Recognition of navel orange image with complex background based on residual network

Xiaoxin Li, Yumei Tan*, Xinxin Lu, Bo Zhang

College of Artificial Intelligence, Hezhou University, Hezhou 542899, China

* Correspondence: myt889@163.com

Abstract. In order to solve the problem that the detection effect of navel orange recognition in complex background is poor, we propose a navel orange recognition method based on residual network. In this study, the navel orange classification dataset was constructed and labeled, and the performance of five classic convolutional neural network models on this dataset was evaluated, including AlexNet, Improved LeNet, SqueezeNet, ResNet-18, GoogLeNet. The results present significant accuracy obtained by the ResNet-18 model, with accuracy of 98.27%, which is more suitable for navel orange image recognition in complex background.

1. Introduction

Artificial picking navel orange has the problems of high labor intensity, low efficiency and high cost. The establishment of automatic, intelligent and modern picking system is of great significance to ensure timely harvesting of agricultural products, reduce picking costs and increase agricultural income [1]. The robot can realize the picking function, mainly because it can accurately identify the fruit in the complex agricultural environment. Traditional fruit recognition is mainly realized by processing color model or gray histogram [2]. Based on machine vision and support vector machine (SVM), Peng et al. [3] used Otsu segmentation algorithm to segment the fruit image, and used SVM classifier to classify and recognize the target. Yu et al. [4] proposed a method of navel orange recognition based on wavelet transform and Otsu threshold denoising.

In recent years, convolutional neural network (CNN) has made significant progress in image classification and recognition. Some scholars try to apply CNN model to the identification of fruits and vegetables [5-8], such as apple, grape and strawberry. However, the research on navel orange detection algorithm in complex agricultural environment is relatively less, especially in the case of illumination and occlusion, traditional algorithm algorithms have the problems of poor recognition rate of navel orange due to serious interference of background noise, and there is no public navel orange image dataset available. For this reason, we build a dataset of navel orange classification images, and evaluate the performance of five CNN models on this dataset. The experimental results indicated the effectiveness and feasibility of the proposed method, which provides a new revolution for navel orange automatic detection.

2. Materials and methods

2.1. Building dataset

The construction of navel orange dataset includes three steps: data collection, labelling images and dataset segmentation.
Data collection: The navel orange images were collected in the real agricultural environment by digital cameras and smart phones. It should be noted that these images are captured under different lighting conditions, different scene backgrounds and different angles. Then, all images are processed in a unified format, and the image size is fixed to $256 \times 256$ pixels, and the sample of navel orange is shown in Table 1.

| Label | 0 | 1 | 2 | 3 | 4 |
|-------|---|---|---|---|---|
| Sample | ![Sample 0](image1) | ![Sample 1](image2) | ![Sample 2](image3) | ![Sample 3](image4) | ![Sample 4](image5) |
| Description | mature and no occlusion | mature and partial occlusion | immature and no occlusion | immature and partial occlusion | bad |

Labelling images: As shown in Table 1, according to color, maturity, defect and occlusion, the navel orange images are divided into five categories and marked with numbers as 0, 1, 2, 3 and 4.

Dataset segmentation: Because the original samples collected are too few, we get a dataset with 3835 navel orange images through data augmentation methods, such as clipping, flipping, mirror, etc. Finally, the dataset was divided into 80% training set and 20% test set, and the description as shown in Table 2.

| Label | Training | Test | Total |
|-------|----------|------|-------|
| 0     | 735      | 183  | 918   |
| 1     | 745      | 187  | 932   |
| 2     | 635      | 158  | 793   |
| 3     | 540      | 135  | 675   |
| 4     | 414      | 103  | 517   |

2.2. CNN model

2.2.1. Residual block

With the deepening of network layers, it also brings problems such as difficult training, gradient disappearance and network degradation, which makes the deepening of network have no significant effect on performance improvement. He et al. [9] proposed residual network (ResNet) based on VGG model [10] and introduced the residual block as shown in the Figure 1.

![Figure 1. Residual block](image6)

From Figure 1, it can be seen that a shortcut connection is added, thus the learning process will change from learning features directly to adding some features on the basis of the previously learned features to obtain better features. In this way, a complex feature $H(x)$, which was previously learned...
layer by layer independently, now become such a model $H(x) = f(x) + x$, where $x$ is the feature at the beginning of the shortcut connection, and $f(x)$ is the residual. Therefore, the learning goal changes from learning complete information to learning residuals, the difficulty of learning high-quality features is greatly reduced, which greatly improves the performance of the model.

