scientific reports

Check for updates

OPEN

Apple quality identification and classification by image processing based on convolutional neural networks

Yanfei Li[1,2], Xianying Feng[1,2✉], Yandong Liu[1,2] & Xingchang Han[1,2,3]

This work researched apple quality identification and classification from real images containing complicated disturbance information (background was similar to the surface of the apples). This paper proposed a novel model based on convolutional neural networks (CNN) which aimed at accurate and fast grading of apple quality. Specific, complex, and useful image characteristics for detection and classification were captured by the proposed model. Compared with existing methods, the proposed model could better learn high-order features of two adjacent layers that were not in the same channel but were very related. The proposed model was trained and validated, with best training and validation accuracy of 99% and 98.98% at 2590th and 3000th step, respectively. The overall accuracy of the proposed model tested using an independent 300 apple dataset was 95.33%. The results showed that the training accuracy, overall test accuracy and training time of the proposed model were better than Google Inception v3 model and traditional imaging process method based on histogram of oriented gradient (HOG), gray level co-occurrence matrix (GLCM) features merging and support vector machine (SVM) classifier. The proposed model has great potential in Apple's quality detection and classification.

Abbreviations

ANN	Artificial neural network
ASM	Angular second moment
CNN	Convolutional neural networks
CON	Contrast
COR	Correlation
DA	Discriminant analysis
ENT	Entropy
GLCM	Gray level co-occurrence matrix
HOG	Histogram of oriented gradient
HIS	Hyperspectral imaging
ILSVRC	ImageNet large scale visual recognition challenge
MNF	Minimum noise fraction
NIR	Near-infrared
PLS	Particle least square
SPA	Successive projections algorithm
SVM	Support vector machine
SWIR	Shortwave infrared

Apples are very popular agricultural products with high nutritional value[1]. After years of development, China has become the world's largest apple producer, with apple planting area and yield accounting for more than 50% of the world. One of the important reasons affecting the export of apples is that the quality of the apples is rather spotty. With increased attention for fruits of high quality and safety standards, the demand for automatic, accurate

[1]School of Mechanical Engineering, Shandong University, Jinan 250061, Shandong, China. [2]Key Laboratory of High Efficiency and Clean Mechanical Manufacture of Ministry of Education, Shandong University, Jinan 250061, Shandong, China. [3]Shandong Academy of Agricultural Machinery Sciences, Jinan 250100, Shandong, China. ✉email: fxying@sdu.edu.cn

and fast quality identification continues to grow[2]. The exponential population spurt threatens to reduce levels of food security as time progresses[3,4]. Therefore, defective apples should be precisely detected and automatically weeded out before they are sold in the market.

Bio-molecular sensing technology, hyperspectral imaging techniques, multispectral imaging, and traditional machine vision technology are effective detection methods for detecting quality fruits. Low et al.[5] constructed an electrochemical impedance genosensing platform based on graphene/zinc oxide nanocomposite to enhance sensitivity of plant disease detection. Wijesinghe et al.[6] proposed a detection method based on bio-photonic technology to improve sensitivity of apple diseases. Hyperspectral imaging techniques combine the advantages of traditional spectroscopy technology and imaging technology[7]. Due to the rich data information, hyperspectral imaging technology has developed into the effective method for quality identification of fruit. Zhang et al.[8] presented the detection algorithm based on successive projections algorithm, particle least square discriminant analysis and minimum noise fraction (SPA-PLS-DA-MNF) by using hyperspectral imaging techniques and spectral analysis for the rottenness apple detection with the accuracy of 98%. The automated detection system based on near-infrared (NIR) coded spot-array structured light was proposed recently for detection of defective apples with overall identification accuracy of 90.2% for three categories[9]. Keresztes et al.[10] proposed hyperspectral imaging (HSI) in the shortwave infrared (SWIR) system to detect bruise region for 'Jonagold', 'Kanzi' and 'Joly Red' apples. Partial least squares-discriminant analysis (PLS-DA) model was also employed to discriminate between sound, bruised, glossy and stem regions and reached to 94.4% accuracy. Zhang et al.[11] introduced full-wavelength model to detect blueberry bruising using hyperspectral transmittance imaging with accuracy of 81.2%. With the research of hyperspectral technology, multispectral technology is favored by researchers in the field of fruit quality identification. Unay et al.[12] presented automatic classification system to grade bi-colored apples into two-category and obtained 93.5% overall accuracy by multispectral imaging techniques. Li et al.[13] proposed multi-spectral algorithm to detect early decay in citrus. Zhang et al.[14] proposed image recognition method to inspect damage, insect damage, bruises, decay apples by multispectral imaging with overall detection accuracy of 91.4%. Pontes et al.[15] used Mass spectrometry imaging for orange trees disease detection. A number of imaging models have been effectively used in fruit quality identification, such as reflectance[16], transmittance[17], fluorescence[18], and Raman[19], but the improvement of recognition accuracy is still a challenge. Although hyperspectral and imaging technology has broadened the application range of machine vision, the huge amount of data in hyperspectral images affected the efficiency of detection. The uneven brightness exists in the hyperspectral image, which still interfered with the detection of apple surface defects.

Most quality inspections of fruits can be achieved with traditional machine vision based on color cameras. Li et al.[20] realized the automatic detection of citrus fruit surface defects based on brightness transformation and image ratio algorithm, and achieved 98.9% detection rate. Dubey et al.[21] presented a method using color, texture and shape features from images and reached to 95.94% disease detection accuracy. Moallem et al.[22] introduced a computer vision-based algorithm to identify the defect in apple and obtained accuracy of 92.5% and 89.2% for healthy and defected apples using SVM. Bhargava and Bansal[23] proposed fruit grading system with SVM to grade mono-colored apples into healthy or defected quality categories by textural, geometrical, and statistical features. This system was tested by two datasets with maximum accuracy of 96.81% and 93.00%, respectively. Bhargava and Bansal[24] presented a system to detect the quality of apple, avocado, banana, and orange by fuzzy C-means clustering with accuracy of 95.72% using SVM. But the above methods are only for defect detection, and do not carry out more detailed classification. With color, texture and shape features are often used to identify disease of fruit, the identification rate of those works is highly dependent on feature extraction.

Recently, artificial intelligence methods represented by deep learning, especially convolutional neural networks (CNN), have achieved a series of important results in the field of fruit quality detection and classification. To overcome issues of above methods, some researchers used deep learning methods for quality identification in agriculture. Sun et al.[25] employed pixel-based convolutional neural network to identify early decay peaches and achieved 97.6% detection rate. Barman et al.[26] constructed Self-Structured classifiers by CNN to grade citrus leaf into three categories with training accuracy and validation accuracy of 98% and 99%, respectively. Fan et al.[27] presented a deep learning architecture based on CNN to detect defective apples and obtained training accuracy and testing accuracy of 96.5% and 92%, respectively. However, above references not used mainstream frameworks to verify and compare on his own dataset, and training time of model was not considered. Thus, it is particularly important to find a robust, accurate and fast model for apple quality identification and classification.

The overall goal of present study was to evaluate potential of proposed model based on CNN for the accurate and fast grading of apple quality. The specific objectives of the work were to: (1) develop and train proposed CNN-based identification architecture for apple quality classification using apple samples; (2) compare the overall performance of proposed CNN-based architecture, Google Inception v3 model (mainstream framework), and HOG/GLCM + SVM (traditional method); and (3) evaluate the performance of three models for grading apples into three quality categories using independent dataset.

