

Received April 3, 2022, accepted April 16, 2022, date of publication April 21, 2022, date of current version April 29, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3169159

Improved Residual Network for Automatic Classification Grading of Lettuce Freshness

YANLEI XU¹, YUTING ZHAI, QINGYUAN CHEN, SHUOLIN KONG, AND YANG ZHOU¹

College of Information and Technology, Jilin Agricultural University, Changchun 130118, China

Corresponding author: Yang Zhou (zhouyang@jlau.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 31801753, in part by the Jilin Provincial Science and Technology Department International Exchange and Cooperation Project under Grant 20200801014GH, and in part by the Jilin Province Education Department "13th five-year" Science and Technology Research Planning Project under Grant JJKH20200336KJ.

ABSTRACT To solve the problem of the low efficiency of traditional lettuce freshness classification methods and sample damage, we proposed an automatic lettuce freshness classification method based on improved deep residuals convolutional neural network (Im-ResNet). We built an image acquisition system to obtain the freshness classification dataset of lettuce leaves. For improving the classification accuracy, we developed an image acquisition system for curating the freshness of lettuce leaves. Then, we proposed a novel method that was derived from the existing ResNet-50 (which uses ReLU activation function) known as Improved Residual Networks (Im-ResNet): the new method factored extra convolutional layer, pooling layer, fully-connected layers, and a random ReLU (RRReLU) activation function. We also performed the corresponding experiments using the Im-ResNet network compared with four network architectures (AlexNet, GoogleNet, VGG16 and ResNet50). The experimental results showed that the proposed network had more significant advantages in the recognition accuracy and loss value of lettuce freshness compared with the traditional deep networks. The recognition accuracy of the validation set of the proposed model can reach to 95.60%. Different from the physical and chemical methods, our scheme can automatically and non-destructively classify the freshness of lettuce.

INDEX TERMS Deep learning, residual convolutional neural network, image classification, nondestructive testing, lettuce freshness.

I. INTRODUCTION

Nowadays, lettuce is widely eaten and has become an important part of human diet. This is because lettuce is rich in carotene, antioxidants, vitamins, and other nutrients. In addition, lettuce contains lots of dietary fiber and trace elements, which can enhance the digestion and absorption of protein and fats and improve blood circulation of the stomach and intestines [1]. Generally, the origin of lettuce is far away from major retail centers. The freshness of picked lettuce would decline rapidly due to physiological deterioration and microbial degradation. Especially, it should be noted that the nitrate naturally contained in lettuce would be transformed into nitrite during storage, which is harmful to human health. Therefore, the freshness classification of lettuce are of great significance to consumers [2].

We can conclude from recent researches that the detection methods of vegetable freshness mainly include leaf spectral values, chemical parameter methods, radio frequency (RF)

and electronic nose sensor methods. Xie *et al.* proposed a hyperspectral image detection method for spinach freshness. They found the three wavelengths with the highest recognition rate using genetic algorithm, and constructed the spinach database. Finally, the freshness detection of spinach was completed based on the database [3]. Lunadei et al conducted-on monitoring fresh-cut spinach leaves based on multispectral images [4]. The chemical method is an important mean for detecting the freshness of vegetables. Qiu *et al.* analyzed the correlation of ChlF parameters from three types of leafy vegetables and obtained the freshness and storage time of the vegetable [5]. Moreover, Sensors can also be applied to detect the freshness of vegetables. Le *et al.* designed a radio frequency sensor to monitor the freshness of packaged vegetables based on temperature and humidity [6]. Xu applied electronic nose technology to the study of spinach freshness. They used image and odor information of spinach during storage to detect and analyze the freshness [7]. Huang *et al.* proposed an electronic nose and multi-sensory data fusion technology. Acquiring the image and odor information of the samples using the camera and electronic nose, this method

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea F. Abate¹.

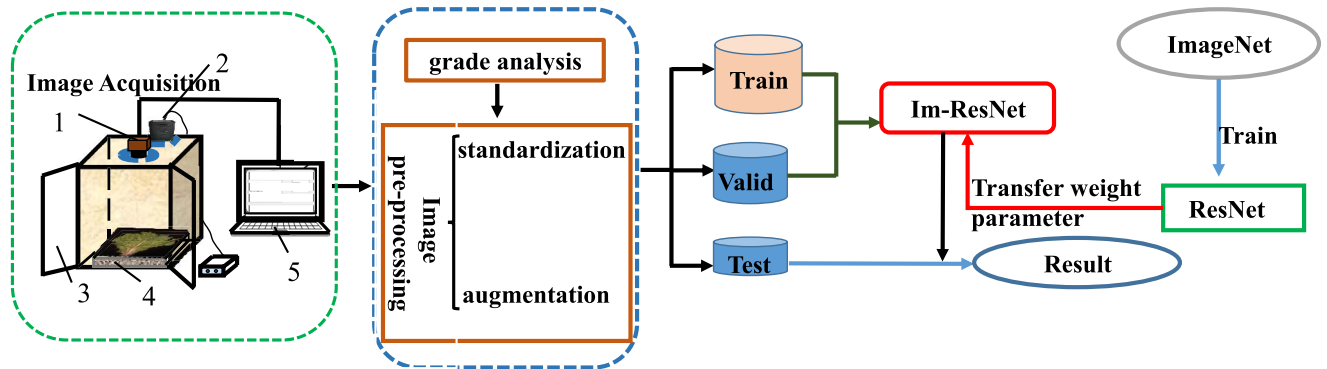


FIGURE 1. The schematic diagram of lettuce freshness recognition. The green dotted box on the left side of the schematic corresponds to the image acquisition system. 1. Camera, 2. Bromine tungsten lamp, 3. Closed camera, 4. Loading platform, 5. Personal computer.

can quickly and non-destructively detect the freshness of spinach during storage [8]. Although the above detection methods all show their own advantages, chemical methods would cause external damage to vegetable leaves, and the cost of equipment such as sensors and spectrometers is expensive, and radio frequency methods need to design complex circuits. Therefore, it is very urgent and meaningful to propose a better alternative for detecting the freshness of vegetable.

With the development of deep learning, it plays the increasingly important role in the field of image recognition. Currently, target detection in images is almost achieved through deep learning methods [9]–[11]. Xia *et al.* constructed a multi-scale input hierarchical fusion convolutional neural network architecture to classify lettuce light stress. The average accuracy rate can reach 89% [12]. Amin *et al.* used VGG16 with a modified classifier to study the defect classification of jujube. The proposed model can achieve an overall classification accuracy of 96.98 % [13]. As we all know, the deep residual convolutional neural network (ResNet) is an excellent deep network, which has been widely studied and used in practice. Arthur *et al.* used ResNet in Tomato's external defect detection. They trained the model separately by training the last layer of ResNet, using feature extraction and fine-tuning, and finally achieved an average recognition accuracy of 94.6% [14]. In addition, Zhou *et al.* used a three-layer adaptive network to replace the fully connected layer of ResNet. Then the improved ResNet50 was used to classify the quality of broccoli heads and showed better performance [15]. Therefore, the deep network model can achieve fruit and vegetable classification for different tasks.

