International Journal of Agriculture and Food Sciences Research
ISSN: 3006-7685 (Print), ISSN: 3006-3736 (Online) | Volume 2, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijafsr.v2n1.02
Journal homepage: https://wepub.org/index.php/IJAFSR/index

Research on Apple External Quality Grading Method Based on DXNet Model

Chao Li

Shaanxi Fruit Industry Development Co., Ltd., Shaanxi 712000, China

ABSTRACT

In the wave of agricultural modernization, intelligent apple selection line technology is a key link in improving fruit quality and production efficiency. There are various methods for grading the external quality of apples, some of which focus on detecting single external features of apples, such as color, size, shape, texture, and visual defects; Other methods are based on deep learning techniques to grade the appearance quality of apples. However, these manual methods for extracting single external features of apples have certain limitations in terms of grading accuracy, and cannot comprehensively and objectively evaluate the external quality of apples. To improve the accuracy and stability of the lossless grading method for apples, this paper proposes a Yan'an apple grading method based on the DXNet model. The convolutional blocks of traditional convolutional neural networks are extracted as feature extractors for apple images, and the learned feature maps are subjected to global max pooling and global average pooling, respectively. The two one-dimensional vectors after pooling are concatenated and input into a classifier mainly composed of fully connected layers for classification.

KEYWORDS

DXNet model; Classification of External Quality of Fruits

1. INTRODUCTION

In the wave of agricultural modernization, the intelligent fruit selection line technology, as a key link in improving fruit quality and production efficiency, is undergoing unprecedented changes. The following is a review of the current research status of intelligent fruit selection line technology for fruits.

(1) Technical Status and Development Trends: Currently, researchers are analyzing and exploring intelligent fruit sorting production lines at home and abroad to address the problems of low recognition accuracy and long processing time for individual fruit information in fruit sorting. These studies combine industrial robots and visual sorting technology, proposing methods such as using convolutional neural networks (CNN) and spectral techniques to analyze the chemical properties of fruits. These technologies have demonstrated practical value in fruit defect detection and maturity sorting.

(2) Application case: For example, the 4.0 fruit post harvest intelligent selection equipment developed by Shenzhen Daochuang Intelligent Innovation Technology Co., Ltd. can sort and package fruits based on differences in size, sugar content, coloring, and other factors. This device has been applied in multiple core fruit production areas, including nearly 10 provinces, municipalities, and autonomous regions such as Yunnan, Guangxi, Hubei, and Xinjiang, with a cumulative detection and packaging of millions of tons of various fruits.

(3) Technical advantages: These intelligent sorting technologies not only improve sorting efficiency, but also reduce manual errors. For example, Daochuang Intelligent's equipment can non-destructive detect quality issues inside fruits, such as mold heart disease, sugar heart disease, etc., while also quickly detecting appearance defects.

(4) Shortcomings: Currently, there are various methods for grading the external quality of Apple. Some methods focus on detecting single external features of apples, such as color, size, shape, texture, and visual defects. Other methods are based on deep learning techniques to grade the appearance quality of apples. For example, Liao Wei et al. constructed a green apple random forest recognition model, which achieved a recognition accuracy of 90% for green apples. Li Huihui and others focused on the study of apple surface texture, attempting to explore the correlation between texture and sweetness. Zhao Juan et al. extracted and segmented apple images, used computer vision technology to extract apple defects, and compared them with manual identification results. They found that the accuracy of defect detection was as high as 92.5%. Li Long et al. used texture and gradient features to identify apple scars, with a correct recognition rate of 95% for 80 samples.

However, these manual methods for extracting single external features of apples have certain limitations in terms of grading accuracy, and cannot comprehensively and objectively evaluate the external quality of apples. To overcome this issue, researchers have begun exploring deep learning based grading methods. Xue Yong et al. used the GoogLeNet deep transfer model to detect defects in apples, achieving a recognition accuracy of 91.91% for the test samples. Wang Dandan et al. proposed an apple target recognition method based on a region based fully convolutional network, with a recognition accuracy of 95.1%. Zhang Lichao and others improved the LeNet5 model and applied it to apple variety recognition with an accuracy of 93.7%. Wu Xing and others used the YOLOv3 model to detect apples, with an average accuracy of 94.69%.

To improve the accuracy and stability of non-destructive grading methods for apples, this paper proposes a deep learning based Yan'an apple grading method. Extract convolutional blocks from traditional convolutional neural networks as feature extractors for apple images, perform global max pooling and global average pooling on the learned feature maps, concatenate the pooled two one-dimensional vectors, and input them into a classifier mainly composed of fully connected layers for classification.

2. RESEARCH ON APPLE EXTERNAL QUALITY GRADING METHOD BASED ON DXNET MODEL

2.1. Models and Materials

2.1.1. Models

The DXNet model adopts multi convolutional neural network fusion technology to enhance the performance of image processing tasks. This model integrates multiple convolutional neural networks, each focusing on specific features of the image, and generates a more comprehensive feature representation through feature fusion. DXNet can provide accurate results when performing visual recognition and classification tasks, and demonstrates significant advantages in processing complex visual data.

