Chapter 10
An Apple Grading Method Based on Improved VGG16 Network

Lin Kaiyan, Zhang Haoping, Liu Chang, Si Huiping, Wu Junhui, and Chen Jie

Abstract Apple grading is an important link of improving the commercial value of apples. This paper proposed an improved convolutional neural network M-VGG16 based on very deep convolutional networks (VGG16) to improve the accuracy and efficiency of apple grading. A convolutional layer of the Conv3, Conv4 and Conv5 was, respectively, deleted compared with VGG16, then a 1 * 1 convolution was connected in parallel with the three groups of convolutions. All convolutional layers were activated by ReLU function. A batch normalization layer was added between each convolutional and activation function layer. Under same dataset, the M-VGG16 network classification prediction accuracy rate is 96.3%, which is higher than AlexNet and VGG16. Compared with traditional machine learning methods, the classification effect of deep learning method has more significant advantages. Therefore, the apple grading method based on deep learning has the advantages of higher grading efficiency and stronger generalization ability, and it provides a new solution for apple classification.

10.1 Introduction

Fruit is a major agricultural product in the world. As the largest fruit production and processing country in the world, China is known as the *kingdom of fruit*. Fruit processing has developed to be the third pillar agricultural industry in China with obvious advantages and competitiveness in agricultural product processing industry [1]. Improving the quality of fruits is a key factor for enhancing the competitiveness in the international market. Drawing on the experience of developed countries, fruit grading is one of important ways to increase production value.

Apple classification and rating were usually handled with visual screening which was laborious and time consuming. With the development of machine learning technology, deep learning methods taking advantage of fruit image datasets to mine deep features and had brought new research ideas for solving traditional fruit grading

L. Kaiyan · Z. Haoping · L. Chang · S. Huiping (✉) · W. Junhui · C. Jie
Modern Agricultural Science and Engineering Institute, Tongji University, Shanghai, China
e-mail: sihuiping@tongji.edu.cn

L. C. Jain et al. (eds.), *Smart Communications, Intelligent Algorithms and Interactive Methods*, Smart Innovation, Systems and Technologies 257,
https://doi.org/10.1007/978-981-16-5164-9_10

problems [2–5]. Great progress had been made in the field of deep learning methods applied in fruits grading. Sa et al. [6] proposed a method for fruit detection utilizing deep convolutional neural networks and developed an object detector called faster region-based CNN (faster R-CNN) to detect and classify seven types of fruits. This method not only improves accuracy but also allows for rapid deployment. Muresan et al. [7] proposed a deep neural network that was capable of identifying 60 types of fruits. And the paper introduced a high-quality dataset named Fruit-360 that contained popular fruits for obtaining a good classifier. The network finally reached an accuracy of 96.3% on the test set. Kestur et al. [8] presented a *MangoNet* based on architecture of semantic segmentation. Compared with variant architectures of fully convolutional networks (FCN), the *MangoNet* had better detection robustness because it could eliminate the impact of illumination conditions, occlusion in the detection process. The network met the requirements for detection and counting in mangoes yield estimation. In conclusion, the reason why the deep network is faster and more accurate than traditional classification is to improve the architecture in new application scenarios.

With convolutional networks becoming more popular in the field of computer vision, a number of works have been carried out to improve the original architecture. In this paper, we developed a supervised algorithm based on VGG16 in the bid to adapt to the particular fruit classification. The algorithm named modified VGG16 (M-VGG16) could solve the problem of overfitting caused by less data or complicated model in the area of deep learning. Experimental results showed that the M-VGG16 need less time than Alexnet and VGG16 in the training period, but achieving a much higher prediction accuracy.

10.2 Material Preparation

10.2.1 Apple Grading Standards

The apple grading standards showed in Table 10.1 were provided by the *National Standard of the People's Republic of China for Fresh Apples* (GB/T10651-2008) [9]. The defects of apple mainly include stab, crush and insect fruit. According to the main features, apples are divided into such four grades as superior grade, first grade,

Table 10.1 The standards of apple grading

Features	1st grade	2nd grade	3rd grade	4th grade
Fruit diameter (mm)	$\geq$100	$\geq$85	$\geq$70	<70
Fruit shape	Regular	Slight defect	Defective	Deformity
Red color degree (%)	$\geq$90	$\geq$80	$\geq$55	<55
Defective area (cm^2)	0	0	0	$\geq$0

Fig. 10.1 Apple images of four grades

(a)1st grade. (b) 2nd grade. (c) 3rd grade. (d) 4th grade.

second grade and malformed fruit, which are abbreviated as 1st grade, 2nd grade, 3rd grade and 4th grade, respectively.

