

# High performance vegetable classification from images based on AlexNet deep learning model

Ling Zhu<sup>1,2,3</sup>, Zhenbo Li<sup>1,2,3\*</sup>, Chen Li<sup>1,2,3</sup>, Jing Wu<sup>1,2,3</sup>, Jun Yue<sup>4</sup>

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China;

2. Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture, Beijing 100083, China;

3. Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, Beijing 100083, China;

4. College of Information and Electrical Engineering, Ludong University, Yantai 264025, China)

**Abstract:** Deep learning techniques can automatically learn features from a large number of image data set. Automatic vegetable image classification is the base of many applications. This paper proposed a high performance method for vegetable images classification based on deep learning framework. The AlexNet network model in Caffe was used to train the vegetable image data set. The vegetable image data set was obtained from ImageNet and divided into training data set and test data set. The output function of the AlexNet network adopted the Rectified Linear Units (ReLU) instead of the traditional sigmoid function and the tanh function, which can speed up the training of the deep learning network. The dropout technology was used to improve the generalization of the model. The image data extension method was used to reduce overfitting in the learning process. With AlexNet network model used for training different number of vegetable image data set, the experimental results showed that the classification accuracy decreases as the number of data set decreases. The experimental verification indicated that the accuracy rate of the deep learning method in the test data set reached as high as 92.1%, which was greatly improved compared with BP neural network (78%) and SVM classifier (80.5%) methods.

**Keywords:** vegetable classification, deep learning, Caffe, AlexNet Network, ImageNet

**DOI:** 10.25165/j.ijabe.20181104.2690

**Citation:** Zhu L, Li Z B, Li C, Wu J, Yue J. High performance vegetable classification from images based on AlexNet deep learning model. *Int J Agric & Biol Eng*, 2018; 11(4): 217–223.

## 1 Introduction

China has become the world's largest vegetable market, its vegetable production and sales accounted for more than 50% of the global market proportion. But now most of the picking and sorting of vegetables still rely on manual operation with a large amount of labor force and low work efficiency, which seriously affects the development of commercialization of vegetable products. The use of image classification and identification technology can realize automatic picking and grading of vegetables, which has a relatively wide application value.

At present, some progress has also been made in the classification of vegetables. Huo et al.<sup>[1]</sup> extracted texture features of vegetable images, classification of vegetable images with SVM classifier, the accuracy rate of classification about 80%. He et al.<sup>[2]</sup> proposed a method of automatic classification of fruits and vegetables based on visual characteristics, extracted the visual characteristics of fruits and vegetables. Automatic fruit and vegetable recognition and classification were realized by SVM classifier. The proposed method in the supermarket agricultural products database has made a high recognition rate. On the basis of these studies, in order to improve the accuracy of classification,

this paper put forward a method of deep learning.

Deep learning<sup>[3-7]</sup> as a new machine learning algorithm has been proposed in recent years and breakthroughs have been made in image recognition, computer vision and other fields. Compared with the traditional manual design feature method and deep learning method, the data features extracted by the deep learning method can better characterize the rich internal information of the massive data, and can automatically learn representative features without manual extraction. Deep learning, while changing the traditional machine learning method, also changes our understanding of human perception, and a breakthrough has been made in the field of image classification and other applications. Krizhevsky et al.<sup>[8]</sup> applied the deep learning method to the field of image classification and greatly improved the image classification accuracy rate. Ciresan et al.<sup>[9]</sup> used the deep of the neural network to achieve the classification and identification of traffic signs, and improved the classification criteria for commonly used images. Tan et al.<sup>[10]</sup> designed a deep learning neural network for the recognition of fruit and vegetable pathological images. Based on the propagation analysis of network errors, a parameter learning method for elastic momentum was proposed and the identification experiment of fruit pathology image was carried out with apple as an example, which significantly improved the accuracy of fruit and vegetable disease recognition in learning network. At present, deep learning networks are mainly Stacked AutoEncoder, Restricted Boltzmann Machine, Deep Belief Network and Convolutional Neural Network (CNN), and so on. Among them, the CNN<sup>[11-15]</sup> is the most significant in the image classification task.

The development of CNN in image classification task is more and more mature. Gong et al.<sup>[16]</sup> proposed a method of plant leaf classification based on CNN. The experimental results on the

**Received date:** 2017-07-04    **Accepted date:** 2018-04-03

**Biographies:** Ling Zhu, Master candidate, research interests: computer vision, Email: 1023291660@qq.com; Chen Li, Master candidate, research interests: computer vision, Email: 1378406289@qq.com; Jing Wu, Master candidate, research interests: computer vision, Email: 306930180@qq.com; Jun Yue, PhD, Professor, research interests: computer vision, Email: 1755978665@qq.com.

**\*Corresponding author:** Zhenbo Li, PhD, Associate Professor, research interests: computer vision. College of Information and Electrical Engineering, China Agricultural University, No. 17, Qinghua East Road, Beijing 100083, China. Tel: +86-18701586839, Email: zhenboli@126.com.

