An Automatic Recognition Method of Fruits and Vegetables Based on Depthwise Separable Convolution Neural Network

ZHAN Yuhui¹, CHEN Mengyao¹, CHEN Yuefen¹*, LUO Zhaoqian¹, Luo Yao¹, Li Kedi¹

¹School of Electronic & Information Engineering, Taizhou University, Taizhou 318000, China
*Corresponding author’s e-mail: tzc_cyf@tzc.edu.cn

Abstract. Traditional fruit and vegetable classification is mostly based on manual operation, which is inefficient. Deep convolution neural network shows excellent performance in feature learning and expression. In this paper, an automatic recognition system of fruits and vegetables based on deep convolution neural network is designed. By using depthwise separable convolution instead of the traditional standard convolution, a neural network is constructed with less parameters, which is suitable for equipment with limited resources. A small data set including 12 kinds of common fruits and 8 kinds of common vegetables is established for training and testing through network download and physical shooting. The experimental results show that the recognition accuracy reaches 95.67%.

1. Introduction
China is a big agricultural country in the world, with high output of fruits and vegetables, which have a high demand in people's daily life, and the market scale of fruit and vegetable industry continues to grow. At present, when fruits and vegetables are sold, most of them use manual input of all kinds of fruit and vegetable codes for manual weighing and pricing, resulting in a lot of waste of human resources. In some large supermarkets, there are self-service fruit and vegetable weighing equipment, but customers also need to choose the corresponding item or picture from the picture library, then the machine will automatically price after confirmation. Due to the variety of fruits and vegetables, customers still spend a long time to choose the right item. With the rise of artificial intelligence, people's lives begin to move towards intelligence. There is an urgent need for a system that can replace manual work and has the function of automatically identifying the categories of fruits and vegetables, so as to promote the automation of agricultural products trading and fruit and vegetable sales.

With the rapid development of machine vision technology, it has also been widely used in the detection of agricultural products. In this technology, image of objects can be obtained by camera instead of human eyes, and object features are extracted by means of image processing technology, and recognition is realized by various classifiers. The increasing maturity of image processing technology and the continuous development and optimization of neural network, support vector machine(SVM), deep learning algorithm provide a better guarantee for the accuracy of recognition. Bolle[1] proposed in 1996 that any number of fruits and vegetables can be recognized by extracting the colour, texture and shape features of the image. Koslowski[2] extracted the colour and texture features of the image, and used SVM to classify the fruit image. ZHANG Z.C.[3] proposed a fruit and vegetable classification algorithm based on multi-feature fusion in the BOF model, which uses the CLBP feature and surf feature of the image, generates the feature dictionary based on the K-means
clustering algorithm, and uses the SVM classifier to realize the recognition. Arivazhagan[4] used the H and S statistical histograms in HSV colour space as colour features and the autocorrelation matrix in wavelet domain as texture features for fruit and vegetable recognition. YANG J.P.[5] studied Gabor features of fruit and vegetable images, used PCA and FLD to reduce dimensions, and finally used error correction SVM as classifier. YU L.[6] also studied the feature extraction, feature reduction and classifier of fruit and vegetable images. HUANG W. Y.[7] designed a supermarket fruit and vegetable recognition system based on embedded android development board based on HSV colour histogram feature and LBP rotation invariant pattern texture feature. TAO H.W[8] and others achieved the classification and recognition of 46 kinds of fruits and vegetables by extracting texture features and colour features of fruit and vegetable images and fusing the two features. WANG S.P.[9] extracted the area, colour and shape features of the object and used SVM classifier to recognize the fruit, but the recognition result of the algorithm is not ideal for the fruit with similar features.

