Research on Defective Apple Detection Based on Attention Module and ResNet-50 Network

Lei Zhao[1], Zhenhua Li[2], Qinjun Zhao[1](✉), Wenkong Wang[1], Rongyao Jing[1], Kehua Du[1], and Shijian Hu[1]

[1] University of Jinan, Jinan 250022, China
cse_zhaoqj@ujn.edu.cn
[2] Shan Dong Cereals and Oils Detecting Center, Jinan 250012, China

Abstract. In defective apple detection, stem and calyx are easily confused with defects, and the detection accuracy of defective apples is lower. In order to solve these problems, this paper proposes a defective apple detection algorithm based on attention module and ResNet-50 network. CAM attention module and LeakyReLU activation function are used to optimize ResNet-50 network, which is named as C-ResNet-50 network. During network training, we use the cosine attenuation learning rate method, which effectively reduces the oscillation of training loss and accelerates the speed of network convergence. After the training and validation of the C-ResNet-50 network, the detection accuracy of defective apples reaches 97.35%, which is 2.33% higher than that of unimproved ResNet-50 network, 3.16% higher than VGGNet network and 4.14% higher than AlexNet network. This proves that the C-ResNet-50 network can improve the accuracy of defective apple detection.

Keywords: Defective Apple Detection · ResNet-50 · Attention Module · LeakyReLU Activation Function

1 Introduction

Apple has become a widely eaten fruit worldwide because of its beautiful flavor and high nutritional value. China has become the largest producer and consumer of apples in the world, and its industry has an important position in the development of part of China's rural economy. Defective apple detection is an important part of apple production and sales. When a batch of apples is mixed with defective apples will affect the sales profit of the whole batch of apples and then affect the economic benefits of fruit farmers. The traditional defective apple detection adopts manual selection method, which not only consumes much manpower but also fails to meet the demand in the detection speed. After a period of time, some image processing methods have been developed [1]. This method has a fast detection speed, but the stem and calyx have similar features to the apple defect in the image, which will reduce the accuracy of defective apple detection.

At present, deep learning has been widely used in the fruit defects detection due to its advantages in feature extraction [2]. By pre-training the model, automatic features

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2024
Published by Springer Nature Switzerland AG 2024. All Rights Reserved
B. Wang et al. (Eds.): ICMTEL 2023, LNICST 535, pp. 298–307, 2024.
https://doi.org/10.1007/978-3-031-50580-5_26

extraction and continuous optimization of features, it can quickly process a large number of data with better performance and higher accuracy. Some researchers analyze apple images and spectral images by deep learning technology to detect Apple defects and have made some research progress [3]. XUE Yong used GooLeNet deep migration model to detect Apple defects, and the test accuracy reached 91.91% [4]. Sovon Chakraborty used CNN network to detect and identify fresh and rotten fruits and obtained an accuracy of 93.72% [5]. Chithra used Kapur's algorithm to sort apples as defective and normal, and the test accuracy reached 93.33% [6]. Nur Alam used hyperspectral imaging (HSI) to detect apple skin detection, and the accuracy reached 94.28% [7].

Although defective apple detection has been developed for a long time, the accuracy still needs to be improved. To improve defective apple detection accuracy, this paper uses CAM, LeakyReLU activation function and ResNet-50 for fusion to get a new network. The new network is named as C-ResNet-50 network, which is used to detect defective apple task. The C-ResNet-50 network is built on the Pytorch framework. The data set is set as training set and validation set according to the ratio of 8:2, that is applied to the training and optimization of the network. Finally, the accuracy of the C-ResNet-50 is compared with AlexNet, VGGNet and ResNet-50.

2 Network Design and Optimization

This paper studies defective apple defection, including network improvement and optimization, data set setting, network training and result analysis.