Swedish leaf dataset show that, the accuracy rate of the algorithm is significantly better than that of the traditional blade recognition algorithm. Hu et al.<sup>[17]</sup> proposed a dynamic gesture recognition method for analyzing static trajectory images using CNN with the normalized image as the input of the CNN model after a series of preprocessing steps. Testing the performance of the proposed method on a self-built database, the experimental results show that the dynamic hand gesture recognition accuracy rate of digital 0-9 is obviously improved. Qu et al.<sup>[18]</sup> proposed a CNN model based visible spectral remote sensing image target recognition method, on the basis of the structure of traditional LeNet-5 network, the ReLu activation function is introduced instead of the traditional sigmoid function and tanh function, the structure of network is adjusted and optimized, to improve the accuracy rate and efficiency of target identification. In summary, it is shown that the CNN has achieved remarkable results in image classification and recognition tasks.

In this paper, based on deep learning model of AlexNet<sup>[19]</sup> of vegetable image target classification, five categories of vegetable images were obtained from the ImageNet<sup>[20-22]</sup> data set, and the data set was divided into training data set and test sets. Since each kind of vegetable has fewer images, the data expansion method is adopted to increase the number of each category of vegetable images, in order to reduce overfitting in the learning process. Using the Caffe<sup>[23,24]</sup> of open source learning framework, the AlexNet network model was used to train the vegetable image data set and tested on the test set. Using AlexNet network model to train different number of vegetable image data set, the experimental results show that the classification accuracy rate decreases as the number of data set decreases. This paper also contrasts with the traditional BP neural network and SVM classifier. Experimental results show that the method used in this paper has a good classification effect.

## 2 Methods and materials

### 2.1 Vegetable image dataset preparing and preprocessing

The dataset used in this paper is taken from ImageNet (<http://www.image-net.org>), which is the largest database of image recognition in the world today, with millions of annotated data. In this paper, five categories of vegetables were trained: broccoli, pumpkin, cauliflower, mushrooms and cucumber. Since the number of images of the five categories of vegetables used in this paper was not enough for learning, therefore, the vegetable image data sets were rotated at 90°, 180° and 270° by method of data expansion. By this method, the image data sets were enlarged by 4 times. Figure 1 shows some images from the image dataset.

In Figure 1, each category of vegetables from left to right is the image of original, rotating 90°, 180° and 270°. Each image size is different in vegetable image data set, in order to neatly arrange the image, the image is processed into the same size in Figure 1 (80×80 pixels). The data expansion method is suitable for training images and test images. In the training phase, the increase in data can produce additional training samples, thereby reducing the impact of overfitting. In the test phase, the increase in data helps to improve the classification accuracy rate.

### 2.2 Caffe Deep Learning Framework

Caffe was developed by Berkeley Vision and Learning Center, created by Jayang Qing, leaded and completed by Evan Shelhamer. Caffe is a deep learning tool developed in C++, with fast computing, modular and scalable, and support from open source communities. Currently in the field of computer vision Caffe is a relatively popular toolkit with a lot of expansion.



Note: The images from top to bottom are: Broccoli, Cauliflower, Cucumber, Mushroom, Pumpkin.

Figure 1 Five categories of vegetable data expansion

The framework of Caffe consists of five components: Blob, Solver, Net, Layer, and Proto. Solver is responsible for the deep of network training, each Solver contains a training network object and a test network object. Each network consists of several layers. The Feature map for each Layer's input and output is represented as Input Blob and Output Blob. Blob is the structure in which Caffe actually stores data, it is an indefinite dimension of the matrix, it is generally used to represent a straightened four-dimensional matrix in Caffe. The four dimensions correspond to batch size, channel number of feature map, height of feature map, and width of feature map respectively. Proto is based on Protobuf open source project of Google, it is similar to XML data exchange format, users only need to define the data members of the object in the format, which can be used to serialize and deserialize the objects in multiple languages. In Caffe, the structure of the network model is defined, stored and read.

### 2.3 AlexNet model construction and model training

AlexNet obtained the first place in the 2012 ImageNet image classification competition. This is quite good for traditional machine learning classification algorithms. It proves that the effectiveness of CNN in complex models, implementing on the GPU makes the training get results within an acceptable time range, promoting the development of the deep of learning. The network structure of the AlexNet model is shown in Figure 2.

AlexNet network has 8 layers, the first five layers are the convolutional layers, the latter three layers are the full connected layers. It can be seen from Figure 2 that the back of the first, second and fifth convolutional layers is the pooling layer, the last layer is the softmax layer (output layer). In the first convolutional layer (conv1) and the second convolutional layer (conv2) directly followed by the Response-normalization layer, which is norm1,

norm2 layer.

Each feature map of the convolutional layer is almost always obtained by combining the multiple feature maps calculated by the upper layer. The main function of the convolutional layer is feature extraction. The calculation process of the convolutional layer is as follows.

$$x_n^l = \sum_{i \in M_n} x_i^{l-1} * k_{in}^l + b_n^l \quad (1)$$

where,  $x_n^l$  represents the  $n^{\text{th}}$  feature map of layer  $l$ ;  $M_n$  represents a

set of feature maps selected from the input feature maps;  $k_{in}^l$  represents the  $i^{\text{th}}$  element in the  $n^{\text{th}}$  convolution kernel of layer  $l$ ;  $b_n^l$  represents the  $n^{\text{th}}$  offset of layer  $l$ ; “\*” represents the convolution process.

The AlexNet network has five convolutional layers, taking broccoli as an example, visualizing the output characteristics of convolutional layer (only show the first 36 images) as shown in Figure 3.

