The Alexnet-ResNet-Inception Network for Classifying Fruit Images

Wenzhong Liu[a,b,*]

[a]School of Computer Science and Engineering, Sichuan University of Science & Engineering, Zigong, 640000, China;
[b]Key Laboratory of Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things, Zigong, 640000, China;
*To whom correspondence should be addressed. E-mail address: newwenzhong@hotmail.com.

Abstract

Fruit classification contributes to improving the self-checkout and packaging systems in supermarkets. The convolutional neural networks can automatically extract features through directly processing the original images, which has thus attracted wide attention from researchers in terms of fruit classification. However, it is difficult to achieve more accurate recognition due to the complexity of category similarity. In this study, the Alexnet, ResNet, and Inception networks were integrated to construct a deep convolutional neural network named Interfruit, which was then utilized in identifying various types of fruit images. Afterwards, a fruit dataset involving 40 categories was also constructed to train the network model and to assessits performance. According to the evaluation results, the overall accuracy of Interfruit reached 92.74% in the test set, which was superior to several state-of-the-art methods. To sum up, findings in this study indicate that the classification system Interfruitr ecognizes fruits with high accuracy and has a broad application prospect. All data sets and codes used in this study are available at https://pan.baidu.com/s/19LywxsGuMC9laDiou03fxg, code: 35d3.

Keywords: Fruit classification; Alexnet; ResNet; Inception

1. Introduction

In the food industry, fruit represents a major component of fresh produce. Fruit sorting not only helps children and those visually impaired people to guide their diet(Khan and Debnath, 2019), but also assists the supermarkets or grocery stores in improving the self-checkout, fruit packaging, and transportation systems. Fruit classification has always been a relatively complicated problem, as a result of their wide variety and irregular shape, color and texture characteristics(García-Lamont et al., 2015). In most cases, the trained operators are employed to visually inspect fruits, which requires that, these operators should be familiar with the unique characteristics of fruits and maintain the continuity as well as consistency of identification criteria(Olaniyi et al., 2017). Given the lack of a multi-class automatic classification system for fruits, researchers have begun to employ Fourier transform near infrared spectroscopy, electronic nose, and multispectral imaging analysis for fruit classification(Zhang et al., 2016). However, these devices are expensive and complicated in operation, with no high overall accuracy.

The image-based fruit classification system requires only a digital camera, and can achieve

favorable performance, which has thus attracted extensive attention from numerous researchers. Typically, this new solution adopts wavelet entropy, genetic algorithms, neural networks, support vector machines, and other algorithms to extract the color, shape, and texture characteristics of fruits for recognition (Wang and Chen, 2018). For fruits that have quite similar shapes, color characteristics become the criteria for the successful fruit classification(Yoshioka and Fukino, 2010). Nonetheless, these traditional machine learning methods require the manual feature extraction process, and feature extraction methods may be redesigned in calibrating parameters(Yamamoto et al., 2014). For example, for apple and persimmon images that are very similar in color and shape, the traditional methods can hardly accurately distinguish between them. To solve this problem, a computer vision-based deep learning technology is proposed(Koirala et al., 2019). Notably, deep learning is advantageous in that, it directly learns the features of fruit images from the original data, and the users do not need to set any feature extraction method(Kamilaris and Prenafeta-Boldú, 2018). Convolutional neural networks stand for the earliest deep learning methods used for identifying fruits, which adopts numerous techniques, such as convolution, activation, and dropout(Brahimi et al., 2017). However, the deep learning methods have not been widely utilized to classify many categories of fruits, and the classification accuracy is still not high(Rahnemoonfar and Sheppard, 2017).

To enhance the recognition rate of deep learning for fruits, a deep learning architecture named Interfruit was proposed in this study for fruit classification, which had integrated the AlexNet, ResNet, and Inception networks. Additionally, a common fruit dataset containing 40 categories was also established for model training and performance evaluation. Based on the evaluation results, Interfruit's classification accuracy was superior to the existing fruit classification methods.

2. Materials and Methods

2.1 Data set

Altogether 3,139 images of common fruits in 40 categories were collected from Google, Baidu, Taobao, and JD.com to build the image data set, IntelFruit **(Figure 1)**. Each image was cropped to 300*x*300 pixels. **Table 1** shows the category and number of fruit pictures used in this study. For each type of fruit images, 70% images were randomly assigned to the training set, while the remaining 30% were used as the test set. The as-constructed model was trained based on the training set and evaluated using the test set.

