IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 20, 2021, accepted June 29, 2021, date of publication July 5, 2021, date of current version July 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3094802

Apple Leaf Disease Recognition and Sub-Class Categorization Based on Improved Multi-Scale Feature Fusion Network

YUANQIU LUO, JUN SUN, JIFENG SHEN, XIAOHONG WU, LONG WANG, AND WEIDONG ZHU
School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China

Corresponding author: Jun Sun (sun2000jun@sina.com)

This work was supported in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions under Grant PAPD-2018-87, in part by the Project of Faculty of Agricultural Equipment of Jiangsu University under Grant 4121680001, and in part by the Synergistic Innovation Center of Jiangsu Modern Agricultural Equipment and Technology under Grant 4091600030.

ABSTRACT Apple diseases cause a lot of economic losses to fruit growers in China. Early diagnosis and accurate recognition of apple diseases can control the spread of disease and reduce production costs. However, the significance of disease characteristic of apple leaves in complex environment is relatively weak, and the fine-grain among different diseases of apple leaves is high, and the conventional feature extraction methods will lose the discrimination information. To solve these problems, an apple disease classification model based on multi-scale feature fusion is proposed in this paper. Firstly, the information flow of conventional residual network (ResNet) was improved to achieve efficient information circulation through changing the position of batch normalization and rectified linear unit (ReLU). Secondly, in order to solve the problem of serious loss of information in ResNet downsample, the channel projection and spatial projection of downsample were separated. Lastly, the 3 × 3 conv in ResBlocks was replaced by pyramid convolution, and the dilated convolution with different dilation rate was introduced into pyramid convolution to enhance the output scale of feature maps and improve the robustness of the model. The optimized model was verified on the dataset of this paper, and the optimized model had stronger anti-noise ability and better robustness, excellent learning effect and fast convergence speed. The classification accuracy on the original dataset is 94.24%, and that on the preprocessed dataset is 94.99%. The results demonstrate that the optimal model has a high accuracy, which can provide a reference for the prevention and control of apple leaf diseases.

INDEX TERMS Apple disease, multi-scale feature fusion, deep learning, ResNet, classification.

I. INTRODUCTION

In recent years, apple planting area in China has increased year by year, from 0.679 million km^2 to 2.243 million km^2, and the total yield have increased from 2.2752 million tons to 38.49 million tons [1]. China ranks first in the world in terms of yield and planting area. However, the yield per unit area of apple is still significantly lower than the world's average. The occurrence of serious apple diseases is an important reason for the low yield. In general, the recognition of apple diseases is realized by experts through simple observation [2]. Due to the lack of experts, fruit growers have to judge symptoms according to their own observation, which is time-consuming and difficult, because it requires rich experience to correctly identify diseases [3]. The above problems show that the establishment of apple disease intelligent recognition system is extremely urgent. However, the field background is complex, the early characteristics of the disease are not obvious, and the fine-grain of apple diseases is high, which increases the difficulty of disease recognition [4]. Identifying the disease leaf accurately between complex environment and healthy leaves is the key to the control of apple leaf diseases [5].

To solve the above problems, scholars have been developed committing to develop some methods that automatically, quickly and accurately identify apple diseases [6]. Lali *et al.* [2] found that texture and color features play an important role in apple disease classification. However, each apple disease spot has its own characteristics, and the shape

The associate editor coordinating the review of this manuscript and approving it for publication was Hiram Ponce.

characteristics are not very useful for disease classification. Based on this, they proposed to combine LBH features and color features to classify apple diseases. Omrani *et al.* [7] proposed a classification model of apple leaf disease based on support vector regression, and the classification performance of support vector regression (SVR) model based on radial basis function (RBF) and polynomial function were compared. The results showed that the performance of SVR based on radial basis function was better. Sun [8] combined image processing and support vector machine to classify apple leaf diseases. Firstly, the disease image was processed by gray transformation; secondly, the disease image was segmented by setting the threshold, and the features were extracted by gray level co-occurrence matrix and principal component analysis; finally, the parameters were optimized by particle swarm optimization algorithm, and the accuracywas 96.969% by support vector machines (SVM). Dubey and Jalal [9]used K-means clustering to extract the region of interest, then segmented the disease according to color, texture and shape, combined different features, and finally used SVM to classify, the accuracy was 95.94%. Saeed *et al.* [10] used pretained VGG19 to extract deep features, and combined the features from the fully connected layers 6 and 7 with a PLS-based parallel fusion method. Then, the best features were selected by PLS project method and classified by the ensemble baggage tree. The accuracy of the model reached 90.1% in Plantvillage. Sharif *et al.* [11] used the optimized weighted segmentation method to extract the citrus lesion spots, and extracted the best features based on the hybrid method consisted of PCA score, entropy and covariance vector. The selected features were fed to M-SVM for classification, and the accuracy of the method on the citrus disease data set reached 97%.

Most of the above literatures are based on color and texture to identify apple disease, which are not conducive to the promotion of apple leaf disease identification methods. Compared with the conventional machine vision technology, the research based on deep learning (DL) can achieve parallel processing of mass data, and the feature extraction based on DL is completed in the training process without manual operation before training [12]. In recent years, the breakthrough of convolutional neural network (CNN) accelerates the development of DL related researches [13]. CNN is a kind of deep neural network and has many applications [14], including many complex tasks, such as image classification, segmentation and object detection. Driven by the breakthrough of CNN related research, DL has emerged a series of powerful architectures, such as AlexNet [15], GoogLeNet [16], Residual network (ResNet) [17], DensNet [18]. With the development of CNN, there are more effective methods for apple leaf disease identification. Liu *et al.* [19] removed part of the fully connected layer, added the pooling layer and introduced the initial structure of GoogLeNet to AlexNet, and proposed a new apple leaf disease classification model based on CNN model. Compared with the standard AlexNet model, the optimized model had smaller parameters and faster convergence speed. In order to improve the classification performance on imbalanced datasets, Zhong and Zhao et al. [20] proposed three methods to train DenseNet: regression, multi-label classification and focus integration loss function. The experimental results showed that the performance of the three methods were better than that of the traditional cross entropy loss. Yu *et al.* [21] proposed DCNN based on region of interest (ROI) perception to classify apple diseases. Firstly, ROI sub-network was designed to separate leaves, then VGG was used to classify diseases, and the accuracy rate was 84.3%. To classify apple leaf diseases accurately, Yan *et al.* [22] changed the first two layers of VGG full connection layer into global average pooling and BN, which improved the classification accuracy of VGG network, reduced the amount of parameters, and the accuracy rate was 99.01%. Rubab *et al.* [23] used pretrained Alexnet and Vgg19 to extract features and transferred features into k-nearestneighbor (KNN), M-SVM and DT for classification. The performance of M-SVM was the best, and the accuracy reaches 96.9%. Nasir *et al.* [24] used fine-tuning and pretraining Vgg19 to extract features, and combined with the contour features extracted by pyramid histogram of oriented gradients (PHOG). Then "relevance-based" optimization technique was used to select the best features from fused vector for classication. The accuracy of this method is 99.6%, which is better than previous techniques.