ResNet-50 network is optimized using the CAM and LeakyReLU activation function, which is defined as the C-ResNet-50 network. CAM is the Channel Attention Module. Feature channel weight is obtained by using CAM. Through the application of that weight, the network can greatly capture meaningful information and improve the accuracy of defective apple detection. The activation function increases the nonlinearity of the network. ResNet-50.

2.1 ResNet-50 Network

The ResNet-50 network consists of four network layers, convolutional layer, pooling layers and full connection layer. Each network layer is composed of a different number of residual modules. That is 3, 4, 6 and 3, respectively. Each residual module is composed of three convolutional layers, which have convolution kernels of different sizes.

Residual Module. The residual module consists of a convolution module and a direct connection module (see Fig. 1). When the output of the convolutional module is 0, $H(z) = z$, this solves the problem of decreasing accuracy caused by the increase of network depth. The direct connection module is directly completed by simple connection, without introducing additional parameters, which reduces the computing burden and improves the network accuracy [8]. The output of the residual module is as follows.

$$H(z) = F(z) + z \tag{1}$$

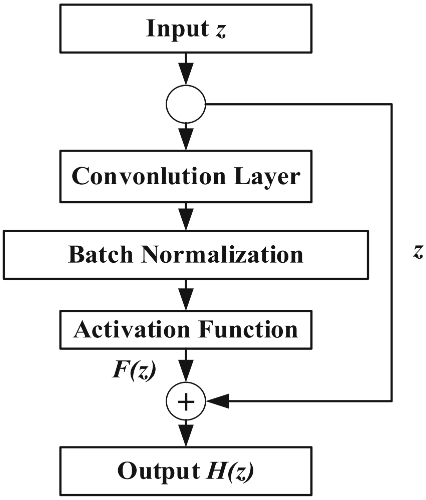

Fig. 1. The network structure of the residual module

Convolution Layer. The function of the convolution layer is to extract different features of the input. It contains multiple convolution kernels. Each neuron element in the convolutional layer is linked to multiple neurons in the previous layer. When the convolution kernel works, it regularly sweeps the input features, summates matrix elements multiplication in the receptive field and superimposes the deviation quantities.

Batch Normalization Layer. The Batch Normalization layer is as follows. 1. It adjusts data distribution and speeds up network training and convergence. 2. Prevent gradient explosion and gradient disappearance. 3. Prevent overfitting.

The specific calculation process of the Batch Normalization layer is as follows. Firstly, the mean value of samples is calculated. Then, the variance of samples is calculated and standardized. Finally, linear transformation and migration. The calculation formula of the Batch Normalization layer is as follows.

$$\mu = \frac{1}{n}\sum\nolimits_{i=1}^{n} z_i \tag{2}$$

$$\sigma^2 = \frac{1}{n}\sum\nolimits_{i=1}^{n} (z_i - \mu)^2 \tag{3}$$

$$(z_i)_{norm} = \frac{z_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{4}$$

$$y_i = \alpha(z_i)_{norm} + \beta \tag{5}$$

where: μ is the average value, n is the number of elements, z_i is the value of each element, σ^2 is the variance, ε is the deviation, $(z_i)_{norm}$ is the normalized value of each element, α is the parameter of the linear transformation, β is the offset, and y_i is the output of the batch normalization layer.

Activation Function. ReLU activation function is adopted in ResNet-50 network. When the input $x \leq 0$, the output $F(x)$ is 0. When the input $x > 0$, the output $F(x)$ is still x (see Fig. 2).

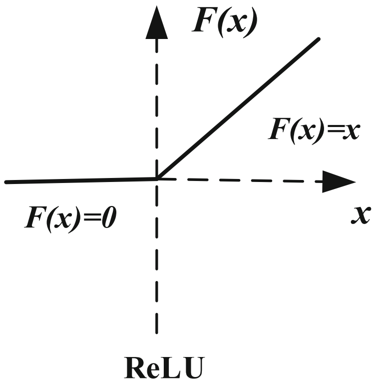

Fig. 2. ReLU activation function

The formula of ReLU activation function is as follows.

$$F(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{6}$$

The ReLU activation function uses simple mathematical operations to make the network have good sparsity, which can reduce the amount of computation and reduce overfitting. However, ReLU activation function has some shortcomings, such as, when the input $x < 0$, the output $F(x)$ is 0, which leads to the output is always 0. Neuron no longer learns in the future learning, which is the "Dead Neuron" phenomenon.

