Nondestructive detection of yellow peach quality parameters based on 3D-CNN and hyperspectral images

Shipu Xu¹, Yong Liu¹, Wenwen Hu¹, Yingjing Wu¹, Sijia Liu², Yunsheng Wang¹, Chang Liu³

¹Agricultural Information Institutes of Science and Technology, Shanghai Academy of Agricultural Sciences, Shanghai, 201403, China
²Department of Railway Transportation, Shanghai Institute of Technology, Shanghai 201418, China
³School of Information Engineering, Nanchang Hangkong University, Nanchang Jiangxi, 330038, China

Corresponding author: YUNSHENG WANG (e-mail: wys188@163.com), CHANG LIU (e-mail: 70202@nchu.edu.cn)

Abstract. In recent years, efficient and accurate detection of fruit quality has become an important means of non-destructive testing of fruit quality. In order to solve the problems of cost, efficiency and accuracy in the non-destructive testing of yellow peach quality. This paper proposes a new method for simultaneous detection of yellow peach hyperspectral multiple quality parameters. This method uses the method of extracting and recombining wavelengths at equal intervals in the full-band spectrum instead of the characteristic wavelength selection method. It can make the hyperspectral image contain all-band spectral information. This method uses 3D-CNN to replace the original regression modelling method, which improves the accuracy of model prediction. This method uses the shared network convolutional layer method to perform multi-task learning on the sugar content and hardness of yellow peaches, and finally realizes the simultaneous detection of multiple quality parameters of yellow peaches.

1. Introduction
In recent years, hyperspectral imaging technology can detect the quality of fruits more efficiently and accurately because it integrates the image and spectrum information of the measured object. It has become an important means of non-destructive inspection of fruit quality. At present, the internal quality hyperspectral detection indicators of yellow peaches are mainly concentrated in the three categories of sugar content, firmness and water content.

Existing research mainly focuses on a single quality parameter of a certain type of fruit, such as Mo et al. [1] used hyperspectral absorption images to establish a partial least square regression model of apple internal sugar content, and obtained the correlation coefficient and root mean square error of the optimal model. It is 0.876 and 0.514°Brix. Guan Xiaomei et al. [2] proposed a method to optimize the number of partial least squares factors to improve the predictive ability of the sugar content model, reduce the complexity of the model, and improve the prediction correlation coefficient and root mean square error to 0.871 and 0.604°Brix. Ekramirad et al. [3] compared the apple firmness prediction models established by partial least squares, least squares support vector machines and multiple linear...
regression methods, and concluded that the correlation coefficient of the apple firmness prediction model established by partial least squares reached 0.92. Crichton et al. [4] used spectral reflectance information to estimate the water content of apple slices, and established a prediction model for the spatial average spectral reflectance curve using the partial least squares method. In the previous literature [5], scholars have also studied the simultaneous detection of multiple quality parameters, but they are all combined with multiple models after modelling a single quality parameter, and they have not really achieved relying on the same information and the same modelling. Multiple quality parameters of the method are tested simultaneously.

Hyperspectral analysis technology includes two steps: characteristic wavelength selection and regression modelling. Dong et al. [6] used continuous projection algorithm (SPA) and uninformmed variable elimination algorithm (UVE) to extract characteristic spectra from the full spectrum, and used partial least squares regression (PLS), least square support vector machine (LSSVM) and inverse BP algorithm builds a model. The model predicts the sugar content, firmness, and water content of apples. The results show that the comprehensive judgment ability of the SPA-LSSVM model is better than other models. Zhang et al. [7] used competitive adaptive re-weighted sampling (CARS), SPA, random leapfrog algorithm (RF), and CARS-SPA, CARS-RF combined algorithms to extract effective wavelength results from apple hyperspectral images in the prediction of apple sugar content. This paper uses partial least squares and least squares support vector regression algorithms to establish different prediction models based on the selected effective wavelength. The results show that the least squares support vector regression model based on the 10 effective wavelengths selected by CARS-SPA has the best prediction effect. The feature wavelength selection method of the existing prediction model is complicated, and the extracted feature spectrum is relatively concentrated, and the whole band information is lost. Traditional algorithms such as multiple linear regression, partial least square regression, and least square support vector machines are still used in the selection of prediction algorithms. There is still much room for improvement in the prediction accuracy of the model.