Materials and methods
This study is in compliance with relevant institutional, national, and international guidelines and legislation.

Build dataset.
It is important that richer information features of the apple data were constructed for the model to identify targets. 'Yantai Red Fuji' apples were purchased form RT-MART supermarket in Jinan, Shandong Province. In order to increase the accuracy and robustness of recognition model, the time, angle and light intensity of the image collection were different. In total, 3600 original images (a resolution of 3120×4160) were obtained by the same mobile camera (13 mega-pixels, F-Stop = f/2, Exposure time = 1/50 s, ISO speed = 151,
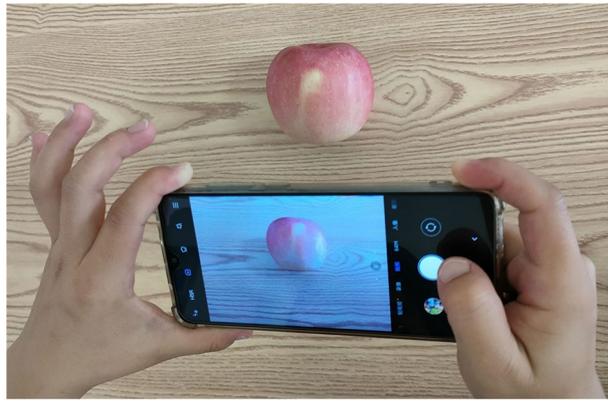
Figure 1. The process of images capturing using mobile camera.

Focal length = 4 mm, Redmi Note 4X, China)[28], which were then divided into three categories of premium, middle and poor grade. The process of images capturing using mobile camera was shown in Fig. 1.

The problem of insufficient size of the training dataset has been solved by techniques of data augmentation[29]. In the case of insufficient dataset, a direct and effective way of data enhancement technology can increase the diversity of training samples, improve the robustness of the model, and avoid overfitting. Changing the training samples can reduce the dependence of the model on some attributes or features using data augmentation technology. The more valuable data based on existing dataset were created by data augmentation strategy. As a result, the performance and robustness of model were improved. In order to reduce the calculation time of data augmentation, the pixels of the original pictures were zoomed out images with a quarter of the original images (a resolution of 780 × 1040). A few sample images in each class were shown in Fig. 2.

In this work, several augmentation techniques without changing the semantics of the images were applied in each scaled-down image such as increasing salt and pepper noise, increasing Gaussian noise, flips, rotation, brightness, and darkness operation. The noise density of salt and pepper noise was 0.3. The operation of increasing Gaussian noise prevented effectively the neural network from fitting all the features of the input image. The manipulation of horizontal flips was used. The percentage of brightness and darkness was respective 1.5 and 0.9. The manipulation of rotation was used to each sub-image, which could generate other five sub-images at 60°, 120°, 180°, 240°, and 300°. Finally, 36,000 apples sub-images were obtained. Augmented sub-images were shown in Fig. 3. The training and validation sets were independent and randomly sampled form 36,000 apple sub-images dataset with proportion of 80% and 20% (28,800 for training, 7200 for validation).

Overall identification architecture. A typical convolutional neural network (CNN) consists of input layer, convolutional layer, full connection layer and output layer. The advantage of convolution is local receptive fields and shared weights, rather than the way of all neurons are connected in the artificial neural network (ANN). In this way, the training parameters of the network were substantial decreased[30]. A CNN-based identification architecture, which was composed of an input layer, 6 convolutional layers (convolution and pooling operations), 2 full connection layers and an output layer, was developed for apple quality recognition. The specific configurations of the proposed model were shown in Table 1.

In the first convolutional layer, in order to acquire high-level features, convolutional kernel shape was 5 × 5, the number of convolution kernels were 8, the stride of the convolution kernel was 1. After the convolution operation, the size of the input image did not change, but the dimension was increased from 3 to 8 because of the convolution mode of SAME PADDING was selected. During the convolution operation, the original features of the input images were not lost using SAME PADDING. In order to reduce the dimensions of images, pooling layer kernel shape was 3 × 3, the stride of the pooling kernel was 2. Inspired by the Hebbian theory[31], convolutional kernel shape of convolution layer 3 was 1 × 1. The convolution kernel (1 × 1) could connect highly correlated features in the same spatial location but different channels. This leaded to a large difference in feature information between adjacent pixels. Therefore, pooling was not used in this layer.

Network updating process. The proposed model was trained using the error back propagation algorithm, which was divided into two processes. The flowchart of the updating process for proposed model was shown in Fig. 4.

Feed-forward propagation. When it comes to convolution operations, it was inseparable from the concept of discrete convolution in mathematics. Discrete convolution was defined as Eq. (1).

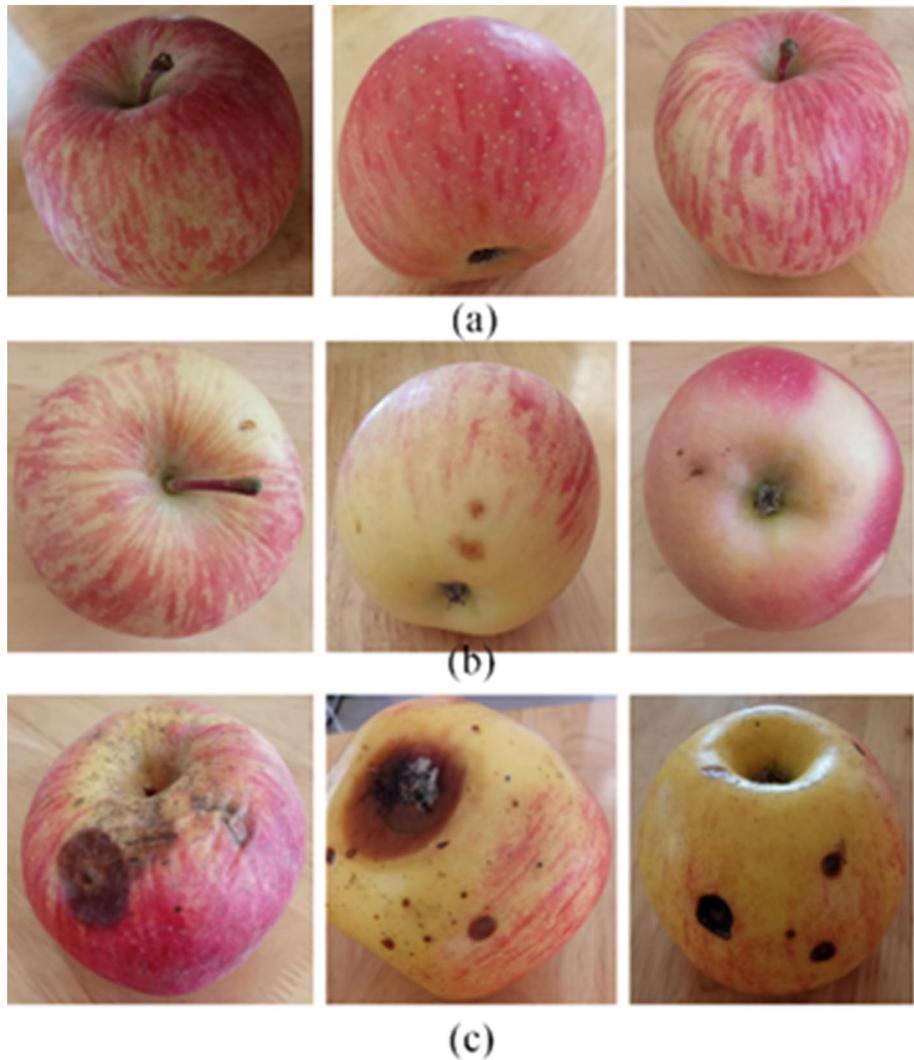$$y(n) = \sum_{i=-\infty}^{\infty} x(i) \cdot h(n-i), \tag{1}$$

Figure 2. A few of images of premium (**a**), middle (**b**) and poor (**c**) grade apples.

where $x(n)$ and $h(n)$ represent discrete sequences respectively. $y(n)$ represents a new sequence obtained by convolution.