2.2 Convolutional Layer

The convolutional neural networks are a variant of deep networks, which automatically learn simple edge shapes from raw data, and identify the complex shapes within each image through feature extraction. The convolutional neural networks include various convolutional layers similar to the human visual system. Among them, the convolutional layers generally have filters with the kernels of 11 × 11, 9 × 9, 7 × 7, 5 × 5 or 3 × 3. The filter fits weights through training and learning, while the weights can extract features, just similar to camera filters.

2.3 Rectified Linear Unit (ReLU)Layer

Convolutional layers are linear, which are unable to capture the non-linear features. Therefore, a rectified linear unit (ReLU) is used as a non-linear activation function for each convolutional layer.ReLU suggests that, when the input value is less than zero, the output value will be set to zero. Using the ReLU, the convolutional layer is able to output the non-linear feature maps, thereby reducing the risk of overfitting.

2.4 Pooling Layer

The pooling layer is adopted for compressing the feature map after the convolutional layer. The pooling layer summarizes the output of the neighboring neurons, which reduces the activation map size and maintains the unchanged feature.There are two methods in the pooling layer, i.e. maximum and average pooling. In this paper, the maximum pooling (MP)method was adopted. Typically, the MP method remains the maximum pooling area, and it is the most popular pooling strategy.

2.5 ResNet and Inception Structure

The general convolutional neural networks tend to overfit the training data and have poor performance on the actual data. Therefore, the ResNet and Inception layer was used to solve this problem in this study.The Deep Residual (ResNet) network changes several layers into a residual block. Besides, the ResNet Network solves the degradation problem of deep learning networks, accelerates the training speed of deep networks, and promotes the faster network convergence.

In addition, the Inception structure connects the results of convolutional layers with different kernel sizes to capture features of multiple sizes. In this study, the inception module was integrated into one layer by several parallel convolutional layers. Notably, Inception reduces the size of both modules and images, and increases the number of filters. Further, the module learns more features with fewer parameters, making it easier for the 3D space learning process.

2.6 Fully Connected and Dropout Layer

Fully connected layer (FCL) is used for inference and classification. Similar to the traditional shallow neural network, FCL also contains many parameters to connect to all neurons in the previous layer. However, the large number of parameters in FCL may cause the problem of overfitting during training, while the dropout method is a technique to solve this problem. Briefly, the dropout method is implemented during the training process by randomly discarding units connected to the neural network. In addition, the dropout neurons are randomly selected during the training step, and its appearance probability is 0.25. During the test step, the neural network is used without dropout operation.

2.7 Model Structure and Training Strategy

In this study, the convolutional neural network, **IntelFruit**, was constructed to classify fruits (**Figure 2**). According to **Figure 2**, the input image with a size of 227×227×3 was fed into the IntelFruit network. IntelFruit was a stack architecture integrating AlexNet + ResNet + Inception, which consisted of an AlexNet component, a ResNet component, an Inception component, and

three fully connected layers. Notably, the last fully connection layer played a role as a classifier, which calculated and output the scores of different fruits.

To minimize errors, the Adam optimizer was also employed in this study, which was superior in its high computational efficiency, low memory requirements and great suitability for large data or many parameters. The learning rate of the Adam optimizer was set to a constant of 1×10e-4, and CrossEntropyLoss was used as a cost function. Thereafter, the as-proposed model was trained and tested end-to-end on the i7-8750H processor, with 32 GB of running memory and the operating system of WIN 10 x64.

2.8 Metrics of Performance Evaluation

The prediction performance of classifier was evaluated by two metrics, including accuracy (***Acc***), average ***F1-score.*** To be specific, the metrics were defined as follow:

$$Accuracy = \frac{N_P}{N_{total}} \tag{1}$$

$$AvgF1-score = \frac{\sum F1-score}{N_{total}} \tag{2}$$

Where N_p is the number of all correctly classified pictures , N_{total} is the number of all pictures. Average ***F1-score*** was calculated using the method average = “weighted” of the sklearn.metrics package

3. Result and Discussion

3.1 Loss and Accuracy Rate Curve

In terms of time and memory consumption in model training, the loss vs. accuracy curve is an effective feature. **Figure3.A** presents the loss rate curve of IntelFruit on training and test sets in 200 iterations. Clearly, the loss curve of the test set was similar to the training set, with lower errors at epoch 73, indicating the high stability of IntelFruit. **Figure 3.B** illustrates the accuracy curves of the training and testing sets. A low error rate was achieved at epoch 79, suggesting that IntelFruit effectively learned data and might serve as a good model for fruit recognition.