The above literatures focus on image level rough classification, and these models are difficult to apply in complex environment. The fine-grained categorization is to identify the sub-categories under the large class. Compared with traditional classification task, the difference and difficulty of fine-grained categorization task is that the granularity of image category is finer, and the tasks are more challenging. There are little minor differences in Figure 1, so apple leaf disease classification belongs to fine-grained classification. Although convolution neural network promotes the research on agricultural diseases, there are few reports on fine-grained categorization in agricultural diseases. In the field of DL, most research used bounding box or landmark annotations to assist fine-grained categorization tasks, which is expensive, such as CUB-200 and CelebA [25]. In the field of agriculture, there are few studies on fine-grained crop disease, and most researchers still focus on coarse-grained categorization. Yang *et al.* [26] proposed

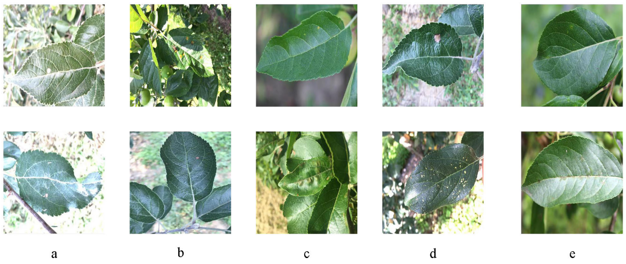


FIGURE 1. The difficulties of the dataset.

LFC-Net to study the fine-grained classification of tomato leaf diseases. LFC-Net was divided into three parts: The Location network detected high-informative regions in the image, the Feedback network guided the optimization iteration of Location network, and the Classification network used informative regions proposed for classification. Compared with the traditional CNN, the proposed model had obvious advantages, and the accuracy was 99.7%. Yang *et al.* [27] proposed NASNetLarge fine-grained classification model based on attention mechanism. The model used the informative regions of images to recognize disease, and the F1-score reached 93.05%. Hu *et al.* [28] proposed MDF-ResNet for fine-grained identification. The model can fuse species, coarse-grained diseases and fine-grained disease features to identify diseases. To identify Maize leaf diseases, Mingjie LV [29] proposed DMS-Robust AlexNet. The model combined with dilated convolution and multi-scale convolution, and the accuracy was 98.62%. The background of the dataset in above literature is simple, and a picture has only one leaf. However, the background is complex and mixed with healthy leaves in the actual environment, which increases the difficulties of classification.

This paper took ResNet-50 as the basic network and proposed an improved ResNet model based on multi-scale feature fusion. With the improvement on the conventional ResNet information flow mode, the efficient information flow can be realized. The downsample was improved to reduce the information loss and increase the translation invariance of the model. The 3 × 3 conv of ResBlocks was replaced by the pyramid convolution, and the dilation convolution was added to the pyramid convolution to increase the scale output of the network and extract multi- scale information. The improved ResNet model was applied to the classification of apple leaf diseases, which provided reference to the improvement of apple disease intelligent diagnosis app.

Our main contributions can be summarized as follows: (i) We propose a CNN framework based on ResNet for fine-grained categorization of apple leaf diseases, which can be applied to other leaf disease data sets and provide reference for intelligent control of apple leaf diseases; (ii) The information flow mode, downsample and 3 × 3 conv in ResBlock are improved to accelerate convergence speed, reduce information loss and extract multi-scale characteristics. Compared with the pre-trained neural network, the model has better robustness and achieves high accuracy of apple leaf health/diseases image classification.

The rest of this paper is organized as follows. Section Materials And Methods introduces the main experimental dataset and our proposed method based on improved information flow, improved downsample and multi-scale fusion network. Section Model Training introduces experimental equipment and training parameters. The experimental results and analysis of the results of different networks are described in Section Results and Analysis. Finally, this paper concludes in Section Conclusion.

II. MATERIALS AND METHODS

A. DATASET AND DATA PREPROCESSING

The dataset were obtained from FGVC7 [30] (https://www.kaggle.com/c/plant-pathology-2020-fgvc7) and Baidu AI Studio (https://aistudio.baidu.com/aistudio- /datasetdetail/11591). All images were collected in natural environment, with uneven illumination, small disease spot and big background noise. As shown in Table 1, there are five kinds apple disease leaf pictures and one kind apple healthy leaf pictures. Alternaria boltch, grey spot and mosaic were obtained from Baidu AI studio, and healthy leaf, scab and rust were from FGVC7.

TABLE 1. Information of the database images.

Class	Disease name/health	Images number	Source
0	alternaria boltch	520	Baidu AI studio
1	grey spot	520	Baidu AI studio
2	mosaic	520	Baidu AI studio
3	health	520	FGVC7
4	rust	520	FGVC7
5	scab	520	FGVC7

The dataset in this paper has the following difficulties in training process: (i) the dataset shown in Figure 1(a) has different shooting angles and uneven illumination of leaves, which increases the difficulty of disease spot identification; (ii) rust and grey spot disease pictures are showed in Figure 1(b). The healthy leaves in the background are not processed in the collection process of these pictures, and the disease pictures are mixed with healthy leaves, which increase the difficulty of classification; (iii) as shown in Figure 1(c), the early disease spots are very small and difficult to be recognized; (iv) Figure 1(d) is grey spot and mosaic disease. The disease spots on these figures are too similar with the background, which increases the difficulty of disease classification; (v) Figure 1(e) are healthy leaf and scab disease. The similarities between them are very high and it is difficult to recognize and classify them.

The dataset was randomly divided into training set and test set according to the ratio of 4:1. To avoid the influence of different image pixels on classification results, all images were adjusted to 512 × 512 pixels. In order to avoid over-fitting, this study preprocessed the dataset. As shown in FIGURE 2, the datasets were augmented by changing brightness, rotation and adding gaussian noise in this paper. The brightness of original image changes randomly to the original brightness of 0.5-1.5 in changing brightness, and the original image is rotated randomly by 90 degrees, 180 degrees and 270 degrees in rotation. All augmentations are based on the ImageEnhance function in Pillow.

