Research Article

Detection and Classification of Early Decay on Blueberry Based on Improved Deep Residual 3D Convolutional Neural Network in Hyperspectral Images

Shicheng Qiao,[1,2] Qinghu Wang,[1] Jun Zhang,[1] and Zhili Pei ⓘ [1]

[1]*Inner Mongolia University for Nationalities, College of Computer Science and Technology, Tongliao 028043, China*
[2]*College of Information and Electric Engineering, Shenyang Agricultural University, Shenyang 110866, China*

Correspondence should be addressed to Zhili Pei; zhilipei@imun.edu.cn

Recently, the automatic detection of decayed blueberries is still a challenge in food industry. Early decay of blueberries happens on surface peel, which may adopt the feasibility of hyperspectral imaging mode to detect decayed region of blueberries. An improved deep residual 3D convolutional neural network (3D-CNN) framework is proposed for hyperspectral images classification so as to realize fast training, classification, and parameter optimization. Rich spectral and spatial features can be rapidly extracted from samples of complete hyperspectral images using our proposed network. This combines the tree structured Parzen estimator (TPE) adaptively and selects the super parameters to optimize the network performance. In addition, aiming at the problem of few samples, this paper proposes a novel strategy to enhance the hyperspectral image sample data, which can improve the training effect. Experimental results on the standard hyperspectral blueberry datasets show that the proposed framework improves the classification accuracy compared with AlexNet and GoogleNet. In addition, our proposed network reduces the number of parameters by half and the training time by about 10%.

1. Introduction

Blueberries are popular worldwide for their excellent flavor and high nutritional value [1]. Most of blueberries used for fresh consumption are hand-picked and transported over long distances. Damage during transportation will accelerate fruit decay and reduce overall quality [2]. Therefore, it is important to identify rotten blueberries from healthy blueberries to remove low-quality blueberries from the fresh blueberry supply chain [3].

Under the current industrial standard, the internal decay of blueberry is usually judged by the human touch or by observing the dark rotten tissue of blueberry [4, 5]. The decayed tissue of blueberry becomes darker and more obvious, similar to black, and easier to observe with the naked eye. However, it takes a lot of manpower and time to identify the degree of decay, and it will become inaccurate after several hours of continuous inspection [6]. In addition, the inspection efficiency is very low, and the inspection of early decay is not accurate. The development of hardness measurement method accelerates the detection process of fruit quality evaluation and makes fruit classification more accurate, including blueberry hardness and texture analyzer [7], tomato acoustic pulse response measurement [8], and peach inspection method based on frequency resonance [9]. These methods can provide more accurate hardness measurement, but many hardness measurement technologies need direct touch with fruits, which may cause blueberries to be damaged.

Some researchers at home and abroad used nondestructive testing technology such as machine vision and Hyperspectral Imaging to detect fruit disease or maturity [10] and achieved some excellent results. Georgina et al. [11] used machine vision technology to extract 14 types of features such as color, shape, and texture of citrus and then used Classification And Regression Trees(CART), naive Bayes

(NB), and multilayer perceptron (MLP) to detect citrus canker, black spot, and sclerosis. Lewers et al. [12] used machine vision technology to detect pomegranate disease, where K-means and threshold segmentation methods were used in the experiment to extract the lesion area of pomegranate, and the discrete wavelet transform is adopted to get a set of visual features of the lesion as the input vector of the support vector machine (SVM) model so as to identify pomegranate disease. Lorente [13] uses the hyperspectral imaging system to obtain the hyperspectral images of sound, slight-decayed, moderate-decayed, and severe-decayed peaches, the threshold segmentation method to detect the disease area of peaches, and then the successive projections algorithm (SPA) to extract six characteristic wavelengths, establishes the partial least squares discriminant analysis (PLS-DA) model to identify the disease, and further improves the identification rate of the rotten peaches. Wang et al. [14] also uses hyperspectral imaging technology to obtain the spectral data of the region of interest, where five characteristic wavelengths are extracted by the permutation test method, and the multiple partial least squares regression discriminant analysis model is used to detect the citrus-rot disease caused by fungal infection. Wang et al. [15] used hyperspectral imaging technology to obtain apple spectrum data. Firstly, the threshold segmentation method is used to segment Apple lesion area and extract hyperspectral data; then, the successive projection algorithm is adopted to extract three characteristic wavelengths from the full wavelength; finally, an improved linear discriminant analysis combined with the support vector machine and BP artificial neural network model to detect apple disease. Liu et al. [16] used hyperspectral imaging technology to detect and distinguish the crack, peel spots, malformation, hidden damage, and normal fruit of nectarine. Ten characteristic wavelengths were extracted and the top ten principal component values were obtained by principal component analysis. The disease areas of Nectarine were extracted by threshold segmentation. Finally, the principal component value and six texture indexes (mean, contrast, correlation, energy, homogeneity, and entropy) are fused to establish the ELM model to detect and distinguish the external defect samples and intact samples.

Hyperspectral imaging technology covered the range of 420–1000 nm was employed to detect the nectarine fruit in the literature [17]. 400 RGB images were acquired through a total of 400 samples, which included four types of defective features and sound features. After acquiring hyperspectral images of nectarine fruits the spectral data were extracted from region of interest (ROI). Using Kennard Stone algorithm, all kinds of samples were randomly divided into training set (280) and testing set (120). First of all, according to the calculation of partial least squares regression (PLSR), 10 wavelengths at 497 nm, 534 nm, 657 nm, 677 nm, 696 nm, 709 nm, 745 nm, 823 nm, 868 nm, and 943 nm were selected as the optimal sensitive wavelengths (SWs), respectively. Subsequently, the image of the 876 nm wavelength was selected as the feature image; then, principal component analysis (PCA), Sobel edge detector, and region growing algorithm were carried out among defective and normal

nectarines to extract the defective region. Moreover, ten principal components (PCs) were selected based on PCA and seven textural feature variables (mean, contrast, correlation, energy, homogeneity, and entropy) were extracted by using gray level co-occurrence matrix (GLCM), respectively. Finally, the ability of hyperspectral imaging technique was tested by using the extreme learning machine (ELM) models. The ELM classification model was built on the basis of the combination of PCs and textural features. The results show the correct discrimination accuracy of defective samples was 91.67 %, and the correct discrimination accuracy of normal samples was 100%. The research revealed that the hyperspectral imaging technique is a promising tool for detecting defective features in nectarine which could provide a theoretical reference and basis for design in the classification system of fruits in further work [18–20].