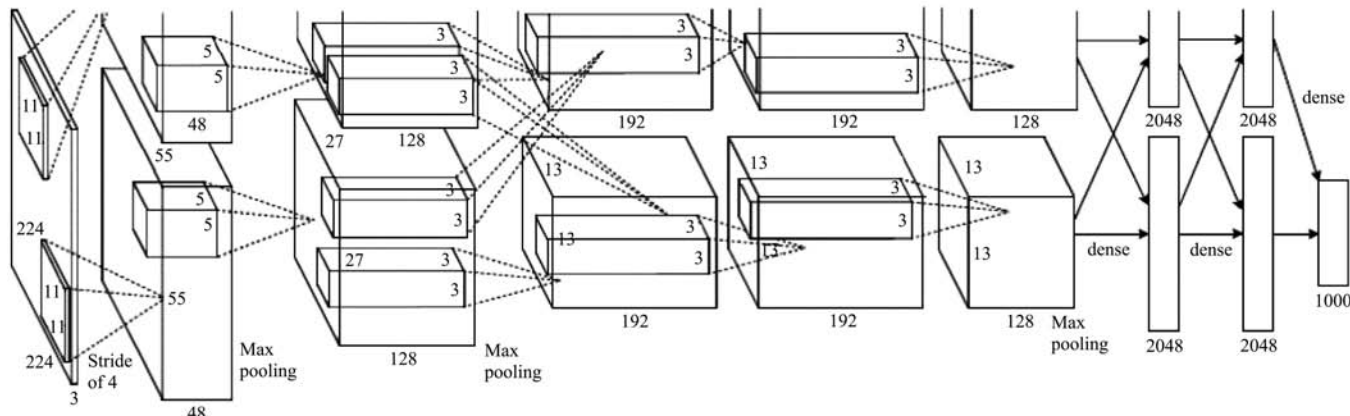


Figure 2 Structure of AlexNet network<sup>[8]</sup>

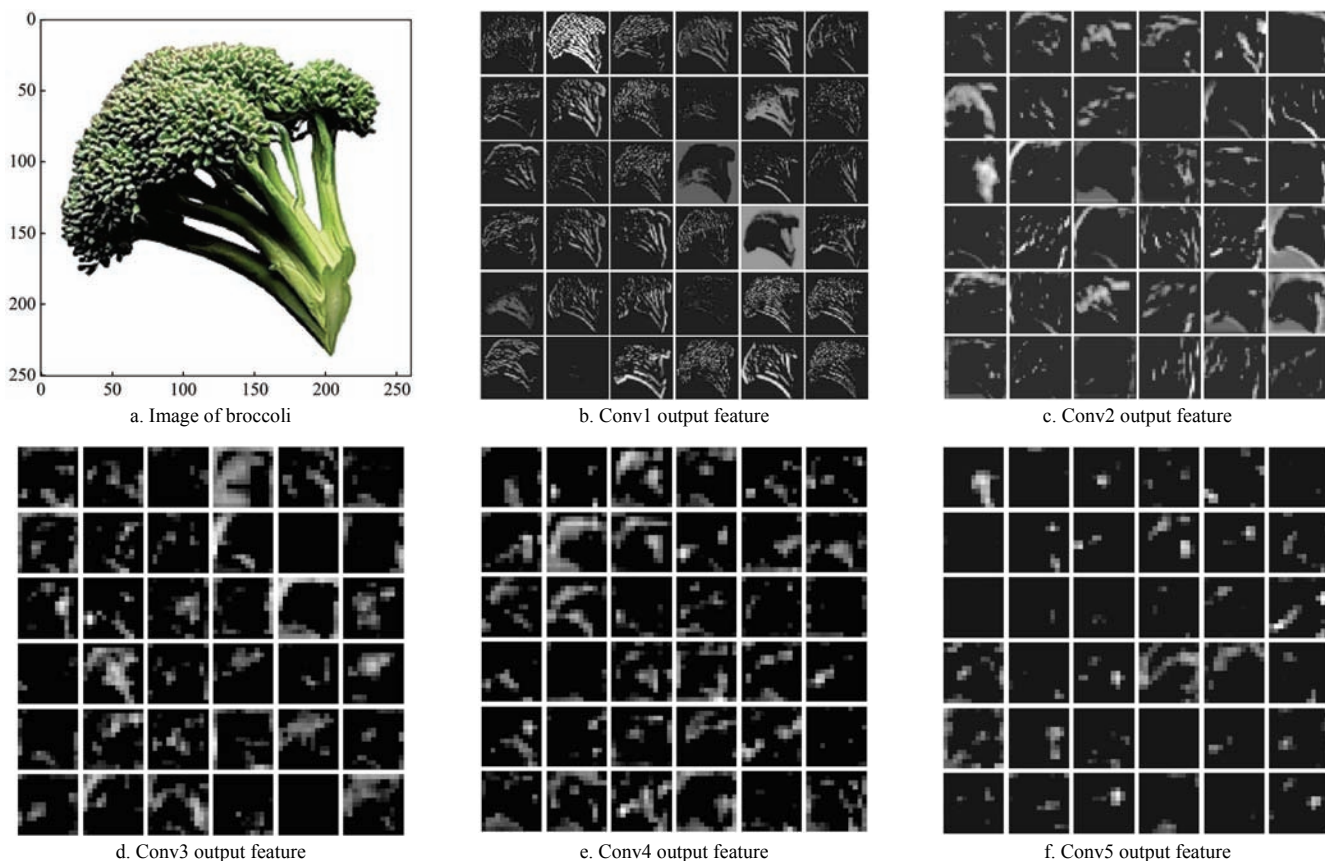


Figure 3 Visual convolutional layers output feature of broccoli image

Figure 3 shows that the contour feature of the image near the input layer is clear and close to the shape of the original image. When the number of layers increases, the features gradually change into blur and abstract, it is possible to speculate that the layers of the model represent the characteristics of different degrees of abstraction.