B. IMPROVED INFORMATION FLOW THROUGH THE NETWORK

ResNet was specifically created by many residual blocks (ResBlocks) to allow information through the network convenient. The information flow of original ResBlocks is shown

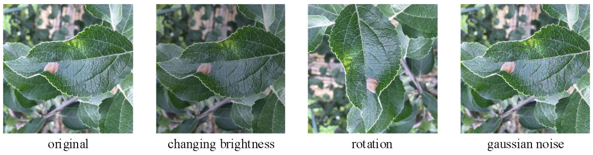
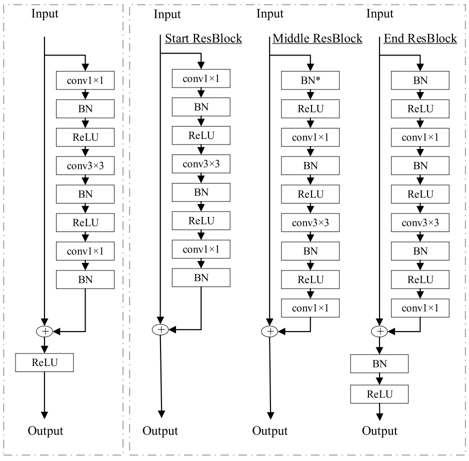


FIGURE 2. The preprocessing of the dataset.



(a) original

(b) Improved ResBlocks

FIGURE 3. (a) Original ResBlocks; (b) Improved ResBlocks (* is ineffective in the first middle ResBlock).

in FIGURE 3 (a). There are three convolution layers (two 1 × 1 conv and one 3 × 3 conv), three batch normalization (BN) [31] and three ReLU activation function. The arrow without operation represents the most direct path of information flow. Observing FIGURE 3(a), it can be found that the Relu in the main propagation path may have a negative effect on the propagation of information by resetting the nzegative signal to zero [32]. In addition, there is no complete signal standardization in ResBlocks, and all BN are used for branches. With the increasing of depth, the information on the main propagation path becomes more unnormalized.

The improved information flow is shown in FIGURE 3(b), and ResBlocks in each layer are divided into three stages: one Start ResBlocks, several Middle ResBlocks and one End ResBlocks. BN and ReLU were added before each branch after residual connection to standardize the information; the first BN in the first Middle ResBlock was eliminated because the signal was normalized at the Start ResBlock; BN and ReLU were added at the end of each layer to standardize the complete signal.

C. IMPROVED DOWNSAMPLE

The original downsample of ResNet is shown in FIGURE 4(a). The input is mapped to the output channel through 1 × 1 conv in the original projection shortcut. Only 1 × 1 conv with stride 2 can make the spatial matching of input and output. However, in the process of spatial projection, the 1 × 1 conv with stride 2 will directly ignore 75% information, and add the remaining 25% signal with branch signal, causing information loss, and introducing lots of noises, which will have a negative impact on the network.

The proposed downsample is presented in FIGURE 4(b). The spatial projection is disentangled from channel projection. The spatial projection is performed by 2 × 2 max pooling and the channel projection is performed by 1 × 1 conv. The improved downsample reduces information loss and increases the translation invariance of the network.

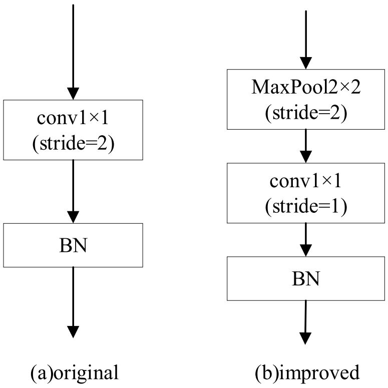


FIGURE 4. (a) original downsample; (b) improved downsample.

D. DILATED CONVOLUTION

Dilated convolution [33] can increase the receptive field of the model by adding a cavity into the standard convolution kernel. Compared with the original ordinary convolution, the dilated convolution has one more parameter: dilation rate, which refers to the number of cavities between the points of the convolution kernel. The ordinary convolution with dilation rate 1 and 3 × 3 conv with dilation rate 2 are respectively shown in FIGURE 5(a) and FIGURE 5(b). As can be seen from FIGURE 5(b), a 3 × 3 conv with dilation rate 2 can be regarded as a 5 × 5 conv, and the value without dot is zero. The parameters of convolution don't increase, but the receptive field of convolution become larger, and the scale output of convolution is changed.

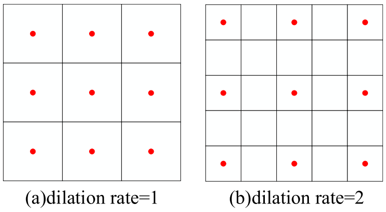


FIGURE 5. The receptive field at different dilation rates: (a) Dilation rate =1; (b) Dilation rate =2.

E. PYRAMIDAL CONVOLUTION

The background of the dataset is complex and noisy. It may not meet the demand that extracting single-scale features from the network. It is necessary to extract multi-scale information for analysis. In order to extract the information on

TABLE 2. Improved Resnet architectures.

Stage	Output	Improved ResNet 50
starting	112×112	3×3,64, s=2 3×3,64, s=1 3×3,64, s=1
0	56×56	$\begin{bmatrix} 1\times 1,64 \\ \begin{bmatrix} (3\times 3, \mathrm{D}=4,16)\times 4 \\ (3\times 3, \mathrm{D}=3,16)\times 3 \\ (3\times 3,\mathrm{D}=2,16)\times 2 \\ 3\times 3,\mathrm{D}=1,16 \end{bmatrix} \\ 1\times 1,256 \end{bmatrix}\times 3$
1	28×28	$\begin{bmatrix} 1\times 1,128 \\ \begin{bmatrix} (3\times 3, \mathrm{D}=4,32)\times 4 \\ (3\times 3, \mathrm{D}=3,32)\times 3 \\ (3\times 3,\mathrm{D}=2,32)\times 2 \\ 3\times 3,\mathrm{D}=1,32 \end{bmatrix} \\ 1\times 1,512 \end{bmatrix}\times 4$
2	14×14	$\begin{bmatrix} 1\times 1,256 \\ \begin{bmatrix} (3\times 3, \mathrm{D}=4,64)\times 4 \\ (3\times 3, \mathrm{D}=3,64)\times 3 \\ (3\times 3,\mathrm{D}=2,64)\times 2 \\ 3\times 3,\mathrm{D}=1,64 \end{bmatrix} \\ 1\times 1,1024 \end{bmatrix}\times 6$
3	7×7	$\begin{bmatrix} 1\times 1,512 \\ 3\times 3,512 \\ 1\times 1,2048 \end{bmatrix}\times 3$
	1×1	Global avg pool 6-d fc
#params		25.12×10^6

different scales of the figure, the convolution kernels of different sizes were stacked together and pyramid convolution (PyConv) was proposed in [34]. The number and size of convolution kernels in PyConv are set according to the task requirements, which make PyConv flexible and scalable. In the conventional ResNet resblocks, feature extraction mainly relies on 3 × 3 conv. The 3 × 3 conv in ResNet ResBlocks is replaced by PyConv in this paper, which improves the recognition ability of the network. In the original PyConv, the dilation rate is fixed at 1 and the receptive field may not meet demand in the training process. As shown in FIGURE 6, in order to obtain higher scale information, PyConv with different dilation rates are proposed to expand the receptive field of the model and increase the scale output of the model without increasing the parameters of the model (compared with increasing the size of convolution kernel).