The convolution of the two-dimensional discrete function $f(x, y)$ and $g(x, y)$ was defined by Eq. (2).

$$f(x,y) \otimes g(x,y) = \sum_{i}^{\infty} \sum_{j}^{\infty} f(i,j) \cdot g(x-i, y-j). \tag{2}$$

The output of convolution layer neuron were defined as Eq. (3).

$$f(x) = act\left(\sum_{i,j}^{n} x_{ij}\theta_{(n-i)(n-j)} + b\right), \tag{3}$$

where act represents activation function. x_{ij}, θ and b represents i row and j column of pixels, kernel shape and offset value, respectively.

The output of convolution layer and pooling layer were obtained by Eqs. (4) and (5), respectively.

$$O_j = \phi\left(\sum_{Z} X_Z \theta_j + b_j\right), \tag{4}$$
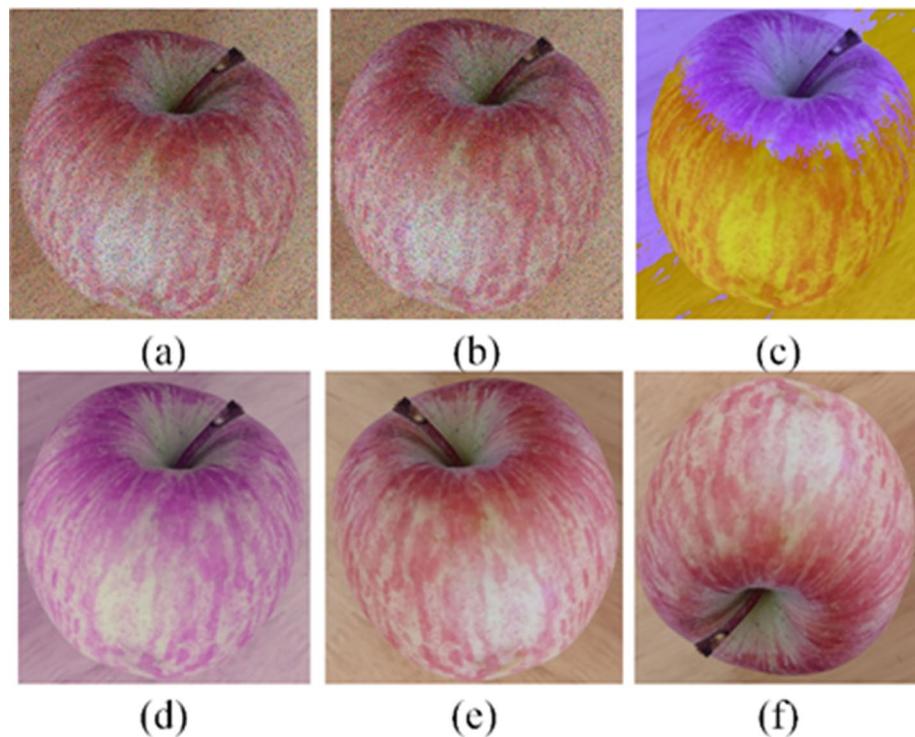
$$S_j = f(P(O_j) + b_j), \tag{5}$$

Figure 3. Representative augmented images: (**a**) increasing salt and pepper noise; (**b**) increasing Gaussian noise; (**c**) brightness; (**d**) darkness; (**e**) flips; (**f**) rotation at 180°.

Layer	Operation	Kernel shape	Number of Kernel	Stride	Number of neurons	Number of training parameters	Number of connections
Conv1	Convolution	5×5	8	1	–	208	8,998,912
	Pooling	3×3	–	2	–	24	865,280
Conv2	Convolution	3×3	16	1	–	1168	12,633,088
	Pooling	3×3	–	2	–	48	432,640
Conv3	Convolution	1×1	32	1	–	1168	1,470,976
Conv4	Convolution	3×3	64	1	–	18,496	50,013,184
	Pooling	3×3	–	2	–	192	432,640
Conv5	Convolution	3×3	64	1	–	46,460	31,204,160
	Pooling	3×3	–	2	–	240	115,200
Conv6	Convolution	3×3	64	1	–	466,144	6,644,736
	Pooling	3×3	–	2	–	192	23,040
Fc1	Full connection	–	–	–	256	590,080	590,080
Fc2	Full connection	–	–	–	256	65,792	65,792
Output	Softmax	–	–	–	3	–	–

Table 1. Detailed configurations and properties of the proposed model.

where S_j represents pooling output of the j feature. f, P and b represent activation function, down-sampling function and offset value, respectively.

After the processing of the pooling layer, a series of feature maps were obtained. Take out the pixels in order from the feature maps and arranged them into a vector. This method was called rasterization. The definition of rasterization was shown as Eq. (6).

$$O_k = [x_{111}, x_{112}, \ldots, x_{11n}, x_{121}, x_{122}, \ldots, x_{12n}, \ldots, x_{1mn}, x_{2mn}, \ldots, x_{jmn}], \tag{6}$$

where O_k represents rasterized vector. Finally, the rasterized vectors were input to the fully connected layer and the classification results were obtained.
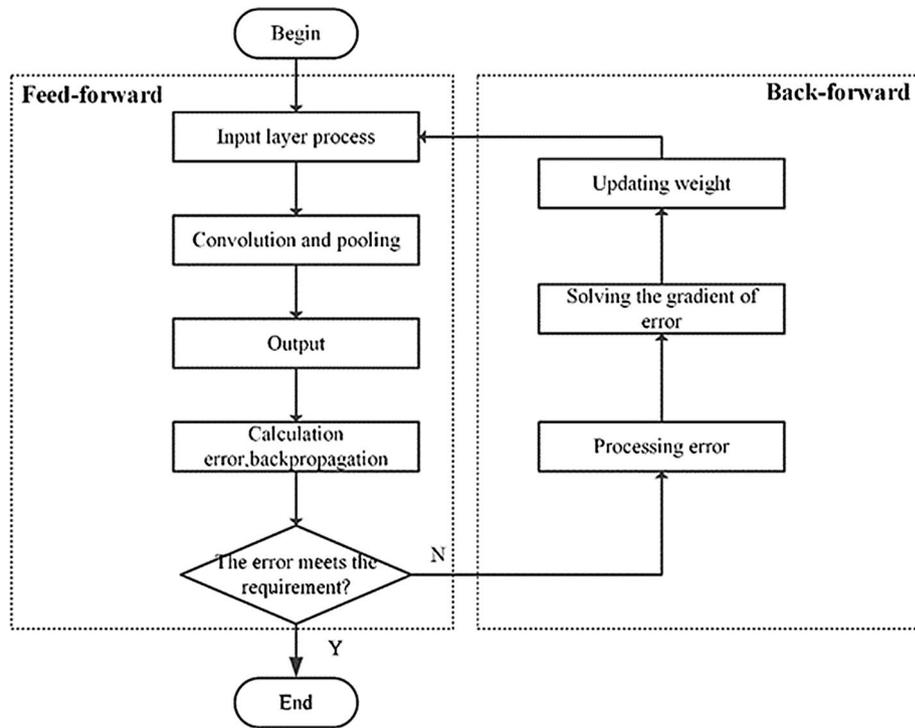
Figure 4. Flowchart of the algorithm for proposed apple identification model.