As above mentioned, due to the development of image processing technology, it has been widely used in fruit and vegetable recognition. However, the extraction of fruit and vegetable image features mainly relies on artificial experience, and generally extracts the geometric, texture and colour features. But there are many kinds of fruits and vegetables, and there is no unified feature for all kinds of fruits and vegetables at present. Therefore, the artificial feature extraction has some limitations. In the current era of artificial intelligence, deep learning technology has shown superior performance in target recognition and classification. Many researchers have tried to apply deep learning theory to the field of agricultural product detection [10-13]. In this paper, deep learning technology is applied to the recognition of common fruits and vegetables to improve the recognition rate. The image data in this paper comes from physical shooting and network download, and the number is limited. In order to alleviate the phenomenon of training over fitting caused by small samples, some data enhancement techniques are used. But generally speaking, the deep learning convolutional neural network model has a large amount of parameters and calculation, which has a high requirement for equipment. Therefore, in order to reduce the training parameters and improve the training speed, this paper uses the depthwise separable convolution neural network to get a model with less parameters and calculation, which is suitable for equipment with limited resources. and classifies 20 kinds of fruits and vegetables. The model is used to train and identify 12 kinds of fruits and 8 kinds of vegetables.

2. System design

The overall design of this system includes the acquisition of fruit and vegetable images, data enhancement, the design and training of depthwise separable convolution neural network, and the recognition of test images, as is shown in Figure 1.

![Figure 1. System block diagram](image)
2.1 Image acquisition and data enhancement
At present, there is no unified image data set of fruits and vegetables at home and abroad. The image set used in this experiment consists of three parts: (1) physical photos from supermarkets, vegetable farms and other fruit and vegetable sales occasions; (2) fruit and vegetable image data set from the kaggle database (https://www.kaggle.com/mbkinaci/fruit-images-for-object-detection) (3) from Baidu picture download. In the experiment, 12 kinds of common fruits, including banana, blueberry, cherry, green apple, red fuji, kiwi fruit, lemon, fresh litchi, orange, pitaya, strawberry and plum, and 8 kinds of common vegetables, including carrot, cucumber, eggplant, green pepper, red pepper, tomato, potato and red onion, were collected. Figure 2 is the schematic diagram of some samples used in this paper.

Figure 2. Sample diagram of fruits and vegetables

Due to the small scale of the image data, it is easy to appear "over fitting" phenomenon during training, which leads to the weak generalization ability of the model and affects the accuracy of the test. Therefore, this paper uses some geometric transformations of the image, including: (1) randomly rotate the image in the range of 40 degrees; (2) randomly translate the image in the horizontal direction, the degree of translation is not more than 0.2 times of the original width; (3) randomly translate the image in the vertical direction, the degree of translation is not more than 0.2 times of the original height; (4) randomly cut the image. By the above transformations, images are enhanced to alleviate the over fitting phenomenon. All samples are labeled, and the label value of each type of fruit and vegetable is set to 0 ~ 19 during storage, and is converted to the one hot coding form for training and testing of the model, that is, the label value of each type of fruit and vegetable is a vector with a length of 20. \( y = \{a_0, a_1, a_2, ..., a_{18}, a_{19}\} \). For category i, only \( a_i = 1 \), other values are 0. The labeled samples are divided into training set, verification set and testing set.

2.2 Design of depthwise separable convolution neural network
In the traditional convolution operation, the convolution kernel performs convolution operation with each channel of the input feature map at the same time, while the depthwise separable convolution decomposes the standard convolution operation into depthwise convolution and 1*1 pointwise convolution. In other words, each convolution kernel is only responsible for one channel, and the number of convolution kernels is equal to that of input channels. Then, through pointwise convolution, the feature map is fused in the depth direction to generate a new feature map. The neural network model constructed in this paper is shown in Figure 3, including five depthwise separable convolution layers, five maximum pooling layers, one flatten layer and two fully connected layers.
The input image is 128 * 128 * 3 colour image, the number of channels in the five convolution layers is 16, 32, 64, 128, 256, the convolution core size is 3 * 3, the padding mode is "same", the activation function is "relu function", the size of maximum pooling layer is 2 * 2, the neuron number of the full connection layer is 100, the output layer has 20 neurons and uses softmax function to calculated the probability of the 20 categories. The output size of each layer in the network and the number of weights required for training are shown in Table 1. It can be seen from Table 1 that there are 457971 training parameters in neural network where we use depthwise separable convolution. If the ordinary convolution layer is used to realize the neural network with the same structure, 804328 parameters needs to be trained, which implies the training parameters are only 57% of the ordinary convolution layer while using depthwise separable convolution layer, greatly reduces the training parameters and shortens the training time.