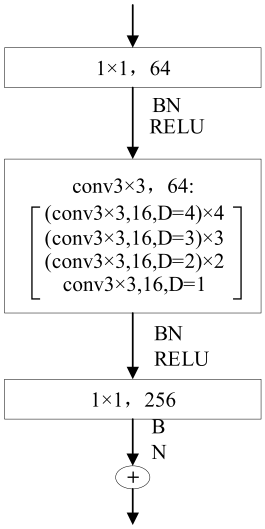


FIGURE 6. Improved ResBlock.

F. MODLE STRUCTURE

The optimized model structure based on multi-scale feature extraction is shown in FIGURE 7, and the model is divided into six stages. The first stage is the beginning stage, in which the dimension of input image is increased by a 7 × 7 conv to prepare for feature extraction of 0-layer. The middle four stages are the 0-3 layers of ResNet, which perform feature extraction and reduce the size of the feature map.

The specific parameters of each layer of the model are observed in Table 2. Compared with conventional ResNet, the zero to two layers of the ResNet were improved and PyConv composed of 3 × 3, 5 × 5, 7 × 7 and 9 × 9 conv with dilation rate of 1, 2, 3 and 4 were selected in this paper. The number of input and output channels of each convolution kernel in PyConv is 1/4 of that of conventional ResBlocks. The third layer of the optimized model is the same as that of the conventional ResNet, which integrates the multi-scale information extracted from 0-2 layers. The last stage is the classification stage, where the full connection layer is selected as classifier. Kaiming normal distribution is chosen to initialize convolution kernels, and large-scale convolution kernels are replaced by multiple cascaded

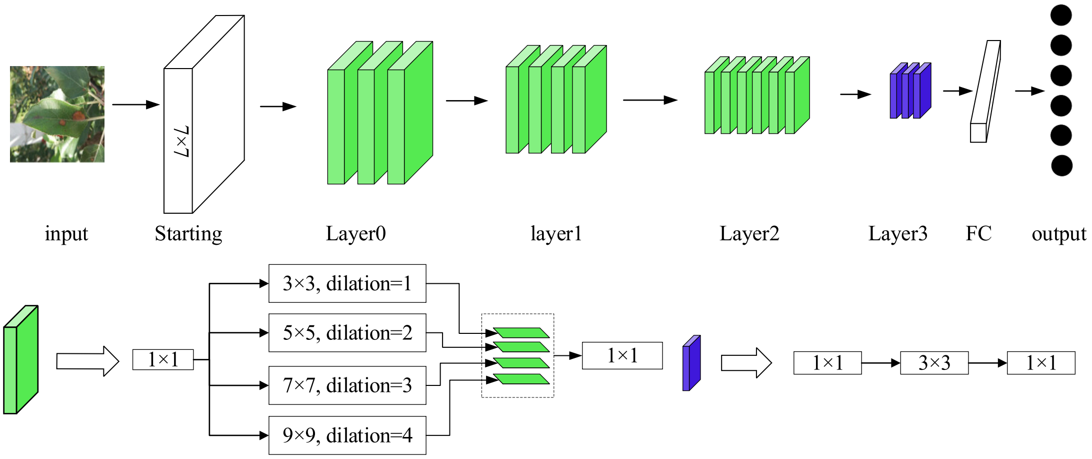


FIGURE 7. Model structure.

3×3 conv. The effects of different information flow, different downsample, no dilated convolution and dilated convolution on the performance of the model were compared in this paper, and the optimized model was compared with conventional ResNet and other improved model.

III. MODEL TRAINING

A. EXPERIMENTAL PLATFORM

The experimental software is Ubuntu 18.04 LTS 64-bit system, using python as deep learning framework. The hardware is configured with 16GB of memory, equipped with Intel i7-7700k CPU and GTX1080Ti 11GB graphics card.

B. EXPERIMENTAL PARAMETERS

The training set and test set were divided into several batches by batch training and the batch size was set to 16. The test set were tested, accuracy and loss were saved in the log after the training set had been trained. After many experiments, the Stochastic Gradient Descent (SGD) was used to optimize the model. The momentum was set to 0.4, and the weight decay was set to 0.0015. In the experiment, the number of epochs was set to 100, and the initial learning rate was set to 0.1. The learning rate updating strategy with segmented exponential decay was adopted, and the learning rate was reduced to 10% of original after every 30 epochs.

IV. RESULTS AND ANALYSIS

A. THE EFFECTS OF DIFFERENT INFORMATION FLOW

In training process of the model, the accuracy is used to evaluate the performance of the model. The accuracy is computed by the eq. (1) [35].

$$\text{Accuracy} = \frac{\text{correctly classified number of data}}{\text{total number of data}} \times 100\% \quad (1)$$

Two experiments with different information flow are carried out in Table 3. The accuracy is increased by 2.40% after the improvement, and the accuracy is greatly improved in Table 3.

TABLE 3. Validation accuracy rate (%) comparison results of different information flow.

Information flow mode	Accuracy (%)	params
ResNet original information flow	91.83	25.12×10^6
improved information flow	**94.23**	25.12×10^6

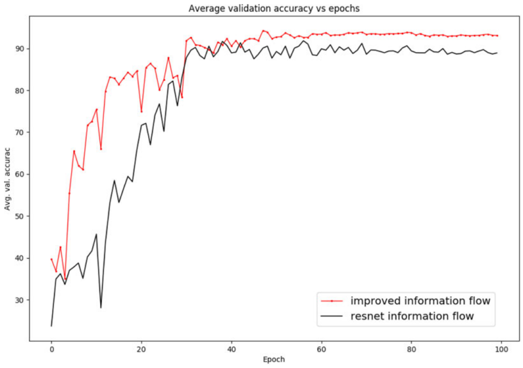


FIGURE 8. Validation accuracy curves of different information flow.