Back-ward propagation. Backward propagation was mainly the propagation of errors. The error vector δ_k of rasterization was defined as Eq. (7).

$$\delta_k = [\delta_{111}, \delta_{112}, \ldots, \delta_{11n}, \delta_{121}, \delta_{122}, \ldots, \delta_{12n}, \delta_{1mn}, \delta_{2mn}, \ldots, \delta_{jmn}]. \tag{7}$$

The vector error of the pooling layer and convolution layer were shown as Eqs. (8) and (9), respectively.

$$\Delta_k = \{\Delta_1, \Delta_2, \ldots, \Delta_m\}, \tag{8}$$

$$\Delta_p = F(\Delta_k), \tag{9}$$

where m and p represent the number of pooling and convolution, respectively. Δ_k and Δ_p represent vector error of the pooling layer and convolution layer, respectively. F represents Up-sampling Function.

The weight update of a certain region C in the convolutional layer q was calculated by Eq. (10).

$$\frac{\partial E}{\partial \theta_q} = rot180\left(\left(\sum_p O_p\right)_j rot180(\Delta_q)\right), \tag{10}$$

where E, θ_q represent error function, weight, respectively. $rot180$ represents the matrix that was rotated 180°. O_p, Δ_q represent Pooling output, sum of all bias gradients, respectively. The final propagation error Δ_p was defined as Eq. (11):

$$\Delta_p = \left(\sum_{q \in C} \Delta_q rot180(\theta_q)\right) \cdot O_p. \tag{11}$$

Apple quality identification using other two methods. In order to compare the performance of apple quality identification by proposed CNN-based architecture, the Google Inception-v3 model[32] was used for quality identification under the same dataset. In this model, the fully connected layer was replaced by global average pooling for reducing the computational complexity. The main contribution of Google Inception-v3 was the Inception module. Inception v3 was the most classic and stable model of Google Net, it contained 10 inception modules. The accuracy of the model was improved by increasing the depth and width of the network and reducing parameters in Inception module. The structure of Inception was composed of convolution operations corresponding to 1×1, 3×3, and 5×5 convolution kernels and pooling operations corresponding to 3×3 filters, which increased the adaptability of the network to scale. The structure diagram of Inception was shown in Fig. 5.
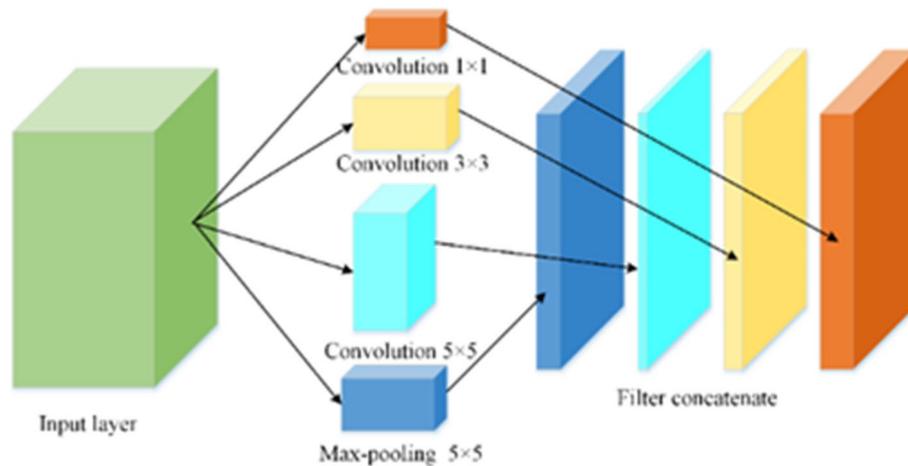
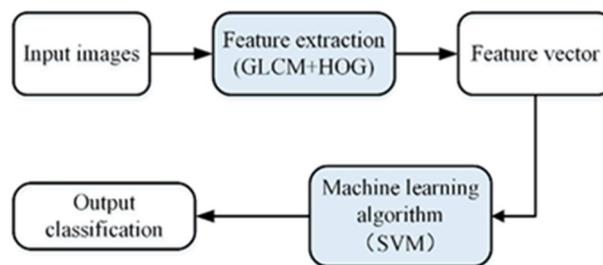**Figure 5.** Diagram of internal structure of inception module.

Figure 6. Flowchart of traditional method.

Similarly, a traditional method was applied for apple quality identification in this study. The traditional method was the work of converting images data from two-dimensional gray space to target pattern space. The result of classification was that the image was divided into several subareas of different categories according to different attributes. Generally, the difference in properties between different image regions after classification should be as large as possible, and the internal properties of the regions should be stable. The flowchart of traditional method was shown in Fig. 6.

The comprehensive information of the image gray level related to direction, adjacent interval, and amplitude of change were reflected by the GLCM of apple image, which were the basis for analyzing the local patterns of the image and arrangement rules. The texture description method of GLCM studied the spatial dependence of gray levels in image texture[33].

The apple images features were extracted by calculating and counting the gradient direction histogram of the local area of the images using HOG method. In order to improve the robustness of HOG features to change in illumination, square root Gamma compression was used to achieve the normalization. The normalized images were convolved using one-dimensional discrete differentiation $[-1,0,1]$ to obtain the gradient component in the horizontal and vertical direction. According to the horizontal vertical gradient of the current image pixel, the gradient amplitude and gradient direction of the pixel were obtained, and the gradient direction histogram was also constructed.

Results and discussions

Experimental details and results of proposed CNN-based architecture.
The methods of parameters selection were used in a variety of ways in the literatures of training CNN model. However, some basic principles still need to be observed in the parameter setting. In this work, Cross entropy function was selected as loss function. Adam Optimizer was selected as optimizer since Adam algorithm made the update of weights and offsets more stable. The size of the input images was set to $208 \times 208 \times 3$. The maximum number of training step was set to 3001 taking into account the total number of data sets and the number of layers of the architecture. In the training process, the learning rate is too large, which makes the network unable to converge, and the learning rate is too small to make the function converge slowly[34]. Therefore, learning rate was set to 0.0001 in this work. The training batch size was selected as 20.