In the abovementioned research studies, whether machine vision technology or hyperspectral imaging technology is used, the disease areas of citrus, pomegranate, and other medium-sized fruits need to be separated from the normal areas. Because the color characteristics of the disease areas are obviously different with that of normal areas, the disease areas can be easily separated by threshold segmentation. However, the skin color of blueberry is darker, and the color characteristics of its normal area and disease area are similar, so it is difficult to segment blueberry disease effectively by using the conventional threshold segmentation method [15]. With the development of intelligent signal processing technology, using the convolutional neural network (CNN), we can overcome the abovementioned shortcomings [21]. CNN has a very prominent performance in machine vision tasks by using the local receptive field model to simulate human brain image processing. For example, the two-dimensional convolutional neural network (2D-CNN) is used to mine the spatial features of the principal component band, and the spectral features are fused by the feature fusion technology to classify the images [21]. In the two-channel CNN method, one-dimensional convolutional neural network is used for spectral feature information, while the spatial feature information is extracted, fused, and classified by 2D-CNN, which also achieves good results. However, there is a disadvantage in these methods: before extracting features, principal component analysis (PCA) and other methods should be used to select the principal component band to reduce the dimension, otherwise too many parameters will be introduced, which is difficult to train and optimize the deep network.

The advantage of 3D-CNN convolution kernel in extracting hyperspectral image features is that the spectral information and spatial information are extracted synchronously, which gives full play to the advantages of 3D hyperspectral image [22]. The 3D-CNN feature model directly extracts the spectral spatial features of hyperspectral image end-to-end, which has better classification effect than 2D-CNN features. Spectral spatial-based residual network introduces the residual structure into the 3D-CNN network and uses two 3D convolution kernels of spectral and spatial features to extract deep features, which can improve the recognition accuracy for mildew blueberries [23].

There are some problems in the existing 3D-CNN model, for example, the number of network layers is generally shallow, hyperparameter optimization is time-consuming and laborious, and the accuracy needs to be further improved. To solve the abovementioned problems, the traditional hyperspectral 3D convolution method is improved to obtain the deep features with stronger representation, and it combines the tree structured Parzen estimator (TPE) adaptively and selects the super parameters to optimize the network performance [23]. In addition, aiming at the problem of few samples, this paper proposes a novel strategy to enhance the hyperspectral image sample data, which can improve the training effect.

The contributions of this article are summarized as follows:

(1) An improved Deep Residual 3D Convolutional Neural Network is proposed. The input image of the model is the original hyperspectral image, no dimensionality reduction method is needed, and the image space and spectral characteristics are retained. The extracted features are more representative of hyperspectral images. It makes full use of spectral and spatial 3D correlation information instead of just their separate and independent feature information.

(2) It can avoid introducing excessive parameters, prevent overfitting, and improve computing efficiency; compared with 2D-CNN, 3D-CNN is more suitable for hyperspectral image processing tasks.

(3) Rich spectral and spatial features can be rapidly extracted from samples of complete hyperspectral images using our proposed network. This combines the tree structured Parzen estimator(TPE) adaptively andselects the super parameters to optimize the network performance. In addition, aiming at the problem of few samples, this paper proposes a novel strategy to enhance the hyperspectral image sample data, which can improve the training effect.

2. Blueberry and Its Hyperspectral Imaging Features

Since this study uses hyperspectral imaging mode to detect rotten areas of blueberries, this section needs to introduce blueberries and their hyperspectral imaging functions. Blueberry is a typical climacteric fruit. In the process of maturity, the physical and chemical properties of the inside of the fruit are constantly changing, the color is gradually changing from green to blue or dark purple, and the picking period of blueberry is relatively concentrated. Figure 1(a) shows fresh blueberries on fruit trees. Because the temperature in picking season is high in summer, the fruits are easy to soften or even brown after picking. In the process of transportation, storage, and sales, they are also prone to rot and disease. Because the picking time of fruit is one of the key factors that lead to the taste of fruit, picking the fruit in advance will lead to too stiff and sour taste, affect the flavor and value of the fruit, and it is difficult to meet the eating requirements; picking the fruit too late will lead to over-

ripeness, be easy to deteriorate, and be inconvenient for storage, so it is not easy to carry out subsequent processing. Figure 1(b) shows the mildewed blueberry. Therefore, sorting blueberries after picking is of great significance to increase the added value of blueberries.

Hyperspectral imaging integrates image processing and spectroscopic techniques to obtain the hyperspectral 3D cube data (hypercube). Hyperspectral data cube is not really images that represent spatial 3D. Strictly speaking, the hyperspectral image should be a 2.5D image data. In terms of images, most of the digital images usually are RGB (red, green, and blue) images, which are made up of three basic colors. That is to say, an RGB image can be divided into red, green, and blue components, and each component can generate a gray image. In digital images, this grayscale image is composed of a 2D data matrix, and each data in the matrix is commonly referred to as a pixel. For example, a 256×256 RGB image, its actual data storage size is $256 \times 256 \times 3$, where 3 represents its three RGB components. If these 3 components are extended to hundreds or thousands of continuous bands, such as 100 continuous bands, the data of the image will be expanded to $256 \times 256 \times 100$, and this 100 is the expansion of the spectrum, which makes the image add rich spectral information. The x and y of a hyperspectral image represent its image in the pixel dimension. If you take a point from the image dimension, this point can be connected in the spectral dimension to get the spectrum at this point.

3. Deep Residual 3D Convolutional Neural Network

3.1. 3D Convolutional Neural Network. 2D-CNN, as a classic deep learning in image processing, has outstanding performance in a variety of machine vision tasks, such as image classification, object detection, and dense captioning tasks [24–26]. The advantage of 2D-CNN is that the features can be directly extracted from ordinary images to complete end-to-end processing. The structure of 2D-CNN is shown in Figure 2, where N is the size of the convolution kernel in the convolution layer and L is the number of output channels of the convolution layer. The convolution process can be written by the following equation:

$$V_{1j}^{xy} = f\left(\sum_m \sum_{h=0}^{H_{1-1}} \sum_{w=0}^{W_{1-1}} K_{1jm}^{hw} V_{(1-1)m}^{(x+h)(y+w)} + b_{1j} \right), \quad (1)$$

where m is s the number of channels; H_{1-1} and W_{1-1} are the size of the convolution kernel, respectively; k and b are the linear coefficients.

Each channel needs to train a convolution kernel when performing 2D convolution processing. If 2D-CNN is used directly in the hyperspectral image classification task, a large number of parameters will be introduced into the calculation because of the many channels in the hyperspectral image. Too many parameters not only make the network more prone to overfitting and affect the accuracy but also greatly reduce the training speed and calculation efficiency of the network.