According to Table 10.1, instead of a single feature, a multi-feature fusion method was beneficial to improve the grading accuracy. Firstly, if one apple has defective area, it will be classified in the 4th grade. If one apple has none defective area, it will be classified according to other characteristics. Secondly, if the other three characteristics are judged to be same rank, the fruit will be classified in this rank. If the above characteristic judgment results are inconsistent, the grade corresponding to the worst characteristic parameter shall prevail.

10.2.2 Data Acquisition

The datasets used in this paper came from two sources. One was to take photos of apples purchased from the fruit wholesale market. Another one came from Web search engine and open-source database. For the first one, images were acquired using the RGB camera. For the second source, images with complete outlines and clearly visible features were selected as part of the dataset.

A total of 3000 apples pictures were collected with each grade having 750 pictures so as to maintain the balance of samples in each category. The image set used for training consisted of 2100 images. The test set was made of 900 images. Examples of apples of different grades are shown in Fig. 10.1.

10.2.3 VGGNet

The VGGNet had confirmed depth was beneficial for the classification accuracy in visual tasks. It utilizes 3 * 3 convolutional layers which are the smallest receptive field. In spite of a large depth, the number of weights is decreased significantly due to a stack of small convolutional filters [10]. Two excellent-performance models (VGG16 and VGG19) were released for accelerating further research. In this paper, VGG16 was chosen as baseline because VGG19 was better suited to complex tasks.

The network structure of VGG16 is shown in Fig. 10.2. The input of the convolutional network is fixed size 224 * 224 * 3 RGB images. Compared with previous work, rather than relatively large receptive fields, 3 * 3 convolution kernels are used

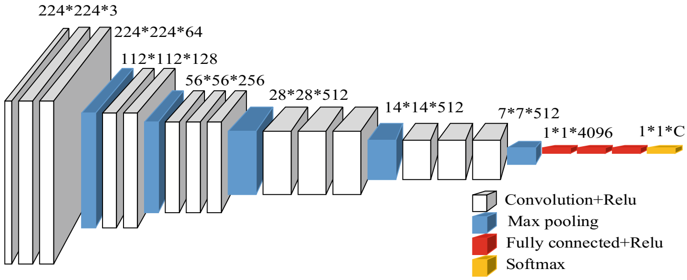


Fig. 10.2 The network architectures for VGG16

in the whole net. Therefore, the VGG16 increases the number of convolutional layers, making the convolutional neural network deeper and the prediction result more accurate. Furthermore, VGG16 has good generalization ability and show excellent effect when applied to various data types [11].

During training of convolutional neural network, small datasets of labeled images and too complicated models are easier to cause overfitting [12]. Therefore, it was essential for preventing overfitting to collect larger datasets or introduce better techniques.

In our experiment, the size of datasets is small and the resolution of apple images is relatively low. If the network has considerably deep layers, feature information would be lost seriously, resulting in disappearance of the gradient and unsuccess in updating weights of the back propagation. Despite deep convolutional neural networks having attractive qualities, it is obvious to be expensive for a relatively simple task requiring large memory and long training time. The depth VGG16 is moderate, and it is suitable for tasks with fewer samples. In the paper, the VGG16 was selected as the basic network model.

10.3 Improvement of VGG16 Architecture

As we have known, it is difficult to train deeper neural networks. A residual learning framework is easier to optimize the degradation problem. With *shortcut connections* to skip one or more layers, a deep residual network could address the notorious problem of vanishing/exploding gradients. The number of parameters is directly determined by convolution kernels size in a convolutional neural network. A big kernel has big perceptual field for recognizing large objections. However, the disadvantage is that the parameters would be redundant [13–15]. Therefore, 1 * 1 convolution filters are often used in terms of network structure design. It performs a linear combination of pixels on different channels, with result in increasing the linear expression and changing the dimension of the feature map. The feature map will be reduced by reducing the dimensionality of 1*1 kernel, and the parameter amount will be reduced accordingly.