Compared with traditional chemical analysis methods, near-infrared spectroscopy analysis technology is a reliable, fast and non-destructive technology. It takes less time and is suitable for continuous fruit quality testing. This paper proposes a new method for simultaneous detection of yellow peach hyperspectral multiple quality parameters. This method uses the method of extracting and recombining wavelengths at equal intervals in the full-band spectrum instead of the characteristic wavelength selection method. It makes the hyperspectral image contain all-band spectral information. This method uses a three-dimensional convolutional neural network (3D-CNN) to replace the original regression modelling method. This method improves the prediction accuracy of the model, and uses the shared network convolutional layer method to perform multi-task learning on the sugar content, firmness and water content of yellow peaches. Finally, the simultaneous detection of multiple quality parameters of yellow peaches is realized.

2. Materials and Methods

2.1 Yellow peach quality parameter collection and index determination

The experimental object of this research is Jinxiu yellow peach, from the yellow peach base in Wanzhang Village, Qingcun Town, Fengxian District, Shanghai, at 30°55′30″ north latitude and 121°32′36″ east longitude. After removing the damage and internal rotten yellow peaches found in the sampling process, 90 bagged yellow peaches and 118 unbagged yellow peaches were finally obtained as test samples, all of which were normal fruits without obvious defects and damage. In this experiment, the ATAGO PAL-1 digital display brix meter was used to measure the sugar content of yellow peach, and the top GY-4 digital display fruit firmness meter was used to measure the fruit firmness.
2.2 Yellow peach hyperspectral image acquisition and sample expansion

The hyperspectral sensor used in this experiment is the FieldSpec HandHeld2 portable visible light-near infrared spectrometer from the US ASD company, the wavelength range of the measurement spectrum is 325~1075nm and the hyperspectral reflection image of yellow peach is used.

Refer to the literature [7, 11, 12] to select the equator as the incident position of the light source, and select 4 reflection images up, down, left, and right of the target area 0-10 from the light source irradiation point as the focus position of the yellow peach. The hyperspectral images of 245 yellow peach samples are collected, and the process of image collection, denoising and sample expansion is as follows:

1) Through the scale of the obtained hyperspectral data in the wavelength range of the instrument, the signal-to-noise ratio is low, so the separated spectral data is removed, and more than 450~900nm images are finally selected for analysis.

2) Use the SG smoothing algorithm to finely process the hyperspectral raw data, and shield the light intensity information of the relative wavelength.

3) For the yellow peach, extract 4 hyperspectral reflection images of 0-10 incident areas from the light source irradiation point in the full wavelength range of 450~900nm.

4) The hyperspectral image obtained in the process 3) is captured at intervals of 10 segments at 450~900nm, and regenerated into a new yellow peach hyperspectral image to expand the original hyperspectral image. Finally, 9800 11×11×45 yellow peach hyperspectral images were obtained, of which 11×11 was the displacement of the single and double brightness maps, containing 45 spectral bands information in total.

2.3 Multi-task learning model based on 3D-CNN

2.3.1 Principle of 3D-CNN

Convolutional Neural Network (CNN) is one of the most successful areas of deep learning algorithm applications. CNN include 1D-CNN, 2D-CNN and 3D-CNN. 1D-CNN are mainly used for sequence data processing, 2D-CNN are often used for image text recognition, and three-dimensional convolutional neural networks are mainly used for medical image and video data recognition.

When a new image is given, CNN does not know exactly which parts of the original image these features should match, so it will try every possible position in the original image, which is equivalent to turning this feature into a filter. This matching process is called convolution operation, which is the origin of the name of convolutional neural network. In order to effectively reduce the amount of calculation, another effective tool used by CNN is called "Pooling". Pooling is to shrink the input image, reduce pixel information, and only retain important information. The operation of pooling is also very simple. Normally, the pooling area is 2*2 in size. And then converted to the corresponding value according to certain rules, such as taking the maximum value (max-pooling) and average value (mean-pooling). It uses this value as the resulting pixel value.

FIGURE 1 shows the max-pooling result of the 2*2 pooling area in the upper left corner. This article takes the maximum value max (0.77, -0.11, -0.11, 1.00) of this area as the result of pooling, as shown in the FIG. 1.
Fig. 1 The process of pooling

Maximum pooling retains the maximum value in each small block, which is equivalent to retaining the best matching result of this block (because the closer the value is to 1, the better the match). It does not specifically focus on which place in the window is matched, but only whether there is a match. By adding the pooling layer, the image is reduced, which can greatly reduce the amount of calculation and reduce the machine load.