(a)

(b)

FIGURE 1: Comparison of blueberries with different levels. (a) Fresh blueberries on fruit tree; (b) mildewed blueberries.
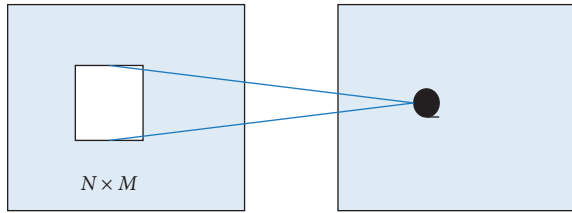

$N \times M$

FIGURE 2: The structure of 2D-CNN.

Usually, in order to solve this kind of problem, scholars take dimension reduction as preprocess before inputting hyperspectral image. For example, they use the PCA method to extract 3 principal component channels in the hyperspectral image, use random PCA (randomized PCA, R-PCA) to keep 10 or 30 principal component channels, and then use 2D-CNN for classification. Since 2D-CNN only performs convolution operations in space and simple linear operations in the spectral dimension, the obvious disadvantages of this type of method is that it will cause the loss of spectral data, which will affect the recognition results [27].

Differing from 2D-CNN, the 3D-CNN convolution structure is shown in Figure 3, where N and M are the plane of the convolution kernel in the convolution layer and spectral dimensions, and L is the number of output channels of the convolution layer. The model can be described as follows:

$$V_{1j}^{xyz} = f\left(\sum_m \sum_{h=0}^{H_{1-1}} \sum_{w=0}^{W_{1-1}} \sum_{r=0}^{R_{1-1}} K_{1jm}^{hwr} V_{(1-1)m}^{(x+h)(y+w)(z+r)} + b_{1j} \right),$$

(2)

where R_{1-1} is the spectral dimension of convolution kernel.

The 3D-CNN algorithm, which has one more convolution kernel dimension R_{1-1} than the 2D-CNN, can solve the above problem because it has the following advantages:

(1) The input image is the raw hyperspectral image, without the need to use the dimension reduction method, and the image space and spectral features are preserved.

(2) The extracted features are more representatives of hyperspectral images. 3D-CNN is different from 2D-CNN. Instead of plane convolution, it performs convolution operations in both spatial and spectral dimensions to extract the features of the "spectral" combination of hyperspectral images. It makes full use of spectral and spatial 3D correlation information instead of just their separate and independent feature information.

(3) It can avoid introducing excessive parameters, prevent overfitting, and improve computing efficiency. Assuming that the size of the convolution kernel is 3, the number of hyperspectral channels is 200, and the number of output channels is 32, the first 2D-CNN operation requires $3 \times 3 \times 200 \times 32 = 57600$ parameters, and 3D-CNN operation requires $3 \times 3 \times 3 \times 1 \times 32 = 864$ parameters.

Therefore, compared with 2D-CNN, 3D-CNN is more suitable for hyperspectral image processing tasks. However, as the network structure deepens, the vanishing gradient problem will appear, which can affect the training effect of deep neural networks, so introducing the residual error structure is particularly critical.

3.2. Residual 3D-CNN Structure. In deep learning, the deeper the network structure, the more accurate the extracted features and the better the classification results. However, as the network structure continues to deepen, gradients will diffuse or explode during the backpropagation process, resulting in bad effect on network training. After the residual structure is proposed, due to the existence of shortcuts, the gradient is more easily and effectively propagated, which is good to solve the problem. In order to build a deeper network structure, this paper also introduces residuals into 3D-CNN and designs the residual 3D convolution structure block.

According to the design rules for the size of the convolution kernel in 2D-CNN, several consecutive 3×3 convolution kernels have the same field of view as the large convolution kernel and contain fewer parameters and fewer more complex nonlinear features. Research results show that
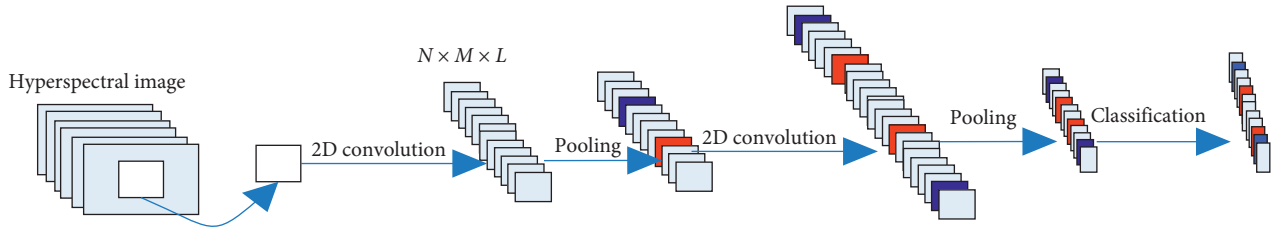
FIGURE 3: The structure of 3D-CNN.

the $3 \times 3 \times 3$ small convolution kernel is the optimal choice for the spatiotemporal feature learning of video input. In addition, many algorithms for CT 3D image detection also use $3 \times 3 \times 3$ convolution kernels and have achieved good results. Because hyperspectral images and video and the CT images have plane image information and similar 3D data structures, as a reference, this paper designs all the convolution kernel structures used for spectral feature extraction in the network to the size of $3 \times 3 \times 3$.

The residual convolutional structure block is shown in Figure 4. In the residual structure of this paper, there are two forms of shortcut, one is the identity residual block, whose input and output dimensions remain the same, as shown in Figure 4(a). The other is the convolutional residual block, which has different input and output dimensions. The purpose of the design is to change the number of channels. The shortcut of the convolutional form uses $l \times l \times l$ convolution kernel, which will not introduce a large number of parameters, as shown in Figure 4(b). The deepening or complication of the network structure will necessarily introduce some additional hyperparameters, such as the size of the convolution kernel of each convolutional layer and the number of channels, so these hyperparameters need to be selected more reasonably.

In order to improve the calculation efficiency, the network does not directly perform a convolution operation with the size of 3 on each convolutional layer input but uses a bottleneck structure, which will effectively reduce the number of parameters and computational complexity. Assume that there are 256 features as inputs, and if only $3 \times 3 \times 3$ convolution operations are performed, $256 \times 3 \times 3 \times 3 \times 256 = 1769472$ convolution operations must be performed; if the bottleneck structure is adopted, then only $(256 \times l \times l \times l \times 64) + (64 \times 3 \times 3 \times 3 \times 64) + (64 \times 1 \times 1 \times 1 \times 256) = 143360$ convolution operations are performed. The bottleneck structure is used in NIN, GoogleNet, and ResNet [12, 28]. This structure can effectively reduce the computational complexity and enhance the nonlinear expression ability of the network to a certain extent.