2.1.2. Materials

Fuji apples grown locally in Yan'an City, Shaanxi Province, were captured using the built-in camera of the Samsung s24 phone in a well lit environment during the day. The camera was a Sony IMX363 model (48 million pixels), and 3000 images were taken in supermarkets and surrounding orchard areas in Yan'an City, including the top, side, and bottom of the apples. Ensure that the data sources are diverse and widely distributed, in line with real scenarios. A total of 12000 apples were collected, and 12000 images were obtained as the initial dataset. According to the indicators in Table 3-4, the collected apple images were divided into four levels: A, B, C, and D, with each level containing 3000 images. The collected dataset contains a large amount of noise interference, including apples located at the edge of the image, interference from other objects in the image, or multiple apples in one image. Grade the apple with the highest proportion in the image. Apple images are affected by factors such as background and lighting during collection, which can interfere with the classification of apple appearance. 30% of the 12000 original apple images were used as test samples, 20% as validation samples to evaluate the training status of the model, and the remaining 6000 images were enhanced to obtain 12000 images as the training set. The number of images at each level in each dataset was the same. The experimental platform consists of two parts: computer hardware and development platform. The computer is configured with Intel Core I78750HCPU, 16GB of memory, and GTX1060 graphics card. This application development should be based on the J2EE platform, with the main code developed using Java programming language and server-side Java technology.

2.1.3. Experimental site

The operation of the model is carried out at the Big Data Center on the 16th floor of Shaanxi Guoyun Valley in Xianyang City, Shaanxi Province; Applied in a large-scale processing production line in Yan'an City, Shaanxi Province.

2.2. Method

Preprocess the collected Yan'an apple images, and input the preprocessed image data into a feature extractor for learning. Use a classifier to classify the learned feature maps and obtain grading results. The performance of the model is measured using the grading accuracy curve and loss function value curve on the validation set, as well as the grading accuracy and F1 on the test set.

2.2.1. Data preprocessing

Deep learning belongs to the category of supervised learning. Compared with traditional computer vision methods, it can automatically learn features through data-driven models, eliminating the step of manually searching for features. Deep networks learn more expressive features through data. Therefore, a large number of samples are required to train neural network models, making data augmentation of existing images a crucial step in model training. Expanding the dataset can improve the generalization ability of the model and reduce the risk of overfitting. The data augmentation adopts traditional digital image processing methods, with the specific steps of rotating the existing image within a 70 ° angle range, moving it vertically and horizontally by 20%, scaling it by 80% to 120%, flipping it vertically and horizontally, and rotating it counterclockwise by 20 ° for cropping transformation. All augmentation methods are randomly performed with the same probability. Normalize the enhanced data to obtain an input image of 224 pixels by 224 pixels, and perform data augmentation on four levels of apples simultaneously to ensure that the input data remains unchanged. When each round of data is read into the video memory, the framework generates 32 enhanced images, totaling 64 images for the current round of training, ensuring that the total number of images participating in the training is 15000. The partially enhanced image is shown in Figure 1. It can be seen from Figure 1 that after completing data augmentation, there have been significant changes in the size and angle of the apple.

2.2.2. Multi Convolutional Neural Network Fusion

The DXNet model consists of three parts: data preprocessing, feature extractor, and classifier. Two convolutional blocks are used to extract image features, and the feature maps learned by the network are summarized by concatenating the two feature tensors according to their dimensions. The concatenated one-dimensional vector is input into the classifier for classification and output the results. Utilize data augmentation methods to enhance image data and input it into the DenseNet121 model and Xception model. Using convolutional layers from two networks to learn apple images, the

1024 × 7 × 7 feature maps and 2048 × 7 × 7 feature maps obtained after training are respectively subjected to global average pooling and global max pooling to replace the traditional fully connected method. Global average pooling performs global average pooling on each feature map output by the last convolutional layer, which calculates the average value of all pixels. After forming a feature point, these feature points are combined into the final feature vector, and global max pooling calculates the maximum value of all pixels in each feature map. After pooling 1024 feature maps and 2048 feature maps, 1024 and 2048 data points can be obtained, respectively. After connecting the data points, one-dimensional feature vectors with lengths of 1024 and 2048 can be obtained, respectively. Global average pooling can significantly reduce the number of parameters, reduce the network size, and alleviate the problem of large estimation variance caused by limited neighborhood size in feature extraction. After feature extraction is completed, two feature vectors are input into the classifier

In the middle. The Concatenate layer, as the top of the classifier, can produce the effect of feature fusion. The feature information extracted by two convolutional feature networks is fused, and the features of the image itself increase while the information under each feature remains unchanged. That is, two one-dimensional feature vectors are concatenated according to the vector dimension to obtain a one-dimensional vector with a length of 3072, which is input into the first fully connected layer. The number of nodes in the first fully connected layer is set to 1024, and the activation function uses

