

Implementation of Small Sample Citrus Fruit Classification Detection Based on Improved ResNet18 Residual Network

Hai Li¹, Hua Qing¹, Shihui Wu¹, Haiqing Bai¹, Bo Lin²

¹Chongqing College of Humanities, Science and Technology, Chongqing 400000, China

²Isuzu (China) Engine Co., Ltd, Chongqing 400000, China

Abstract: Citrus is an important means of production, and traditional citrus classification using manual labor is time-consuming, labor-intensive, and costly. This article proposes a new residual block structure based on the introduction of pre activated residual block mode to build the ResNet18 network model, establish a small sample database of 4000 citrus fruits in four categories and use this network model for classification detection. The average accuracy of the final classification detection of ResNet18 after optimization reached 93.764%, the training time of the model was reduced by 41.47%, and the improved model showed improvements in both training speed and accuracy.

Keywords: Resnet; classification detection.

1. Introduction

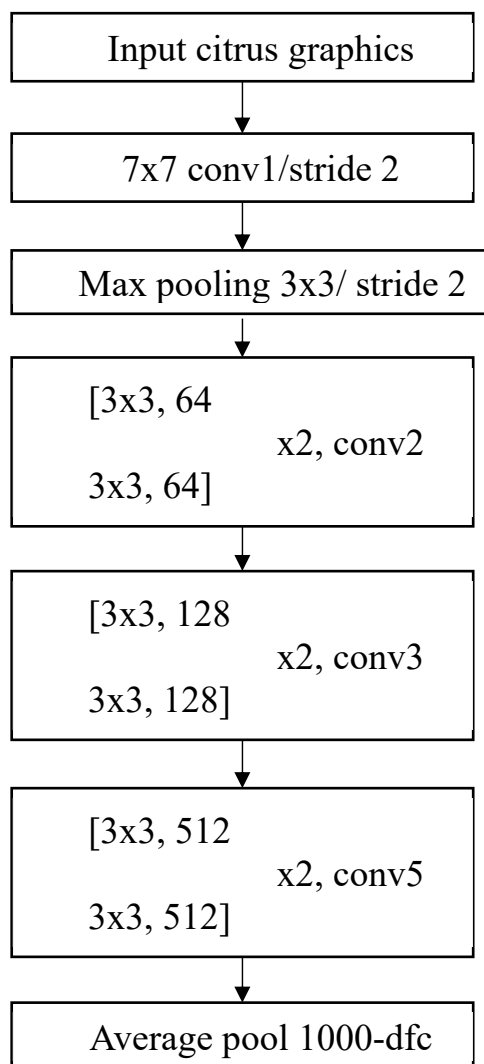
Citrus, as one of China's major economic crops and export fruits, surpassed apples for the first time to become the largest fruit in China in 2018. In 2019, the citrus area in China was 41.72 million mu, with a yield of 45.84 million tons, both in terms of area and quantity, it ranks first in the world, and Chinese citrus accounts for nearly one-third of the world's citrus [1]. Therefore, achieving classification and detection of citrus fruits is of great significance.

Shi Yu et al. proposed the E-Res2Net model for the classification of citrus leaves, by combining the SE attention mechanism with the Res2Net proposed by GAOS et al. based on Res Net, the model quality was improved, and the final classification accuracy reached 86.04%. Yue Yang, Zhang Wei and others designed an adaptive VGG16 network (Ad-VGG16) based on the breast cancer risk prediction problem and used the cross validation mechanism to predict the risk of breast cancer data set [2-5]. Jiang Jinhui adopted the shushang2 citrus method that combines SAM and VGG16, which effectively saves time on dataset creation and model training [6].

However, the neural network models used in the appeals are relatively complex, due to the large number of parameters, this article only improves the residual block structure of ResNet, performing certain optimizations on residual blocks that have been pre activated, the accuracy of citrus classification is significantly improved while keeping the number of network parameters basically unchanged.