The pooling layer is also called the down sampling layer, this

layer produces the sampling result of the input feature map. The pooling layer does not change the number of feature map, but change the size of the feature. The calculation of the pooling layer is as follows.

$$x_n^l = \beta_n^l \text{down}(x_n^{l-1}) + b_n^l \quad (2)$$

where,  $\beta_n^l$  represents the  $n^{\text{th}}$  multiplication of the layer  $l$ ;  $\text{down}(\cdot)$  represents the pooling function.

In this paper, the pooling layer adopts max pooling operation, the operation follows the first norm1, norm2, and the fifth convolutional layer (conv5). Max pooling is to find the max value in each area, and finally extract the main features in the original feature map. The calculation of the max pooling is as follows.

$$P_n = \max_{i \in R_n} C_i \quad (3)$$

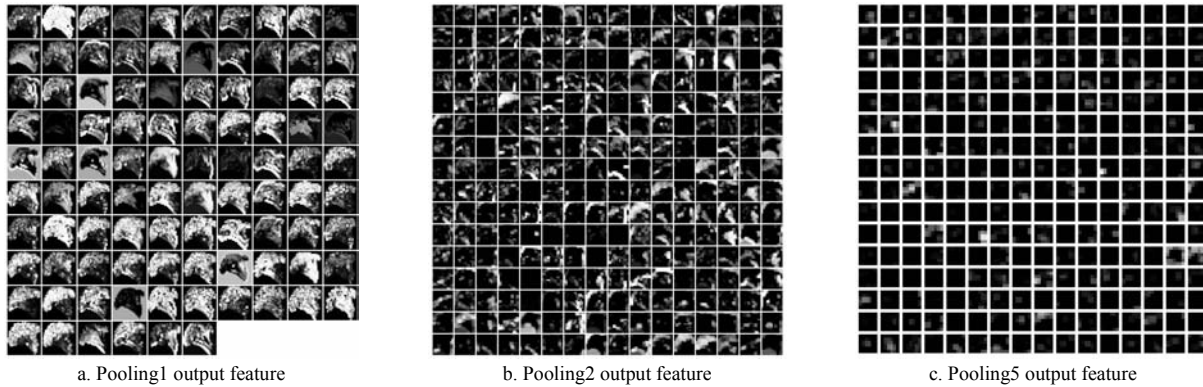


Figure 4 Visual pooling layers output feature of broccoli image

The input of the pooling layer is derived from the previous convolutional layer, the main role is to provide a strong robustness, and reduces the number of parameters. The pooling layer generally has no parameters, therefore, when the back propagation, only the input parameters are derived, there is no need to update the weights.

The output function followed by each conv layer and the full connected layer is the ReLU. The output function of the AlexNet network uses the ReLU instead of the traditional sigmoid function and the tanh function, the calculation formula for ReLU is as follows.

$$f(x) = \max(0, x) \quad (4)$$

ReLU belongs to the unsaturated nonlinear function. For the ReLU function, if the input is greater than 0, the output is equal to

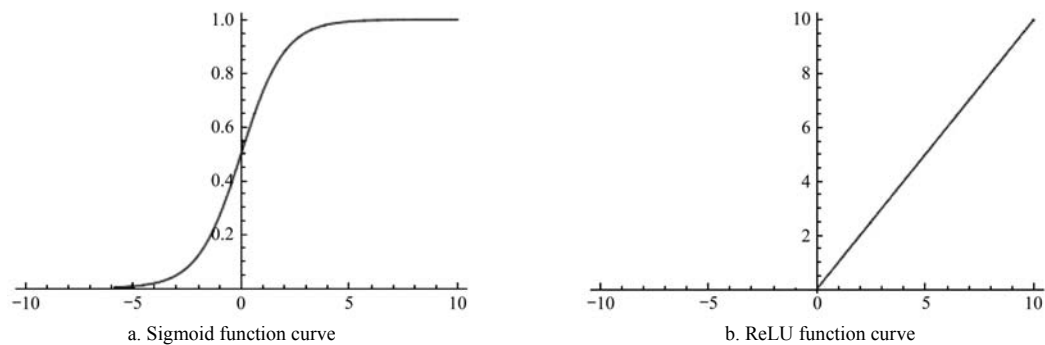
where,  $R_n$  represents the  $n$ th pooling area in the feature map;  $C_i$  represents the  $i^{\text{th}}$  pixel value of  $R_n$ .

By max pooling operations, the feature vectors of the convolutional layer output are reduced and the invariance of the feature propagation is improved. It also can reduce the risk of overfitting. The output characteristics of the pooling layers are shown in Figure 4.

the input, otherwise the output is 0. And when using the ReLU function, the output does not tend to saturate as the input increases gradually. The sigmoid function and the ReLU function curves are shown in Figure 5.