In addition, a batch normalization layer (BN) is introduced after each convolutional layer. BN can effectively prevent vanishing gradient and gradient explosion. Although it introduces additional calculations, it can make the overall convergence rate of the model faster. It is worth noting that the network uses ELU (exponential linear units) instead of ReLU (rectified linear unit) as the nonlinear activation function. Although the ReLU function has very good characteristics and is widely used, when its existence input is negative, the derivative will become 0 and no longer change, which will lead to the problem that neurons die and will never be activated. To solve this problem, the ELU function presents a "Soft saturation" state at the part of less than 0, making the derivative not become 0, thus keeping the neuron alive [29].

4. Detection and Classification Based on 3D Deep Residual Model

The input of the network is a 3D data matrix in 3D-CNN, which is obtained by taking a pixel in the original image as the center and its size as $S \times S \times L$, where L is the number of hyperspectral image channels and S is the size of plane dimension. However, the amount of calculation and recognition accuracy introduced by different sizes of the visual field range are also different. According the tradeoff among the accuracy rate, operation efficiency, and other factors, this paper finally fixed the dimension to 7×7 in the multispectral image.

As we all know, the deep learning model has the two optimization tasks. One is the optimization of internal parameters, such as the allocation of weights in neural networks; the other is the optimization of hyperparameters, such as the structural parameters and learning rate of neural networks. The optimization of hyperparameters has always been a difficult point in deep learning, such as the number of channels and the size of the convolution kernel in equation (2); in addition, there are also choices for weight initialization methods, regularization methods, and different training methods. Setting these parameters requires rich training experience, professional knowledge, and a large number of experiments. Therefore, the TPE algorithm is introduced for adaptive hyperparameter optimization, which is used to quickly select the suitable hyperparameters, and it is more time saving and labor saving compared to manually adjusting the hyperparameters. In addition, the training effect is also better.

It is assumed that $\lambda_1, \lambda_2, \cdots, \lambda_n$ represent the hyperparameters selected in the model; $\Lambda_1, \Lambda_2, \cdots, \Lambda_n$ represent the selection domain of each hyperparameter; then, the hyperparameter selection domain space of the model is defined as $\Lambda = \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_n$. When k-fold cross-validation method is used for hyperparameter $\lambda \in \Lambda$, the optimization problem of hyperparameters can be expressed as the follows:
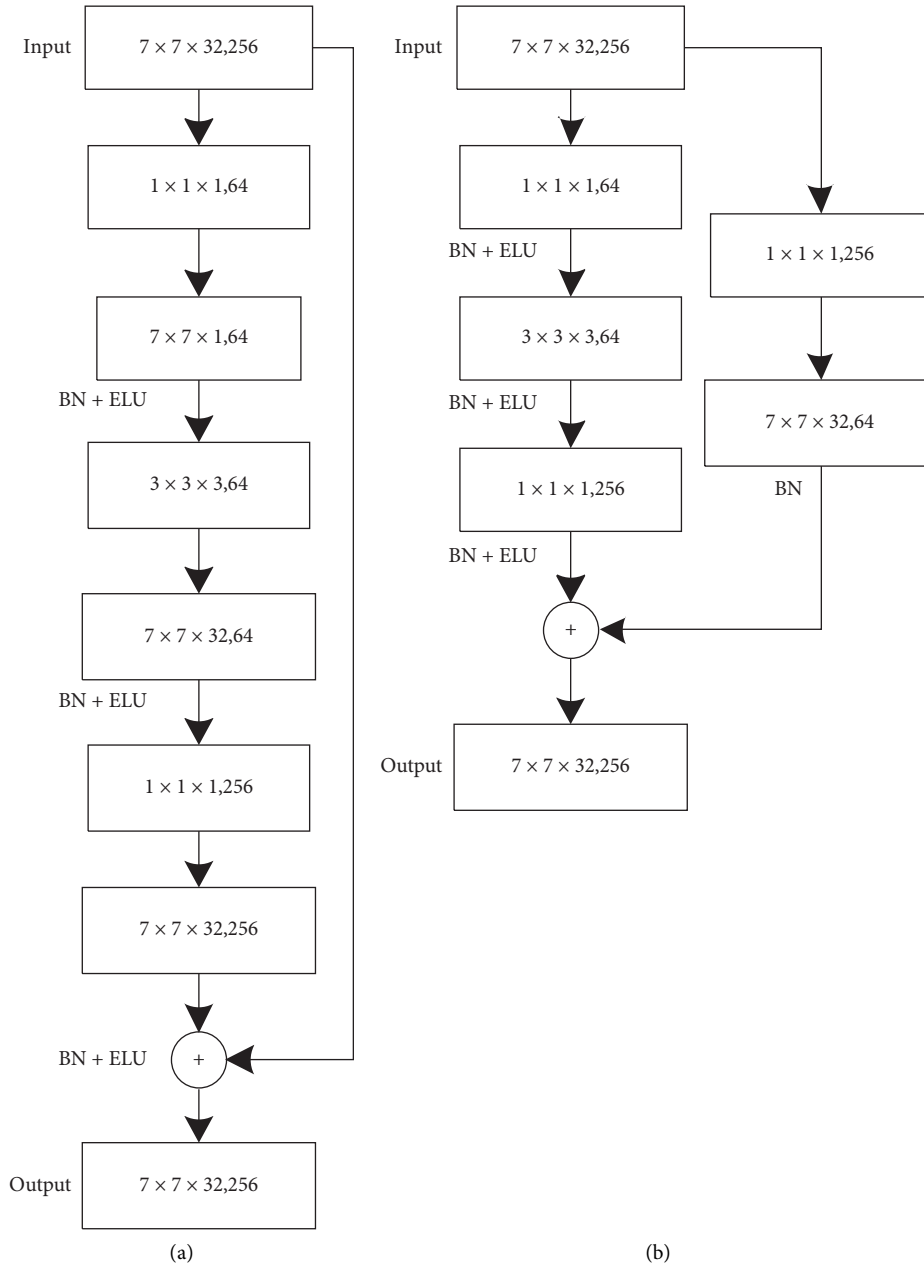
FIGURE 4: Residual convolutional structure. (a) Identity residual block; (b) convolutional residual block.

$$f(\lambda) = \frac{1}{k} \sum_{i=1}^{k} L\left(\lambda, D_{train}^{i}, D_{validation}^{i}\right), \qquad (3)$$

where $L(\cdot)$ is the loss function in training and D_{train} and $D_{validation}$ are denoted as samples in the training set and validation set, respectively.

Recently, the most commonly used hyperparameter optimization methods are still manual search and grid search (violent search), but their efficiency is extremely low, so hyperparameter optimization has always been a very tedious process.