2. Optimized ResNet Network Model

Traditional convolutional neural networks such as LeNet, AlexNet, VGG, and GoogleNet all suffer from degradation as the number of layers increases after reaching a certain level. But with the release of ResNet in 2015, ResNet was the first to propose the concept of residual connections, allow data to enter certain network levels and directly input to deeper network levels. Its topology structure can efficiently train very deep neural networks [7].The ResNet18 network structure is as follows:

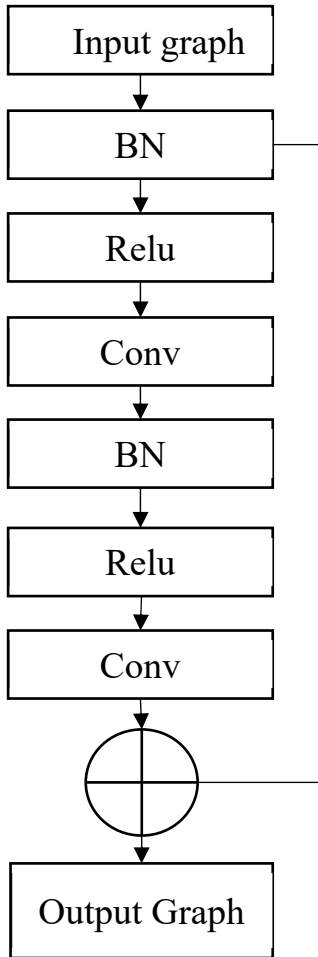


ResNet18 has a total of 18 layers, with each convolutional block followed by a batch normalization and a Relu activation function. Finally, citrus classification was performed using the Softmax function after global average pooling and

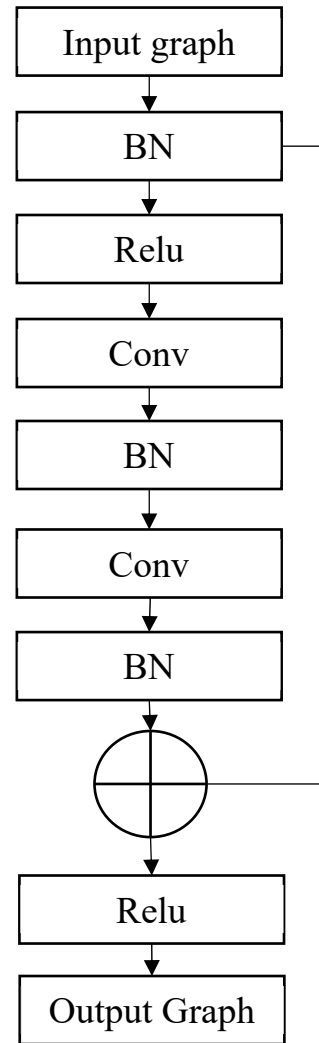
flattening.

2.1. Improvement of incomplete blocks

In 2022, Xia Minggui, Tian Ruijun, and others used residual units with pre activated structures to solve the potential problems of gradient vanishing and exploding in the original residual units, and ultimately achieved good results. Its structure is as follows:



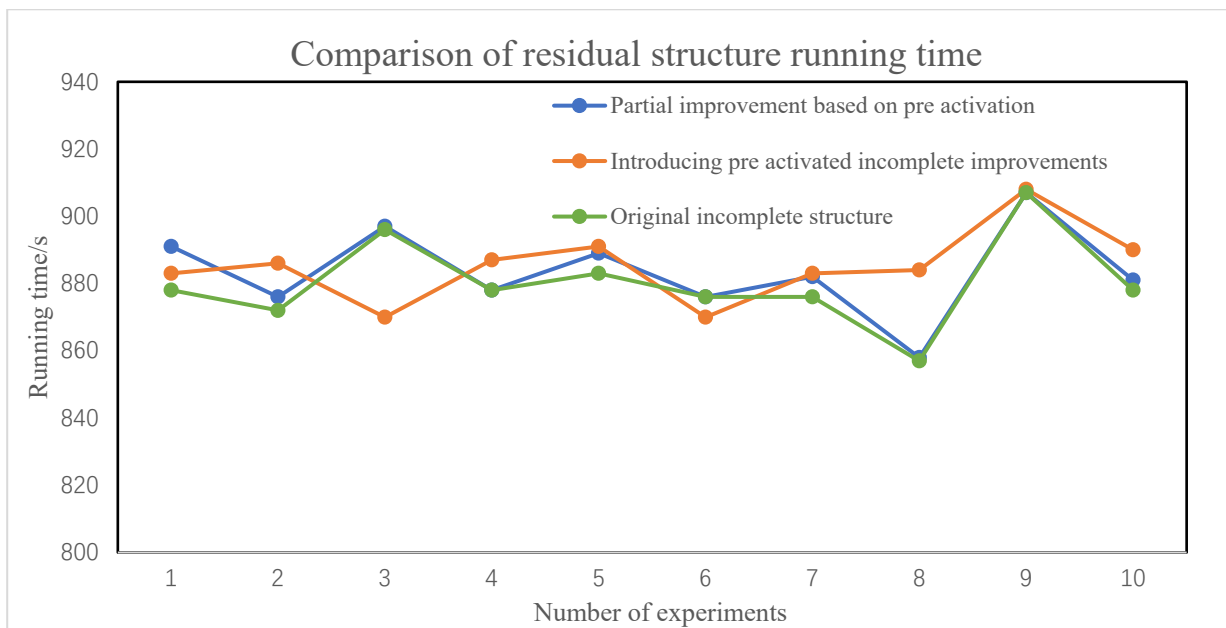
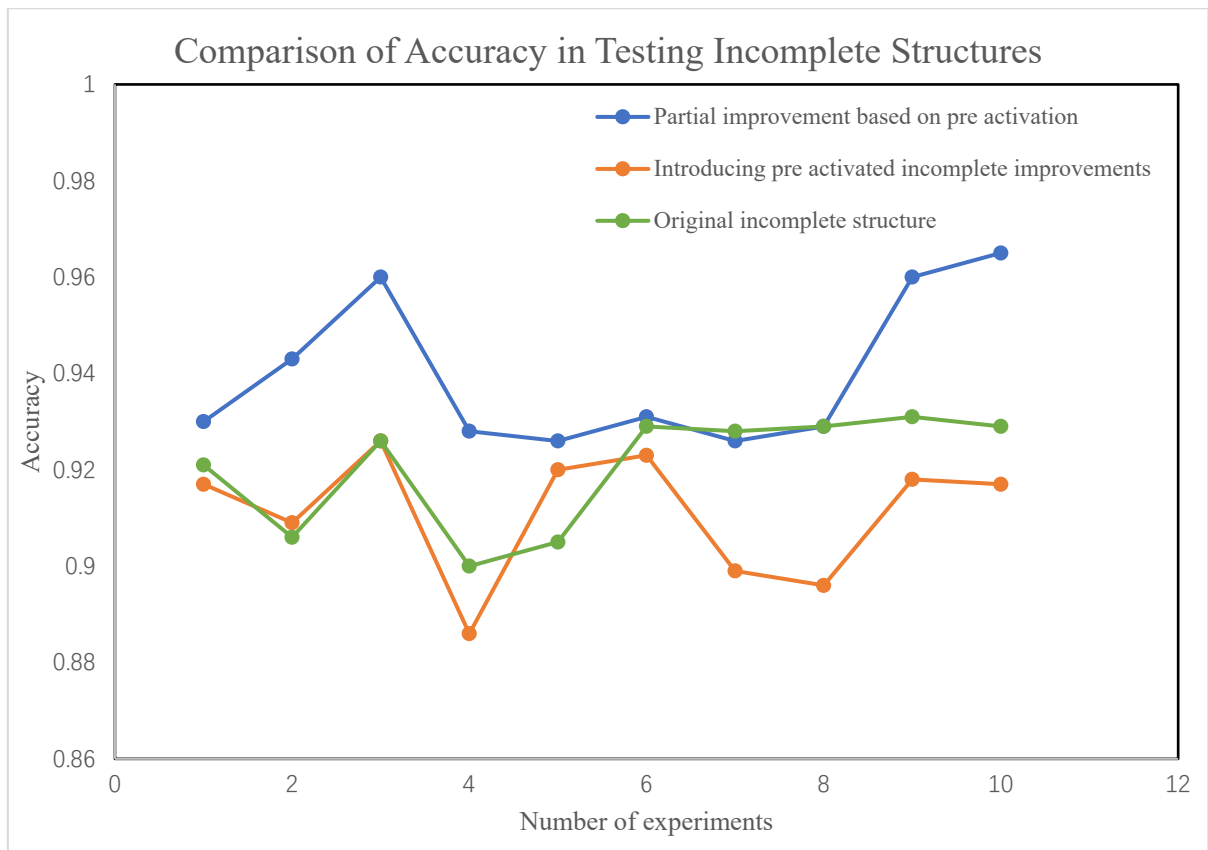
But we believe that batch normalization should still be performed after the second convolution, because the next stage is to add the original input image to the computed result, and previous networks have used batch normalization after convolution, so the original image entering the residual block must also be the result of batch normalization. Performing batch normalization after the second convolution will make their data distributions the same, and adding them together can better highlight the role of residual blocks. Then, applying activation functions to the added results can improve the network's expressive power. So we propose the following novel residual block structure:



2.2. Comparison of residual block improvement experiments

Using the original residual block, introducing pre activated residual blocks and the new residual block proposed in this paper for 10 repeated experiments, the experimental results are as follows:

Result analysis: The average accuracy of 10 experiments with the introduction of pre activated residual structure improved to 0.93764, while the average accuracy of 10 experiments with the introduction of pre activated residual structure was only 0.90979. The average accuracy of the original residual structure was 0.91956. The residual re improved structure with pre activated structure has increased the accuracy by 1.8 percentage points compared to the original structure, and by 2.785 percentage points compared to the residual structure with pre activated structure. The highest accuracy of the residual re improved structure with pre activated structure can reach 0.96. The overall performance is significantly improved due to two other residual structures.



In terms of running time, these three experiments have similar running times, with an average running time of about 14 minutes and 40 seconds for 10 experiments, and the maximum difference in running time is less than 30 seconds. So, from an overall perspective, the introduction of a residual improvement structure with pre activation structure significantly improves the accuracy of citrus classification while keeping the number of network parameters basically unchanged and the training time basically the same.

3. Experiment

3.1. Data collection and classification

Three fifths of the dataset used comes from Kaggle, and

two fifths is collected by oneself, including 1000 high-quality products, 1000 first-class products, 1000 second-class products, and 1000 inferior products. Each level is divided into a test set and a training set according to a 9:1 ratio. The training set is further divided into a validation set and a training set in a ratio of 2:8. The validation set does not participate in backpropagation, and the test set does not participate in training.

Citrus classification: Based on market research and personal life experience, citrus fruits are mainly classified into superior fruits, first-class fruits, second-class fruits, and inferior fruits based on color, decay, and defects (referring to spots, damage, and dents). Citrus fruits with bright and uniform color, deep color, overall orange or red color, and

slight defects are considered superior fruits, as shown in Figure 1. Citrus fruits with bright color, overall yellow color, and deep color but obvious defects are classified as first-class fruits, as shown in Figure 2. Citrus fruits with obvious green

color are classified as second-class fruits, as shown in Figure 3. Citrus fruits with rotten fruit and green main body but severe damage are classified as inferior fruits, as shown in Figure 4 [8-9].



Figure 1. Superior fruits



Figure 2. First-class fruit



Figure 3. Second-class fruit



Figure 4. Inferior fruits

3.2. Experimental considerations

The images in the dataset are divided into a training set and a testing set according to a certain proportion, and then the images in the training set are normalized to meet a normal

distribution. The image size of all training sets is adjusted to a uniform size, and then the training set images are further divided into a validation set and a training set according to a certain proportion. The validation set does not participate in backpropagation. The total training set images are input into

the ResNet network model for training, and then backpropagation is performed according to the loss function. The parameters are changed to save the best training parameter results, and the best parameter results are put into the testing set for testing. During testing, the testing data also needs to be normalized according to the mean and variance of the training set.