In this work, we first proposed an improved ResNet50 (Im-ResNet) deep network scheme for lettuce freshness classification. Then, the collection of image data task can be completed by low-cost imaging equipment. The proposed model can be trained and validated using the collected data. Finally, the experimental results prove that the classification system based on the deep network has the advantages of high recognition accuracy, low cost, and no damage to lettuce, which shows an attractive application prospect.

II. MATERIALS AND METHODS

The freshness classification method of lettuce proposed in our research consists of five parts: data acquisition, grade analysis, image pre-processing, model construction and model testing, as described in Figure 1. First, the images of lettuce can be acquired from an image acquisition device. Next, we performed freshness grade analysis and image pre-processing after transmitting the lettuce images to the server. The processed data would be divided to three parts: training set, validation set and test set. Then, we loaded the trained ResNet50 weights and parameters on the ImageNet dataset into the newly constructed Im-ResNet model. We trained the model on the training set and validation set to obtain the optimal weights and parameters. Finally, we used the trained Im-ResNet model on test set to obtain the lettuce freshness grade.

A. DATA ACQUISITION

In order to collect experimental data, our research team planted lettuce in the experimental greenhouse of our university. It should be noted that we took glass lettuce as the sample material in our work. The experiment was conducted in two batches from August 25 to October 30, 2019, and from August 30 to October 28, 2020, respectively. Besides, we planted the lettuce at approximately 14×18 cm when it reached 4-6 leaves. Then, the fresh lettuce samples were stored in a constant temperature and humidity equipment (The actual equipment is shown in Figure 2).

In our experiment, we used a Canon CMOS HD SLR camera (aperture value of $f/5.6$, IOS speed of 800, and exposure time of $1/10$) to collect the morphological information of lettuce, because it could achieve the non-invasive and high-throughput data collection. Besides, this camera has the advantages of high resolution, low cost, and small size, which is very suitable for our research. We took out samples three times a day at 8:00 a.m., 4:00 p.m., and 9:00 p.m. to photograph the front and back sides of the lettuce. The photograph action was carried out for 7 consecutive days, and the whole process of lettuce from fresh to rotten was recorded.



FIGURE 2. The storage conditions and environment of lettuce.

The constant black background and lettuce image center are all conducive to deep learning. Therefore, the bottom of the image acquisition device was set as a black background, and the lettuce to be inspected was placed on it. To obtain the comprehensive information of the lettuce samples, we also rotated each lettuce by 90°, 180° and 270°, and collected the information using the camera. After image acquisition, we screened out 1939 JPG lettuce images with a resolution of 2304 × 3456 to construct the lettuce dataset.

Referring to some reports [16]–[19], lettuce moisture would decrease due to the transpiration process, chlorophyll degradation with time, and browning of the roots. Therefore, we considered the changes in the appearance of lettuce in terms of days. Under the help of experts from the College of agriculture in our university, the lettuce leaves were classified into four quality classes. The lettuce images of the four grades were stored in four different folders, labeled 0 folder for grade A, labeled 1 for grade B, labeled 2 for grade C, and labeled 3 for grade D. The appearance and quality standards for each grade are shown in Table 1. It should be noted that grade C was considered the lowest standard for acceptance for consumption or sale, and the lettuce grade D was considered a waste product. The leaf images numbers of the four grades were 438, 586, 415, and 500 respectively.

B. IMAGE PREPROCESSING

Here, in order to facilitate the later model training, the bicubic interpolation method was used to uniformly adjust the image size to 224 × 224 [20]. In addition, the values of the three RGB (red, green, and blue) color channels were normalized with (1). The values of R, G, and B from each channel subtracted the average of all values of that channel, respectively. The differences divided the standard deviation of all values of that channel as the output.

$$X_a = \frac{X - \mu(R, G, B)}{\sigma(R, G, B)} \tag{1}$$

TABLE 1. Appearance and quality standards for each grade.

Grade	Time(day)	Water content (%)	textured	color	Weight(g)
A	1	≥80%	very clear	Bright green, lustrous	11-14g
B	2-3	70%-80%	relatively clear	Green, matte	9-12g
C	4-5	50%-70%	not clear	Dark green, matte	4-8g
D	6-7	≤50%	fuzzy	Black green, matte	1-2g

where X and μ are the input image matrix and the mean value of each channel of the image, respectively. The σ is the standard deviation, and X_a represents the normalized result of X .

Regularization technologies are commonly used to prevent over fitting under the case of insufficient training data. The data augmentation refers to the process of creating new samples similar to the training set, which can be regarded as a regularization technique [21]. Data augmentation can generate more data based on limited amount of data. The robustness of the model can be improved by increasing the number and diversity of samples. In addition, randomly changing the samples can also reduce the dependence on certain attributes and improve the generalization ability of the model. Lettuce images are classified according to the color changes of roots and leaves in the classification process. Therefore, the mirror rotation, random cropping and image sharpening operations are used to increase the dataset. The results of these different augmentation methods and the original dataset are shown in Figure 3. After augmentation operation, the number of images reached 2500 for each of the four categories. We divided the enhanced dataset into training set, validation set and test set and specified 60% of the images as training images, 30% as validation images, and the remaining 10% as test images.

C. IM-RESNET ARCHITECTURE

The depth of convolutional neural network (CNN) plays a major role in the classification accuracy and detection performance. The deeper the CNN, the smaller the classification error. However, the deeper neural networks would lead to difficult training. In general, pre-training CNN mainly includes AlexNet [22], GoogleNet [23], VGG16 [24] and ResNet50 [25]. Lettuce images were inputted four pre-trained models, and the detection results are shown in Table 2. From the training results, ResNet50 has higher accuracy than AlexNet, VGG16, and GoogleNet. The ResNet50 accuracy can reach 88.54% and 86.23% on the training and validation sets, respectively. The ResNet50 network solves the degradation problem by introducing a deep residual learning framework. ResNet50 convolution module and the mapping module contain two forms of shortcuts, which can perform 1 × 1 convolutions and then sum and directly sum.