The TPE algorithm is a sequential model-based global optimization algorithm (Smoa). The Smoa algorithm uses the previous hyperparameters to recommend the next hyperparameters through optimization criteria. Different Smoa algorithms use different optimization criteria. TPE algorithm takes expected improvement (EI) as optimization criterion. After each iteration, the algorithm returns the hyperparameter selection of the best EI. In this way, by continuously recommending hyperparameters with the best EI standard, the algorithm can find the optimal hyperparameter faster than grid search. Compared to the random forest algorithm, TPE adopts 2 probability distributions to simulate the posterior probability, which has better modeling strategies and advantages in hyperparameter optimization. The types of hyperparameters can be integers and

continuous real numbers, for example, the number of neurons uses integers and the dropout ratio uses continuous real numbers, and the optimization method of the classifier can use SGD, RMSProp, Adam, etc.

4.1. Structure of Our Model. The network input first is proposed by a convolution layer with the convolution kernel of $l \times l \times 7$ and the step size of $l \times l \times 2$ and a maximum pooling layer with the kernel of $l \times l \times 3$ and the step size of $l \times l \times 2$. The purpose is to reduce the number of channels and improve the operation efficiency. Then, two groups of residual structural units are designed, where each unit is composed of two convolutional residual structural blocks. The first group of residual structural unit is set to $l \times l \times 3$, whose purpose is to extract and fuse spectral features; the second group of residual structural unit is set to $3 \times 3 \times 3$, which is used to extract the spectral features of the hyperspectral image. Finally, a $7 \times 7 \times 1$ global pooling layer and a fully connected layer (FC) are used for classification; each hidden layer uses the strategy in the literature [28] to initialize the convolution kernel parameters and regularize the specification term L_2 of 1×10^{-4}. The activation function is expressed as the exponential linear unit, and the Adam optimizer is selected to train our model in experiment.

4.2. Training Process and Algorithm Framework. According to the structure of hyperparameters which is manually initialized, a search space of hyperparameters is defined for automatic adjustment. There are nearly 10,000 possibilities in the search space. The algorithm and TPE algorithm use the same dataset to search for 50 iterations. 100 epochs are used in training operation. Finally, their recognition accuracy rate is obtained. The hyperparameter with the highest accuracy rate is selected as the hyperparameter of the network.

In this paper, the Softmax layer is used as a classifier. Because it is superior to other classifiers such as support vector machine (SVM) when dealing with multiclassification problems, it has a wide application in deep learning. Its function is defined as follows:

$$S_i = \frac{e^{V_1}}{\sum\limits_{i}^{c} e^{V_1}}, \tag{4}$$

where V_i is the output value of classifier in class i; c is the number of class; and S_i is the relative probability.

The algorithm calculates the relative probability for the output value of each class, and the class with the highest relative probability is the classification results.

For pixel-level classification in hyperspectral images, the overall steps can be divided into 3 steps:

Step 1: a patch region with a size of $7 \times 7 \times L$ from the hyperspectral image is extracted as the network input, and the class label of the central pixel is extracted as the object class, where L is the number of channels of the original hyperspectral image.

Step 2: the basic structure of feature extraction is our improved 3D residual convolution structure, and its schematic diagram is shown in Figure 4. The TPE algorithm is adopted to optimize hyperparameters, which can realize end-to-end hyperspectral "spectrum" feature extraction.

Step 3: the network is trained using crossentropy loss and backpropagation; finally, the detection and classification results are obtained. The Softmax layer turns the output of the deep network into a probability distribution, where the distance between the predicted probability distribution and the real probability distribution can be calculated by crossentropy.

5. Experiment Results and Analysis

5.1. Hyperspectral Curve Analysis. As we all know, there is noise interference between the mildew region and the sound region of blueberry in the wavelength range of 400–450 nm. In order to not affect the accuracy of subsequent detection, the spectral data of this waveband range is removed. In addition, the spectral reflectance of the blueberry mildew area in the visible band (450–760 nm) is slightly higher than that of the sound area. In the near infrared band (760–1000 nm), the spectral reflectance of the sound region is higher than that of the mildew region [30]. The reason for the difference of spectral reflectance between the blueberry mildew area and sound area is that the color of the blueberry mildew area is slightly different from that of sound area, and the main components and physical and chemical properties of the blueberry mildew area are changed due to the decay of blueberry disease so that the spectral reflectance is changed. Therefore, the spectral data of 450–1000 nm range were used to establish a training and testing dataset so as to detect the mildewed blueberry. The Hyperspectral Imaging System is used to collect spectral images and is shown in Figure 5.

5.2. Dataset. Training a deep learning network requires a large number of image samples, but the collected blueberry data is often insufficient in practical application. In order to obtain more data so that the deep learning model has strong generalization ability, the obtained blueberry hyperspectral images are expanded. The MATLAB software was used to perform angle rotation, scale transformation, mirror transformation, and adding noise to expand the number of obtained images. Finally, the image is reshaped to the same size 256×256. These images are divided into the training set and the testing set, whose number is shown in Table 1.

5.3. Parameter Setting. The network parameter settings proposed in this paper are as follows: depth = 40, growth_rate = 12, bottleneck = True, reduction = 0.5, batch size is set to 16, learning rate is set to 0.001, and maximum number of iterations is set to 10,000 times; in order to improve optimization efficiency, the ADAMDAM optimization algorithm is adopted. This optimization method is performed using an improved stochastic gradient descent
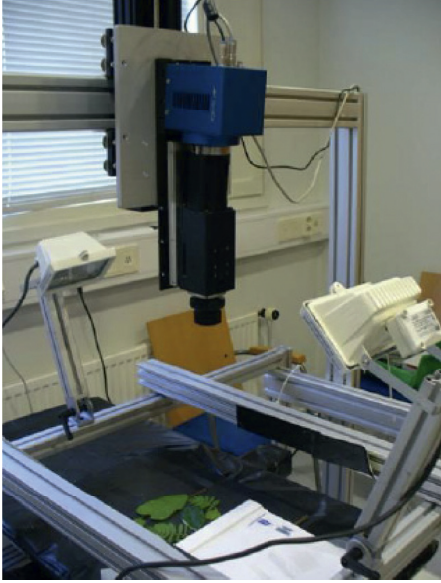
FIGURE 5: Hyperspectral Imaging System for blueberry classification.

TABLE 1: Dataset of blueberry on different situations.

	Sound	Slight-decayed	Moderate-decayed	Severe-decayed
Training	15820	7821	3575	6951
Testing	526	425	358	398
Total	16346	8246	3933	7349

algorithm, which can iteratively update the neural network weights based on the training data.