The paper proposed a new network called M-VGG16 based on the VGG16 network, the shortcut connections unit is shown in Fig. 10.3. Although the shortcut connections unit added extra parameters, the entire network still had lower complexity and could be easily trained. Therefore, the 1 * 1 kernel could not only provide the previous layer information to the later layer network, but also reduce the parameters in the network and shorten the training time. The overall framework for the M-VGG16 is depicted in Fig. 10.4.

A convolutional layer of the Conv3, Conv4 and Conv5 is, respectively, deleted compared with VGG16, and then a 1 * 1 convolution is connected in the three groups of convolutions. The channels of the three 1 * 1 convolution kernels are 128, 256 and 512, respectively. The output feature maps of the maxpooling layer and that of the 1 * 1 convolution are stacked in series with "concatenate" way so that subsequent network layers could obtain the information of all the previous layers and avoid gradient disappearance. All convolutional layers are activated by ReLU function. Batch normalization (BN) layer is placed between each convolutional layer and activation function layer. The BN changes the distribution of the input value of each layer to standard normal distribution with standardization method so that making the input value fall in the area where the nonlinear function is more sensitive for the input value.

The feature information of different grades of apples is highly similar, so the above improvement method could reduce the loss of feature information and increase the accuracy of model recognition.

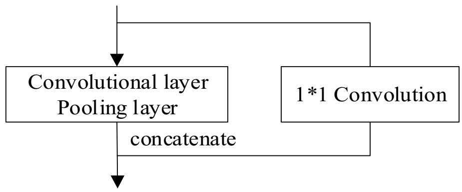


Fig. 10.3 The shortcut connections unit

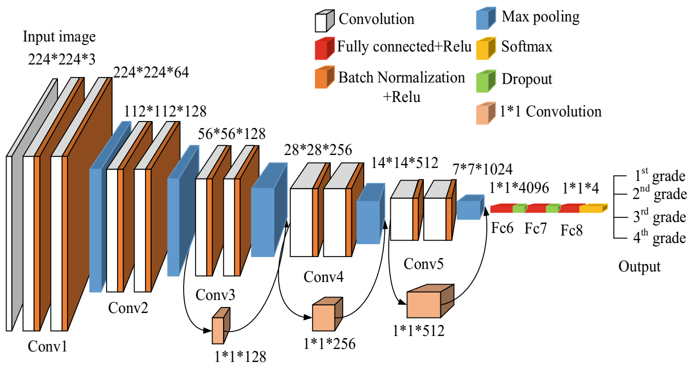
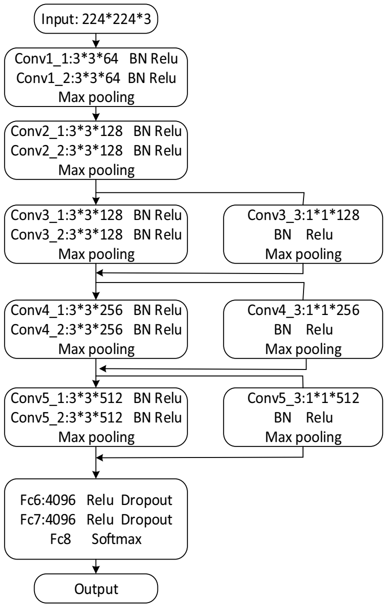


Fig. 10.4 The network architectures for M-VGG16

Fig. 10.5 M-VGG16 model parameter diagram



The parameters of each layer of the M-VGG16 network model are shown in Fig. 10.5 in detail. The initial input image size is 224 * 224 * 3. The final output is a four-dimensional vector, in which each value in represents the probability of one picture belonging to responding category.

10.4 Experiment of Apple Grading

10.4.1 Hardware Platform

In this experiment, all networks were trained and tested on a deep learning experiment platform that had a NVIDIA P100 GPU, Intel Core i7 processor and 60 GB memory running on an Ubuntu 16.04 Linux operating system. Open-source libraries such as OpenCV and PIL were applied in image preprocessing. TensorFlow is used to create network model.