The accuracy curves of the validation set in training process is shown in FIGURE 8. The x-axis is the number of training epochs, and the y-axis is the corresponding training accuracy. Compared with the improved model, the accuracy curve of the conventional ResNet information flow model rose slowly at the beginning of training. The ReLU in the main propagation path of ResBlocks will have a negative impact on the information flow by zeroing the negative signal, and there is no BN in the propagation path to normalize the complete signal. The positive signal (after BN) in the path may be reset to zero under the influence of ReLU (after residual connection), which will affect the network performance. After many epochs, the network started to adjust the weight and output the signal which was not affected by ReLU. In addition, the complete signal became more unnormalized with the increasing of depth, which was not conducive to learning. The improved flow has higher accuracy and faster convergence speed at the beginning of train on our dataset. The horizontal comparison curve shows that the improved model needs less than 15 epochs to outperform the best accuracy of ResNet on all first 30 epochs. These results show that the improved information flow has the advantages of accelerating the learning process, accelerating the convergence of training, and significantly reducing the training epochs.

B. THE EFFECTS OF DIFFERENT DOWNSAMPLES

The classification accuracy of different downsample is shown in Table 4. The accuracy of the downsample using conventional ResNet is the lowest, only 90.83%. The accuracy of improved downsample with 3 × 3 max pooling is 94.23%, and that with 2 × 2 max pooling is 94.23%. Compared with conventional downsample, the performance of downsample is improved. Some of the disease spots in our dataset are small and rare. The conventional downsample uses a 1 × 1 conv for spatial projection and channel projection, which will directly ignore most of the information, reduce the recognition rate of disease spots and cause serious loss of information. It is proved in Table 4.

TABLE 4. **Validation accuracy rate (%) comparison results of different downsample.**

Dowmsample mode	Accuracy (%)	params
ResNet downsanple	90.83	25.12×10^6
downsample with 3×3 max pool	94.23	25.12×10^6
downsample with 2×2 max pool	**94.23**	25.12×10^6

FIGURE 9 shows the accuracy curves of test set in training process. In the early stage of training, the accuracy of the model with conventional downsample is significantly lower than that of the two improved models, and the accuracy of the two improved models has little difference. After 60 epochs, all models reach convergence, and the convergence interval of the improved downsample with 2 × 2 max pooling is significantly better than other models. The optimized accuracy of the improved models with 2 × 2 max pooling and 3 × 3 max pooling is same, but there is a certain gap in the convergence interval, which may be due to that overlapping pooling will reduce the recognition rate of small spot. Improved downsample reduces the loss of information and increases the translation invariance of the network.

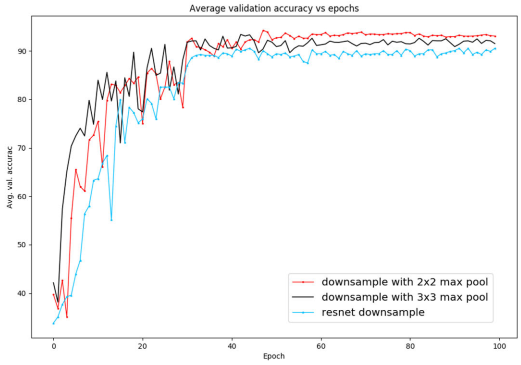


FIGURE 9. **Validation accuracy curve of different downsample.**

C. THE EFFECTS OF DILATED CONVOLUTION

Dilated convolution can change the scale information of the feature map without increasing the computational complexity of the model. Table 5 shows the classification performance of the model before and after adding dilated convolution. The accuracy of the model using conventional convolution is 92.31%, and after adding dilated convolution, the accuracy of the model is improved by 1.92%, reaching 94.23%, which indicates that the performance of the model is improved by adding dilated convolution. Increasing the receptive field can reduce the influence of background noise and increase the recognition rate of spots on the disease leaves.

TABLE 5. **Validation accuracy rate (%) comparison results of different convolution mode.**

Convolution mode	Accuracy (%)	params
PyConv without dilation convolution	92.31	25.12×10^6
Improved ResNet	**94.23**	25.12×10^6

The accuracy of each epoch in training process is shown in FIGURE 10. After 40 epochs, the accuracy of the model with dilated convolution is generally better than that of the model with conventional convolution, especially in the complete convergence interval of the model. The conventional convolution in PyConv is replaced by dilated convolution, which increases the scale output of the model without changing the complexity of the model and improves the classification performance of the model in our dataset.

D. COMPARISON WITH RESNET VARIANTS

In order to verify the classification performance of the optimized model, the optimized model is compared with the conventional ResNet and other improved models on our dataset in this paper. The results are shown in Table 6. In the original dataset, the accuracy of ResNet, ResNeXt [36], iResNet[32], PyConvResNet[34] and the optimized model are 86.38%, 88.46%, 89.42%, 90.71% and 94.23%. In the preprocessed dataset, the accuracy of ResNet, ResNext, iResNet, PyConvResNet and optimized model are 91.75%, 92.75%, 92.87%, 92.71 and 94.99%. The classification accuracy of improved ResNet is higher than that of other training models in the original dataset and preprocessed dataset.

FIGURE 11 is the accuracy curve in training process, FIGURE 11 (a) is the accuracy curve of the original dataset, and FIGURE 11 (b) is the accuracy curve of the preprocessed dataset. In the early stage of training process, the optimized model can achieve higher accuracy faster within less epochs, and the model's accuracy is higher within the same number of epochs. After a certain number of epochs, the model reaches convergence, and the performance of the optimized model is obviously better than resnet and resnet

TABLE 6. Validation accuracy rate (%) comparison results of different model on dataset.

Model	Original dataset (%)	Preprocessed dataset (%)	params
ResNet 50 [17]	86.38	91.75	25.56×10^{6}
ResNeXt 50 [36]	88.46	92.75	25.0×10^{6}
iResNet 50 [32]	89.42	92.87	25.56×10^{6}
PyConvResNet 50 [34]	90.71	92.71	24.85×10^{6}
Improved ResNet 50	**94.23**	**94.99**	25.12×10^{6}

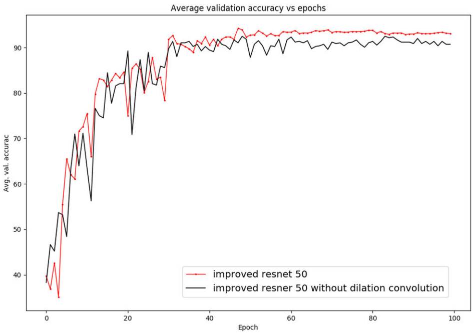


FIGURE 10. Validation accuracy curve of different convolution mode in ResBlocks.

TABLE 7. Validation accuracy rate (%) comparison results of different model on dataset.

Method	Original dataset (%)	Preprocessed dataset (%)
VGG16	91.99	93.07
MobilenetV2	90.06	91.67
Efficientnet	90.38	92.27
our	**94.23**	**94.99**

variants. In our dataset, the conventional ResNet extracts single-scale features which is easily affected by noise. The optimized model based on multi-scale features fusion has stronger anti-noise ability and can reach higher accuracy. As shown in Table 4 that after preprocessed, the classification accuracy of improved ResNet is improved by 0.75%, and that of other models is improved by 2%-5%, which indicates that the robustness of improved ResNet is better. The above results show that the optimized model has better anti-noise ability, faster convergence speed and higher accuracy than other models, which proves the superiority of ResNet based on multi-scale feature fusion in apple diseases classification.