The input of the network is a 3D data matrix with the size of $S \times S \times L$, where L is the number of hyperspectral image channels and S is the field of view. The computation complexity has very close relation with the size of fields of view, so its size needs to be further experimentally determined.

In order to verify the generalization ability of our algorithm, all datasets are divided into three parts: dataset 1, dataset 2, and dataset 3. Figure 6 shows the accuracy of running 10 epochs on different blueberry hyperspectral datasets with different input sizes S. It can be found that the larger the input size, the faster the accuracy rate of the algorithm rises before 3 epochs and the faster the model can converge. The time taken to train 10 epochs and the time spent on testing with different input sizes are shown in Figure 6. It can be found that the larger the size of input, the longer the training time spends. Since the input of larger size converges faster than the input of smaller size, it also requires more training and testing time. Therefore, according the tradeoff between the recognition performance and calculation efficiency, the input size of the hyperspectral image is fixed as to $7 \times 7 \times L$.

5.4. Quantitative Evaluation Indexes. In order to evaluate the performance of our proposed model, the FPPI is adopted

as evaluation standard, which focuses on the frequency of occurrence of FP (False Positive). For the mildew detection rectangle obtained for each image, the evaluation criteria used in this paper are Detection Rate (DR) and False Positive Per Image (FPPI), and the relationship is as follows:

$$DR = \frac{TP}{(TP + FN)}, \tag{5}$$

$$FPPI = \frac{FP}{(FP + TN)}, \tag{6}$$

where TP represents the number of positive samples detected correctly; TP + FN represents the number of all positive samples in the picture; and FP + TN represents the number of false positives. In addition, overall accuracy (OA), average accuracy (AA), and kappa coefficient(K) are also selected as quantitative evaluation indexes [31].

5.5. Qualitative and Quantitative Comparison Analysis. In order to better verify the performance of our proposed algorithm in this paper, AlexNet [32], GoogleNet [33], 3D-CNN [34], and ResNet [35] are selected as comparison models; the accuracy of the four algorithms is given from the corresponding paper and open source, and the accuracy is provided by 5 independent tests. The overall classification accuracy is the ratio between the prediction accuracy and the total number on all test sets. The average classification accuracy is the ratio between the correct prediction of each class and the total number of each class, and finally the average value of all class accuracy is taken; kappa coefficient represents the proportion of error reduction, and its calculation is based on the confusion matrix.

In this paper, the neural network AlexNet has 8 layers; the first 5 layers of the convolution layer extract the image features and use the pooling layer to reduce the dimension of the image features; multiple convolutions make the image features become more abstract from the concrete, which can better characterize hyperspectral images. As shown in Table 2, with the increase of the number of iterations, the accuracy of the network has been increasing to 100%. In fact, due to the lack of hyperspectral blueberry image, there is an overfitting situation in the process of training. The overfitting will cause all moderate-decayed blueberries to be classified into severe-decayed blueberries when the trained network is adopted to classify sound, slight-decayed, moderate-decayed, and severe-decayed blueberries.

When the number of iterations of the network reaches 200, the fitness of the training model is not very high. When classifying the blueberry hyperspectral images, the network cannot classify the blueberry correctly. When the sound blueberry hyperspectral images are input into the network for recognition after the network training is completed, more than 50% blueberries are classified as sound and more than 40% are classified as decayed blueberries, but the sound probability is greater than the decayed probability, so it can be judged as sound conditions, and the purpose of accurate classification can be achieved.
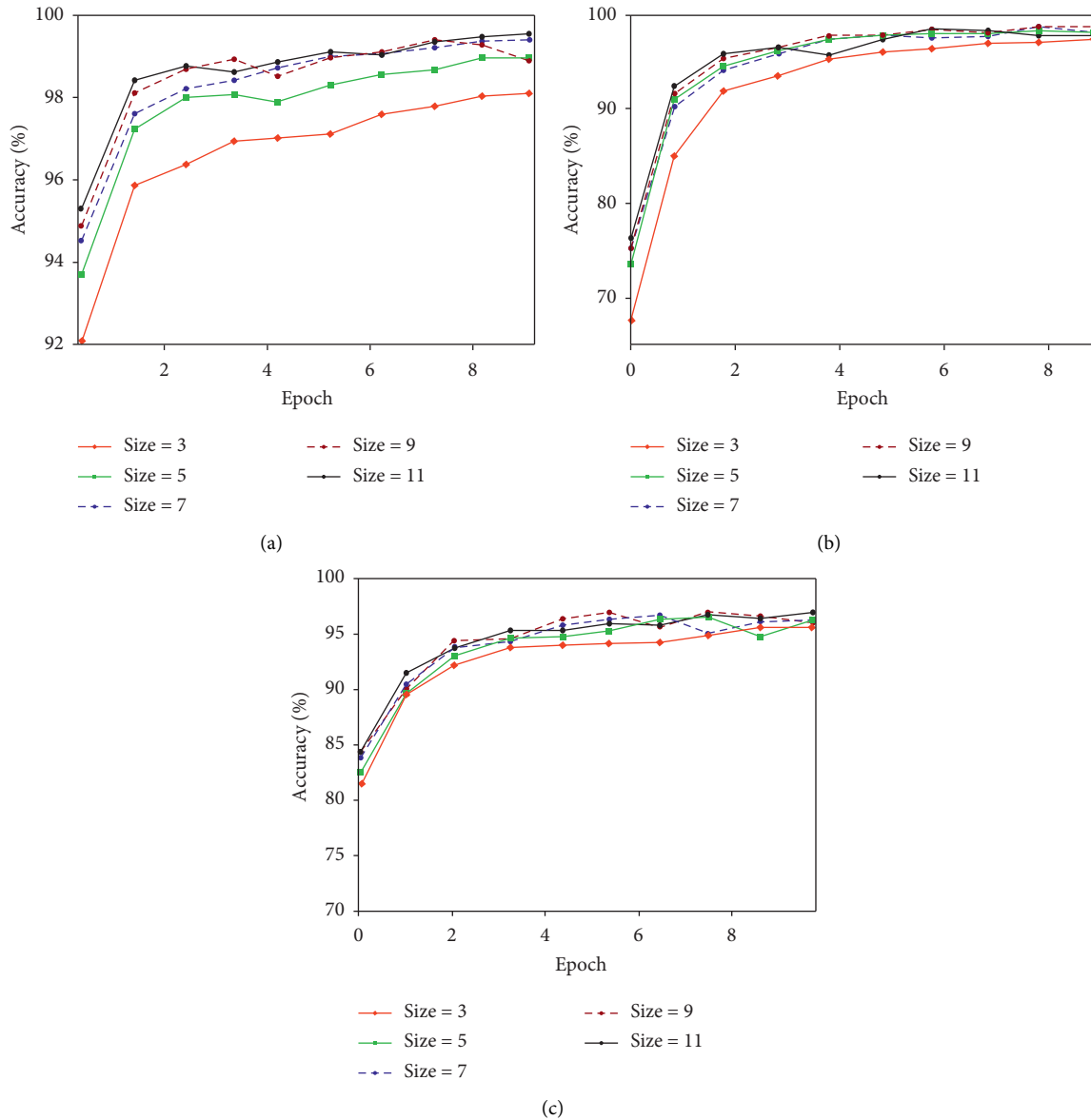
(a)

(b)

(c)

FIGURE 6: Accuracy of different epochs on different datasets. (a) Dataset 1; (b) dataset 2; (c) dataset 3.