10.4.2 Experiment Procedure

As shown in the training flowchart (Fig. 10.6) of the experiment, the process of forward propagation was to extract features, and they were mapped to classification

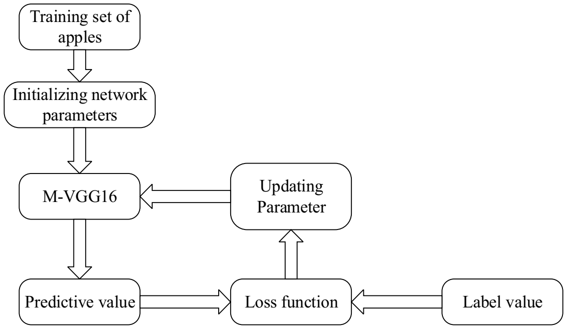


Fig. 10.6 M-VGG16 training flowchart

results by full connection layer. In the process of back propagation, the errors or loss between the predicted value and the label value was calculated. Here are the specific training steps.

1. Setting input layer size, the kernel size, number and step size of each network layer.
2. Initializing the network parameters according to the standard Gaussian distribution. The mean and variance of parameter distribution are 0 and 1, respectively.
3. Performing the network training in forward propagation. The input images pass through each layer of the network. The prediction result is outputted finally.
4. Calculating the error between the predicted value and the label value and updating the parameters with the gradient descent algorithm.
5. Repeating the above operation to stop when reaching the set training times and saving the model parameters.

10.4.3 Training Parameters

The paper utilized Adam (adaptive moment estimation) optimization method for training. It was an improved stochastic gradient descent algorithm with the advantage that the learning rate could be adjusted for parameters of different frequencies during network training. Weighted sum of cross entropy error function and complex loss were calculated as loss function of the network in this paper.

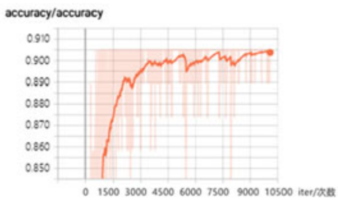
$$\text{Loss} = -\frac{1}{n}\sum_{p=1}^{n}\sum_{i=1}^{n}[y_i * \log(y_p) + (1 - y_i) * \log(1 - y_p)] + \frac{\lambda}{2}\sum_{i=1}^{n}(y_i - y_p)^2 \tag{10.1}$$

where n is the total number of input samples, y_i is the label corresponding to each sample, y_p is the probability of predicting the input sample as y_i, λ is the weight of the complexity loss.

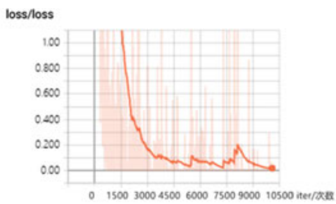
Network hyper parameters needed to be set before training the network. The learning rate has a crucial influence on whether a model could be successfully trained. The learning rate for this experiment was set to 0.0001, and batch size represented the number of samples for training each time. An epoch meant that all data in the training set is calculated. In this experiment, the batch size was set to 64 and epoch was set to 200. In addition, the random inactivation probability of the dropout layer was 0.5.

10.4.4 Experimental Results

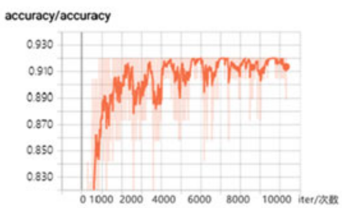
In this experiment, we utilized the accuracy curve and the loss function curve with corresponding actual score as evaluation index for apple classification. AlexNet and VGG16 were selected to compare with the proposed algorithm. Figure 10.7 depicts



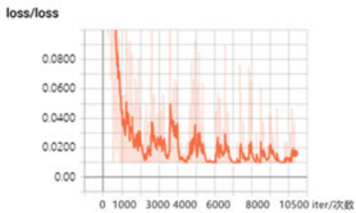
(a)Accuracy curve of AlexNet.