3.2 Confusion Matrix

In this work, the proposed deep learning network IntelFruit was trained on the fruit dataset. Afterwards, the model was evaluated on the test set, which showed good performance. **Figure 4** presents the confusion matrix of the classification results, where each row represents the actual category, while each column stands for the predicted result. In addition, the number (*m-th* row and *n-th* column) indicated the number of actual instances to the *m-th* label and predicted to be the *n-th* label.

The performance of the classifier was visually evaluated based on the results, and highlighted classes and features of the network model were also determined. IntelFruit obtained a high recognition rate. Typically, the best classified fruits were Grape_Black and Pineapple with different shapes, colors and characteristics from other fruits. As clearly observed from **Figure 4**,

among the 40 categories, 67 images were incorrectly predicted as the 23 category, and the remaining 856 images were correctly predicted. Therefore, the best classified fruit Grape_Black and Pineapple etc. achieved an accuracy of 100%. By contrast, the worst classified fruits were Apricot and Plum, with low accuracy. According to the above results, the IntelFruit model was able to better identify different fruits.

3.3 Comparison of Classification Performance

To evaluate the effectiveness of these models, the as-proposed method was compared with the existing methods for modern deep learning. The models were evaluated on the test set by the accuracy rate, and avg ***F1-score*** (**Table 2**). In **Table 2**, the model IntelFruit achieved lower false positive and false negative rates, which demonstrates the effectiveness. For the fruit dataset involving 40 categories, the accuracy value of the proposed model was 92.74%, which was subsequently compared with of Alexnet, GoogLeNet and ResNet18. The accuracy values of these three methods were as follows, 83.97% for Alexnet, 84.83% for GoogLeNet, and 75.52% for ResNet18. In addition, **Table 2** clearly illustrated that the avg F1-score of the proposed model was 96.23%, which was also superior to the existing models. In the case of fruit recognition, IntelFriu was more effective than those previous methods, revealing the superiority of the proposed AlexNet-ResNet-Inception network. In general, the intelFriut model with the highest recognition rate has promising application value in the food and supermarket industries.

Noticeably, IntelFruit was associated with many advantage, it ushered in a new method to classify 40 different types of fruits simultaneously. The high-precision results showed that, convolutional neural networks might also be used to achieve high performance and faster convergence, even for the smaller data sets. This model captured images to train the model without preprocessing the images to eliminate the background noise and the lighting settings. This model showed excellent performance in the evaluated cases; however, it was linked with some difficulties in some cases. For instance, for categories Apricot and Plum, some categories were easily confused with others due to the insufficient sample sizes, leading to false positives or lower accuracy.

4. Conclusions

It is quite difficult for supermarket staff to remember all the fruit codes, and it is even more difficult to sort the fruits automatically if no barcodes are printed on the fruits. In this work, a novel deep convolutional neural network named intelFriut is proposed, which is then used to classify the common fruits and help supermarket staff to quickly retrieve the fruit identification ID and price information. IntelFriut is an improved stack model that integrates AlexNet, ResNet and Inceptiont, with no need for extracting color and texture features. Furthermore, different network parameters and DA techniques are used to improve the prediction performance of this network. Beside, this model is evaluated on the fruit dataset intelFriut, and compared with several existing models. The evaluation results show that the intelFriut network proposed in this study achieves the highest recognition rate, with the overall accuracy of 92.74%, which is superior to other models. Taken together, findings in this study indicate that the network combining AlexNet, ResNet and Inception achieves higher performance and has technical validity. Therefore, it can be concluded that, intelFriut is a novel and highly computational tool for fruit classification with broad application prospects.

Acknowledgements

This work was partially supported by the Opening Project of Key Laboratory of Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things (No. 2018WZY02).

References

Brahimi, M., Boukhalfa, K., Moussaoui, A., 2017. Deep learning for tomato diseases: classification and symptoms visualization. Applied Artificial Intelligence 31, 299-315.