The sub-images from the dataset need to be processed and recognized by the learning model before model training. Figure 7 showed that two batches of images were randomly generated by preprocessing (label 0, label 1, label 2, label 1). Label 0, label 1, and label 2 represent premium grade, middle grade, and poor grade, respectively.
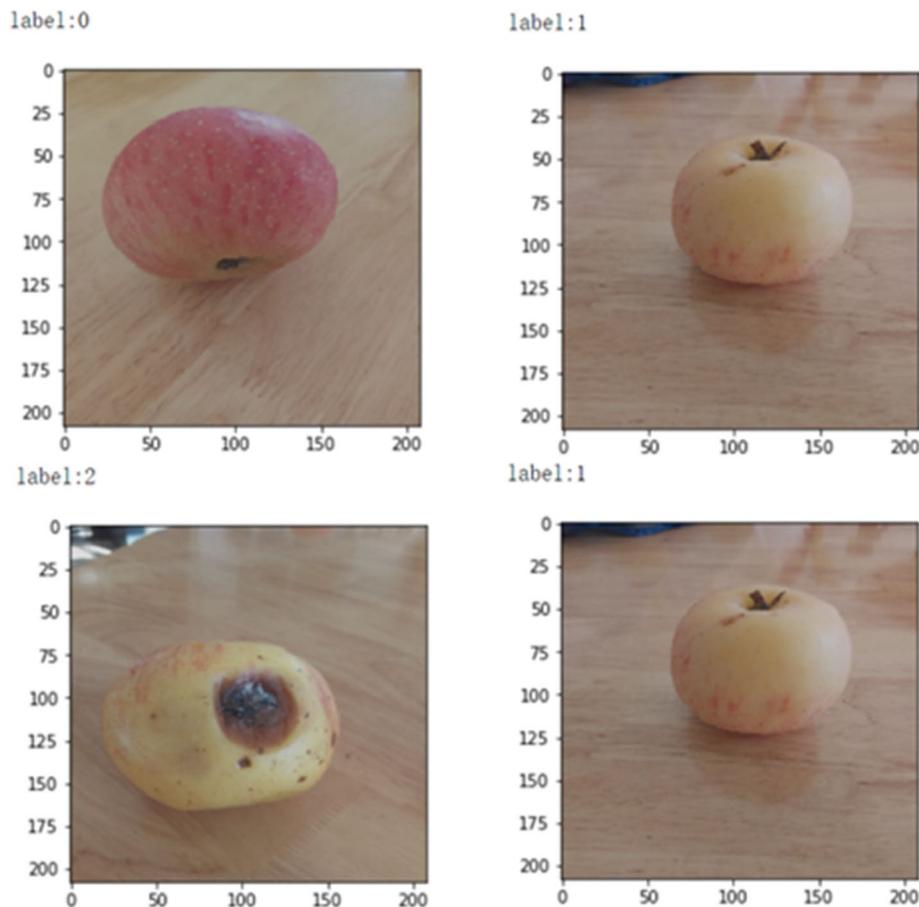
Figure 7. Representative preprocessed sub-images for proposed model.

After preprocessing, the proposed CNN-based architecture was implemented by TensorFlow (1.8.0 CPU only) on Windows system with Intel i7-10700@2.9 GHz and 16 GB RAM using Python Language and was visualized by TensorBoard (1.8.0). The proposed model was fully shown in Fig. 8.

The training and validation accuracy curves of the proposed model were shown in Fig. 9. The whole training process achieved satisfactory results by optimizing weights and bias values at each step. The training time of proposed model was 27 min. The accuracy curves of training and validation sets increased exponentially at 1000 steps and held steady around 96% and 93% after about 2000 steps, respectively. It showed that there is no or slight overfitting in proposed model. After 3001 steps, the trained proposed model and it all parameters were saved. It also indicated that the recognition accuracies in training and validation sets reached their maximum at the 2590th and 3000th step (99% and 98.98%), respectively. The corresponding losses were 0.554 and 0.589 for training and validation, respectively. The training results demonstrated that the proposed architecture has a great potential for apple quality identification.

Performance of Google Inception v3 model for apple quality identification. Although Google Inception v3 model has achieved very good recognition results on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), the process of training all the parameters of the model were relatively time-consuming with the huge training datasets applied. A trained Google Inceptionv3 model was downloaded from Github. In order to save time and prevent overfitting, the parameters of the convolutional layers and the pooling layers were not changed and only the last layer of the model was trained during the training process. Parameters setting for Google Inception v3 model were similar to proposed model. The greatest accuracy generated by Google Inceptionv3 model in training and validation were 92% and 91.2% at the 3000th and 2700th step, respectively, as shown in Fig. 10. In addition, the training time for Google Inception v3 model was 51 min. Although the accuracy curves of the Google Inception v3 model fluctuated less than the proposed model, the proposed model can achieve much better accuracy in apple quality identification than the Google Inception v3 model.

Performance of traditional methods for apple quality identification. In order to achieve the effect of image enhancement, the gray range of the sub-images were changed by weighted mean method. After the grayscale processing, bilateral filtering was applied in this work which reduced sharp changes and noise of image grayscale. Bilateral filtering was a non-linear filter[35] that retained the relationship between pixels in the spatial
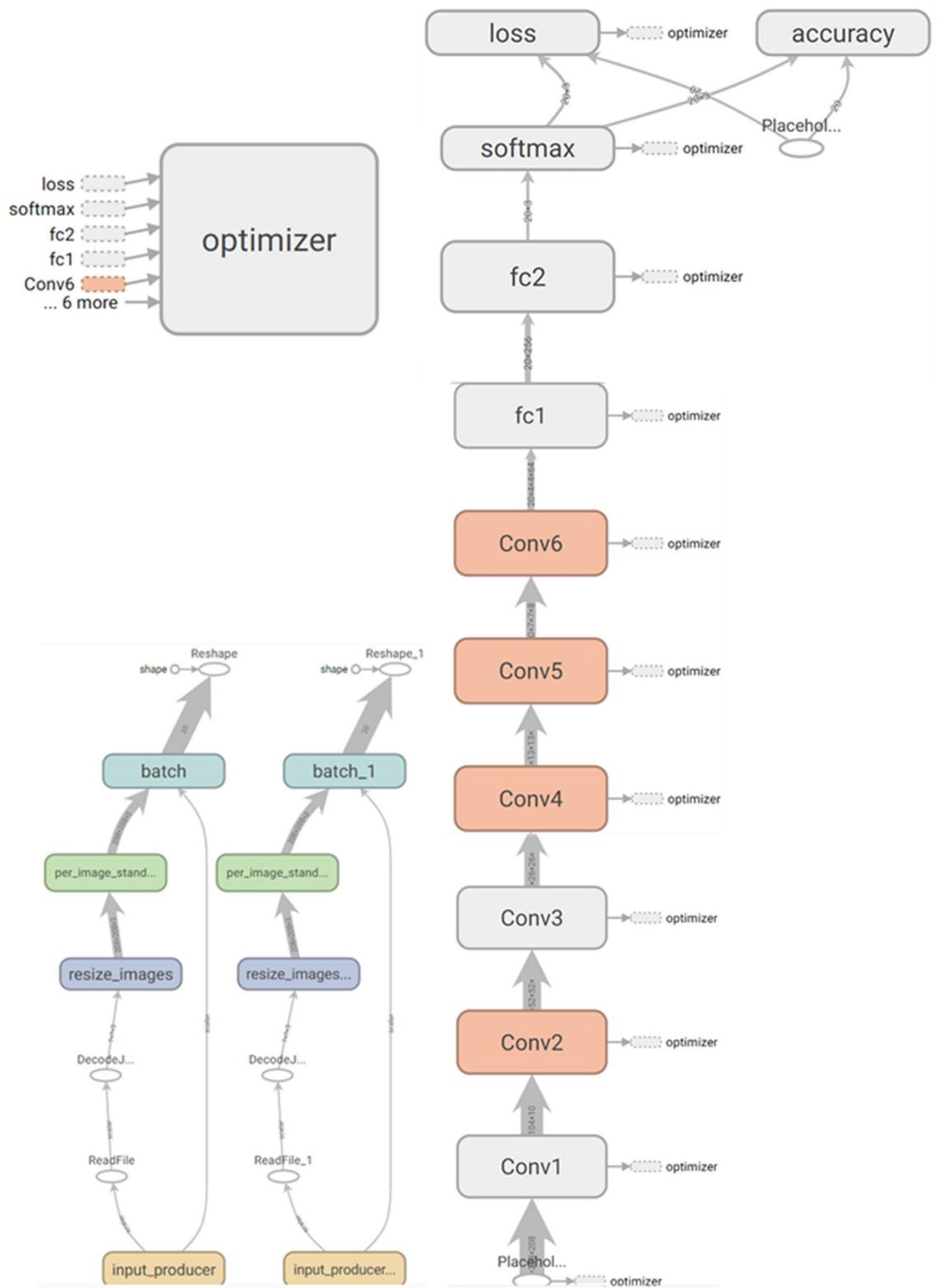
Figure 8. Visual architecture of the proposed model.

distance at the time of sampling and that increased the correlation between pixels to maintain edge features. In the preprocessing of the apple sub-images, the neighborhood diameter was set to 90, sigmaColor and sigmaSpace were set to 75. Bilateral filtering could preserve the detailed contour information of the apple. The results of preprocessing were shown in Fig. 11.