E. COMPARISON WITH OTHER MODELS

TABLE 7 presents the classification accuracy of VGG16, MobilebetV2, Efficientnet and our models on the original dataset and preprocessed dataset. The results indicate that our model based on improved multi-scale feature fusion network outperform other common models by 1-4% in term of accuracy. This demonstrates that the multi-scale feature fusion network has stronger feature extraction ability and allows the models to focus on the key feature of disease.

The model based on multi-scale feature fusion network achieves the highest accuracy of 94.23% and 94.99%, and provides the best categorization perform on our dataset. Thus, our model can be used in subsequent experiment for apple disease image classification.

F. CONFUSION MATRIX

The specific classification results of different models on original dataset will be discussed in this section. The results of different models are shown in TABLE 8. Accuracy, Precision, Recall and F1-score are used to evaluate the performance of the model. The Precision, Recall and F1-score can be expressed by the following formulas, respectively,

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

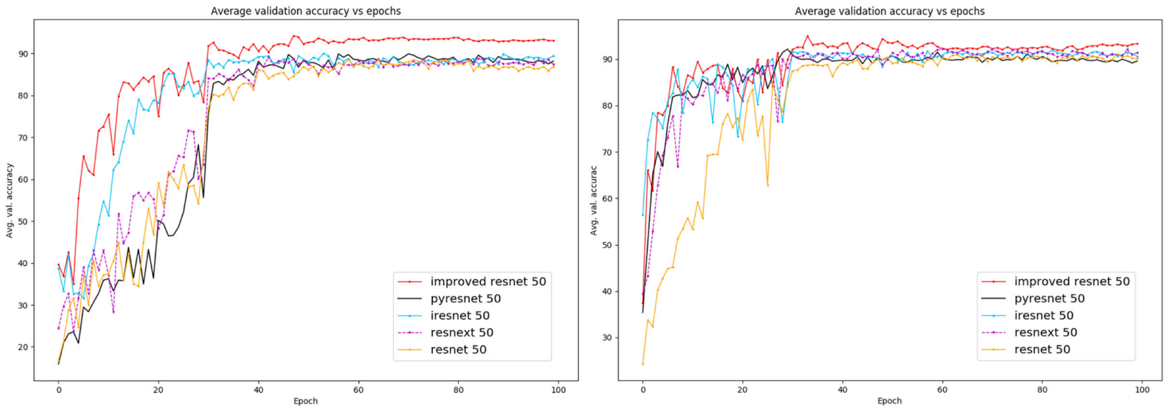
$$F1 = 2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{4}$$

where TP is true positive, FN is false negative and FP is false positive.

Improved ResNet has higher Precision, Recall and F1-score in TABLE 8. FIGURE 12 is the confusion matrix of different models, Y-axis is real label and X-axis is prediction label. 0-5 in the coordinate label refers to alternaria boltch, grey spot, health leaf, mosaic, rust and scab for the convenience of drawing. Taking FIGURE 12(a) as example, 90 in matrix grid means that 90 of 104 health pictures in test set are predicted as health pictures. The optimal model reaches an accuracy level of 94.23% on the test set of 624 images, of which 588 images are correctly classified. And it has the optimal performance in all models with 94.75% precision, 94.23% recall and 94.49% F1-score. In the comparison model, alternaria boltch, grey spot and health leaf are the most easily misclassified, especially the healthy leaves in the disease pictures are extracted as category features. Improved ResNet misclassifies only one health class, and predicts three images of diseases as health class, which indicate that the misclassification rate between disease class and health class is decreased. In addition, the recognition rate

TABLE 8. Results comparison of different methods.

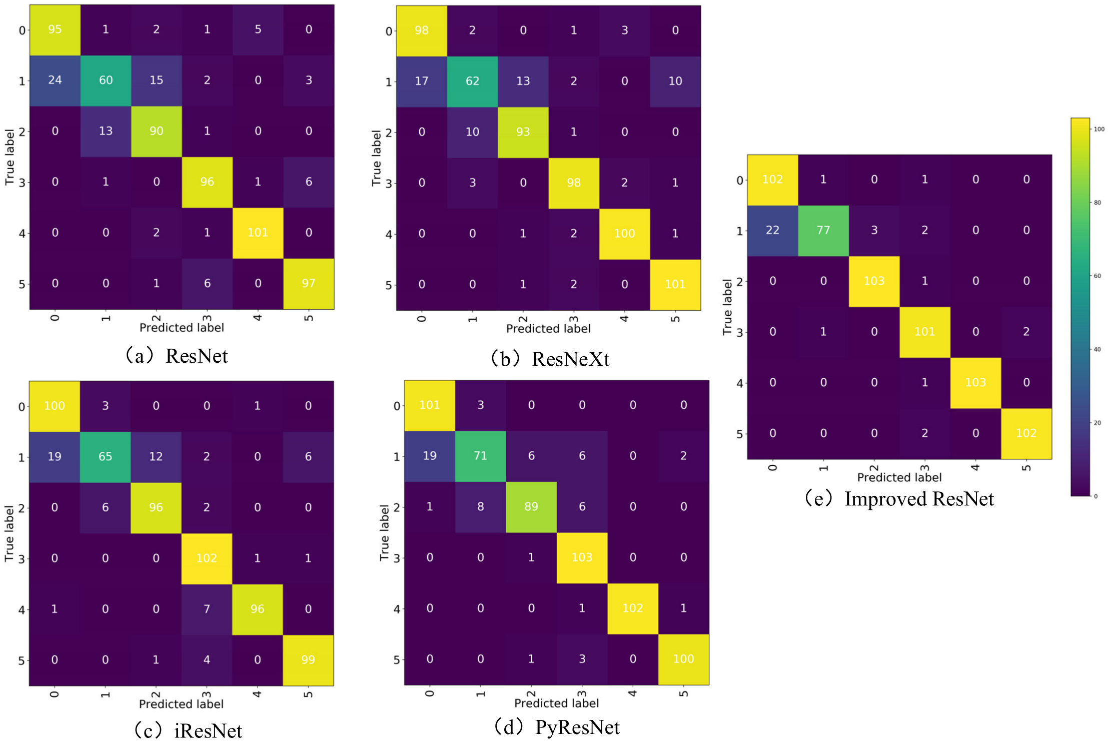
Method	Accuracy/%	Precision/%	Recall/%	F1-score/%
ResNet	86.38	86.21	86.38	86.29
ResNeXt	88.46	88.15	88.46	88.30
iResNet	89.42	89.63	89.42	89.56
PyConvResNet	90.71	90.91	90.71	90.81
Improved ResNet	**94.23**	**94.75**	**94.23**	**94.49**



（a） original datasets accuracy

（b） pre-processed datasets accuracy

FIGURE 11. (a) Validation accuracy rate (%) on original dataset; (b)Validation accuracy rate on preprocessed dataset.