3.3. Loss function design

Using the cross entropy loss function, measure the similarity between the predicted probability distribution and the true probability distribution by calculating the cross entropy between the two. The function model is as follows:

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^K y_{ic} \log(p_{ic})$$

N represents the number of samples, K represents the number of categories, Pic represents the probability that the i-th sample belongs to category c, $y_{ic} = \{0, 1\}$. If the i-th sample belongs to category c, then the corresponding position's y_{ic} is taken as 1, otherwise it is taken as 0.

3.4. Network architecture construction

The network structure is basically built according to the ResNet18 network, and the residual blocks are replaced according to different schemes. When training the network, the Batch size is 32. At the same time, the Softmax structure of ResNet18, which was originally fully connected to the network, was cancelled. Using the ResNet18 network structure, after training 2-3 times, the accuracy of the validation set can achieve a relatively good result. The use of the Softmax structure can significantly reduce the probability of other categories when a large positive input value appears, making the network prone to falling into a local optimum. Therefore, considering the overall situation, the Softmax structure was eliminated [9].

Referring to the paper "Research on Clothing Image Style Recognition Based on Improved ResNet Network and Transfer Learning", modifications were made to the ResNet18 network by changing the first layer structure as shown in Figure 5:

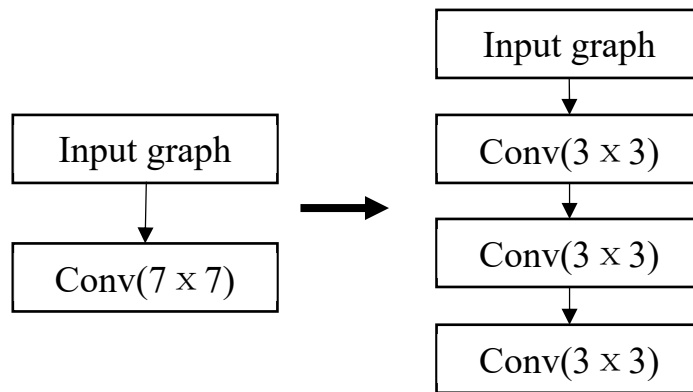


Figure 5. ResNet18 Network First Layer Structure Improvement

Using a combination of three 3x3 convolutional kernels instead of one 7x7 convolutional kernel can maintain the receptive field and output feature map size unchanged in the network [11]. The receptive field is the region of influence of a unit on the input of a feature map output by a certain layer in a neural network [9]. The calculation formula is as follows:

$$F(i) = (F(i + 1) - 1) \times \text{Stride} + \text{Ksize}$$

According to the formula, it can be calculated that using a combination of three 3 × 3 convolutional kernels to replace one 7 × 7 convolutional kernel can maintain the same receptive field size, and the hierarchical structure of three 3x3 convolution kernels can increase the depth of the network, capture style features at multiple scales, introduce more nonlinearity, reduce overfitting, and improve the performance of the network [9].

We randomly divided the dataset into a test set and a training set, and then trained it 20 times using the ResNet network. Finally, we conducted testing on the test set, which was a round of experiments. Then, we conducted 10 rounds of comparative experiments on the two ResNet network structures. In each round of comparative experiments, the training and testing sets of the two ResNet network structures were consistent.

Experimental analysis shows that after using the first layer improvement of the network, both the accuracy of the test set

and the training time have been significantly optimized in each experiment. In each comparative experiment, experiments using roll in improvement have shown significant improvements in both the training time and the accuracy of the test set compared to the original experimental scheme without improvement. Using convolutional blocks for improvement, the average accuracy of the test set is 0.95245 with a variance of 1.43125×10^{-4} , while the original plan without improvement has an average test set of 0.91 with a variance of 1.958×10^{-4} after 10 experiments. Compared to the original scheme, the convolutional improvement scheme has increased accuracy by 4.245 percentage points and is also more stable. In terms of training time, the original scheme had an average training time of 2527.7s and a variance of 2689.79s². The improved convolution scheme had an average training time of 1479.4s and a variance of 331.6s². The improved convolution scheme reduced the training time by an average of 17.17min and was also more stable in terms of training time. So overall, the convolution improvement scheme is significantly optimized.