A three-dimensional convolution is a three-dimensional filter that computes a low-dimensional feature representation from three dimensions (x, y, z), and the output is a three-dimensional convolution space. It is very useful in video event detection, 3D medical image pictures and more. Of note, its use is not limited to three-dimensional space, but can be applied to two-dimensional input as well, as shown in FIG. 2.

Fig. 2 Three-dimensional convolutional neural network (3DCNN)

Three-dimensional data (or spatial data) has a variety of expressions, such as voxel images, three-dimensional point clouds, and RGB-D images. These information carriers add depth information in three-dimensional space based on the expression of planar image information, so that the three-dimensional data points are uniquely determined in space, which greatly enriches the dimension of acquiring spatial information. The voxel image can visually observe the shape and posture of the three-dimensional object. In addition, it is easier to transplant the processing and analysis method of the ordinary two-dimensional image onto the voxel image than the three-dimensional point cloud data and the RGB-D image. At the same resolution, 3D point cloud data occupy less spatial resources and are easy to express complex textures. RGB-D images simultaneously record RGB images and depth images through RGB-D cameras. RGB images contain object surface color and texture information. Depth images contain spatial shape information of objects. By combining 2 images to detect and identify 3D models. Due to the portability of voxel images, therefore, this paper uses voxel images as a carrier for three-dimensional data.

The complete 3DCNN in this experiment consists of 10 layers and consists of two modules: automatic feature learning module and tumor classification module [8, 10]. The feature extraction module is composed of 8 layers of networks and can be divided into 4 pairs, each of which is followed by a down-sampling layer. We use the 3D convolution kernel as the feature extractor to extract the training volume data space information. The following formula represents a three-dimensional convolution operation:
In the above formula, the $w_{ki}^l$ represents the kernel in which the 3-dimensional volume space $h_{ki}^{l-1}$ is convoluted. And the $w_{ki}^l(m,n,t)$ represents the weights of each voxel in the convoluted kernel. The output value of the corresponding feature space node is:

$$h_i^l = \sigma(\sum_k w_{ki}^l u_{kl}^l + b_i^l) \quad (2)$$

The entire 3DCNN network model is shown in FIG. 3 [9].

In this paper, a three-dimensional convolutional neural network is introduced into the detection of hyperspectral data of yellow peach. On the basis of the original two-dimensional network that can only extract features of single-band brightness images, it increases the capture of hyperspectral full-band information to improve the final network.

2.3.2 Principles of multi-task learning

Multi-task learning is a machine learning method as opposed to single-task learning. In the field of machine learning, the standard algorithm theory is to learn one task at a time, that is, when the output of the system is a real number. The complex learning problem is first decomposed into theoretically independent sub-problems, and then each sub-problem is learned separately, and finally a mathematical model of the complex problem is established by combining the learning results of the sub-problems. Multi-task learning is a joint learning in which multiple tasks are learned in parallel, and the results affect each other.

The early research work of multi-task learning originated from the study of an important problem in machine learning, that is, the problem of inductive bias. The process of machine learning can be regarded as the process of analyzing the empirical data related to the problem and summarizing the model that reflects the nature of the problem. The function of inductive bias is to guide the learning algorithm how to search in the model space. The performance of the searched model will be directly affected by the inductive bias, and any learning system that lacks inductive bias cannot be effective. Different learning algorithms, such as decision trees, neural networks, support vector machines, etc. It has different inductive biases. People need to manually determine which learning algorithm to use when solving practical problems. In fact, they choose different ones subjectively. A very intuitive idea is whether the process of determining the inductive bias can also be completed automatically through the learning process, that is, the idea of "learning to learn" is adopted. Multi-task learning just provides a feasible way for the realization of the above ideas, that is, using the useful information contained in
related tasks to provide a stronger inductive bias for the learning of the task concerned. It was shown in the FIG. 4.

![Fig. 4 Principles of multi-task learning](image)

The sugar content and firmness of yellow peaches are equal to the quality parameters of yellow peaches. They jointly determine the quality of yellow peaches. At the same time, there are correlations among various quality parameters. Therefore, this paper introduces the multi-task learning method into the research on the quality parameter detection of yellow peaches. By sharing the convolutional layer of the three-dimensional convolutional neural network, a network is designed to complete the detection of sugar content and firmness at the same time, which not only reduces the training cost, but also reduces the task's overload fitting risks, while improving prediction accuracy compared to traditional models.