2.2 CAM Attention Module

CAM attention module (see Fig. 3) consists of the pooling part, MLP fully connected network and sigmoid function. Channel data are obtained by global average pooling and global maximum pooling, respectively for input features. These pass through the shared MLP network and then add up to get new channel data. Finally, this data is converted from 0 to 1, which becomes the weight of each feature channel via the sigmoid function. The optimized feature data is obtained by multiplying the input feature and the feature channel weight. The CAM attention module outputs the optimized data. The output formula of CAM attention module is as follows.

$$W_C = \sigma\{MLP[MaxPool(F)] + MLP[AvgPool(F)]\} \tag{7}$$

$$F' = W_C \otimes F \tag{8}$$

where: F is the CAM module input, $MaxPool$ is the global max-pooling, $AvgPool$ is the global average-pooling, MLP is the fully connected network, σ is sigmoid function, W_C is the weight, and F' is the output of CAM module.

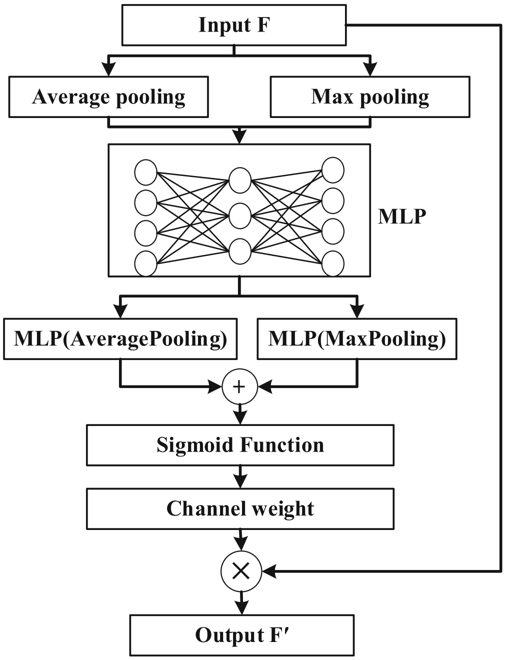

Fig. 3. CAM attention module

2.3 Activation Function Improving

In order to solve the "Dead Neuron" phenomenon, this paper adopts LeakyReLU activation function for improvement. When $x \geq 0$, the output remains x. When $x < 0$, the output is ax (a is a custom number). With this improvement, there is still an output when the input is less than 0. The improvement of activation function can maintain the normal update of neurons while keeping less computation. The formula for the LeakyReLU activation function is as follows.

$$F(x) = \begin{cases} x & x > 0 \\ \alpha * x & x \leq 0 \end{cases} \tag{9}$$

LeakyReLU activation function (see Fig. 4).