Comparison with other network structures:

	Test set accuracy
Introducing pre activated residuals to improve ResNet network	0.9376
L-SE-VGG network	0.8875

4. Summarize

At present, the research and development achievements in key components of fruit picking devices in China are still insufficient, especially in terms of intelligent and efficient, environmental adaptability, and coordination and cooperation ability during the picking process. There is still significant room for improvement. Given the urgent demand for citrus picking in the industry, this article aims to explore and design an innovative citrus picker with automatic detection, classification and recognition functions, which can reduce labor and economic costs and improve picking efficiency. Based on the external characteristics of citrus fruits and their growth environment, a flexible end effector device for cutting and clamping was designed to mimic the manual picking of citrus fruits. This device ensures the avoidance of obstacles, achieves precise and non-destructive grasping, and successfully completes the citrus fruit picking process. Then, according to the built-in classification detection algorithm model, citrus fruits can be classified and processed. Due to the improvement of the ResNet18 first layer network in programming, this design has shortened the time and improved the accuracy of classification and recognition. Therefore, updating the classification recognition detection network model can promote the development of citrus industrialization and scientific management of orchards. The scientific and rational design of fruit picker structure is of great significance for improving the intelligence level of citrus picking and citrus agricultural production.

5. Funding Statement

This work was supported by the Chongqing Natural Science Foundation Upper-level Project (CSTB2023NSCQ-MSX1069); Science and Technology Research Project of Chongqing Municipal Education Commission (KJZD-K202401801) Chongqing College of Humanities and Technology School Level Project (CRKZK2024002); Chongqing College of Humanities and Technology School Level Project (CRKZK2021003); Science and Technology Project of Chongqing Municipal Education Commission

(KJQN202101801); Chongqing Municipal Education Commission Scientific Research Technology Project Youth Project (KJQN202001802);

Disclosures. The authors declare no conflicts of interest.

Data availability. No data were generated or analyzed in the presented research.

References

- [1] Zhang Chenyu, Nie Shuping, Zhang Hongzhen, etc Design of an Automated Citrus Picking Robot [J] *Agricultural Machinery Use and Maintenance*, 2023, (09): 10-15
- [2] Xiong Longye Research on Classification, Recognition, and Localization Methods of Mature Citrus Fruits in Natural Scenes [D] Chongqing University of Technology, 2020
- [3] Pan Hong Research on Multimodal Medical Image Recognition of Thyroid Nodules Based on Deep Learning [D] Zhejiang University of Traditional Chinese Medicine, 2022
- [4] Xu Pingting, Zheng Jiaqi A Strong Dog Detection and Reminder System Based on Deep Learning Algorithms [J] *Software*, 2023
- [5] Yang Hang Identification algorithm and research of ID card number [D] Chongqing University of Posts and Telecommunications, 2017
- [6] Xiang Chao Research on Human Posture Analysis of Video Images in Unmanned Aerial Vehicle Systems [D] Nanjing University of Aeronautics and Astronautics, 2020
- [7] Xia Mingguai, Tian Ruijun, Jiang Huiyu, etc Research on Clothing Image Style Recognition Based on Improved ResNet Network and Transfer Learning [J] *Journal of Textile Engineering*, 2024, 2 (01): 12-20
- [8] Fan Liming Research on the Classification Method of Cervical Cell Pathology Images Based on Deep Learning [D] Xiangtan University, 2021
- [9] A. Faraji, M. Noohi, S. A. Sadrossadat, A. Mirvakili, W. Na and F. Feng, "Batch-Normalized Deep Recurrent Neural Network for High-Speed Nonlinear Circuit Macromodeling," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 11, pp. 4857-4868, Nov. 2022, doi: 10.1109/TMTT.2022.3200071