Table 1. Comparison of training parameters between ordinary convolution and depthwise separable convolution

| Layers               | Layer1 | Layer2 | Layer3 | Layer4 | Layer5 | Fc1  | Fc2  | total  |
|----------------------|--------|--------|--------|--------|--------|------|------|--------|
| ordinary convolution | 448    | 4640   | 18496  | 73856  | 295168 | 409700 | 2020 | 804328 |
| depthwise separable convolution | 91     | 668    | 2400   | 8896   | 34176  | 409700 | 2020 | 457971 |

2.3 Training
The purpose of training is to continuously learn and adjust the weights according to the training data set, so as to obtain the optimal weights. The training process is shown in Figure 4.
First, a batch of samples are read from the training set as the input of the above neural network, and calculated through the forward operation of the network, in which the activation function used is the Relu function: \( f(x) = \max(0, x) \). Finally, the predicted values calculated by the model are converted into probability values belonging to each category by softmax function as follows:

\[
\hat{y}_k (k = 1, 2, \ldots, N) \text{ is the predicted value of the forward calculation of the model, } p_i \text{ is the output probability that the current input belongs to the category } i, \text{ and } N \text{ is the number of categories.}
\]

There is an error between the output value of convolutional neural network and the label value (real value). Different weights in the neural network will produce different errors. Therefore, we need to introduce a loss function to measure the error between the output value and the label value. In this paper, the cross entropy loss function is used, which is shown as follows:

\[
\text{loss} = - \sum_{i=1}^{N} p'_i \log p_i
\]

Where \( p_i \) is the output probability value of the model prediction for the fruits of category \( i \), while \( p'_i \) is the ideal probability value, and the loss value of each epoch is the average value of all samples. The error back propagation is used to adjust the weights of the network to find the optimal solution of the model, that is, the optimization of the network. The common optimization methods include gradient descent method, momentum optimization method and adaptive learning rate optimization method. In this paper, Adam optimization is used.

3. Experimental results and analysis

Based on Keras 2.3.1 and Tensorflow 2.0.0, the depthwise separable convolution neural network model is constructed and trained. In the experiment, there are 2284 original pictures in the test data set and 8472 original pictures in the training data set, 20% of which are used for verification in the training process. After the image processing described in Section 2.1 and data expansion, each batch of training samples is generated, and the number of images in each batch is set to 20. The learning rate is set to 0.001 during training. After 10 epoch training, the model achieves 98.81% accuracy in the verification set. We use the model for prediction, as shown in Figure 5, it can be seen that the prediction result is consistent with the result of manual judgment. The model is applied to the prediction of the test data set, and the overall accuracy is 95.67%. In order to further analyze the prediction results of each fruit and vegetable, the accuracy, recall rate and F1-value of each fruit and
vegetable are given in Table 2, which shows that the recognition of most fruits and vegetables in this experiment has achieved a high accuracy.

![Figure 5. Schematic diagram of test results](image)

| Fruit/Vegetable  | Precision | Recall | F1-Score | Support |
|-----------------|-----------|--------|----------|---------|
| blueberry       | 0.98      | 0.96   | 0.97     | 104     |
| cherry          | 0.99      | 0.96   | 0.98     | 157     |
| banana          | 0.88      | 0.96   | 0.93     | 127     |
| green apple     | 0.95      | 0.96   | 0.96     | 109     |
| red fuji        | 0.98      | 0.92   | 0.95     | 105     |
| kiwi            | 0.99      | 0.97   | 0.98     | 97      |
| lemon           | 0.95      | 0.94   | 0.95     | 111     |
| orange          | 0.99      | 0.95   | 0.97     | 106     |
| plum            | 0.96      | 0.94   | 0.96     | 94      |
| Pitahaya        | 0.99      | 0.94   | 0.96     | 106     |
| strawberry      | 0.91      | 0.96   | 0.94     | 107     |
| Cucumber        | 0.90      | 0.98   | 0.94     | 118     |
| carrot          | 0.96      | 0.95   | 0.94     | 165     |
| litchi          | 0.91      | 0.95   | 0.94     | 108     |
| eggplant        | 0.96      | 0.94   | 0.96     | 106     |
| onion           | 1.00      | 0.98   | 0.96     | 106     |
| potato          | 0.98      | 0.93   | 0.97     | 116     |
| red pepper      | 0.96      | 0.98   | 1.00     | 138     |
| green pepper    | 0.96      | 0.98   | 0.98     | 107     |
| tomato          | 0.97      | 0.94   | 0.94     | 111     |

