameters, the neuron will no longer activate any data. If this happens, the gradient of this neuron will always be 0. This requires careful parameter adjustment and setting an appropriate small learning rate, and the occurrence of this problem will be reduced. In addition, because the re Lu function only retains the data of X and other data is "squeezed" to 0, there may be distinguishable information in the compressed data, and the compression amplitude is very large, which makes it completely unrecoverable.

3 Experimental results and analysis

In this experiment, 5000 picture samples are collected in different sizes, and the pixels need to be adjusted to 128 × 128 size, then randomly cut to 60 × The purpose is to increase the diversity of samples, then input the processed images into the network for operation, and configure the training samples and test samples according to the ratio of 0.8:0.2. The framework used for running is the pytorch framework, which inputs a batch of pictures for processing every iteration.

3.1 Dataset

The experimental data were scratch, insect bite, decay, brown spot, fruit stem and fruit calyx, 1000 pieces of each type. The ratio of training set to test set is 0.8:0.2. Two groups of experiments were carried out: the first group constructed a complete data set and trained directly. The second group expands the data set, rotates the training picture, and trains the expanded data.

3.2 Model training and results

Table 1 shows the test results of direct training. From the table, CAS can see that with the increase of the number of iterations, the parameters of the network model will be updated continuously, and the final classification accuracy will converge. When the number of iterations reaches 30, it will reach stability. From the test results on the trained network model, it can be seen that the accuracy is 0.9. It has stronger creativity, higher accuracy of test results and more stable model.

Tab. 1. Classification results

Epoch	5	10	15	20	25	30	35
ACC	10.06%	30.48%	60.53%	80.96%	88.26%	93.23%	92.36%
Loss	17.5	10.6	2.3	1.5	0.3	0.1	0.1

4 Conclusion

In China, artificial fruits are used for classification, which is time-consuming, laborious and inefficient. In order to overcome the defects and shortcomings of traditional classification methods and improve the efficiency and accuracy of fruit classification, this paper studies the method based on a variety of feature extraction algorithms and improved convolution neural network to realize fruit detection and classification, and the accuracy of classification is 92.23%.

References

1. Zhang Y , Wang S , Ji G , et al. Fruit classification using computer vision and feedforward neural network[J]. Journal of Food Engineering, 2014, 143:167-177.
2. Dubey S R , Jalal A S . Detection and Classification of Apple Fruit Diseases Using Complete Local Binary Patterns. Third International Conference on Computer & Communication Technology. IEEE Computer Society, 2012.
3. Feng B , Wang M . Computer Vision Classification of Fruit Based on Fractal Color. transactions of the chinese society of agricultural engineering, 2002.
4. Rashidi M , Seyfi K . Classification of Fruit Shape in Cantaloupe Using the Analysis of Geometrical Attributes. World Journal of Agricultural Sciences, 2007(6).
5. F Garcia, Cervantes J , Lopez A , et al. Fruit Classification by Extracting Color Chromaticity, Shape and Texture Features: Towards an Application for Supermarkets. IEEE Latin America Transactions, 2016, 14(7):3434-3443.
6. Vasumathi M T , Kamarasan M . An Effective Pomegranate Fruit Classification Based On CNN-LSTM Deep Learning Models. Indian Journal of Science and Technology, 2021, 14(16):1310-1319.
7. Dandavate R , Patodkar V . CNN and Data Augmentation Based Fruit Classification Model. 2020 Fourth International Conference on I-SMAC. 2020.