2.3.3 Design of the same inspection network for hyperspectral multi-variety parameters of yellow peach

Based on the above analysis, the network model mainly includes three shared three-dimensional convolutional layers, one three-dimensional pooling layer and three separate fully connected layers. The network input is a 11×11×45 three-dimensional hyperspectral image of a yellow peach. The first layer is a convolution layer, the size of the convolution kernel is 3×3×5, and the number of channels is 6; the second layer is a pooling layer, pooling The window size is 2×2×2; the third layer is a convolutional layer, the convolution kernel size is 1×1×7, and the number of channels is 16; the fourth layer is a convolutional layer, and the convolution kernel size is 3×3×5. The number of channels is 32. The ReLu activation function is selected, and the prediction information of sample sugar content and firmness is output after three fully connected layers.

3. Results and Analysis

3.1 Measured values of quality parameters of yellow peach

We measure the sugar content and firmness quality parameter information of 245 yellow peaches and the value distribution is shown in Table 1.

| Quality parameters | Minimum | Maximum | Mean  | Standard deviation |
|--------------------|---------|---------|-------|--------------------|
| Brix(°Brix)        | 6.24    | 16.32   | 11.71 | 2.84               |
| Firmness (Kg·cm⁻²) | 3.01    | 11.23   | 4.61  | 1.02               |

3.2 Label data processing and data set division

According to the measured values of quality parameters of yellow peaches, the range of sugar content is 6.24°–16.32°Brix, which is divided into 11 categories. The interval of each type is 1°Brix. The firmness range is 3.01–11.23 kg·cm⁻², which are divided into 11 categories. The interval of each type is 0.5 kg·cm⁻². We collect 245 hyperspectral images of yellow peaches, process 9800 hyperspectral images data of yellow peaches through the related process, and divide the data set by SPXY method,
of which 80% (7840 hyperspectral images) samples are used as training set and 20% (1960 hyperspectral images) samples are used as test set.

3.3 Evaluation index of yellow peach quality parameter detection algorithm

The prediction accuracy (ACC) and the related coefficient (R), root mean square error (RMSE) and relative analysis error (RPD) between the sample measurement value and the predicted value are selected as the evaluation index of the algorithm prediction accuracy. The calculation is shown in the formula (3)-(6):

\[ ACC = \frac{f_0}{f} \times 100\% \]  
\[ R = \frac{\sum_{i=1}^{n}(\bar{y}_i - y_i)^2}{\sum_{i=1}^{n}(\bar{y}_i - \bar{y}_n)^2} \]  
\[ RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{n-1}} \]  
\[ RPD = \frac{SD}{RMSE} \]

In the formula (3): \( f_0 \) represents the number of samples with correct prediction; \( f \) represents the total number of samples.

The \( n \) represents the number of samples; \( y_i \) represents the measured value of the \( i \)-th sample; \( \bar{y}_i \) is the predicted value of the \( i \)-th sample; \( \bar{y}_n \) is the average of the measured values of the \( n \) samples. The closer the value of \( R \) is to 1, the better the predictive ability of the model.

The \( n \) represents the number of samples; \( y_i \) represents the measured value of the \( i \)-th sample; \( \bar{y}_i \) is the predicted value of the \( i \)-th sample. The smaller the RMSE value, the better the predictive ability of the model.

\[ RPD = \frac{SD}{RMSE} \]

The SD represents the standard deviation of the sample. The larger the RPD value, the better the predictive ability of the model and the higher the accuracy.

3.4 Predicted results of yellow peach quality parameters

The change trend of the prediction accuracy obtained from Table 3 is shown in FIG. 5. It can be seen from FIG. 5 that the accuracy of sugar content prediction is significantly reduced when the interval is 11 bands, resulting in a significant decrease in the accuracy of joint prediction. The intensive extraction of less than 10 bands has little effect on the final prediction results, but the intensive extraction causes the model training and testing time to increase exponentially. Considering the cost and practicability of model training, we choose to extract and reorganize the original hyperspectral image with 10 bands apart.