After preprocessing, the local and structural features of the apple images were extracted using GLCM and HOG methods, respectively. In order to improve recognition efficiency, four texture parameters of Angular
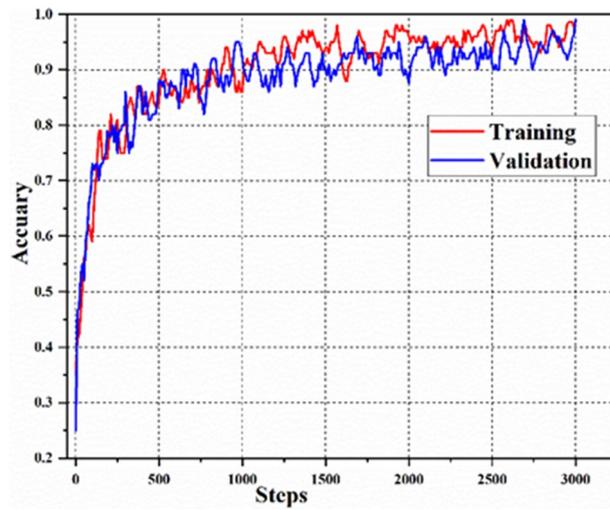
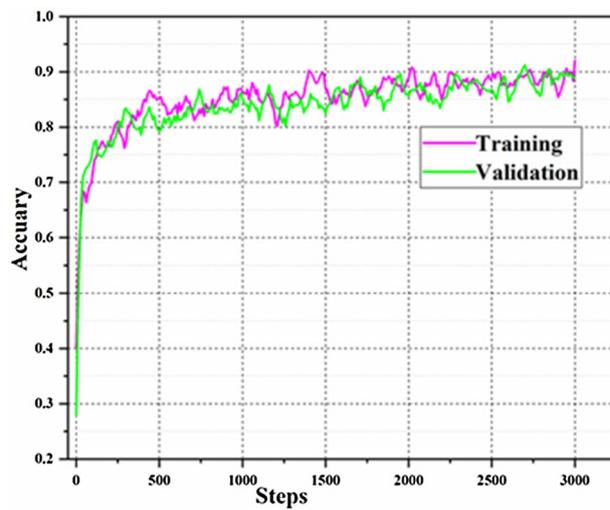**Figure 9.** Training and validation accuracy curves of the proposed model.

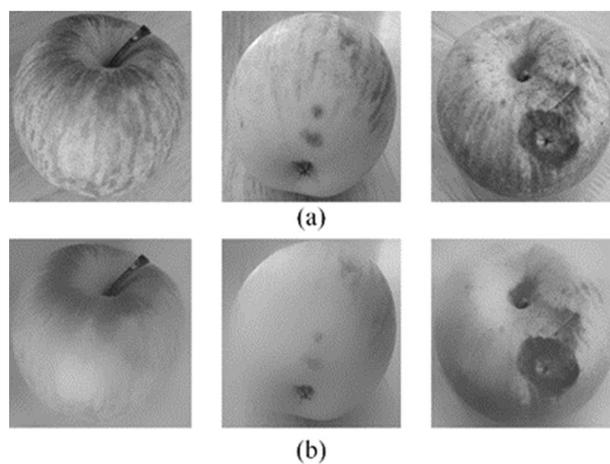**Figure 10.** Training and validation accuracy curves of the Google Inception v3 model.

Figure 11. Representative preprocessed sub-images for traditional method: (**a**) grayscale; (**b**) sub-images processed by bilateral filtering.
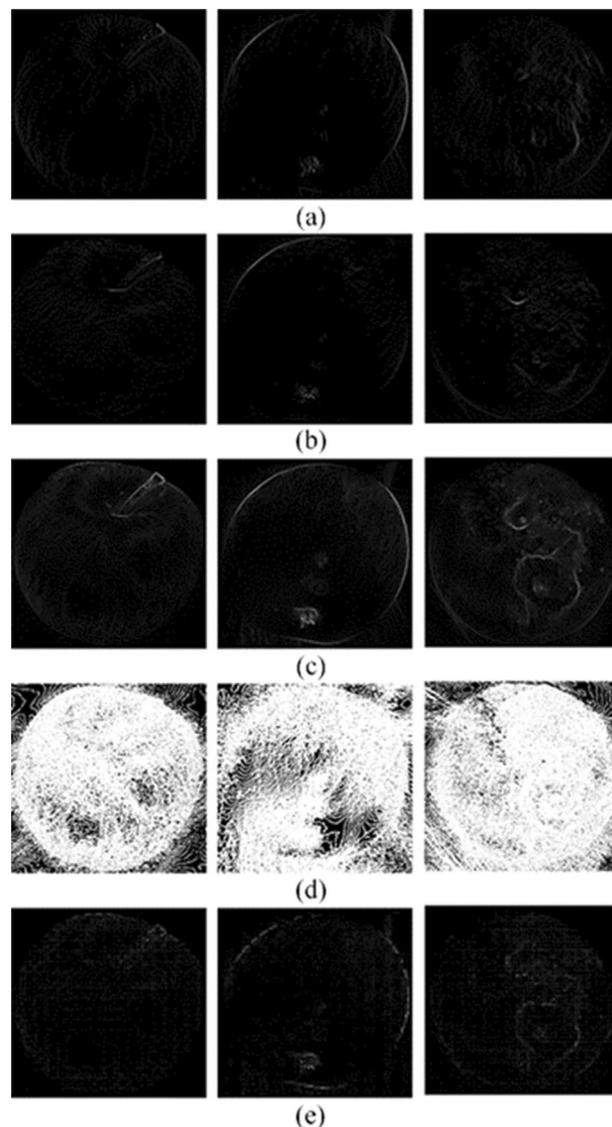
Figure 12. The results of structural feature extraction using HOG for apple images: (**a**) x-direction gradient; (**b**) y-direction gradient; (**c**) gradient magnitude; (**d**) gradient direction; (**e**) HOG features.

Second Moment (Asm), Entropy (Ent), Contrast (Con) and Correlation (Cor) were adopted as texture features. Asm was used to calculate the uniformity of the images. Ent described the amount of information of the apple images. Con reflected the clarity of the images and the depth of the textures. Cor measured the similarity of the gray levels of the images in the row or column direction. Their average and variance were used as the local extracted features.

Before HOG features extraction for apple images, the appropriate apple images block size need to be selected. The setups of the block size in this work were referred to Zhao et al.[36] providing a practical guidance for image identification. The blocks size was 4×4. If the block size was too large, the feature extraction was missing and the feature expression was blurred. If the block size was too small, the excess of useless interference information was collected with computational complexity increasing. With the parameters were selected, the horizontal gradient map and vertical gradient map of the apple images were calculated. After calculation, the images were divided into cell units and a histogram of gradient directions were constructed. A block was composed of 4×4 cells and the normalized gradient histogram was obtained within the block. The HOG feature of the image was obtained by concatenating the features of all blocks. Extraction of structural features using HOG were shown in Fig. 12.