ReLU. ReLU can create a certain sparsity in the network and reduce the interdependence of parameters, alleviating the occurrence of overfitting problems. This fully connected layer adopts L2 regularization to further increase the sparsity of parameters and reduce the risk of model overfitting. Then connect a batch normalization layer, which re normalizes the activation values of the previous layer on each batch, ensuring that the standard deviation of the output of the fully connected layer is close to 1 and the average is close to 0. During the training process of the deep neural network, the input of each layer of the neural network is kept as evenly distributed as possible. It can avoid the problem of gradient vanishing, accelerate the training speed of the model, and improve the generalization ability of the model. Input the normalized fully connected layer parameters into a fully connected layer consisting of 5 nodes, set the Softmax function as the activation function for multi classification tasks, and output the classification results corresponding to each category. The calculation formula for the Softmax function is:

$$S_{\mathrm{i}} = \frac{e^{i}}{\sum_{j} e^{j}}$$

S_i-image being the i-th level

E^j- The upper level function value of the i-th level in the image

j-level number

Input the test set into the trained fusion model for grading. To ensure the credibility of the test set images, data augmentation is no longer performed on the test set. Only the images that need to be graded are normalized, and the pixel values are adjusted to floating-point numbers within 0-1. This means that the model introduces nonlinear elements to accelerate the inference speed of the neural network. The final results obtained after grading are compared and analyzed with the final results obtained by a single model.

The advantages of integrating multiple convolutional network models to improve recognition accuracy and stability include:

(1) Unlike traditional models that use global max pooling or traditional fully connected methods, DXNet's global average pooling operation exhibits strong robustness against spatial variations in the

input and can effectively solve various noise problems in the dataset. Meanwhile, DXNet includes global max pooling, which can effectively preserve texture features and improve grading accuracy.

(2) Multiple convolution kernels of different sizes can extract features of different scales based on the size of the apple, and increasing the number of convolution kernels appropriately can enable the neural network to extract texture features at a lower level, thereby improving classification accuracy.

(3) Multiple models can improve the overall stability of the network and maintain good performance for different inputs.

2.2.3. Results and Analysis

The number of parameters that need to be optimized and the depth of the network are the main factors affecting the overall efficiency of the network. By comparing the parameters that need to be optimized for each model, the number of layers in the network topology, the time required for one iteration of the model, and the time required to complete the grading of the test set, the efficiency of the fusion model is compared with that of the classical convolutional neural network. DXNet requires optimization of parameters and a large number of network layers. The training round takes a longer time due to the parameters being passed between two convolutional blocks. However, compared with other models that perform well, DXNet's completion time for testing set grading is still within an acceptable range. The time is controlled within two digits, and the effect is good and meets practical needs. The efficiency comparison between each model and DXNet is shown in Table 1.

Table 1. F1 values of each model on four categories of apple images

Model	A	B	C	D
VGG16	0.8510	0.8560	0.9499	0.9252
ResNet50	0.9158	0.9022	0.9377	0.9310
Inception_V3	0.8730	0.8554	0.9237	0.9315
DenseNet121	0.9387	0.9447	0.9747	0.9644
Xception	0.9402	0.9245	0.9692	0.9518
DXNet	0.9771	0.9807	0.9825	0.9664

3. CONCLUSION

In response to the issue of external quality grading in the post-processing of Fuji apples, this paper comprehensively considers multiple indicators that affect the external quality of apples and proposes a deep learning based apple appearance grading method. This method can extract deeper texture features and significantly improve stability and accuracy. The end-to-end modeling strategy can incorporate prior knowledge of manual scoring, significantly improving the accuracy of apple external quality grading. On the basis of comparing and analyzing the structure of classical convolutional neural network models, effective modules in the network structure are extracted for combination design, and the classical convolutional neural network model is improved. The experimental results show that DXNet achieves an accuracy of 97.84% on the test set and has good grading performance, meeting the application requirements of intelligent fruit selection production lines.

REFERENCES

[1] Su Qinghua, Yu Miaomiao, Zhang Yichen, Yang Yichen, Liu Zhiyuan, Ayijiamu•Saipai. A Brief Analysis of ResNet50 [J]. Computer Science and Applications, 2022, 12(10): 2233-2236.

[2] Zeng Deyang, Yang Xinrong. Research on Pneumonia Image Classification Based on Improved Neural Network [J]. Modeling and Simulation, 2024, 13(2): 1374-1380.

[3] Xu Wenzhen, Zhu Yixin, Huang Siyuan, Zhang Zemin, Wang Weiwei. Application of deep learning in the field of fruit and vegetable recognition [J]. Software Engineering and Applications, 2021, 10(3): 329-336.

[4] Han Jijun, Lu Yao, Di Wei, Jiang Long. Research on Power Infrastructure Extraction Method Based on RDU-Net Deep Learning Model [J]. Modeling and Simulation, 2021, 10(2): 435-441

[5] Zhao Lukang, Zhu Xijun. Classification of land use types for vegetation cover [J]. Artificial Intelligence and Robot Research, 2024, 13(2): 388-398.