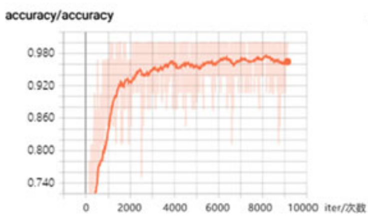
(b)Loss Function curve of AlexNet.



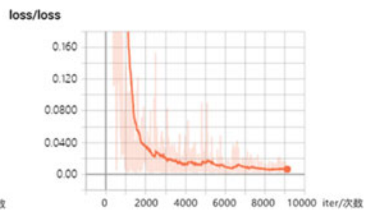
(c)Accuracy curve of VGG16.



(d)Loss Function curve of VGG16.



(e)Accuracy curve of M-VGG16.



(f)Loss Function curve of M-VGG16.

Fig. 10.7 Accuracy and loss function curve of AlexNet, VGG16 and M-VGG16

Table 10.2 Grading results of apples

	AlexNet	VGG16	M-VGG16
Training accuracy (%)	90.5	92.4	97.8
Test accuracy (%)	89.7	88.9	96.3

the accuracy and loss function curve of AlexNet, VGG16 and M-VGG16. Train accuracy in the last point and test accuracy are shown in Table 10.2.

As shown in Fig. 10.7a, c, the training time of VGG16 was longer than that of AlexNet. The main reason is that VGG16 has more convolutional layers than AlexNet and the network depth is deeper; therefore, there are more parameters needed to be learned. Moreover, the training accuracy and test accuracy of VGG16 are 0.924 and 0.889, respectively, in Table 10.2. The accuracy of VGG16 is higher than AlexNet in the training set, but lower in the test set.

M-VGG16 obviously outperformed both AlexNet and VGG16 shown in Table 10.2. The proposed method showed that the training accuracy reached up to 0.978. In test set, it could achieve 0.963 of the maximum accuracy and required less computation time. Although M-VGG16 network has the same number of layers as VGG16, it uses 1*1 convolution layer instead of the 3 * 3 convolution layer, which can not only provide the characteristics of the former network layer to the latter for learning, but also make training time shorter. AlexNet and VGG16 had not achieved good results in the task of apple classification. The former had low classification accuracy due to its simple network structure, while the latter caused overfitting problems due to its complex network structure. It is shown that the proposed method M-VGG16 has moderate complexity and significant generalization.

10.4.5 Apple Grading Based on Machine Learning Algorithms

In the task of apple grading, color is one of the important criteria for distinguishing apples of different grades. The colors of different grade apple vary greatly. It is appropriate to choose the statistical model of color histogram to grade apples. HSV is the most commonly used color model in image processing, which conforms to human perception of color features. Meanwhile, the shape of an apple is also an important criterion for grade judging. The histogram of oriented gradient (HOG) feature can characterize the shape of an object with the help of the gradient or the directional density distribution of the edge. Support vector machine (SVM) is a supervised machine learning algorithm that realizes feature classification by mapping low-dimensional data to high-dimensional feature space and has strong generalization capabilities. Therefore, this paper extracted the HSV color histogram feature and the direction gradient histogram feature separately as the input of the SVM classifier to distinguish apples of different grades. Firstly, the apple images were converted to HSV space. Secondly, quantizing the three HSV channels and extracting the HSV

Table 10.3 Results of apple grading experiments

Method	Index	1st grade	2nd grade	3rd grade	4th grade	Total
HSV + SVM	NS	375	375	375	375	1500
	FD	39	98	18	34	189
	Accuracy (%)	89.6	73.87	95.2	90.93	87.4
HOG + SVM	NS	375	375	375	375	1500
	FD	40	89	10	15	154
	Accuracy (%)	89.3	76.27	97.3	96.0	89.73
M-VGG16	NS	375	375	375	375	1500
	FD	19	27	9	10	65
	Accuracy (%)	94.9	92.8	97.6	97.3	95.7

The all number of samples (NS), the number of false detection (FD)

color histogram feature vector. Finally, the feature vectors are inputted into SVM for training. On the base of the shape feature classification part, the HOG features of each grade of apples were extracted and input into the SVM classifier for training.