### 4. Conclusion

In this paper, an automatic recognition system of fruits and vegetables based on depthwise separable convolution neural network is designed. A small-scale data set including 12 kinds of fruits and 8 kinds of vegetables is established. The training parameters are greatly reduced by using depthwise separable convolution, and the trained model achieves good recognition results in the test set. It shows that the convolution neural network can learn the high-level features of different fruit and vegetable images by using multiple convolution and pooling operations, and can achieve good recognition without manual feature extraction, which is helpful to save the labour cost and promote the automation of fruit and vegetable sales, and is also conducive to the development of unmanned fruit and vegetable supermarket. This paper will further expand the research on automatic recognition of more kinds of fruits and vegetables, and data enhancement technology for data set scale expansion, adaptive algorithms for model parameters of multitask deep convolution neural network, etc.

### Acknowledgments

This work was co-sponsored by the project of the innovation and entrepreneurship training program for college students in 2020 and project of Taizhou University (Grant number:2019PY013).

### Reference

[1] Bolle R M, Connell J H, Haas N, et al. (1996): A produce recognition system. Applications of computer vision. 244-251.
[2] Koslowski M A, Santos F G, Borba G B, et al. (2013) Fruits classification using MPEG-7 descriptors from image patches. In: WVC. IX workshop de visao computational (WVC 2013). Brazil: WVC, 1-6.

[3] ZHANG Z.C, JU Z. (2019) Multi-feature Fusion Fruit and Vegetable Image Classification Based on Bag of Feature Model. Electronic Science and Technology. 13:1-6.

[4] Arivazhagan S, Shebiah R N, Nidhyanandhan S S, et al. (2010) Fruit recognition using color and texture features[J]. Journal of Emerging Trends in Computing and Information Sciences, 1(2):90-94.

[5] Yang J.P. (2011) Research on Fruit and Vegetable Recognition Based on Computer Vision. Dalian University of Technology.

[6] YU Ling. (2015) Research on Feature Analysis of Fruit and Vegetable Automatic Categorization Technology Based on Computer Vision. Southeast University.

[7] HUANG W.Y. (2017) Research on the Identification System of Fruit and Vegetable in Supermarket Based on Embedded Machine Vision. China Jiliang university.

[8] Tao H. W., Zhao L., et al. (2014) Fruits and vegetables recognition based on color and texture features. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 30(16):305-311.

[9] WANG S.P., TANG Z.M., et al. (2010) Recognition Algorithm of Fruits Based on SVM. Journal of Wuhan university of technology. 32(16):44-47.

[10] XIONG J.T., ZHENG Z.H., et al. (2020) Citrus Detection Method in Night Environment Based on Improved YOLO v3 Network. Transactions of The Chinese Society of Agricultural Machinery. 51(4):199-206.

[11] MU L.T., GAO Z.B., CUI Y.J., et al. (2019) Kiwi fruit detection of far-view and occluded fruit based on improved AlexNet. Transactions of the Chinese Society for Agricultural Machinery, 50(10): 24-34.

[12] XUE Y.J., HUANG N., TU S.Q., et al. (2018) Immature mango detection based on improved YOLOv2. Transactions of the Transactions of the CSAE, 34(7): 173-179.

[13] BI S., GAO F., CHEN J.W., et al. (2019) Detection method of citrus based on deep convolution neural network. Transactions of the Chinese Society for Agricultural Machinery, 50(5):181-186.