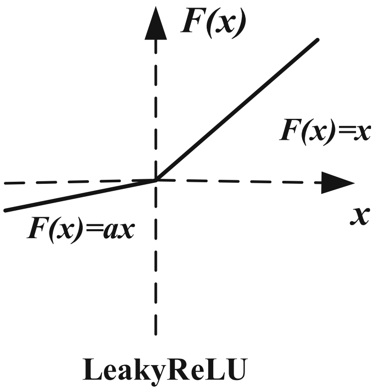

Fig. 4. LeakyReLU activation function

2.4 C-ResNet-50 Network

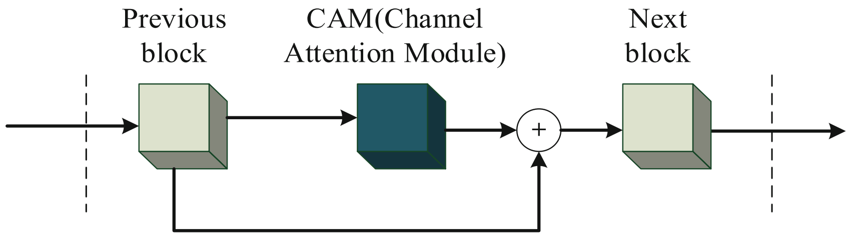

Fig. 5. CAM Attention module joins ResNet-50 network

After each residual module, a CAM attention module is added and a total of sixteen CAM attention modules are inserted. Through the CAM attention module, we can get the characteristic channel weight of the last residual module. Applying this weight to the network can improve network focus on critical information (see Fig. 5).

$$NI = BO + F' \tag{10}$$

where: BO is the previous residual module output, NI is the next residual module input, F' is the CAM attention module output.

3 Experiment and Analysis

3.1 Data-Set Setting

We use 4423 valid images for detective apple detection task. Apples that contain puncture wounds, crushing wounds, sunburn, insect injuries, and cracked apples are defined as defective apples, while apples with no damage or slight damage are defined as normal apples (see Fig. 6).

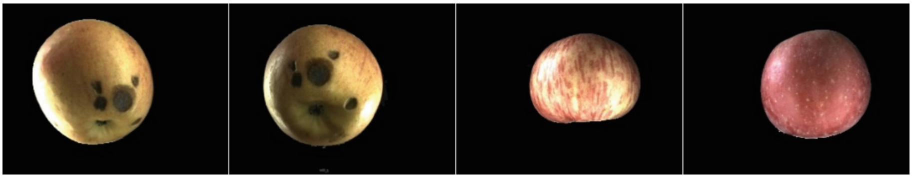

Fig. 6. Defective apple (left 1,2) and normal apple (left 3,4)

In this experiment, specific training was carried out for the classification of stem, calyx and defect to improve the classification accuracy of defective apples. The data set was divided into four categories: 1. apples without defect, stem and calyx, 2. apples having stem and calyx without defect, 3. apples having defect without stem and calyx, 4. apples having defect, stem and calyx. According to the above four classification categories, we classified the data set and sorted out 1125 apple pictures without defect, stem and calyx, 1750 apple pictures having stem and calyx without defect, 798 apple pictures having defect without stem and calyx, 750 apple pictures having defect, stem and calyx. We randomly divided the above data into training set and validation set according to a ratio of about 8:2. The data set for defective apple detection is shown in Table 1.

Table 1. Data set setting

Apple type\Set	Training set	Validation set	amount
without defect, stem and calyx	876	249	1125
having stem and calyx without defect	1371	379	1750
having defect without stem and calyx	639	159	798
having defect, stem and calyx	600	150	750
amount	3486	937	4423

3.2 Training Process

This paper uses cosine attenuation learning rate to training and a 3090 GPU was used to accelerate the training of the model. In the beginning, a large learning rate is used for learning to accelerate the convergence speed. At the end of training, a small learning rate is used to improve the network's convergence accuracy (see Fig. 7). The calculation formula for the learning rate is as follows.

$$lr_t = lr_{min} + \frac{1}{2}(lr_{max} - lr_{min})\left[1 + \cos\left(\frac{T_{cur}}{T_{max}}\pi\right)\right] \tag{11}$$

where: lr_t is the current learning rate, lr_{min} is the minimum learning rate set, lr_{max} is the maximum learning rate set, T_{cur} is the current training times, and T_{max} is the maximum training times.