García-Lamont, F., Cervantes, J., Ruiz, S., López-Chau, A., 2015. Color characterization comparison for machine vision-based fruit recognition, International conference on intelligent computing. Springer, pp. 258-270.

Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. Computers and electronics in agriculture 147, 70-90.

Khan, R., Debnath, R., 2019. Multi Class Fruit Classification Using Efficient Object Detection and Recognition Techniques. International Journal of Image, Graphics and Signal Processing 11, 1.

Koirala, A., Walsh, K.B., Wang, Z., McCarthy, C., 2019. Deep learning–Method overview and review of use for fruit detection and yield estimation. Computers and Electronics in Agriculture 162, 219-234.

Olaniyi, E.O., Oyedotun, O.K., Adnan, K., 2017. Intelligent grading system for banana fruit using neural network arbitration. Journal of Food Process Engineering 40, e12335.

Rahnemoonfar, M., Sheppard, C., 2017. Deep count: fruit counting based on deep simulated learning. Sensors 17, 905.

Wang, S.-H., Chen, Y., 2018. Fruit category classification via an eight-layer convolutional neural network with parametric rectified linear unit and dropout technique. Multimedia Tools and Applications, 1-17.

Yamamoto, K., Guo, W., Yoshioka, Y., Ninomiya, S., 2014. On plant detection of intact tomato fruits using image analysis and machine learning methods. Sensors 14, 12191-12206.

Yoshioka, Y., Fukino, N., 2010. Image-based phenotyping: use of colour signature in evaluation of melon fruit colour. Euphytica 171, 409.

Zhang, Y., Phillips, P., Wang, S., Ji, G., Yang, J., Wu, J., 2016. Fruit classification by biogeography□based optimization and feedforward neural network. Expert Systems 33, 239-253.

Legend

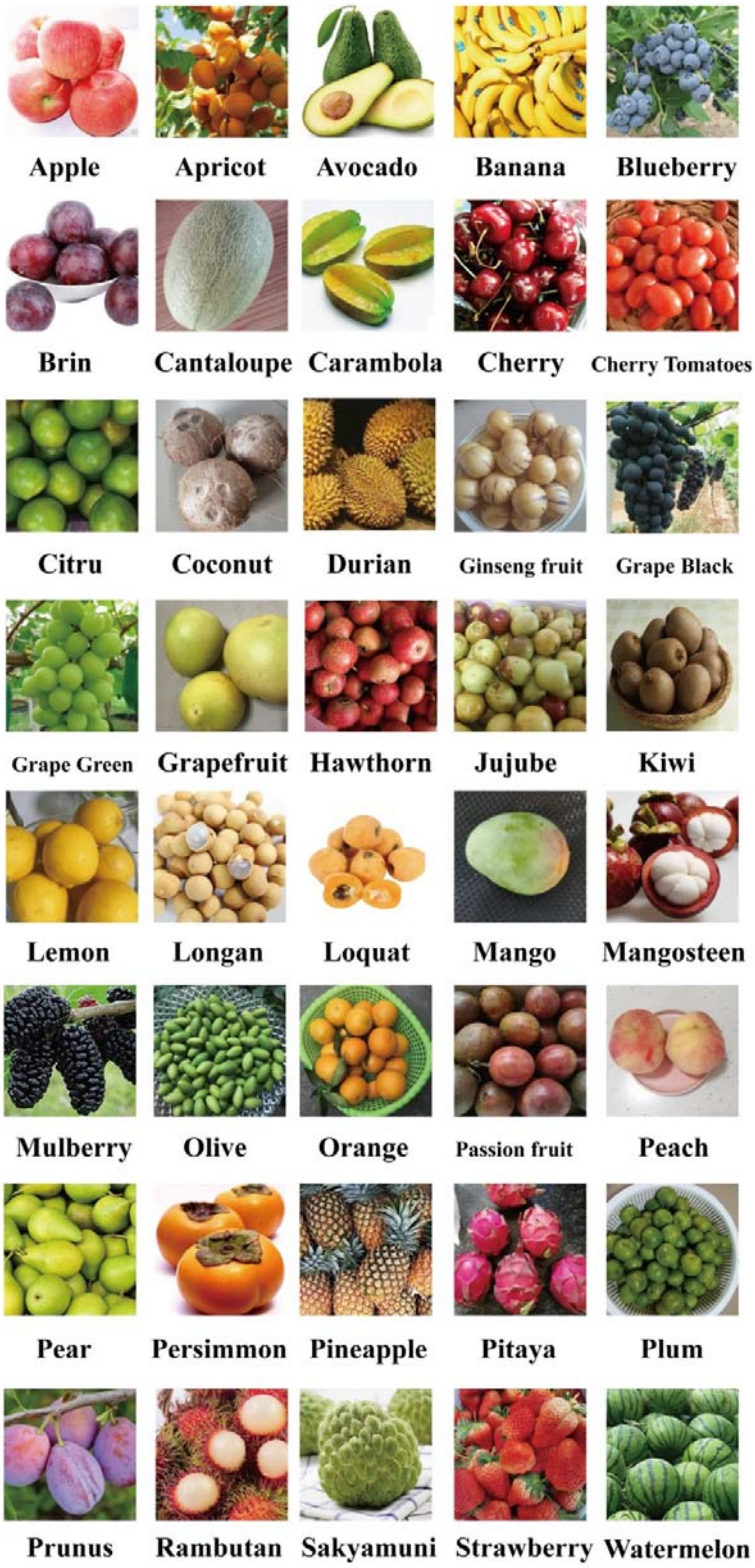
Apple
Apricot
Avocado
Banana
Blueberry
Brin
Cantaloupe
Carambola
Cherry
Cherry Tomatoes
Citru
Coconut
Durian
Ginseng fruit
Grape Black
Grape Green
Grapefruit
Hawthorn
Jujube
Kiwi
Lemon
Longan
Loquat
Mango
Mangosteen
Mulberry
Olive
Orange
Passion fruit
Peach
Pear
Persimmon
Pineapple
Pitaya
Plum
Prunus
Rambutan
Sakyamuni
Strawberry
Watermelon