（a） ResNet

（b） ResNeXt

（e） Improved ResNet

（c） iResNet

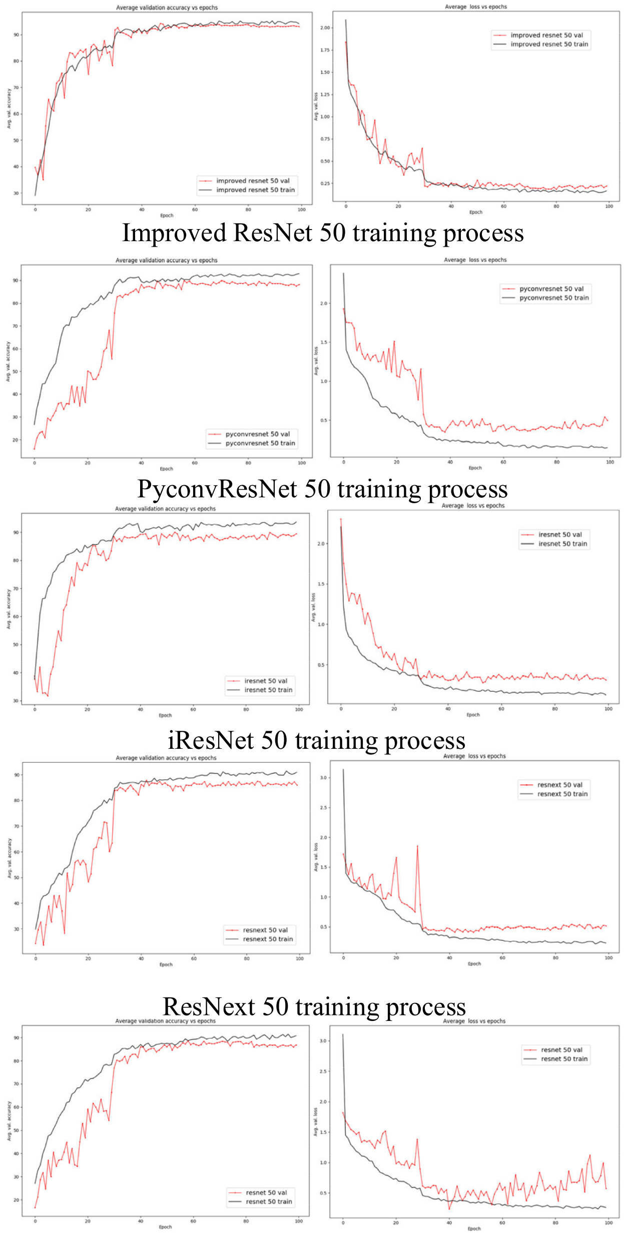
（d） PyResNet

FIGURE 12. Confusion matrix of different models.

of improved ResNet for diseases is significantly improved, and the misclassification for 0, 3, 4 and 5 classes decreases to a certain extent, and the recognition rate of the 1 class disease increases by 5.8%-16.35%. The above results indicate that the feature extraction network based on multi-scale feature fusion has better classification performance and stronger anti-noise ability.

V. CONCLUSION

The multi-scale features fusion ResNet model was adopted to classify apple diseases in this paper. Apple disease dataset was collected in natural environment, with uneven illumination, small disease spot and big background noise. Three improvements were made to ResNet: (i) the information flow in ResBlocks was improved, and improved model had faster convergence speed and better learning effect; (ii) the downsample was improved to improve the translation invariance of the model and reduce the information loss; (iii) the 3 × 3 conv in ResBlocks was replaced by PyConv, and the dilated convolution was added to PyConv. The feature maps were changed from single-scale to multi-scale, which improved the robustness of the model. Compared with other models, the performance of proposed model in apple diseases dataset was better, and proposed model had faster convergence speed and higher classification accuracy. The accuracy of the optimized model was 94.23% in the original dataset,



Improved ResNet 50 training process

PyconvResNet 50 training process

iResNet 50 training process

ResNext 50 training process

ResNet 50 training process

FIGURE 13. The accuracy and loss curves of the training process in TABLE 8 (100 epochs).

and 94.99% in the preprocessed dataset. The model proposed in this paper had good classification performance, fast speed and good robustness for a variety of apple diseases, which provided a theoretical basis to the subsequent establishment of apple disease intelligent recognition system.

APPENDIX

See FIGURE 13.

REFERENCES

[1] Y. Hu, T. Hu, Y. Wang, S. Wang, and K. Cao, "Survey on the occurrence and distribution of apple diseases in China," *Plant Protection*, vol. 42, no. 1, pp. 175–179, Jan. 2016.

[2] M. A. Khan, M. I. U. Lali, M. Sharif, K. Javed, K. Aurangzeb, S. I. Haider, A. S. Altamrah, and T. Akram, "An optimized method for segmentation and classification of apple diseases based on strong correlation and genetic algorithm based feature selection," *IEEE Access*, vol. 7, pp. 46261–46277, 2019.

[3] G. Jun, "Countermeasure of apple fruit diseases in China," *J. Fruit Resour.*, vol. 1, no. 1, pp. 43–44 and 50, 2020.

[4] Z. U. Rehman, M. A. Khan, F. Ahmed, R. Damaševičius, S. R. Naqvi, W. Nisar, and K. Javed, "Recognizing apple leaf diseases using a novel parallel real-time processing framework based on MASK RCNN and transfer learning: An application for smart agriculture," *IET Image Process.*, pp. 1–12, Mar. 2021.

[5] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," *IEEE Access*, vol. 7, pp. 59069–59080, 2019.

[6] M. B. Tahir, M. A. Khan, K. Javed, S. Kadry, Y. Zhang, T. Akram, and M. Nazir, "Recognition of apple leaf diseases using deep learning and variances-controlled features reduction," *Microprocessors Microsyst.*, pp. 1–24, Jan. 2021.

[7] E. Omrani, B. Khoshnevisan, S. Shamshirband, H. Saboohi, N. Anuar, and M. Nasir, "Potential of radial basis function-based support vector regression for apple disease detection," *Measurement*, vol. 55, pp. 512–517, Sep. 2014.

[8] S. Sun, "Classification of apple leaf disease based on image processing and support vector machine," Xi'an Univ. Sci. Technol., Xi'an, China, Tech. Rep., Jun. 2017.