CaffeNet also has 8 layers. The output of each layer is the input of the next layer. The data format has four dimensions in each spectral layer; the first dimension is the number of images, the second dimension is the number of channels, and the third and fourth dimensions are the width and height of images. In deep learning, loss function is often nonconvex, and there is no analytical solution, which needs to be solved by the optimization method. In this paper, the forward algorithm and backward algorithm are called alternately to update the parameters so as to reduce the loss value as much as possible and finally get the local optimal solution. In the process of network iteration, 10-fold crossvalidation is used to verify the performance. It can be seen from this that the accuracy of the network increases rapidly, and the network tends to converge in the process of training and finally reaches 100%. However, due to the lack of data, the increasing number of iterations will lead to overfitting. Because of the huge parameters of the network in the process of overfitting, the data fitting results of the training set are good, but the prediction results of the samples outside the dataset are very poor, where there is a great probability of classification errors. ResNet uses the residual neural network to perform nondestructive detection of blueberries. The detection accuracy rate is up to 90%, and the effect is better. The texture features of the sound blueberry image are obviously different with moderate-decayed and severe-decayed blueberries. It is easy to identify the mildew blueberries using ResNet technology, and the detection effect on the slight-decayed blueberry is poor. The proposed model in this paper is an improved 3D-CNN method for nondestructive detection of blueberries, and its four types of blueberries have better classification performance. Table 3 shows the accuracy under different comparison models. Our proposed algorithm obtains the

TABLE 2: Classification accuracy under different number of iterations.

Number of iteration	Accuracy	Prediction probability
100	0.81	0.84
200	0.87	0.90
300	0.92	0.95
400	0.93	0.96
500	0.95	0.98
1000	0.95	1
1500	0.99	1
2000	0.97	1
3000	0.95	1

TABLE 3: Accuracy under different comparison models.

Class	AlexNet	GoogleNet	3D-CNN	ResNet	Proposed
Sound	85.3	83.4	82.7	97.2	98.25
Slight-decayed	96.3	94.2	89.7	83.1	95.1
Moderate-decayed	69.6	70.4	62.0	84.4	93.6
Severe-decayed	58.0	66.9	65.5	84.4	89.4
OA (%)	77.8	81.9	80.6	85.6	95.2
AA (%)	61.3	64.2	68.3	79.8	91.5
K (%)	74.5	79.3	77.9	84.6	94.6

best classification results, which is 17.2%, 20.2%, and 19.8% higher than GoogleNet in OA, AA, and kappa coefficients, respectively. Compared with GoogleNet, our proposed algorithm greatly improves the classification accuracy. Compared with ResNet, our indicators increased by 13%, 14.4%, and 9.6%, respectively. In other words, our proposed algorithm has the best OA, AA, and kappa coefficients.

In order to analyze the blueberry mold recognition performance of the algorithm proposed in this paper, Figure 7 shows the relationship between the detection rate and FPPI (False Positives per Image). Table 4 is the prediction probability of the blueberries in testing set, which is verified by different models. It can be seen from the experimental results that there is an overfitting situation in GoogleNet and AlexNet. Because the GoogleNet network reaches 22 layers, it can learn a lot of features at the same time, but the amount of training samples in this experiment is relatively small, which also leads to overfitting during network training. Both the ResNet network and the proposed network can accurately identify the decay of blueberry hyperspectral images, but the accuracy of ResNet is not as good as the proposed algorithm in this paper. When FPPI = 1, the detection rate of the proposed detection algorithm is 96.69%, and the best result of the comparison algorithm is the ResNet algorithm, the result is 95.42%, while the detection rates of the GoogleNet, AlexNet, and 3D-CNN are 89.12%, 91.88%, and 92.15%.

5.6. Generalization Performance. This paper tests the classification effect of the trained network on different datasets to verify the generalization ability of the model.
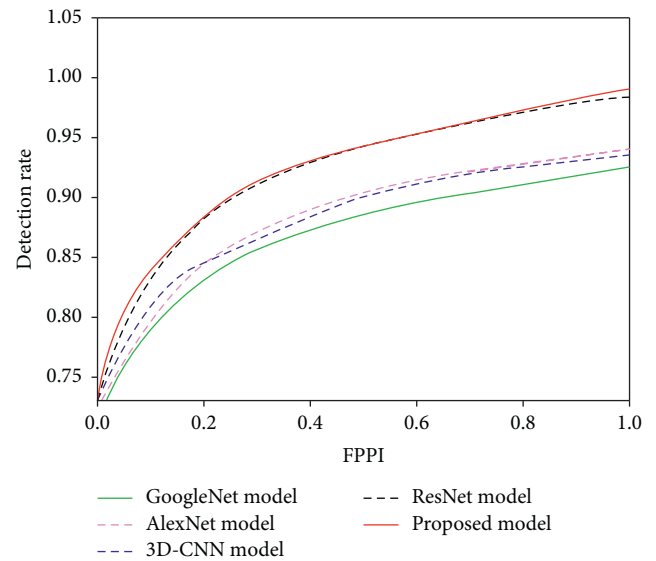

FIGURE 7: Performance comparison in the database.

This paper uses the model trained on blueberry dataset 1 to classify dataset 2, and dataset 3, respectively. The classification layer is different, so the transfer training method is used to replace the classification part of the network model and fine-tune. The parameters of other parts of the network are not updated. The dataset is still divided into 20% training, 10% verification, and 70% test samples. Experiment results are shown in Table 5. It can be found that the hyperspectral classification model has a high accuracy rate for blueberries, which proves that its "spatial spectrum" feature extraction part has a certain generalization ability.

TABLE 4: Detection rate of different models when FPPI = 1.