Figure 1. Categories of IntelFruit data set

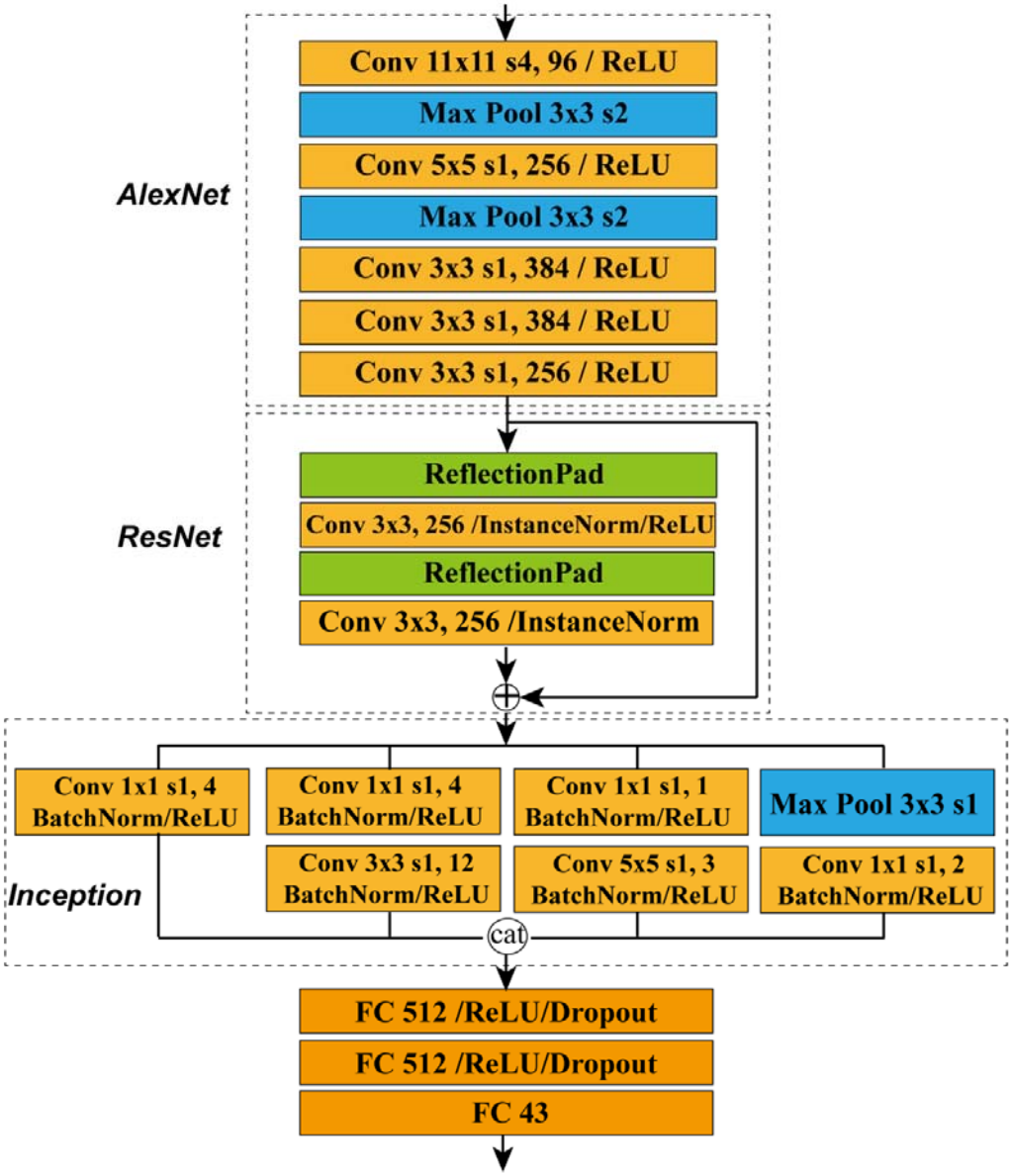

Figure 2. InterFruit Model Structure

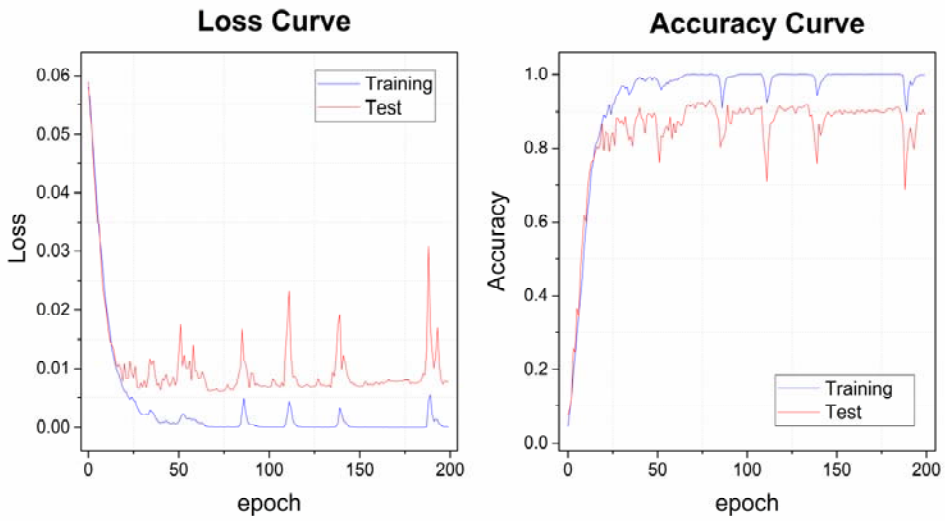

Figure 3. Loss and Accuracy Curves

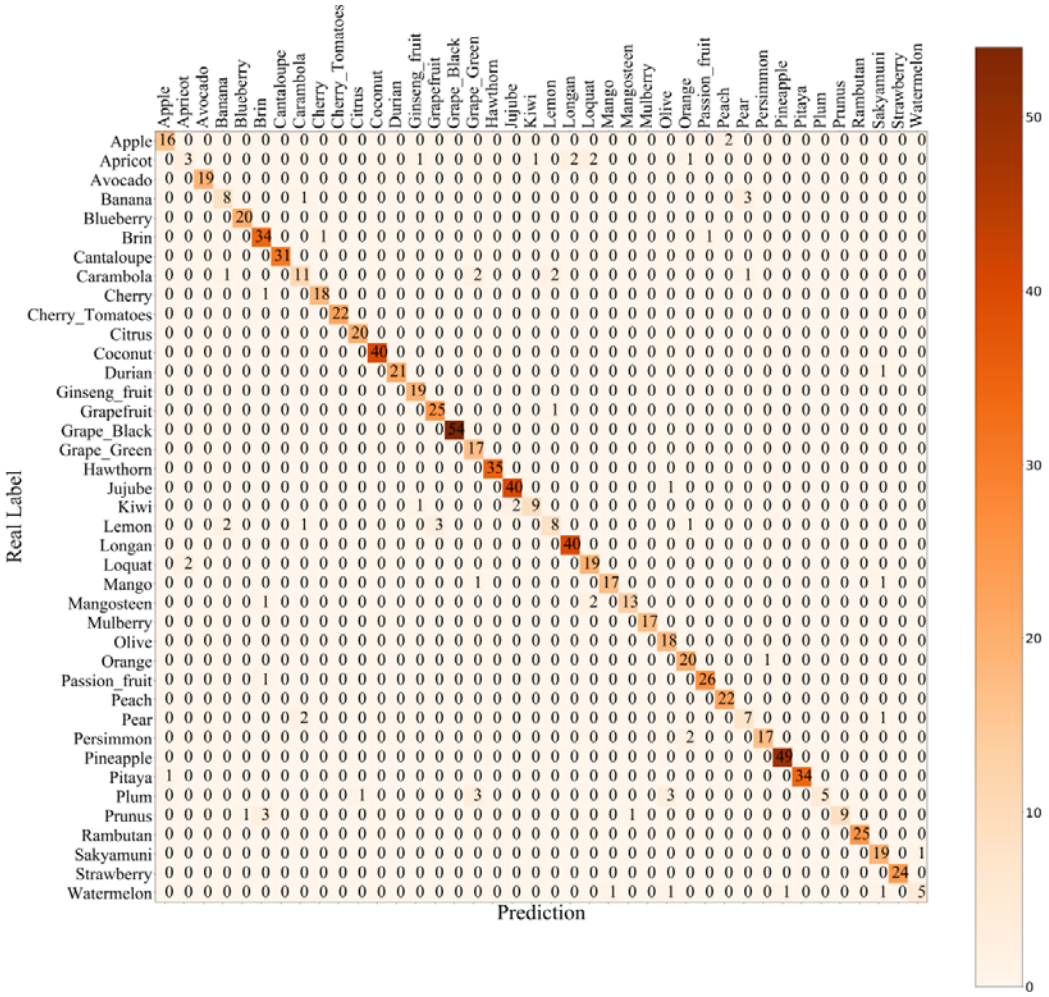

Figure 4. Confusion matrix on the test set

Table 1. Summary of the training and test sets

Label	Category	Number of Training Set	Number of Test set	Total Number
0	Apple	45	18	63
1	Apricot	25	10	35
2	Avocado	47	19	66
3	Banana	28	12	40