Fig. 5 Predictive accuracy of apple quality parameters by sampling recombination at different intervals
3.5 Contrast with traditional methods
The selected comparison algorithms include Uninformed Variable Elimination (UVE), Sequential Projection Algorithm (SPA), Competitive Adaptive Re-weighting (CARS), Random Frog Algorithm (RF); commonly used regression modeling methods include partial least squares regression (PLS), least squares support vector machine (LSSVM), back propagation (BP). The above-mentioned trained three-dimensional neural network is used as the feature extraction method of yellow peach hyperspectral data. Regression analysis was performed on the extracted three-dimensional features and quality parameters such as sugar content and firmness. This paper uses 3D-CNN to extract features for LSSVM regression modeling. The highest correlation coefficient of yellow peach sugar content is 0.827, RMSE is 1.129, and RPD is 1.850. Compared with the best traditional algorithm combination RF+BP, the accuracy is improved by 15.0%. Using 3D-CNN to extract features for BP regression modeling, the highest RP of yellow peach firmness is 0.755, RMSE is 0.691, and RPD is 1.483. Compared with the best traditional algorithm combination, the accuracy of RF+BP is increased by 17.0%. Therefore, the 3D-CNN-based yellow peach hyperspectral multi-quality parameter joint detection model proposed in this paper has better detection results than traditional models.

4. Conclusion
The yellow peach quality parameter detection model built by the three-dimensional convolutional neural network (3D-CNN) increases the extraction of the full-band spectral information, and the detection accuracy of the band information loss caused by the selection of the characteristic band is greatly improved compared with the traditional method. Aiming at the problem of traditional methods cannot achieve the simultaneous detection of multiple quality parameters relying on the same information and the same modelling method, a multi-task learning method is introduced, and multi-task simultaneous detection is achieved by sharing the network convolutional layer. The results show that the method can simultaneously detect the sugar content and Firmness of yellow peaches with high accuracy. Exploring the detection effect of interval band extraction and recombination of hyperspectral images and using fewer bands of hyperspectral images can also achieve prediction accuracy. It shows that this algorithm is suitable for low-precision hyperspectral images and can reduce the cost of image acquisition and model training. The accuracy of the yellow peach multi-quality parameter detection method based on hyperspectral images and 3D-CNN for simultaneous detection of yellow peach sugar content, firmness, and moisture needs to be improved, and feature fusion methods will be introduced to further improve the model.

Acknowledgments
This work was financially supported by the Shanghai Municipal Agricultural Commission Shanghai Agriculture Applied Technology Development Program, China (Grant No.G2015 6-4-1).

References
[1] J Lammertyn, A Peirs, JD Baerdemaeker. Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment[J]. Postharvest Biology and Technology, 2000, 18(2): 121-132.
[2] VA Mcglone, RB Jordan, PJ Martinsen. Vis/NIR estimation at harvest of pre-and post-storage quality indices for Royal Gala apple[J]. Postharvest Biology and Technology, 2002, 25(2): 135-144.
[3] KH Choi, JA Abbott, B Park, CH Choi, KJ Lee. Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of delicious and Gala apples[J]. Transactions of the ASAE, 2003, 46(6): 1721-1731.
[4] Ann Peirs, Ann Schenk, Bart M. Nicolai. Effect of natural variability among apples on the accuracy of VIS-NIR calibration models for optimal harvest predictions[J]. Postharvest biology and technology, 2005, (35): 1-13.
[5] H Bessho, K Kudo, J Omori, Y Inomata, M Wada, T Masuda. A portable non-destructive quality meter for understanding fruit soluble solids in apple canopies[J]. Acta Horticulturae, 2007, (732): 593-597.

[6] BM Nicolai, BE Verlinden, M Desmet, et al. Time—resolved and continuous wave NIR reflectance spectroscopy to predict soluble solids content and firmness of pear[J]. Postharvest Biology and Technology, 2008, 47(1): 68-74.

[7] Gao, L L, Zhu X C, Li C, et al. Improve the prediction accuracy of apple tree canopy nitrogen content through multiple scattering correction using spectroscopy. Agricultural Sciences, 2016(7): 651-659.

[8] Tran D, Bourdev L, Fergus R, et al. Learning spatiotemporal features with 3D convolutional networks[C] //2015 IEEE International Conference on Computer Vision( ICCV), December 7-13, 2015.

[9] Krizhevsky A, Sutskever I, Hinton G. Image net classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6):84-90.

[10] Sun M L, Song Z J, Jiang X H, et al. Learning pooling for convolutional neural network [J]. Neuro-computing, 2017, 224:96-104.

[11] Argyriou A, Evgeniou T, Pontil M. Convex multi-task feature learning [J]. Machine Learning, 2008, 73(3):243-272.

[12] Zhang Y, Yang Q. An overview of multi-task learning [J]. National Science Review, 2018, 5(1):30-43.