Models	AlexNet (%)	GoogleNet (%)	3D-CNN (%)	ResNet (%)	Proposed (%)
Detection rate	89.12	91.88	92.15	95.42	96.69

TABLE 5: Comparison for generalization ability of the model.

Indexes	Dataset		
	1	2	3
OA (%)	98.37	96.35	95.31
AA (%)	97.85	92.61	92.09
K (%)	98.29	96.25	95.22

6. Conclusions

An improved deep residual 3D convolutional neural network (3D-CNN) framework is proposed for hyperspectral images classification so as to realize fast training, classification, and parameter optimization. Rich spectral and spatial features can be rapidly extracted from samples of complete hyperspectral images using our proposed network. This combines the tree structured Parzen estimator (TPE) adaptively and selects the super parameters to optimize the network performance. In addition, aiming at the problem of few samples, this paper proposes a novel strategy to enhance the hyperspectral image sample data, which can improve the training effect. Experimental results on the standard hyperspectral blueberries datasets show that the proposed framework improves the classification accuracy compared with AlexNet and GoogleNet. In addition, our proposed network reduces the number of parameters by half and the training time by about 10%.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] Z. Wang, M. Hu, and G. Zhai, "Application of deep learning architectures for accurate and rapid detection of internal mechanical damage of blueberry using hyperspectral transmittance data," *Sensors*, vol. 18, no. 4, pp. 1126–1135, 2018.

[2] R. Lab, "Machine learning in bioinformatics," *Briefings in Bioinformatics*, vol. 7, no. 1, pp. 86–112, 2009.

[3] D. Lorente, "Selection of optimal wavelength features for decay detection in citrus fruit using the ROC curve and neural networks," *Food and Bioprocess Technology*, vol. 6, no. 2, pp. 530–541, 2013.

[4] S. Bargoti and J. P. Underwood, "Image segmentation for fruit detection and yield estimation in Apple orchards," *Journal of Field Robotics*, vol. 34, no. 6, pp. 1039–1060, 2017.

[5] C. Nandi, B. Tudu, and C. Koley, "A machine vision technique for grading of harvested mangoes based on maturity and quality," *IEEE Sensors Journal*, vol. 34, p. 1, 2016.

[6] J. D. Tygar, "Adversarial machine learning. " internet computing," *IEEE*, vol. 15, no. 5, pp. 4–6, 2011.

[7] N. El-Bendary, "Using machine learning techniques for evaluating tomato ripeness," *Expert Systems with Applications*, vol. 42, no. 4, pp. 1892–1905, 2015.

[8] O. Gupta, "Machine learning approaches for large scale classification of produce," *Scientific Reports*, vol. 8, no. 1, p. 5226, 2018.

[9] M. Rahnemoonfar and C. Sheppard, "Deep count: fruit counting based on deep simulated learning," *Sensors*, vol. 17, no. 4, p. 905, 2017.

[10] M. S. Hossain, M. H. Al-Hammadi, and G. Muhammad, "Automatic fruits classification using deep learning for industrial applications," *IEEE Transactions on Industrial Informatics*, vol. 45, p. 1, 2018.

[11] K. S. Lewers, Y. Luo, and B. T. Vinyard, "Evaluating strawberry breeding selections for postharvest fruit decay," *Euphytica*, vol. 186, no. 2, pp. 539–555, 2012.

[12] S. Cao, "Effect of ultrasound treatment on fruit decay and quality maintenance in strawberry after harvest," *Food Control*, vol. 21, no. 4, pp. 0–532, 2010.

[13] D. Lorente, "Comparison of ROC feature selection method for the detection of decay in citrus fruit using hyperspectral images," *Food & Bioprocess Technology*, vol. 6, no. 12, pp. 3613–3619, 2013.

[14] B. Wang, J. Xue, and S. Zhang, "Detection of decay and disease pear jujube based on hyperspectral imaging technology," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 7, no. 6, pp. 2094–2107, 2013.

[15] P. Liu, H. Zhang, and K. B Eom, "Active deep learning for classification of hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 45, pp. 1–13, 2013.

[16] X. Zhou, "Deep learning with grouped features for spatial spectral classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 1, pp. 97–101, 2017.

[17] X. Yang, "Hyperspectral image classification with deep learning models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 99, pp. 1–16, 2018.

[18] H. Huang, "Spatial-spectral feature extraction of hyperspectral image based on deep learning," *Laser & Optoelectronics Progress*, vol. 54, no. 10, p. 101001, 2017.

[19] S. Liu and Q. Shi, "Multitask deep learning with spectral knowledge for hyperspectral image classification," 2019.

[20] D. Zhang, "Hyperspectral image classification using spatial and edge features based on deep learning," *International Journal of Pattern Recognition & Artificial Intelligence*, vol. 45, 2018.

[21] Y. Chen, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[22] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach," *IEEE Transactions on Geoscience & Remote Sensing*, vol. 54, no. 8, pp. 4544–4554, 2016.

[23] J. Yue, S. Mao, and M. Li, "A deep learning framework for hyperspectral image classification using spatial pyramid pooling," *Remote Sensing Letters*, vol. 7, pp. 7–9, 2016.

[24] H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 99, p. 1, 2017.

[25] W. Zhao, "On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery," *International Journal of Remote Sensing*, vol. 36, no. 13, pp. 3368–3379, 2015.

[26] R. Venkatesan and S. Prabu, "Hyperspectral image features classification using deep learning recurrent neural networks," *Journal of Medical Systems*, vol. 43, p. 7, 2019.

[27] F. Bei, "Semi-supervised deep learning classification for hyperspectral image based on dual-strategy sample selection," *Remote Sensing*, vol. 10, no. 4, pp. 574–585, 2018.

[28] C. Shi and S. Pun, "Multi-scale hierarchical recurrent neural networks for hyperspectral image classification," *Neurocomputing*, vol. 10, 2018.

[29] D. Adam, "Bull. "Convergence rates of efficient global optimization algorithms," *Journal of Machine Learning Research*, vol. 12, no. 10, pp. 3-4, 2011.

[30] S. Yang, "Fuzzy signature-based discriminative subspace projection for hyperspectral data classification," *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, vol. 9, pp. 4196–4202, 2011.

[31] E. Xie, "Unsupervised hyperspectral feature selection based on fuzzy c-means and grey wolf optimizer," *International Journal of Remote Sensing*, vol. 12, 2019.

[32] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, p. 2, 2012.

[33] C. Szegedy, "Going deeper with convolutions," 2014.

[34] Y. Li, "Adaptive batch normalization for practical domain adaptation," *Pattern Recognition*, vol. 80, 2016.

[35] M. E. Paoletti, "Deep pyramidal residual networks for spectral–spatial hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 740–754, 2019.