2.2.2. ResNet
ResNet was built by several stacked residual blocks and developed with many different numbers of layers: 18, 34, 50, 101, 152, and 1202. For all of the above, the ResNet are composed of convolutional, pooling and residual blocks. As shown in Figure 2, the ResNet-18 represents a good compensation between the depth and performance, and this network is composed by one convolutional layer, four stages residual block, one average pooling layer and a fully-connected layer with a softmax. Finally, for saving computing resources and training time, ResNet-18 model was chosen for the development of this study.

![Figure 2. Architecture of ResNet-18 model](image)

3. Results and analysis

3.1. Experimental configuration
The experimental software environment is Ubuntu 16.04 LTS 64 bit system, adopting Caffe deep learning framework and python 2.7 as the programming environment. The hardware environment is 32 GB of computer memory, equipped with Intel® Core™i7-9700@3.0GHz CPU and NVIDIA GTX1080Ti GPU. In the model training process, the initial learning rate is set to 0.005, the stochastic gradient descent (SGD) method is used, and the maximum of iterations is set to 20000.

3.2. Training results
In view of the advantages of classic LeNet, AlexNet, SqueezeNet and GoogLeNet in image recognition, these CNNs were used to verify the effectiveness of the proposed method. As can be seen from the Figure 3, compared with other models, ResNet-18 model achieved the fastest convergence speed and the highest accuracy. After 20000 iterations, the accuracy and loss values of all models tend to be stable, and the accuracy remains above 90%.

![Figure 3. Training results](image)
3.3. Performance comparison of different models

In order to further explore the performance of different models, the recognition accuracy, model size and training time were summarized in Table 3.

Table 3. Performance comparison for different models

| Model           | Depth | Parameter/M | Accuracy/% |
|-----------------|-------|-------------|------------|
| Improved LeNet  | 5     | 3.63        | 90.00      |
| AlexNet         | 8     | 56.88       | 95.07      |
| SqueezeNet      | 11    | 0.72        | 93.87      |
| ResNet-18       | 18    | 11.18       | 98.27      |
| GoogLeNet       | 22    | 5.98        | 98.13      |

From Table 3, it can be seen that the recognition accuracy, parameters and depth of different models are quite different. The accuracy of the shallow Improved LeNet model is only 94.00% with the least number of layers, while the deep models include AlexNet, SqueezeNet, ResNet-18 and GoogLeNet achieved accuracy of 95.07%, 93.87%, 98.27 and 98.13%, respectively. Among of them, the results present significant accuracy obtained by the ResNet-18 model, and compared with GoogLeNet model, the accuracy difference is very small. However, the architecture of GoogLeNet model is more complex than ResNet-18, which is mainly reflected in depth, width and convolution layer. In conclusion, ResNet-18 model is more suitable for navel orange recognition in complex background.

3.4. Test results

In this section, we use the trained model to identify five kinds of unknown samples, as shown in Figure 4. It can be seen that the recognition accuracy of each sample is 100%, which proves the feasibility and effectiveness of ResNet-18 model.

4. Conclusion

In this paper, a method for navel orange detection in complex background based on residual neural network is proposed. The navel orange classification dataset is constructed, and the performance of several convolutional neural network models is evaluated. Experimental results show that ResNet-18 model is better than other models, and it is more suitable for navel orange detection with complex background. In the future, we will enrich the dataset and further improve the performance of the model.
Acknowledgments
This work was financially supported by the project of improving the basic scientific research ability of young and middle-aged teachers in Guangxi Universities (2020KY18016).

References
[1] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(6): 1137-1149.
[2] Muhammad N A, Nasir A A, Ibrahim Z, et al. Evaluation of CNN, Alexnet and GoogleNet for fruit recognition[J]. Indonesian Journal of Electrical Engineering and Computer Science, 2018, 12(2): 468-475.
[3] Peng H, Shao Y, Chen K, et al. Research on multi-class fruits recognition based on machine vision and SVM[J]. IFAC-PapersOnLine, 2018, 51(17): 817-821.
[4] YU Changgeng, LIU Kai. Navel orange recognition based on wavelet transform and Otsu threshold denoising[J]. Journal of South China Agricultural University, 2020, 41(5): 109-114(in Chinese).
[5] Tian Y, Yang G, Wang Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J]. Computers and electronics in agriculture, 2019, 157: 417-426.
[6] Andrushia A D, Patricia A T. Artificial bee colony optimization (ABC) for grape leaves disease detection[J]. Evolving Systems, 2020, 11(1): 105-117.
[7] Lamb N, Chuah M C. A strawberry detection system using convolutional neural networks[C]// 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018: 2515-2520.
[8] Saleem M H, Potgieter J, Arif K M. Plant disease detection and classification by deep learning[J]. Plants, 2019, 8(11): 468.
[9] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
[10] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.