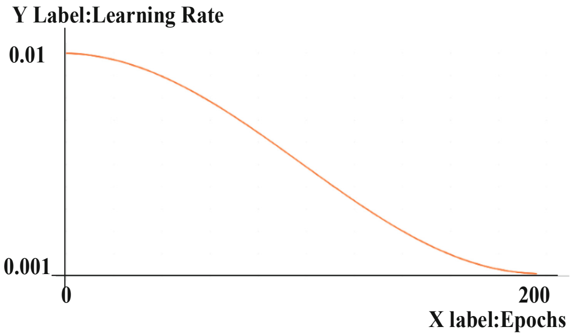

Fig. 7. Learning rate change curve

In order to verify the C-Resnet-50 network accuracy in defective apple detection, this paper selected AlexNet, VGGNet and ResNet-50 together for accuracy comparison.

3.3 Experiment Analysis

In the training process, the cosine attenuation learning rate is compared to constant learning rate. Furthermore, using cosine attenuation learning rate has a faster convergence speed and smaller oscillation (see Fig. 8).

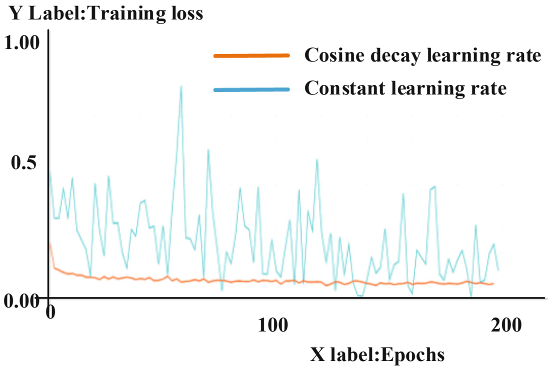

Fig. 8. Training losses

After the training and optimization of the C-ResNet-50 network, we obtained the C-ResNet-50 validation loss diagram and Comparison of C-ResNet-50 and ResNet-50 (see Fig. 9).

After training and verifying several neural networks, we obtained the detection comparison of AlexNet, VGGNet, Resnet-50 and C-Resnet-50 networks (see Fig. 10) and the detection accuracy of the above several networks (see Table 2).

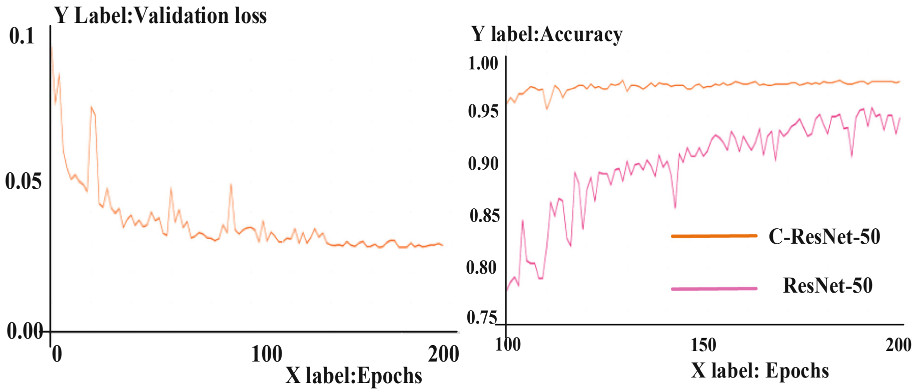

Fig. 9. C-ResNet-50 validation loss (left) and comparison of C-ResNet-50 and ResNet-50 (right)

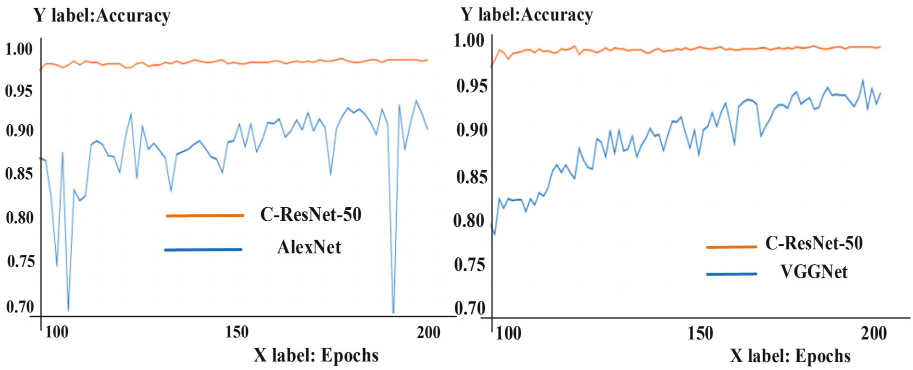

Fig. 10. Comparison of C-ResNet-50 and AlexNet (left) and VGGNet (right)