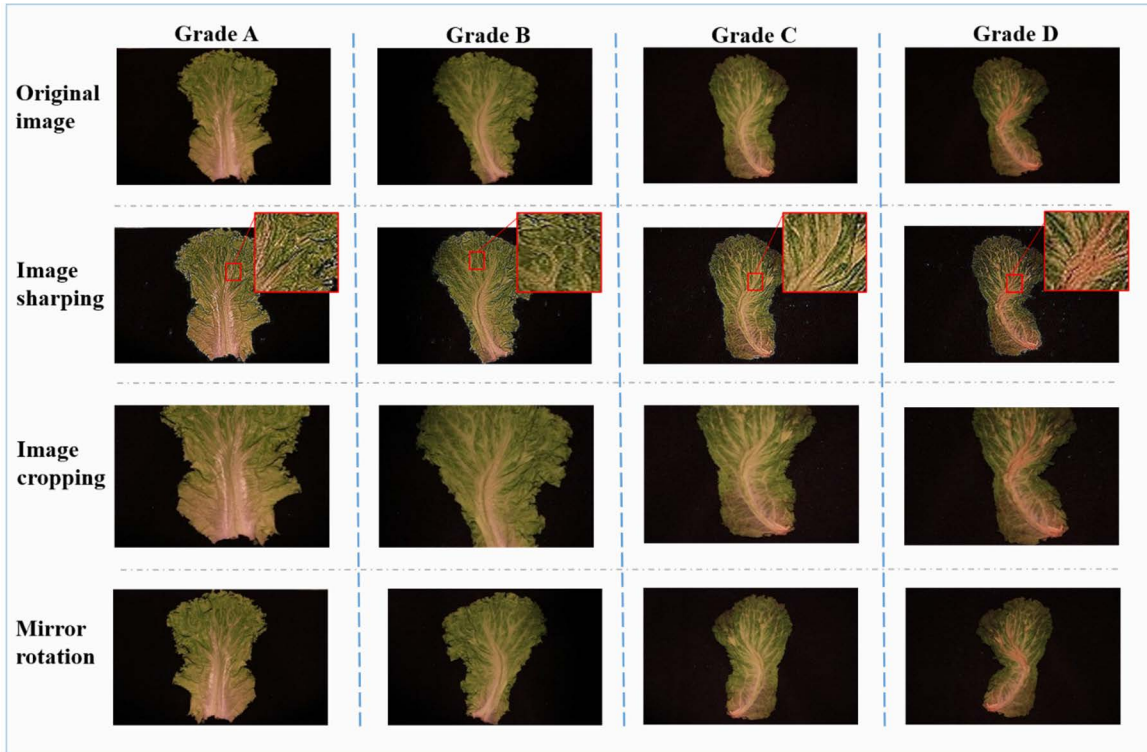


FIGURE 3. The result images corresponding to different augmentation methods.

TABLE 2. The performance comparison of AlexNet, GoogleNet, VGG16 and ResNet50.

Model	AlexNet	VGG16	GoogleNet	ResNet50
Training set accuracy/(%)	82.34	79.6	84.23	88.54
Validation set accuracy/(%)	80.21	78.12	84.5	86.23

This network structure can enhance the ability of extracting features (image texture, shape, and color).

The ResNet50 network relies on only one fully-connected layer after the residual module to achieve the final classification, which is not suitable for inter-species classification such as lettuce freshness grading. Adding a new convolutional layer after the residual module can further refine the lettuce image features to obtain a deeper feature map to achieve accurate classification; the pooling layer can reduce the number of parameters by reducing the feature vector output from the convolutional layer; two fully connected layers increase the number of neurons compared to one fully connected layer, which can improve the non-linear representation and learning ability of the model. Therefore, a new convolutional layer, pooling layer and two fully connected layers can further improve the accuracy of the model. Based on this, we constructed a new network model and named it Im-ResNet. The detailed description of Im-ResNet architecture is shown in Figure 4.

As shown in Figure 4, Im-ResNet model is divided six stages. The network performs a 3×3 zero-padded operation

to maintain the boundary information of an image with an input size of $224 \times 224 \times 3$.

Stage 1: A convolution kernel of size 7×7 slides over the image and performs a dot product at the coverage area with steps size two for feature extraction. As follows,

$$M_2 = \frac{M_1 - F_H + 2P}{S} + 1 \quad (2)$$

where M_1 and M_2 are the sizes of the input image and feature map, respectively. S is the sliding window step, F_H is the convolutional kernel size, and P denotes the boundary fill value.

Stage 2-5 is mainly composed of three blocks, one block of convolution block (Conv Block) and two block of identity block (ID Block).

Conv Block main path includes three convolution layers. Stage 2-5 output feature map sizes are $64 \times 64 \times 256$, $128 \times 128 \times 512$, $256 \times 256 \times 1024$ and $512 \times 512 \times 2048$. The shortcut branch path is a convolution layers with 3×3 convolutional kernel.

ID Block input and output dimensions are same and can be connected in series to deepen the network. The shortcut branch path directly add input to the module.

For deepening the network and decreasing the number of required iterations in the training process, the batch normalization of data is performed after the convolution layer in stage 1-5. In addition, we also construct an activation function Randomized ReLU (RReLU) after batch normalization for increasing the nonlinearity of the neural network model.

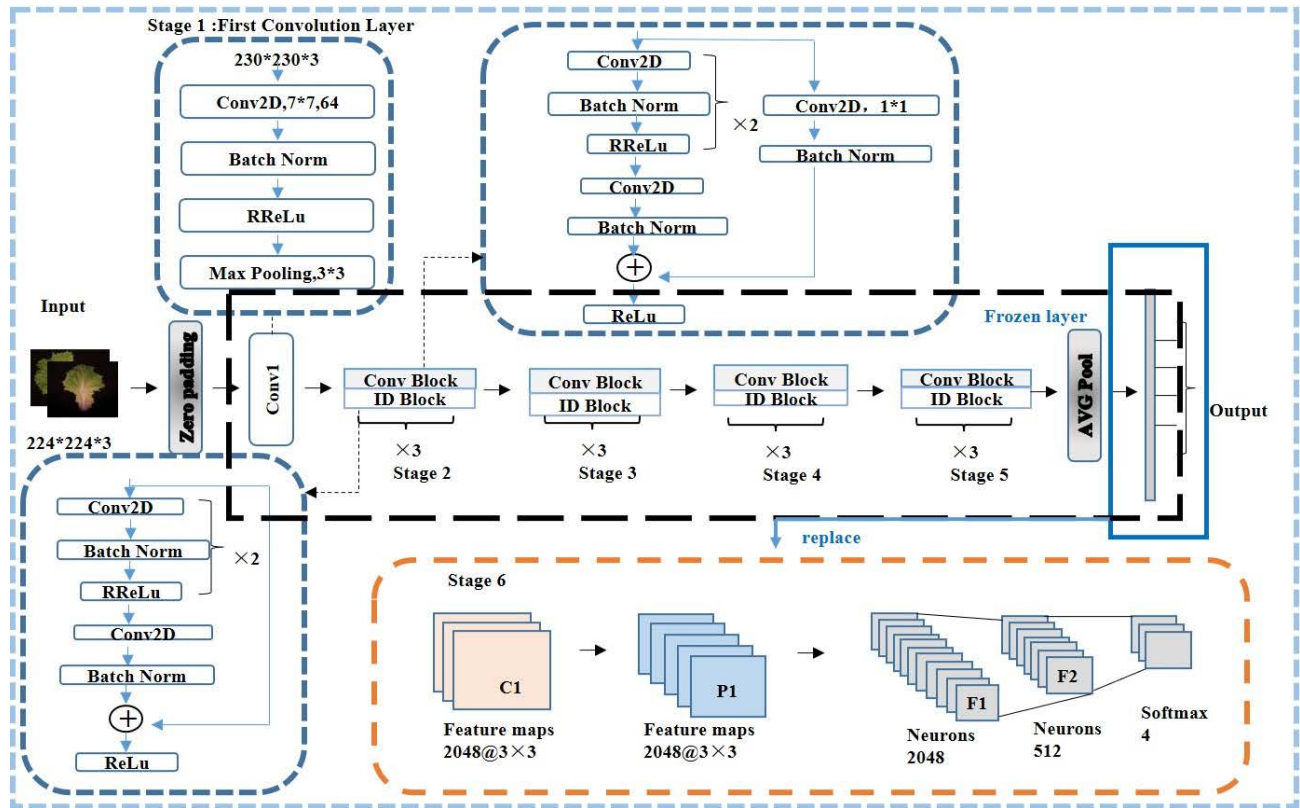


FIGURE 4. The architecture diagram of Im-ResNet.