4	Blueberry	47	20	67
5	Brin	84	36	120
6	Cantaloupe	73	31	104
7	Carambola	42	17	59
8	Cherry	47	19	66
9	Cherry Tomatoes	52	22	74
10	Citrus	49	20	69
11	Coconut	94	40	134
12	Durian	54	22	76
13	Ginseng fruit	46	19	65
14	Grapefruit	62	26	88
15	Grape_Black	127	54	181
16	Grape_Green	41	17	58
17	Hawthorn	84	35	119
18	Jujube	98	41	139
19	Kiwi	31	12	43
20	Lemon	35	15	50
21	Longan	95	40	135
22	Loquat	51	21	72
23	Mango	47	19	66
24	Mangosteen	39	16	55
25	Mulberry	42	17	59
26	Olive	42	18	60
27	Orange	50	21	71
28	Passion fruit	65	27	92
29	Peach	54	22	76
30	Pear	26	10	36
31	Persimmon	45	19	64
32	Pineapple	115	49	164
33	Pitaya	82	35	117
34	Plum	28	12	40
35	Prunus	35	14	49
36	Rambutan	59	25	84
37	Sakyamuni	48	20	68
38	Strawberry	58	24	82
39	Watermelon	24	9	33
	Sum.	2216	923	3139

Table 2. Comparison of Classification Performance

Methods	***Acc***	***F1score***
Alexnet	83.97	91.28
GoogLeNet	84.83	91.79
ResNet18	75.52	86.05
Intelfruit	**92.74**	**96.23**