Table 2. Detection accuracy of defective apple

Number	Network	Accuracy (%)
1	AlexNet	93.21
2	VGGNet	94.23
3	ResNet-50	95.02
4	**C-Resnet-50**	**97.35**

In this experiment, we improved the ResNet-50 network to obtain the C-Resnet-50 network, which performed well in the task of defective apple detection with an accuracy of 97.35%. The accuracy of ResNet-50, AlexNet network and VGGNet were 95.05%, 94.23% and 93.21%, respectively. The accuracy of the C-ResNet-50 network is 2.33% higher than that of unimproved ResNet-50 and 3.16%, 4.14% higher than AlexNet, VGGNet, respectively. All these prove that the C-Resnet-50 network is superior to other convolutional neural networks, and competent for the task of defective apple defection.

4 Conclusion

This paper describes a defective apple detection method. To solve the problem of low accuracy of defective apple detection, the C-ResNet-50 network is proposed, which is a fusion model of CAM attention module, LeakyReLU activation function and ResNet-50 network. In network training, the cosine attenuation learning rate method is adopted to effectively improve the convergence speed and reduce the oscillation of loss. Finally, after validation, the C-ResNet-50 network achieves 97.35% accuracy in the defective apple detection task. Compared with AlexNet, VGGNet and ResNet-50, C-ResNet-50 accuracy increases by 2.33%, 3.16% and 4.14%, respectively. This proves that the C-ResNet-50 network based on attention module and ResNet-50 is competent for the task of defective apple detection.

Acknowledgements. This paper is supported by the Shandong Key Technology R&D Program 2019JZZY021005, Natural Science Foundation of Shandong ZR2020MF067 and Natural Science Foundation of Shandong Province ZR2021MF074.

References

1. Wang, B., Yin, J., et al.: Extraction and classification of apple defects under uneven illumination based on machine vision. J. Food Process Eng **45**(03), e13976 (2022)
2. Abbaspour-Gilandeh, Y., Aghabara, A., Davari, M., et al.: Feasibility of using computer vision and artificial intelligence techniques in detection of some apple pests and diseases. Appl. Sci. **12**(2), 906 (2022)
3. Cheng, C., Zhang, G.T.: Deep learning method based on physics informed neural network with resnet block for solving fluid flow problems. Water **13**(4), 423 (2021)
4. Yong, X.U.E., Liyang, W.A.N.G., et al.: Defect detection method of apples based on GoogLeNet deep transfer learning. J. Agric. Mach. **51**(07), 30–35 (2020)
5. Chithra, P., Henila, M.: Apple fruit sorting using novel thresholding and area calculation algorithms. Soft. Comput.Comput. **25**, 431–445 (2021)
6. Chakraborty, S., Shamrat, F.M.J.M., Billah, M.M., et al.: Implementation of deep learning methods to identify rotten fruits. In: 25th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1207–1212. IEEE, Tirunelveli (2021)
7. Nur Alam, M.D., Thapamagar, R., Rasaili, T., Olimjonov, O., Al-Absi, A.A.: Apple defects detection based on average principal component using hyperspectral imaging. In: Pattnaik, P.K., Sain, M., Al-Absi, A.A., Kumar, P. (eds.) SMARTCYBER 2020. LNNS, vol. 149, pp. 173–187. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-7990-5_17
8. He, J., Jiang, D.: Fully automatic model based on SE-resnet for bone age assessment. IEEE Access **9**, 62460–62466 (2021)