The advantage of RReLU is its fast convergence speed [26]. This function performs the following mathematical operation on each input data,

$$X_b = \max(aX_a, X_a) = \begin{cases} aX_a, & X_a < 0 \\ X_a, & X_a \geq 0 \end{cases} \quad (3)$$

where a is a random number drawn from the uniform distribution $U(l, u)$, $l < u$ and $l, u \in [0, 1]$. As shown in (3), we can obtain a new image matrix X_b through performing the RReLU function on the image matrix X_a .

After the stage 5, a maximum pooling layer of size 3×3 is connected to calculate the maximum value of a particular feature on the image region. The advantage is to reduce the parameters while retaining the main features.

Stage 6: The newly constructed convolutional layer C1, pooling layer P1 and two fully connected layers F1, F2 are shown in the orange dotted box in Figure 4. Here, the filter size of convolutional layer C1 is 3×3 pixels and S is set to 3. The total number of filters is 2048. In order to avoid the feature maps getting smaller and smaller, a zero-patch operation is used. The pixel size and step size of the largest maximum layer P1 are 3×3 and 1, respectively. Then 2 fully connected layers F1 and F2 are connected, with 2048 and 512 neurons, respectively. Because the output of the fully connected layer is unbounded, the softmax function is added at the end of the network. In this way, the predicted result of the model is

transformed into an exponential function, which ensures the non-negativity of the result. The conversion result is divided by the sum of all conversion results so that the result sums to 1. This result conforms to the nature of probability and is very suitable for multi-classification models. S_c represents the probability value of classifying the image X_b into a certain category [27]. Therefore,

$$S_c = \frac{\exp(a_c)}{\sum_{k=1}^K \exp(a_k)} \quad (4)$$

where $c, k \in \{1, 2, 3, 4\}$, a_c is the softmax input value of the c th category, K is the number of categories for classification, and the value of K in our model is 4. Besides, a_k represents the sum of the softmax input values of the four categories. According to the results of the probability values of the four categories, the one with the largest S_c is taken as the final predicted value. The neurons of the fully connected layer are very important for the output of the last layer. These neurons basically constitute the classification layer of the CNN. Therefore, we use softmax with 4 neurons as the last layer activation function of the deep model to achieve four grades of lettuce.

D. IM-RESNET MODEL TRAINING

In order to improve the classification accuracy, Im-ResNet was trained six times on the training set. First, all stages of the model were frozen, and only C1, P1 and F1 and F2 in

the orange dashed boxes were trained, as shown in Figure 4. Then the last convolutional block was unfrozen to retrain the network, and only stage 5 was unfrozen. This process would last until all the convolutional blocks were unfrozen and trained again along with the classifier. At the last time, the convolutional blocks in stage 1-stage 6 were trained.

However, training deep CNN from scratch is a challenging task, because it requires huge amount of data and time for training. Thus, instead of training the ResNet50 from scratch, this study adopted transfer learning to improve the training efficiency. The major objective of transfer learning is to handle the learning tasks on a target domain by utilizing the knowledge extracted from the source domain, when the labeled data in the target domain are not sufficient [28]. Although the image content of the lettuce dataset in this paper is quite different from that of ImageNet, the abstract low-level features such as edges and textures are invariable universal features. In this case, these pre-trained models are very helpful for similar tasks.

E. EVALUATION INDEX

In order to evaluate the performance of the classifier, we used 4 different metrics in the confusion matrix: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Generally, the higher the accuracy, the better the classifier, but if there is an unbalanced sample, the accuracy cannot be used to evaluate it. Therefore, Precision and Recall can evaluate the classification performance on a certain type of sample. In fact, these two indicators are mutually exclusive. F1-score is to find a balance between Precision and Recall. The calculation equation of the evaluation index is as follows,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

In addition to evaluating the performance of the classifier, we also performed visual analysis on the model. The filters or features extracted by the model from the different layers determined the final model accuracy. One way to evaluate the model performance in extracting features is to evaluate the visual patterns of the filter responses. Therefore, we also utilized the Gradient-weighted Class Activation Mapping (Grad-cam) technique to visualize the used areas of random input images to extract feature for prediction of the class of the images. For deep convolutional neural networks, the image information contained in the fully connected layer and the softmax layer is difficult to display in a visual manner. After multiple convolutions to extract features, the convolutional layer of the last layer has accumulated a large amount and valuable the feature map. In the Grad-cam method, the weights corresponding to the feature maps obtained by

gradient global averaging are calculated as follows [29],

$$a_h^K = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^K}{\partial A_{ij}^h} \quad (9)$$

where Z is the number of feature pixels, K is the number of categories for classification, y^K is the fraction corresponding to category K . Besides, A_{ij}^h represents the pixel value at position (i,j) in the h_{th} feature map A . The obtained a_h^K is the sensitivity of category K relative to the h_{th} channel of the feature map. Finally, these weights are weighted under the restriction of the activation function to obtain the pixel points that have a positive influence on the class K to generate the activation heat map. The experiments and analysis on the activation heat map are described in the subsequent section III.B.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

In the experiment, we used Python 3.6 of the PyTorch library to build and train the Im-ResNet network. The operating system of the server (Intel Xeon E5-2680 v4 CPU, Samsung SSD 860 512G hard disk, Kingston DDR4 64GB memory) was Ubuntu 16.04. Besides, NVIDIA TITAN Xp GPU with 12GB of video memory was used to improve the training speed.

Lettuce data was input into model and optimized using the Adam algorithm [30]. During the training process, batch size, learning rate, L2 regularization penalty and dropout [31] parameters all had a great impact on the performance of deep network models. We compared these parameters in the experiment respectively, setting learning rate as 0.01, batch size as 64, L2 regularization parameter as 0.00001, dropout as 60%.

IV. PERFORMANCE ANALYSIS OF TRAINING MODE OF FROZEN PARTIAL CONVOLUTION

In order to prove the advantages of the proposed model, we trained and validated ResNet50 and Im-ResNet models under the training mode of freezing partial convolution layer (in section II.D). The number of training images and validation images were 6000 and 3000 respectively. It should be noted that we trained Resnet50 model on the training dataset for six times. In every training step, the convolution blocks of the network exited the frozen mode one by one. Accordingly, the Im-ResNet was also trained in the same frozen mode for six times. The final classification results of the two models are shown in Table 3 and Table 4, respectively. Observing the Tabs, Im-ResNet and ResNet50 all could obtain the best accuracy under all unfrozen stages. Besides, Im-ResNet model showed better performance than ResNet50 in the term of accuracy and loss.