Using the traditional sigmoid and other functions to back propagation for error gradient, the derivative involved in the division, the calculation is relatively large. But using ReLU, the calculation of the whole process is reduced a lot, which can speed up the training of the network. ReLU will make part of the neuron output is 0, which resulted in the sparseness of the network, and reduce the interdependence of parameters, to alleviate the overfitting problem.

The dropout operation was used on the last two full connected layers. Dropout operation is shown in Figure 6.



Note: The units of the horizontal and vertical axes are all centimeters.

Figure 5 Sigmoid and ReLU function curve

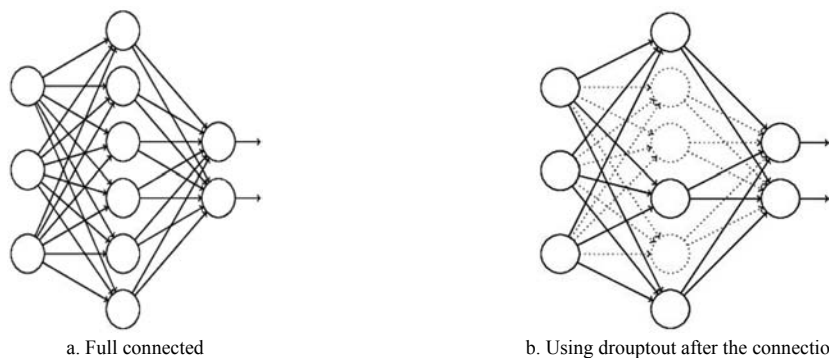


Figure 6 Dropout operation of the AlexNet

In Figure 6, each node represents a neuron. For a certain layer of neurons, by defining the probability to randomly delete some neurons, the probability of discarding parameters for a single neuron per layer is 50%. The dropout effectively prevents overfitting by modifying the structure of the neural network itself.

In order to overcome the problem of overfitting, in training AlexNet network model, first, the learning rate of 0.001 is used to make the network converge as soon as possible. Then reduce the learning rate continuously, and make the network reach the optimal solution. The last of the full connected layer is a softmax layer with 1000 outputs, according to the number of packets of the vegetable image data set, the number of output layer classification is changed from the original 1000 to 5. The traditional gradient descent method is used to train the parameters of the network. The whole learning process in the network gradually reduces the learning rate, to improve the learning speed of the initial network while ensuring that the model can make the loss function reliable convergence.

### 3 Results and discussion

The image data set used in the experiment were expanded by a total of 24 000 images, the training data set accounted for 80% (19 200), the test data set accounted for 20% (4800). The image data set have five categories of vegetables, namely: broccoli, pumpkin, cauliflower, mushrooms and cucumber. The number of training data set and test data set of each category of vegetable image are the same.

Using the test data set to test the classification effect of network. Accuracy rate reaches 92.1%, and loss rate is 0.284.

Figure 7 shows the experiment has a total of 8000 iterations, the iters on the horizontal axis represent the number of iterations. With the increase in the number of iterations, accuracy rate is increasing, loss rate is decreasing, and then tends to be smooth, get the final experimental results.

In order to verify the relationship between the accuracy rate of vegetable image classification and the number of image data set, a total of 10 000, 5000, and 1000 images were randomly selected and trained on the AlexNet network from 24 000 vegetable image data set. Among them, the training data set accounted for 80%, test data set accounted for 20%. The accuracy rates and loss rates of the 10 000, 5000, 1000 vegetable image data sets were 90.7% and 0.305, 86.6% and 0.381, 81.5% and 0.699, respectively.

According to the number of image data set, the number of experimental iterations also changed. When the number of data set is 10000, the accuracy and loss rate of 4000 iterations have been stabilized. And when the data set is 5000 and 1000, it only needs to iterate 2000 times.

Table 1 shows the effect of different numbers of vegetable image data set on image classification accuracy rate and loss rate.

It can be seen from Table 1, when the number of vegetable image data set continue to decrease, the accuracy of classification is decreasing and loss is rising, experiments show that the number of image data set will affect the accuracy of image classification.