[9] S. R. Dubey and A. S. Jalal, "Apple disease classification using color, texture and shape features from images," *Signal, Image Video Process.*, vol. 10, no. 5, pp. 819–826, Jul. 2016.

[10] F. Saeed, M. A. Khan, M. Sharif, M. Mittal, L. M. Goyal, and S. Roy, "Deep neural network features fusion and selection based on PLS regression with an application for crops diseases classification," *Appl. Soft Comput.*, vol. 103, May 2021, Art. no. 107164.

[11] M. Sharif, M. A. Khan, Z. Iqbal, M. F. Azam, M. I. U. Lali, and M. Y. Javed, "Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection," *Comput. Electron. Agricult.*, vol. 150, pp. 220–234, Jul. 2018.

[12] A. Adeel, M. A. Khan, T. Akram, A. Sharif, M. Yasmin, T. Saba, and K. Javed, "Entropy-controlled deep features selection framework for grape leaf diseases recognition," *Expert Syst.*, vol. 1, pp. 1–17, May 2020.

[13] A. Koirala, K. Walsh, Z. Wang, and C. McCarthy, "Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO,'" *Precis. Agricult.*, vol. 162, pp. 219–234, Dec. 2019.

[14] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, 2015, Art. no. 436.

[15] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing coadaptation of feature detectors," *Comput. Sci.*, vol. 3, no. 4, pp. 212–223, Mar. 2012.

[16] 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. (CVPR)*, Jun. 2015, pp. 1–9.

[17] 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.

[18] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[19] B. Liu, Y. Zhang, D. He, and Y. Li, "Identification of apple leaf diseases based on deep convolutional neural networks," *Symmetry*, vol. 10, no. 1, Dec. 2017, Art. no. 11.

[20] Y. Zhong and M. Zhao, "Research on deep learning in apple leaf disease recognition," *Comput. Electron. Agricult.*, vol. 168, Jan. 2020, Art. no. 105146.

[21] H.-J. Yu, C.-H. Son, and D. H. Lee, "Apple leaf disease identification through region-of-interest-aware deep convolutional neural network," *J. Imag. Sci. Technol.*, vol. 64, no. 2, pp. 20507-1–20507-10, Mar. 2020.

[22] Q. Yan, B. Yang, W. Wang, B. Wang, P. Chen, and J. Zhang, "Apple leaf diseases recognition based on an improved convolutional neural network," *Sensors*, vol. 20, no. 12, Jun. 2020, Art. no.3535.

[23] N. Muhammad, N. Bibi, O.-Y. Song, M. A. Khan, and S. A. Khan, "Severity recognition of aloe vera diseases using AI in tensor flow domain," *Comput., Mater. Continua*, vol. 66, no. 2, pp. 2199–2216, 2021.

[24] I. M. Nasir, A. Bibi, J. H. Shah, M. A. Khan, M. Sharif, K. Iqbal, Y. Nam, and S. Kadry, "Deep learning-based classification of fruit diseases: An application for precision agriculture," *Comput., Mater. Continua*, vol. 66, no. 2, pp. 1949–1962, 2021.

[25] Z. Huang and Y. Li, "Interpretable and accurate fine-grained recognition via region grouping," 2020, *arXiv:2005.10411*. [Online]. Available: http://arxiv.org/abs/2005.10411

[26] G. Yang, G. Chen, Y. He, Z. Yan, Y. Guo, and J. Ding, "Self-supervised collaborative multi-network for fine-grained visual categorization of tomato diseases," *IEEE Access*, vol. 8, pp. 211912–211923, 2020.

[27] G. Yang, Y. He, Y. Yang, and B. Xu, "Fine-grained image classification for crop disease based on attention mechanism," *Frontiers Plant Sci.*, vol. 11, p. 2077, Dec. 2020.

[28] W.-J. Hu, J. Fan, Y.-X. Du, B.-S. Li, N. Xiong, and E. Bekkering, "MDFC–ResNet: An agricultural IoT system to accurately recognize crop diseases," *IEEE Access*, vol. 8, pp. 115287–115298, 2020.

[29] M. Lv, G. Zhou, M. He, A. Chen, W. Zhang, and Y. Hu, "Maize leaf disease identification based on feature enhancement and DMS-robust alexnet," *IEEE Access*, vol. 8, pp. 57952–57966, 2020.

[30] R. Thapa, K. Zhang, N. Snavely, S. Belongie, and A. Khan, "The plant pathology challenge 2020 data set to classify foliar disease of apples," *Appl. Plant Sci.*, vol. 8, no. 9, Sep. 2020, Art. no. e11390.

[31] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[32] I. C. Duta, L. Liu, F. Zhu, and L. Shao, "Improved residual networks for image and video recognition," 2020, *arXiv:2004.04989*. [Online]. Available: http://arxiv.org/abs/2004.04989

[33] Z. Lu, Y. Bai, Y. Chen, C. Su, S. Lu, T. Zhan, X. Hong, and S. Wang, "The classification of gliomas based on a pyramid dilated convolution resnet model," *Pattern Recognit. Lett.*, vol. 133, pp. 173–179, May 2020.

[34] I. C. Duta, L. Liu, F. Zhu, and L. Shao, "Pyramidal convolution: Rethinking convolutional neural networks for visual recognition," 2020, *arXiv:2006.11538*. [Online]. Available: http://arxiv.org/abs/2006.11538

[35] U. P. Singh, S. S. Chouhan, S. Jain, and S. Jain, "Multilayer convolution neural network for the classification of mango leaves infected by anthracnose disease," *IEEE Access*, vol. 7, pp. 43721–43729, 2019.

[36] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5987–5995.

YUANQIU LUO received the bachelor's degree from Jiangsu University, in 2019, where he is currently pursuing the master's degree. His research interest includes fine-grained classification of machine vision in agriculture.

JUN SUN was born in Taixing, Jiangsu, China, in 1978. He is currently a Professor and a Doctoral Supervisor with the School of Electrical and Information Engineering, Jiangsu University. He is a Senior Member of the China Society of Agricultural Engineering and a member of Modern Physical Agricultural Engineering Committee of China Agricultural Machinery Society.

JIFENG SHEN was born in Yixing, Jiangsu, China, in 1980. He received the bachelor's degree in computer science and technology and the master's degree in computer application from the Jiangsu University of Science and Technology, in 2003 and 2006, respectively. His research interests include computer vision, pattern recognition, and image processing.

XIAOHONG WU was born in Hefei, Anhui, China, in 1971. His research interests include food nondestructive testing, spectral information processing, and image processing research.

LONG WANG is currently pursuing the master's degree with Jiangsu University. His research interests include image segmentation and classification in agriculture.

WEIDONG ZHU is currently pursuing the master's degree with Jiangsu University. His research interest includes image classification in agriculture.

• • •