The accuracy and loss of different models (Im-ResNet, AlexNet, VGG16, GoogleNet, ResNet50) under various epoch are shown in Figure 5 (A), (B), (C) and (D), respectively. The accuracy and loss of the training set are shown in Figure 5 (A) and Figure 5 (B). As shown in Figure 5 (A), we can find that the accuracy (red curve)

TABLE 3. Comparison of the ResNet50 performance in different condition of convolutional blocks.

Unfrozen	Training set		Validation set	
	Accuracy (%)	Loss	Accuracy (%)	Loss
Fc	92.12	0.41	88.02	0.54
Stage 5+Fc	92.68	0.35	90.99	0.44
Stage 4-5+Fc	93.48	0.29	91.66	0.39
Stage 3-5+Fc	94.34	0.21	92.70	0.35
Stage 2-5+Fc	94.45	0.20	92.70	0.26
Stage 1-5+Fc	95.29	0.12	93.47	0.19

TABLE 4. Comparison of the Im-ResNet performance in different condition of convolutional blocks.

Unfrozen	Training set		Validation set	
	Accuracy (%)	Loss	Accuracy (%)	Loss
C1, P1, F1, F2	92.42	0.31	90.62	0.46
Stage 5+ C1, P1, F1, F2	93.08	0.24	91.99	0.37
Stage 4-5+ C1, P1, F1, F2	93.87	0.19	92.35	0.32
Stage 3-5+ C1, P1, F1, F2	94.44	0.18	92.70	0.33
Stage 2-5+ C1, P1, F1, F2	94.95	0.10	93.75	0.15
Stage 1-5+ C1, P1, F1, F2	95.58	0.08	96.59	0.09

of Im-ResNet is better than other four models. Besides, the accuracy of Im-ResNet would stabilize after the 15th epoch. The Im-ResNet training accuracy reached the highest in the 16th epoch (95.58%), which was 0.29% higher than the accuracy of ResNet50 (95.29%). We also found that ResNet50 loss value was higher than Im-ResNet in the same epoch. Although GoogleNet had improved accuracy by 1.89% and 4.63% compared with AlexNet and VGG16, it was still 11.06% and 11.35% lower than ResNet50 and Im-ResNet, respectively. The two plots Figure 5 (C) and Figure 5 (D) show the accuracy and loss of the validation set corresponding to the five models. Although the accuracy of GoogleNet in the 4th epoch was higher than that of Im-ResNet, the accuracy of GoogleNet fluctuated around 84% and would not increase after the 9th epoch. For Im-ResNet under the 16th epoch, the accuracy of the validation set reached the highest 96.59% with a loss value of 0.09. After 16th epoch, the loss value fluctuated, while the overall performance was still better than that of ResNet50. Therefore, in order to obtain the highest accuracy of the model without over fitting, the training parameters and weights of the 16th epoch were saved.

A. IM-RESNET VISUAL EVALUATION

To further illustrate the feature extraction performance of our model, we showed various filters extracted from different convolutional layers of the proposed model. The partial filters used in the first, middle and last convolutional layers are depicted in Figure 6. We could find that the color features and direction edges were encoded by the first layer of filters. The filter in the middle convolutional layer encoded a simple texture composed of colors and edges. The filter of the last convolutional layer extracted the texture and specific shape of the image. Therefore, we could conclude that the different

filters in our proposed model had the ability to grade lettuce based on image shape, color, and texture.

In addition, the CNN can automatically extract features, and the extracted features can be clearly displayed by visualizing the neural network. Visualization of intermediate activation refers to displaying the output feature maps of each convolutional layer and pooling layer in the network for a given input [32]. As the number of network layers deepens, the image details gradually disappear from the feature map and become more and more abstract. The network extracts a certain part of the lettuce as the target classification information, and human intuitive vision cannot understand the content displayed after the image passes through the activation layer. Although this higher-level content is not observed by the human eye, neural networks can. The first layer of the network is a collection of various edge detectors. At this stage, the activated feature map retains almost all the information in the original image. Figure 7 showed the activation visualization of a given input image in the first stage of the Im-ResNet model. The upper right was the feature of the first layer of convolution, and the lower right was the feature of the first pooling layer. The activation map can be seen that the lettuce was clearly separated from the background. Different channels had different emphasis on image detection. For example, the 6th channel focused more on lettuce texture detection, and the 18th channel focused more on lettuce contour detection. The visual evaluation further proved the advantages of Im-ResNet in feature extraction.

After the input image was processed by convolution, the feature map would be obtained. Especially, as the depth of the network increases, the feature maps became sparser and the extracted features became more representative. Similarly, the gradient-weighted class activation heat map could visually show the features learned by the CNN. In short, when the input image was recognized to be a certain category,