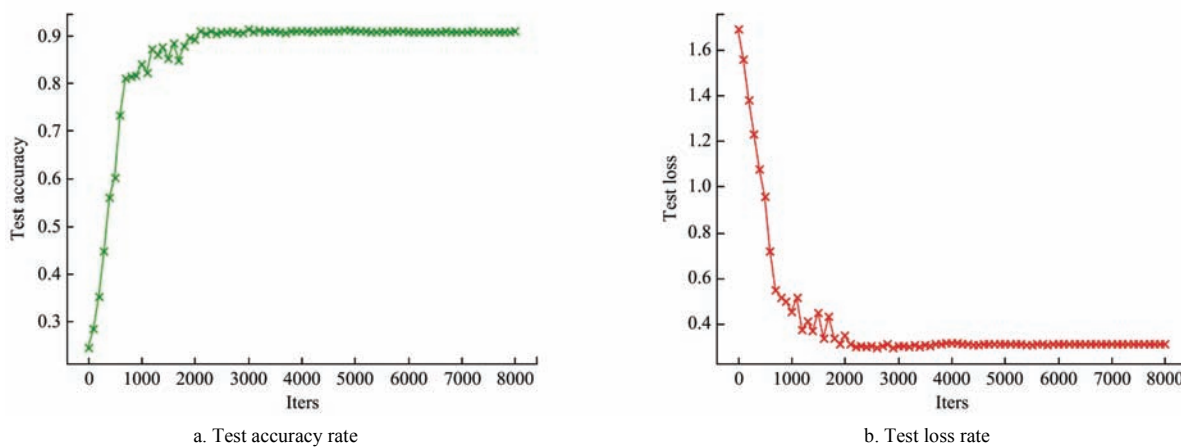


Figure 7 Accuracy and loss rate of 24 000 images

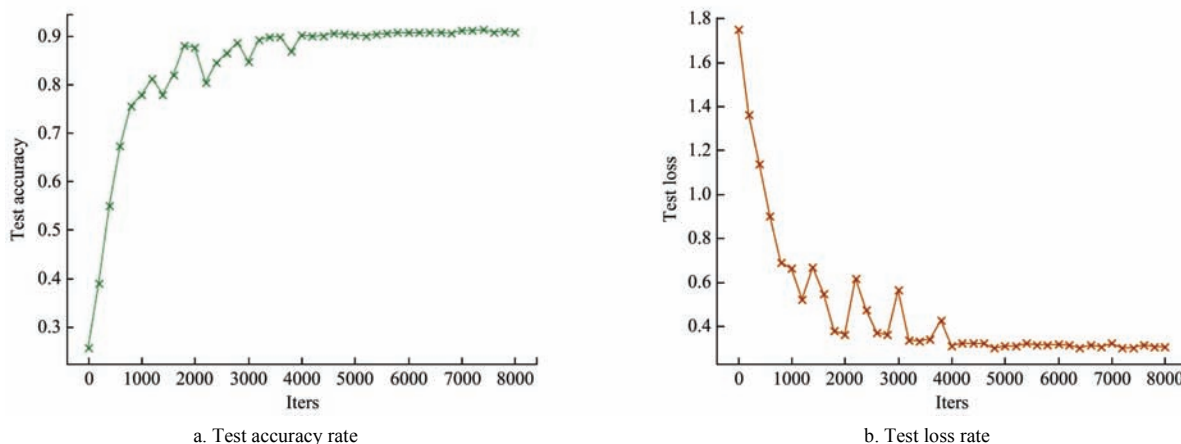


Figure 8 Accuracy and loss rate of 10 000 images

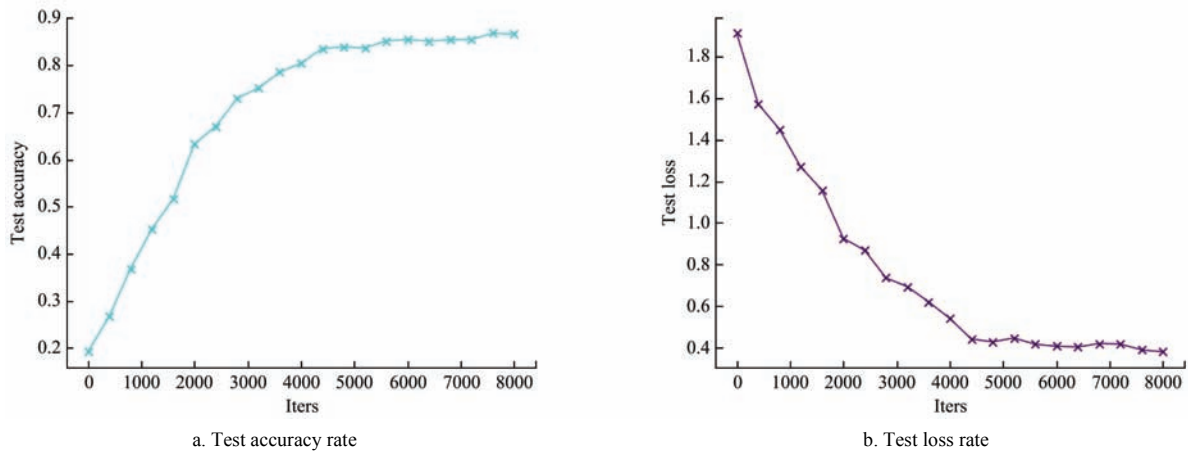


Figure 9 Accuracy and loss rate of 5000 images

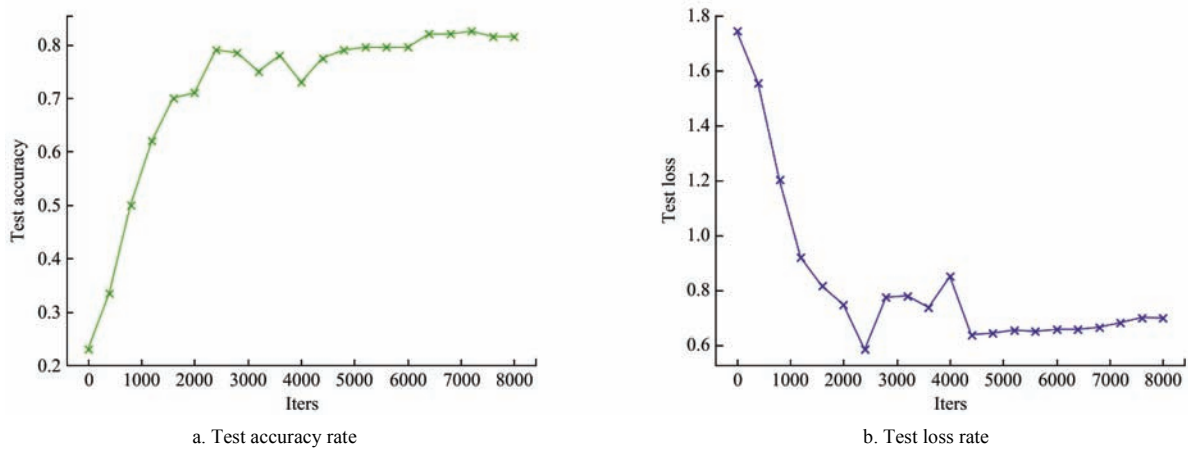


Figure 10 Accuracy and loss rate of 1000 images

**Table 1 Analysis of the number of image data Set and the accuracy of identification**

| The number of image data set | 24000 | 10000 | 5000  | 1000  |
|------------------------------|-------|-------|-------|-------|
| Accuracy/%                   | 92.1  | 90.7  | 86.6  | 81.5  |
| Loss                         | 0.284 | 0.305 | 0.381 | 0.699 |

This paper also makes a comparative experiment with the traditional BP neural network and the SVM classifier. 1000 data set of BP neural network and SVM classifier are selected randomly from the vegetable image data set used in this method. Among them, 800 training data set and 200 test data set, the number of images of each type of vegetables is the same. The two methods are to extract image color and shape features for image classification, in which the color feature using the color moment method, the shape features of the 7 Hu invariant moments method. The characteristic values of Hu in the images are shown in Table 2.

**Table 2 Hu invariant moment characteristic value**

| Image  | H1     | H2     | H3     | H4     | H5      | H6      | H7      |
|--------|--------|--------|--------|--------|---------|---------|---------|
| 1.jpeg | 1.1159 | 0.1179 | 0.0189 | 0.0155 | -0.0002 | -0.0018 | -0.0001 |
| 2.jpeg | 1.3171 | 0.1278 | 0.0084 | 0.0036 | -0.0000 | -0.0007 | -0.0001 |
| 3.jpeg | 1.2760 | 0.1513 | 0.0296 | 0.1434 | 0.0073  | 0.0442  | -0.0025 |
| 4.jpeg | 2.7107 | 0.4582 | 1.8447 | 0.1282 | 0.0440  | 0.0783  | 0.0125  |
| 5.jpeg | 1.6772 | 0.4681 | 0.5045 | 0.1787 | 0.0392  | 0.1084  | -0.0033 |

In Table 2, H1-H7 represent 7 invariant Hu moments. 1.jpeg-5.jpeg are the selected images from each type of vegetable images. For color eigenvalues, the eigenvalues of the nine-order moments of the three channels are extracted, as shown in Table 3.

Traditional SVM classifier is only used for two classification problems, since the image data set has 5 categories, therefore, one-to-one method is used for classification of vegetable images. The experimental results are listed in Table 4.

**Table 3 Color characteristic value**

| Image  | R                  |                  |                    | G                  |                  |                    | B                  |                  |                    |
|--------|--------------------|------------------|--------------------|--------------------|------------------|--------------------|--------------------|------------------|--------------------|
|        | First order moment | Two order moment | Three order moment | First order moment | Two order moment | Three order moment | First order moment | Two order moment | Three order moment |