GLCM and HOG features were merged as the input of SVM classifier[37]. Confusion matrix, a supervised classification learning algorithm, was selected to evaluate the accuracy of the classification results for apple quality. The accuracy can be described as Eq. (12):

$$A = \frac{N_T}{N_V} \times 100\%,$$

(12)

Actual classification	Predicted classification		
	Premium	Middle	Poor
Premium	1879	498	23
Middle	213	1736	451
Poor	21	368	2011

Table 2. Classification results of validation set for traditional method.

Figure 13. The online detection system for apple quality detection.

where A, N_T and N_V represent accuracy of apple quality classification, the number of correct classification, and total number of validation datasets, respectively.

The training time for traditional method was 287 min. After training process, 498 and 23 images of premium apple samples were considered as middle apple class and poor apple class, respectively. 213 and 451 images of middle apple samples were considered as premium apple class and poor apple class, respectively. 21 and 368 images of poor apple samples were considered as premium apple class and middle apple class, respectively. The classification results of validation set using traditional method, which combined GLCM and HOG features and SVM classifier, were shown in Table 2. The SVM classifier was developed based on the GLCM and HOG features to distinguish apple quality with the overall accuracy of 78.14% for validation data set.

Testing and discussions.   Compared with Google inception v3 model and traditional image processing classification method, the proposed model obtained satisfying performance[38]. Therefore, a software was developed for image acquisition using Python and PyQt5 on Windows system. OpenCV and camera's API were integrated into this software by Python language employing to acquire and save images. The weights, biases, and structure of the trained proposed model were saved and converted to a Protobuf format file. this file was loaded the software to realize online detection and classification for apple quality. The online detection system was shown in Fig. 13. Simultaneously, an independent testing dataset (300 apples with 100 for each class) was established to test the performance of the proposed model in online detection system. In the test experiment, four images of every apple were acquired by the camera at different angles. Then these images were predicted and scored by the trained proposed model. Some results of prediction and score were shown in Fig. 14. The quality category (premium apple, middle apple, and poor apple) with the highest total score was considered to be the predicted result. The proposed model demonstrated excellent performance for the separate testing dataset. The performance of proposed model was assessed the assistance of a confusion matrix in Fig. 15. In confusion matrix, 96 premium apples were rightly identified and 4 premium apples were considered as middle apples. 5, 93, and 2 middle apples were shown as premium, middle and poor apples, respectively. 97 poor apples were correctly classified and 3 poor apples were recognized as middle apples. The overall classification accuracy of proposed model for testing set was 95.33%. The trained Google Inception v3 model was also loaded into the software to test the performance with overall accuracy of 91.33% for separate testing dataset. Meanwhile, the trained SVM classifier was tested to distinguish apple quality with the overall accuracy of 77.67% for independent testing dataset.

premium apple (score = 0.999780)
middle apple (score = 0.000215)
poor apple (score = 0.000005)

middle apple (score = 0.977143)
poor apple (score = 0.022758)
premium apple (score = 0.000099)

poor apple (score = 0.975504)
middle apple (score = 0.024341)
premium apple (score = 0.000155)

poor apple (score = 0.999911)
middle apple (score = 0.000086)
premium apple (score = 0.000003)

Figure 14. A few of prediction results for independent testing dataset.
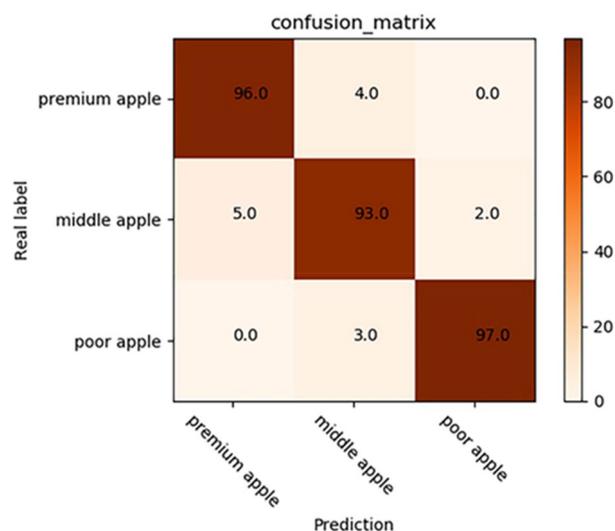
Figure 15. Confusion matrix of proposed model for the separate testing dataset.

Although the detection and classification accuracy will be affected by the complicated working environment such as Apple's moving speed, the number, performance and angle of cameras[39], the detection and classification results obtained by proposed model were superior to spectral imaging technology[9,11,12,40] and traditional machine vision method[22,23]. Feature learning was an advantage of deep convolutional networks over traditional image processing method. Zhang et al.[41] researched blueberry bruising using VGG-16 model (popular architectures). Due to the large number of layers and training parameters of the popular framework, the calculation time did not meet the requirements of Apple's detection and classification. Therefore, a new model for apple detection and classification was proposed in this article.

Conclusions

In this paper, a novel method based on Convolutional Neural Networks (CNN) was proposed and employed for apple quality classification containing disturbing background. Three methods of proposed model based on CNN, Google Inception v3 model (popular architectures) and HOG/GLCM + SVM (traditional imaging process method) were trained and validated to identify apple quality. The proposed model was trained and validated with best training and validation accuracy of 99% and 98.98%, respectively. The greatest accuracy generated by Google Inceptionv3 model in training and validation were 92% and 91.2%, respectively. The SVM classifier was trained based on the GLCM and HOG features to distinguish apple quality with the overall accuracy of 78.14% for validation data set. The proposed model was more acceptable than the other two methods from the accurate results. In addition, the training time of proposed model, Google Inception v3 model and HOG/GLCM + SVM were 27, 51, and 287 min, respectively. The proposed model took the shortest times for training process. Moreover, three methods were tested using independent testing set, obtaining the accuracy of 95.33%, 91.33%, and 77.67%, respectively. The overall results showed that the proposed model has great potential in apple quality detection and classification. The proposed model detects more apple quality attributes including color, size, types, ripe or unripe, and physiological disorders in the future. The proposed model can be further extended to identify more than three categories of apple quality and classify other fruits. The proposed model will be deployed real online sorting equipment to test it performance in the future.