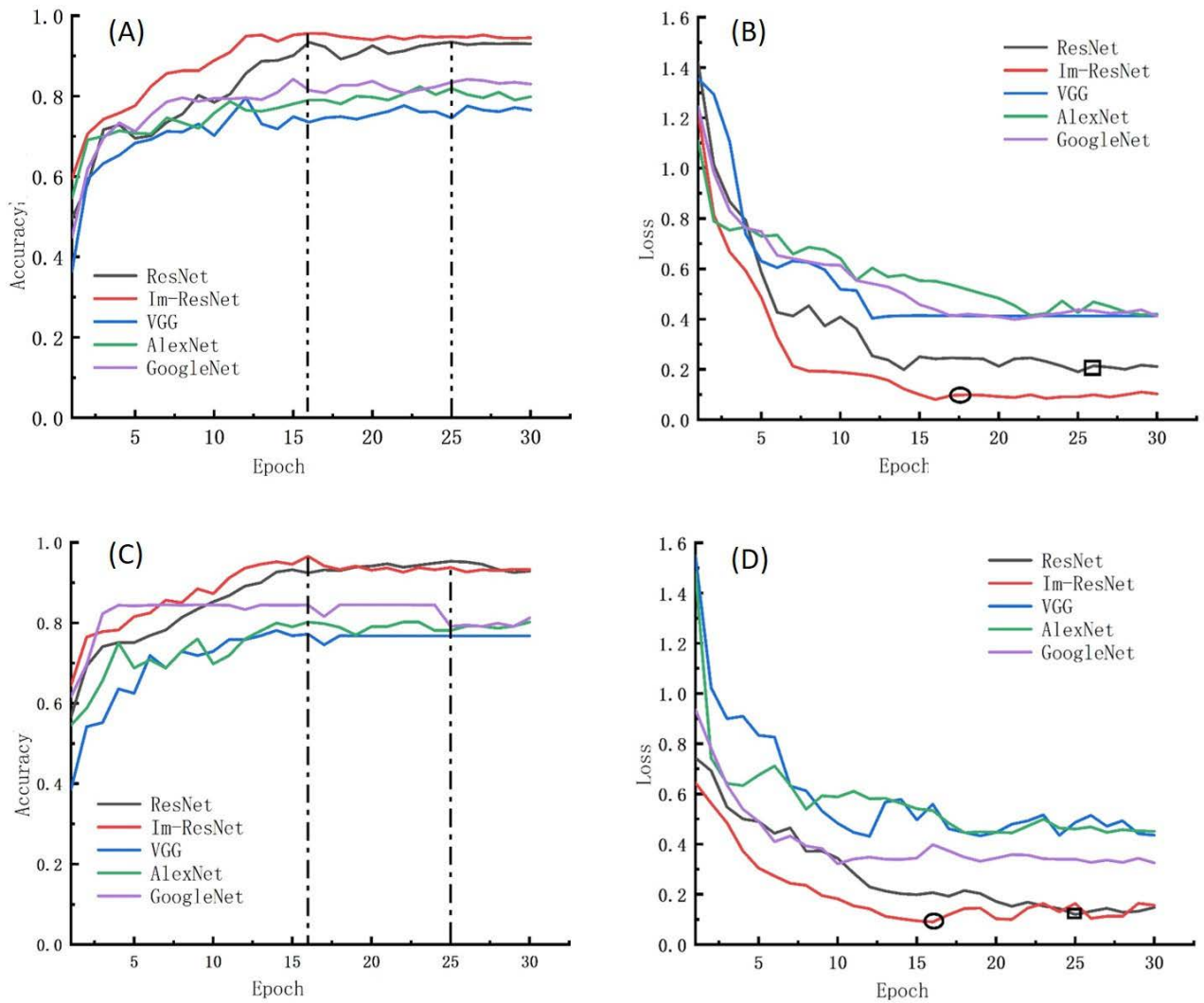


FIGURE 5. Comparison results of Im-ResNet, ResNet50, VGG16, AlexNet and GoogleNet. (A) Training set accuracy, (B) Training set loss, (C) Validation set accuracy, (D) Validation set loss.

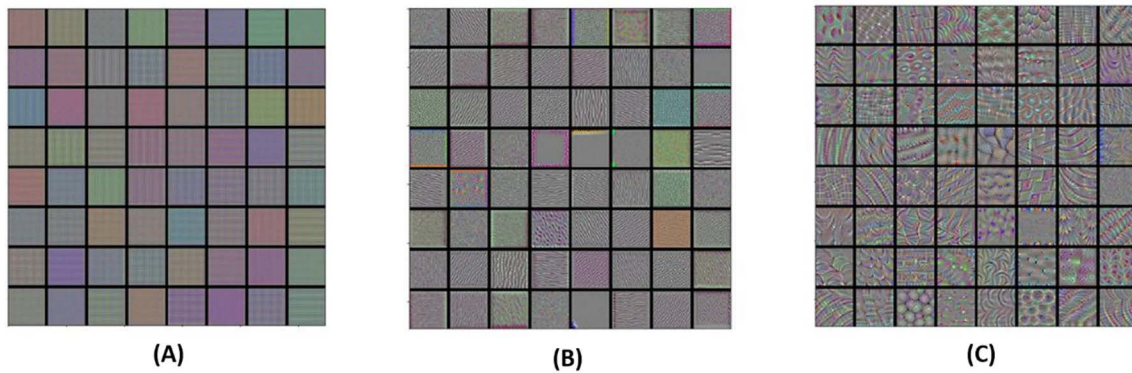


FIGURE 6. The Im-ResNet model filter in different convolutional layers. (A) The first layer of convolution, (B) The middle layer of convolution; (C) the last layer of convolution.

the area pixels in the image that were positively related to this category were displayed with a heat map. The more

sensitive the area in the heat map, the higher the temperature. On the contrary, the less sensitive the area, the lower the

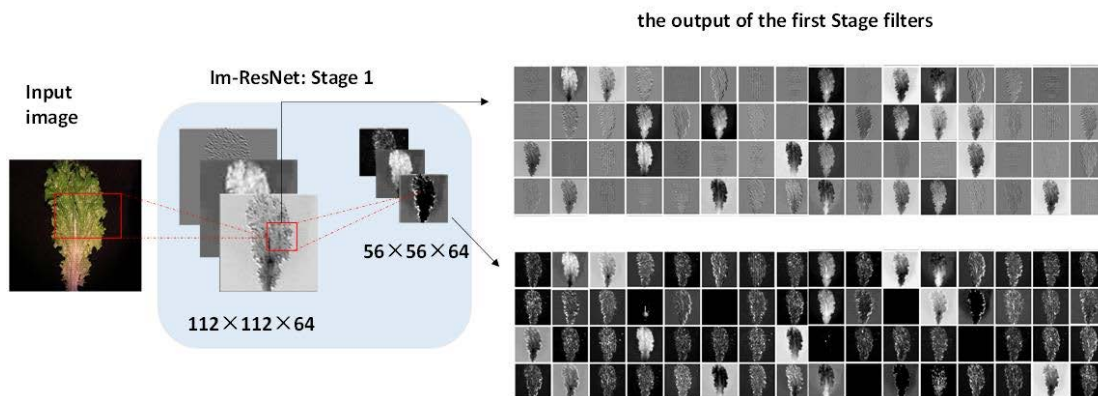


FIGURE 7. Image feature visualization in the first stage of the Im-ResNet model.

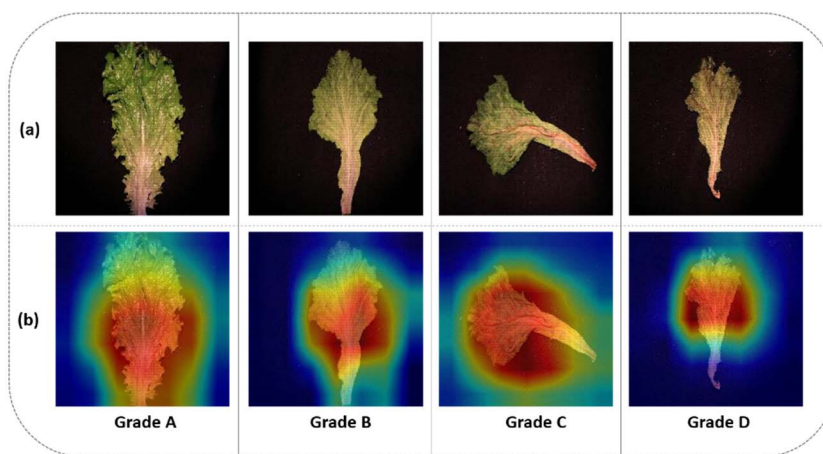


FIGURE 8. Comparison results of original image and Grad-cam visualization. (a) Original images, (b) Grad-cam visualization.

temperature. The comparison results of original image and Grad-cam visualization are shown in Figure 8. Our model could extract the correct features from the correct locations despite rotating the image (grade C) or the leaves wilting and becoming smaller (grade D). These experimental results proved that our model was robust and could identify lettuce freshness of each grade.