| 1.jpeg | 0.4943             | 0.4167           | 0.5281             | 0.0109             | 0.0143           | 0.0309             | -0.0025            | 0.0013           | -0.0010            |
| 2.jpeg | 0.3662             | 0.2890           | 0.4393             | 0.0135             | 0.0187           | 0.0433             | -0.0011            | 0.0047           | 0.0060             |
| 3.jpeg | 0.2781             | 0.4054           | 0.5779             | 0.0284             | 0.0341           | 0.0446             | 0.0111             | 0.0033           | -0.0030            |
| 4.jpeg | 0.3587             | 0.3359           | 0.4045             | 0.1359             | 0.0588           | 0.0660             | 0.0301             | 0.0137           | 0.0089             |
| 5.jpeg | 0.1824             | 0.5480           | 0.5206             | 0.0110             | 0.0680           | 0.0850             | 0.0022             | -0.0007          | 0.0004             |

**Table 4 Analysis of the accuracy of different methods**

| Index      | AlexNet network model | BP neural network | SVM classifier |
|------------|-----------------------|-------------------|----------------|
| Accuracy/% | 92.1                  | 78                | 80.5           |

Table 4 shows the classification accuracy rate of the method used in this paper is obviously higher than that of BP neural network and SVM classifier. As shown in Table 1, the accuracy rate of the deep learning method requires a lot of data support. However, the number of data set has little effect on the accuracy of BP neural network and SVM classifier. So the number of data set used in the deep learning method is different from that of BP neural network and SVM classifier. When the number of data set used is the same (1000), the classification accuracy rate of deep learning method is still higher than that of BP neural network and SVM classifier. The results of comparative experiments verify the effectiveness of the method used in this paper.

#### 4 Conclusions

This paper proposed a high performance method based on deep learning framework for vegetable images classification. The main work is as follows:

Aiming at the demand of automatic picking and sorting of vegetables to target classification, this paper proposes an image classification method based on AlexNet convolutional neural network. 5 kinds of vegetable images are obtained from ImageNet data set, and the data expansion method is adopted to increase the number of image data set, so as to reduce the influence of overfitting. Using deep learning framework Caffe to train your own data set, and then test the classification effect of the network with the test data set. Using AlexNet network model to train different number of vegetable image data set, the experimental results show that the accuracy rate is decreasing as the number of data set decreases. This paper also compared with the traditional BP neural network and SVM classifier, the color and shape features of the image are extracted for classification. Among them, the multi classifier of SVM classifier adopts the one to one method, and the experimental results verify the effectiveness of the proposed method.