References

1. Lu, Y. & Lu, R. Non-destructive defect detection of apples by spectroscopic and imaging technologies: A review. *Trans. ASABE* **60**(5), 1765–1790 (2017).
2. Zhang, B. *et al.* Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *Food Res. Int.* **62**, 326–343 (2014).
3. Koyande, A. K. *et al.* Microalgae: A potential alternative to health supplementation for humans. *Food Sci. Hum. Wellness* **8**, 16 (2019).
4. Ssrab, D. *et al.* Optimum interaction of light intensity and CO_2 concentration in bioremediating N-rich real wastewater via assimilation into attached microalgal biomass as the feedstock for biodiesel production. *Process Saf. Environ. Prot.* **141**, 355–365 (2020).
5. Low, S. S. *et al.* Sensitivity enhancement of graphene/zinc oxide nanocomposite-based electrochemical impedance genosensor for single stranded RNA detection. *Biosens. Bioelectron.* **94**, 365–373 (2017).
6. Wijesinghe, R. E. *et al.* Biophotonic approach for the characterization of initial bitter-rot progression on apple specimens using optical coherence tomography assessments. *Sci. Rep.* **8**(1), 15816 (2018).
7. Lu, Y., Saeys, W., Kim, M., Peng, Y. & Lu, R. Hyperspectral imaging technology for quality and safety evaluation of horticultural products: A review. *Postharvest Biol. Technol.* **170**, 111318 (2020).
8. Zhang, B. *et al.* Detection of early rottenness on apples by using hyperspectral imaging combined with spectral analysis and image processing. *Food Anal. Methods* **8**(8), 2075–2086 (2015).
9. Zhang, C. *et al.* Automatic detection of defective apples using NIR coded structured light and fast lightness correction. *J. Food Eng.* **203**, 69–82 (2017).
10. Keresztes, J. C., Diels, E., Goodarzi, M., Nguyen-Do-Trong, N. & Saeys, W. Glare based apple sorting and iterative algorithm for bruise region detection using shortwave infrared hyperspectral imaging. *Postharvest Biol. Technol.* **130**, 103–115 (2017).
11. Zhang, M., Jiang, Y., Li, C. & Yan, F. Fully convolutional networks for blueberry bruising and calyx segmentation using hyperspectral transmittance imaging. *Biosyst. Eng.* **192**, 159–175 (2020).
12. Unay, D. *et al.* Automatic grading of Bi-colored apples by multispectral machine vision. *Comput. Electron. Agric.* **75**(1), 204–212 (2010).
13. Li, J. *et al.* Fast detection and visualization of early decay in citrus using Vis-NIR hyperspectral imaging. *Comput. Electron. Agric.* **127**, 582–592 (2016).
14. Zhang, B. *et al.* From hyperspectral imaging to multispectral imaging: Portability and stability of HIS-MIS algorithms for common defect detection. *Postharvest Biol. Technol.* **137**, 95–105 (2018).
15. de Moraes Pontes, J. G. *et al.* Mass spectrometry imaging as a potential technique for diagnostic of Huanglongbing disease using fast and simple sample preparation. *Sci. Rep.* **10**(1), 13457 (2020).
16. Lu, Y. & Lu, R. Development of a multispectral structured illumination reflectance imaging (SIRI) system and its application to bruise detection of apples. *Trans. ASABE* **60**(4), 1379–1389 (2017).
17. Pan, L. *et al.* Hyperspectral imaging with different illumination patterns for the hollowness classification of white radish. *Postharvest Biol. Technol.* **126**, 40–49 (2017).
18. Mo, C. *et al.* Fluorescence hyperspectral imaging technique for foreign substance detection on fresh-cut lettuce. *J. Sci. Food Agric.* **97**(12), 3985–3993 (2017).
19. Qin, J., Chao, K. & Kim, M. Raman scattering for food quality and safety assessment. In *Light Scattering Technology for Food Property* (ed. Lu, R.) 387–428 (CRC Press, 2016).
20. Li, J., Rao, X., Wang, F., Wu, W. & Ying, Y. Automatic detection of common surface defects on oranges using combined lighting transform and image ratio methods. *Postharvest Biol. Technol.* **82**, 59–69 (2013).
21. Dubey, S. R. & Jalal, A. S. Apple disease classification using color, texture and shape features from images. *SIViP* **10**(5), 819–826 (2016).
22. Moallem, P., Serajoddin, A. & Pourghassem, H. Computer vision-based apple grading for golden delicious apples based on surface features. *Inf. Process. Agric.* **4**(1), 33–40 (2017).
23. Bhargava, A. & Bansal, A. Machine learning based quality evaluation of mono-colored apples. *Multimed. Tools Appl.* **79**(31–32), 22989–23006 (2020).
24. Bhargava, A. & Bansal, A. Automatic detection and grading of multiple fruits by machine learning. *Food Anal. Methods* **13**(3), 751–761 (2020).
25. Sun, Y., Lu, R., Lu, Y., Tu, K. & Pan, L. Detection of early decay in peaches by structured-illumination reflectance imaging. *Postharvest Biol. Technol.* **151**, 68–78 (2019).
26. Barman, U., Choudhury, R., Sahu, D. & Barman, G. Comparison of convolution neural networks for smartphone image based real time classification of citrus leaf disease. *Comput. Electron. Agric.* **177**, 105661 (2020).

27. Fan, S., Li, J., Zhang, Y., Tian, X. & Huang, W. On line detection of defective apples using computer vision system combined with deep learning methods. *J. Food Eng.* **286**, 110102 (2020).
28. Chai, W. S., Cheah, K. H., Koh, K. S., Chin, J. & Chik, T. Parametric studies of electrolytic decomposition of hydroxylammonium nitrate (han) energetic ionic liquid in microreactor using image processing technique. *Chem. Eng. J.* **296**, 19–27 (2016).
29. Hussain, M. A. I., Khan, B., Wang, Z. & Ding, S. Woven fabric pattern recognition and classification based on deep convolutional neural networks. *Electronics* **9**(6), 1048 (2020).
30. Qian, Y., Dong, J., Wang, W. & Tan, T. Deep learning for steganalysis via convolutional neural networks. In *Proc. SPIE, the International Society for Optical Engineering*, 9409 (2015).
31. Gillett, M., Pereira, U. & Brunel, N. Characteristics of sequential activity in networks with temporally asymmetric Hebbian learning. *Proc. Natl. Acad. Sci. U.S.A.* **117**(47), 29948–29958 (2020).
32. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818–2826 (2015).
33. Liu, Y., Li, Q., Du, B. & Farzaneh, M. Feature extraction and classification of surface discharges on an ice-covered insulator string during AC flashover using gray-level co-occurrence matrix. *Sci. Rep.* **11**(1), 2542 (2021).
34. Bengio, Y. Practical recommendations for gradient based training of deep architectures. In *Neural Networks: Tricks of the Trade* 2nd edn (eds Montavon, G. *et al.*) 437–478 (LNCS, 2012).
35. Jin, C., Kong, X., Chang, J., Cheng, H. & Liu, X. Internal crack detection of castings: a study based on relief algorithm and Adaboost-SVM. *Int. J. Adv. Manuf. Technol.* **108**(3), 3313 (2020).
36. Zhao, R., Wang, H., Wang, K., Wang, Z. & Liu, W. Recognition of bronze inscriptions image based on mixed features of histogram of oriented gradient and gray level co-occurrence matrix. *Laser Optoelectr. Prog.* **57**(12), 98–104 (2020).
37. Liu, J. *et al.* Fuzzy evaluation output of taste information for liquor using electronic tongue based on cloud model. *Sensors* **20**(3), 686 (2020).
38. Rajan, S., Kumar, M., Ansari, M. J., Rao, D. P. & Kaistha, N. Limiting gas liquid flows and mass transfer in a novel rotating packed bed (HiGee). *Ind. Eng. Chem. Res.* **50**(2), 986–997 (2017).
39. Hacking, J. A. *et al.* Improving liquid distribution in a rotating packed bed. *Chem. Eng. Process.* **149**, 107861 (2020).
40. Ashok, V. & Vinod, D. S. Automatic quality evaluation of fruits using probabilistic neural network approach. In *International Conference on Contemporary Computing & Informatics IEEE* (2014).
41. Zhang, M., Wei, L. & Du, Q. Diverse region-based CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **27**, 2623 (2018).

Author contributions

Y.L. and X.F. conceived and designed the study. Y.L. and X.H. analyzed data. Y.L. wrote this manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to X.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.