B. IM-RESNET PERFORMANCE EVALUATION

In order to evaluate the recognition performance of our proposed model, we tested different models using 1000 lettuce images. Our proposed Im-ResNet can the better test accuracy (95.60%) than VGG16 (76.17%), AlexNet (78.56%), GoogleNet (82.98%) and ResNet50 (84.42%).

The Im-ResNet obtained results are demonstrated in Table 5. The average values of our model accuracy, precision, recall and F1-score were 95.60%, 95.70%, 95.60% and 95.61%. Besides, the confusion matrix in Figure 9 showed the distribution of the true and predicted values for the four grades. Observing Figure 9 (B), we could find that grade A had 100% classification accuracy and grade D had 97% classification accuracy, which was higher than the other two

grades (B and C). Observing Figure 9 (A), for grade B and grade C, there were 21 and 9 misclassified images, respectively, and the two grades were confused with each other. For grade C and D, there were 6 and 8 misclassified images respectively, and the two grades were confused with each other. However, all misclassified images distributed near the main diagonal of the confusion matrix. Because the lettuce leaves of grade B were visually more similar to grade C, which means that they were predicted to be close to the correct grade.

In the practical application of deep learning, controlling image acquisition conditions, feature extraction, feature selection and optimal classifiers are the keys to obtaining accurate recognition results. The images used in our work were taken under controlled conditions regarding lighting conditions and camera stabilization and parameter. Besides, the feature extraction and classification steps were completed by the Im-ResNet model. Therefore, the proposed method used the deep convolutional neural network to detect the freshness of lettuce quickly and non-destructively during storage, which not only showed low complexity, but also showed a satisfactory accuracy of 95.60% (see Table 5).

TABLE 5. The performance analysis of Im-ResNet.

Grade	Accuracy (%)	Precision (%)	Recall (%)	F ₁ -score (%)
A	95.60	100	100	100
B	95.60	96.21	91.60	93.85
C	95.60	89.01	94.00	91.43
D	95.60	97.58	96.80	97.18
Average	95.60	95.70	95.60	95.61

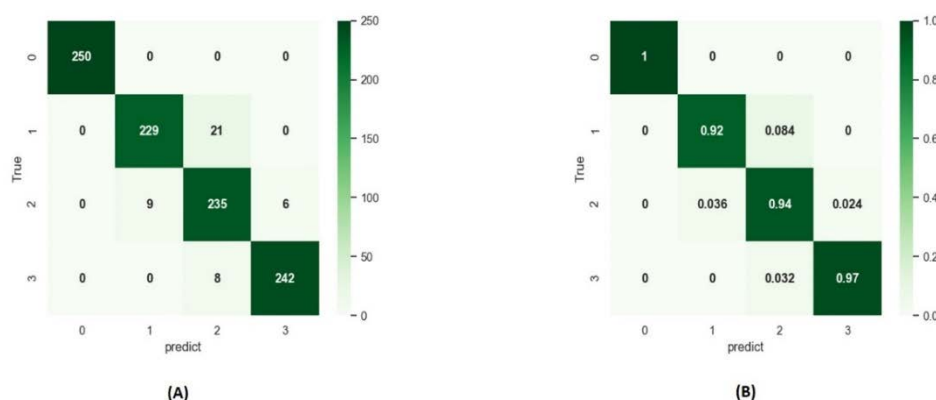


FIGURE 9. Confusion matrix analysis results. (A) Confusion matrix, (B) Normalized confusion matrix.

These results further proved the advantages of Im-ResNet in practical applications.

V. CONCLUSION

In conclusion, we proposed a lettuce classification method based on an improved deep residual convolutional neural network (Im-ResNet). Firstly, a lettuce collection system was built and data collection was completed by this system. Secondly, Im-ResNet was optimized to obtain higher recognition accuracy, and the feature extraction ability of the network was improved by constructing new convolutional layers and pooling layers. In order to prevent over fitting and further improve recognition accuracy, two fully connected layers were used to improve the reconstruction of the fully connected layer of ResNet50. Finally, the experimental results proved that the proposed model had obvious advantages in terms of loss and accuracy compared with the traditional networks ResNet50, AlexNet, VGG16 and GoogleNet. Compared with the traditional grading methods, the Im-ResNet lettuce freshness grading method proposed had many advantages such as fast, high efficiency, undamaged and low cost.

This study would provide technical support for vegetables non-destructive rapid detection and automatic grading. In future research, we will collect more lettuce samples to make the model more robust, and use optimization algorithms to further improve the grading accuracy and prediction ability of the model.

REFERENCES

- [1] Y. J. Jung, A. Padmanabahn, J. H. Hong, J. Lim, and K. O. Kim, "Consumer freshness perception of spinach samples exposed to different storage conditions," *Postharvest Biol. Technol.*, vol. 73, pp. 115–121, Nov. 2012, doi: 10.1016/j.postharvbio.2012.06.005.
- [2] M. Cao, W. Niu, and H. Liu, "Changes in nitrate content during vegetable storage and assessment of food safety for human consumption environment," in *Feb Fresenius Environmental Bulletin*. Berlin, Germany: Fresenius Environ. Bull., Feb. 2021, p. 1457.
- [3] Z. Xie, H. Xu, Q. Huang, and P. Wang, "Spinach freshness detection based on hyperspectral image and deep learning method," *Trans. Chin. Soc. Agricult. Eng.*, vol. 35, no. 13, pp. 277–284, 2019.
- [4] L. Lunadei, B. Diezma, L. Lleó, L. Ruiz-Garcia, S. Cantalapiedra, and M. Ruiz-Altisent, "Monitoring of fresh-cut spinach leaves through a multispectral vision system," *Postharvest Biol. Technol.*, vol. 63, no. 1, pp. 74–84, Jan. 2012, doi: 10.1016/j.postharvbio.2011.08.004.
- [5] Y. Qiu, Y. Zhao, J. Liu, and Y. Guo, "A statistical analysis of the freshness of postharvest leafy vegetables with application of water based on chlorophyll fluorescence measurement," *Inf. Process. Agricult.*, vol. 4, no. 4, pp. 269–274, Dec. 2017, doi: 10.1016/j.inpa.2017.08.001.
- [6] G. T. Le, T. V. Tran, H.-S. Lee, and W.-Y. Chung, "Long-range batteryless RF sensor for monitoring the freshness of packaged vegetables," *Sens. Actuators A, Phys.*, vol. 237, pp. 20–28, Jan. 2016, doi: 10.1016/j.sna.2015.11.013.
- [7] H. X. Xu, "Study on nondestructive detection of freshness of post-harvest spinach based on machine vision and electronic nose," M.S. thesis, Dept. Jiangsu Univ., Zhenjiang, China, 2016.
- [8] X. Huang, S. Yu, H. Xu, J. H. Aheto, E. Bonah, M. Ma, M. Wu, and X. Zhang, "Rapid and nondestructive detection of freshness quality of postharvest spinaches based on machine vision and electronic nose," *J. Food Saf.*, vol. 39, no. 6, Dec. 2019, Art. no. e12708, doi: 10.1111/jfs.12708.
- [9] S. M. Hassan, M. Jasinski, Z. Leonowicz, E. Jasinska, and A. K. Maji, "Plant disease identification using shallow convolutional neural network," *Agronomy*, vol. 11, no. 12, p. 2388, Nov. 2021, doi: 10.3390/agronomy11122388.
- [10] W. Li, W. Sun, Y. Zhao, Z. Yuan, and Y. Liu, "Deep image compression with residual learning," *Appl. Sci.*, vol. 10, no. 11, p. 4023, Jun. 2020.
- [11] Z. Lixin, H. Fadong, and L. Zhengchao, "Image recognition of cotton leaf diseases and insect pests based on transfer learning," *Chin. J. Agricult. Eng.*, vol. 36, no. 7, p. 8, 2020.
- [12] H. Xia, J. Jia, W. Gao, X. Guo, W. Zhang, L. Zheng, and M. Wang, "MFC-CNN: An automatic grading scheme for light stress levels of lettuce (*Lactuca sativa* L.) leaves," *Comput. Electron. Agricult.*, vol. 179, Dec. 2020, Art. no. 105847, doi: 10.1016/j.compag.2020.105847.
- [13] N. Amin, T. G. Amin, and Y. D. Zhang, "Image-based deep learning automated sorting of date fruit," *Postharvest Biol. Technol.*, vol. 153, pp. 133–141, Jan. 2019, doi: 10.1016/j.postharvbio.2019.04.003.