Although the classification of vegetable image data set has achieved some results, compared with the traditional BP neural network and SVM classifier accuracy has improved, but there are still deficiencies can be improved from the following aspects:

(1) Improving vegetable image data set. The image data sets used in this study are set only 5 types of vegetables. On this basis, the number of vegetable species can be increased.

(2) Using feature fusion. The method of feature fusion can be adopted, and merge the features of the first two full connected layers to classify and recognize the vegetable images. The accuracy rate may be improved.

#### Acknowledgements

This research was financially supported by the International Science & Technology Cooperation Program of China (2015DFA00530) and Key Research and Development Plan Project of Shandong Province (2016CYJS03A02).

#### [References]

- [1] Huo Z L, Wu H T, Hua X, Xu Y Y, Zhang Y X. Application of gray level co-occurrence matrix in vegetable species recognition. *Journal of China University of Metrology*, 2015; 26(1): 105–109.
- [2] He J P, Ma Y, Li Q. Fruit and vegetable automatic classification based on appearance feature. *Journal of Chongqing Normal University: Natural Science*, 2016; 3: 115–120.
- [3] Lee S H, Chan C S, Mayo S J, Remagnino P. How deep learning extracts and learns leaf features for plant classification. *Pattern Recognition*, 2017; 71: 1–13.
- [4] Wang P, Li W, Liu S, Gao Z M, Tang C, Ogunbona P. Large-scale Isolated Gesture Recognition using Convolutional Neural Networks. *International Conference on Pattern Recognition. IEEE*, 2016; 7–12.
- [5] Gao X W, Hui R, Tian Z. Classification of CT brain images based on deep learning networks. *Computer Methods & Programs in Biomedicine*, 2017; 138: 49–56.
- [6] Le Q V. Building high-level features using large scale unsupervised learning. In *Acoustics, 2013 IEEE International Conference on Speech and Signal Processing (ICASSP)*, 2013, May. pp.8595–8598.
- [7] Sun Z J, Xue L, Xu Y M, Wang Z. Overview of deep learning. *Application Research of Computers*, 2012; 08. (in Chinese)
- [8] Krizhevsky A, Sutskever L, Hinton G E. Imagenet classification with deep convolutional neural networks. In *Proc. Neural Information Processing Systems*, 2012.
- [9] Dan C, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. 2012; 157(10): 3642–3649.
- [10] Tan W X, Zhao C J, Wu H R, Gao R. A deep learning network for recognizing fruit pathologic images based on flexible momentum. *Transactions of the CSAE*, 2015; 46(1): 20–25. (in Chinese)
- [11] Li Y D, Hao Z B, Lei H. Survey of convolutional neural network. *Journal of Computer Applications*, 2016; 36(9): 2508–2515. (in Chinese)
- [12] Zeng X, Jie L I. Time-frequency image recognition based on convolutional neural network. *Machinery & Electronics*, 2016.
- [13] Zhou T. An image recognition model based on improved convolutional neural network. *Journal of Computational & Theoretical Nanoscience*, 2016; 13(7): 4223–4229.
- [14] Alotaibi M, Mahmood A. Improved gait recognition based on specialized deep convolutional neural networks. *Computer Vision and Image Understanding*, 2017; 164: 103–110.
- [15] Kaixuan Zhao, Dongjian He. Recognition of individual dairy cattle based on convolutional neural networks. *Transactions of the CSAE*, 2015; 31(5): 181–187. (in Chinese)
- [16] Gong D X, Cao C R. Plant leaf classification based on CNN. *Computer and Modernization*, 2014; 4: 12–15.
- [17] Hu J T, Fan C X, Ming Y. Trajectory image based dynamic gesture recognition with convolutional neural networks. *International Conference on Control, Automation and Systems, IEEE*, 2015; pp.1885–1889.
- [18] Qu J Y, Sun X, Gao X. Remote sensing image target recognition based on CNN. *Foreign Electronic Measurement Technology*, 2016; 8: 45–50. (in Chinese)
- [19] Tuama A, Comby F, Chaumont M. Camera model identification with the use of deep convolutional neural networks. *IEEE International Workshop on Information Forensics and Security. IEEE*, 2016.
- [20] Hentschel C, Wiradarma T P, Sack H. Fine tuning CNNs with scarce training data — Adapting ImageNet to art epoch classification. *IEEE International Conference on Image Processing, IEEE*, 2016; pp.3693–3697.
- [21] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015; pp. 1026-1034.
- [22] Ferrari V, Guillaumin M. Large-scale knowledge transfer for object localization in ImageNet. *IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society*, 2012; pp.3202–3209.
- [23] Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E. DeCAF: A deep convolutional activation feature for generic visual recognition. *Computer Science*, 2013; 50(1): 815–830.
- [24] Donahue J, Jia YQ, Vinyals O, Hoffman J, Zhang N, Darrell ET. DeCAF: A deep convolutional activation feature for generic visual recognition. *ICML*, 2014; 50(1): 647–655.