The results are shown in Table 10.3. The overall accuracy based on HSV + SVM and HOG + SVM are 0.874 and 0.8973, respectively. Two machine learning algorithms have achieved relatively good classification results, while the classification accuracy rate based on the convolutional neural network M-VGG16 reaches 0.957. Compared with these two traditional machine learning methods, the classification effect of deep learning methods has more significant advantages.

10.5 Conclusions

This paper studied the apple grading method based on deep learning. In order to solve the problem of overfitting and loss of feature information, this paper proposed an improved neural network M-VGG16. VGG16 was selected as the basic framework to reduce the complexity of the network. 1 * 1 convolution layer was connected in parallel next to the Conv3, Conv4 and Conv5 in the backbone network to reduce the loss of feature information. The three network architectures were tested on the same apple dataset. Comparing with AlexNet and VGG16, the M-VGG16 had the best classification effect and the fastest network convergence rate. It effectively solved the problem of overfitting. Moreover, this article compared the performance of deep learning and traditional classification methods. The conclusion was that the M-VGG16 method based on deep learning has the advantages of higher grading efficiency, stronger generalization ability and automatic feature extraction in apple grading.

Acknowledgements This work was supported by a National Science and Technology Support Project named "Research on Evaluation Index and Standard System of Green Technology" (Project No. 2017YFC0212901).

References

1. Sui, W., Jia, H.: Research on utilization status and classification management of solid waste in fruit processing in China. Trans. Chin. Soc. Agric. Eng. **34**(S1), 172–180 (2018)
2. Huang, C., Fei, J.: Online apple grading based on decision fusion of image feature. Trans. Chin. Soc. Agric. Eng. **33**(1), 285–291 (2017)
3. Gu, Y., Shi, G., Liu, X.: Optimization spectral clustering algorithm of apple image segmentation with noise based on space feature. Trans. Chin. Soc. Agric. Eng. **32**(16), 159–167 (2018)
4. Kumar, A., Kumar, K.: Time-frequency analysis and support vector machine in auto-matic detection of defect from vibration signal of centrifugal pump. Measurement **108**, 119–133 (2017)
5. Jaffar, A., Jaffar, R., Jamil, N.: Photogrammetric grading of oil palm fresh fruit bunch-es. Int. J. Mech. Mechatron. Eng. **9**(10), 18–24 (2017)
6. Inkyu, S., Zongyuan, G., Feras, D.: DeepFruits: a fruit detection system using deep neu-ral networks. Sensors **16**(8), 1222–1228 (2017)
7. Mureşan, H., Oltean. M.: Fruit recognition from images using deep learning. Acta Uni-versitatis Sapientiae Informatica **10**(1), 26–42 (2018)
8. Kestur, R., Avadesh, M., Omkar, N.: Mango Net: A deep semantic segmentation archi-tecture for a method to detect and count mangoes in an open orchard. Eng. Appl. Artif. Intell. **77**, 59–69 (2019)
9. State Administration for Market Regulation, Standardization Administration.: GB/T1065-2008, Fresh apple. China Federation of Supply and Marketing Coopera-Tives (2008)
10. Guo, Z., Chen, Q., Zhang, B.: Design and experiment of handheld near-infrared spec-trometer for determination of fruit and vegetable quality. Trans. Chin. Soc. Agric. Eng. **33**(8), 245–250 (2017)
11. Sapan, N., Bankim, P.: Usage of image processing and machine learning techniques in agriculture—fruit sorting. CSI Commun. **37**(7), 25–27 (2013)
12. Blasco, J., Aleixos, N., Moltó, E.: Computer vision detection of peel defects in Citrus by means of a region oriented segmentation algorithm. J. Food Eng. **81**(3), 535–543 (2007)
13. Xianfeng, L., Weixing, Z.: Apple grading method based on features fusion of size, shape and color. Procedia Eng. **15**, 2885–2891 (2011)
14. Wang, Z., Chi, Q., Guo, W.: Internal quality detection of apples during late development period based on near-infrared spectral technology. Trans. Chin. Soc. Agric. Mach. **49**(05), 348–354 (2018)
15. Devrim, U., Bemard, G., Olivier, K.: Automatic grading of Bi-colored apples by multi-spectral machine vision. Comput. Electron. Agric. **75**(1), 204–212 (2011)