- [14] A. Z. da Costa, H. E. H. Figueroa, and J. A. Fracarolli, "Computer vision based detection of external defects on tomatoes using deep learning," *Biosyst. Eng.*, vol. 190, pp. 131–144, Feb. 2020, doi: [10.1016/j.biosystemseng.2019.12.003](https://doi.org/10.1016/j.biosystemseng.2019.12.003).
- [15] C. Zhou, J. Hu, Z. Xu, J. Yue, H. Ye, and G. Yang, "A monitoring system for the segmentation and grading of broccoli head based on deep learning and neural networks," *Frontiers Plant Sci.*, vol. 11, p. 402, Apr. 2020, doi: [10.3389/fpls.2020.00402](https://doi.org/10.3389/fpls.2020.00402).
- [16] Y. Xu, B. Zhao, Y. Zhai, Q. Chen, and Y. Zhou, "Maize diseases identification method based on multi-scale convolutional global pooling neural network," *IEEE Access*, vol. 9, pp. 27959–27970, 2021.
- [17] A. Majid, "An integrated deep learning framework for fruits diseases classification, CMC-computers," *Mater. Continua*, vol. 71, no. 1, pp. 1387–1402, 2022.
- [18] X. Jian, L. Zhang, Y. Hang, X. Sun, and S. Dai, "Spectral detection for quality and freshness index of main leaf vegetables based on smart cellphone," *Spectrosc. Spectral Anal.*, vol. 39, no. 5, pp. 1524–1529, 2019, doi: [10.3964/j.issn.1000-0593\(2019\)05-1524-06](https://doi.org/10.3964/j.issn.1000-0593(2019)05-1524-06).
- [19] B. Pace, M. Cefola, P. Da Pelo, F. Renna, and G. Attolico, "Non-destructive evaluation of quality and ammonia content in whole and fresh-cut lettuce by computer vision system," *Food Res. Int.*, vol. 64, pp. 647–655, Oct. 2014, doi: [10.1016/j.foodres.2014.07.037](https://doi.org/10.1016/j.foodres.2014.07.037).
- [20] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981.
- [21] J. Shijie, W. Ping, J. Peiyi, and H. Siping, "Research on data augmentation for image classification based on convolution neural networks," in *Proc. Chin. Autom. Congr. (CAC)*, Oct. 2017, pp. 4165–4170.
- [22] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, no. 2, 2012, pp. 1–9.
- [23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1–9.
- [24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [26] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*.
- [27] T. Khan, H. H. R. Sherazi, M. Ali, S. Letchmunan, and U. M. Butt, "Deep learning-based growth prediction system: A use case of China agriculture," *Agronomy*, vol. 11, no. 8, p. 1551, Aug. 2021.
- [28] Z. Kang, B. Yang, S. Yang, X. Fang, and C. Zhao, "Online transfer learning with multiple source domains for multi-class classification," *Knowl.-Based Syst.*, vol. 190, Feb. 2020, Art. no. 105149, doi: [10.1016/j.knosys.2019.105149](https://doi.org/10.1016/j.knosys.2019.105149).
- [29] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, Oct. 2019, doi: [10.1007/s11263-019-01228-7](https://doi.org/10.1007/s11263-019-01228-7).
- [30] H. Xing, G. Zhang, and M. Shang, "Deep learning," *Int. J. Semantic Comput.*, vol. 10, no. 3, pp. 417–439, 2016.
- [31] A. Ashiqzaman, A. K. Tushar, S. Dutta, and F. Mohsin, "An efficient method for improving classification accuracy of handwritten Bangla compound characters using DCNN with dropout and ELU," in *Proc. 3rd Int. Conf. Res. Comput. Intell. Commun. Netw. (ICRCICN)*, Nov. 2017, pp. 147–152.
- [32] K. Hameed, D. Chai, and A. Rassau, "Class distribution-aware adaptive margins and cluster embedding for classification of fruit and vegetables at supermarket self-checkouts," *Neurocomputing*, vol. 461, pp. 292–309, Oct. 2021, doi: [10.1016/j.neucom.2021.07.040](https://doi.org/10.1016/j.neucom.2021.07.040).



YANLEI XU received the B.S. degree in industrial automation from the Changchun University of Science and Technology, China, in 2002, and the M.S. and Ph.D. degrees in signal and information processing and communication and information system from Jilin University, China, in 2005 and 2009, respectively. She is currently a Professor and a Doctoral Supervisor at the College of Information, Jilin Agricultural University. Her research interests include image process, automatic control, and computer application technology.



YUTING ZHAI received the B.S. degree in electronic information science and technology from Jilin Agricultural University, Changchun, China, in 2019, where she is currently pursuing the master's degree with the College of Information Technology. Her research interests include image recognition and deep learning.



QINGYUAN CHEN received the B.S. degree in electronic information science and technology from Jilin Agricultural University, Changchun, China, in 2020, where he is currently pursuing the master's degree with the College of Information Technology. His research interests include image recognition and deep learning.



SHUOLIN KONG received the B.S. degree in electronic information science and technology from Jilin Agricultural University, Changchun, China, in 2020, where he is currently pursuing the master's degree with the College of Information Technology. His research interests include image recognition and deep learning.



YANG ZHOU received the B.S. degree in electronic information science and technology from Jilin Agricultural University, Changchun, China, in 2013, the M.S. degree in information and communication engineering from the Changchun University of Science and Technology, Changchun, in 2016, and the Ph.D. degree in circuit and system from Jilin University, Changchun, in 2019. He is currently a Lecturer with the College of Information Technology, Jilin Agricultural University. His current research interests include imaging, image processing, and deep learning.

• • •