RESEARCH ARTICLE

Leaf Disease Detection Based on Lightweight Deep Residual Network and Attention Mechanism

ZHIYONG XIAO[1], YONGGE SHI[2], GAILIN ZHU[3], JIANPING XIONG[4], AND JIANHUA WU[5]

[1]School of Electronic and Information Engineering, Jiangxi Institute of Economic Administrators, Nanchang 330088, China
[2]Department of Information Engineering, Gongqing Institute of Science and Technology, Jiujiang 332020, China
[3]Zhuhai Unitech Power Technology Company Ltd., Zhuhai 519000, China
[4]Industrial Center, Shenzhen Polytechnic, Shenzhen 518055, China
[5]School of Information Engineering, Nanchang University, Nanchang 330031, China

Corresponding authors: Jianping Xiong (jpxiong@szpt.edu.cn) and Jianhua Wu (jhwu@ncu.edu.cn)

ABSTRACT In today's leaf disease detection, the accuracy of recognition has never been of such importance as it is now. In this aspect, leaf disease recognition method based on machine learning relies heavily on the size of the region of interest and the dispersion of lesions. Professional instrument for leaf disease detection remains a challenging task in accuracy and convenience. A new lightweight model based on advanced residual network and attention mechanism for extracting more accurate region of interest and the lesion, SE-VRNet, was proposed. The proposed SE-VRNet incorporated deep variant residual network (VRNet) and a squeeze-and-excitation (SE) module with attention mechanism, in order to solve the problem that the feature extraction was difficult due to the dispersed location of the leaf disease. The accuracy of top-1 and top-3 obtained by the model SE-VRNet on NewData is 99.73% and 99.98%, respectively, and the accuracy of top-1 and top-3 obtained by the model on SelfData is 95.71% and 99.89%, respectively. The experimental results on the datasets of PlantVillage, OriData, NewData and SelfData were better than other state-of-the-art methods, demonstrating the effectiveness and feasibility of the proposed SE-VRNet in identifying leaf diseases with mobile devices.

INDEX TERMS Leaf disease, identification, deep variant residual network, attention mechanism, squeeze-and-excitation module.

I. INTRODUCTION

Crop leaf diseases are an important biological disastrous problem in agricultural production in the world. For example, plants infected with brown spot will initially have circular brown spots of varying sizes on their leaves. Later, the spots will grow black molds, which will cause the plants to wilt if not treated properly. Only experienced farmers can accurately identify the characteristics of leaf diseases, and then choose the appropriate medicine, and finally use the least amount of pesticides to achieve the therapeutic effect. However, due to the large number of disease types and high outbreak probability, many Chinese farmers are unable to detect the disease in

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

time [1]. At present, China's input of pesticide is still increasing, while European countries began to reduce in the 1980s. Therefore, farmers in China should timely use portable instruments to accurately identify crop leaf diseases at the early stage. Advanced mobile leaf disease detection equipment can reduce pesticide use and increase crop yields. Image processing technology based on deep learning (DL) can provide a new solution for crop leaf disease identification [2]. In recent years, various DL-based network model architectures have been designed and applied to different tasks. For example, Geoffrey Hinton and Alex Krizhevsky designed AlexNet and won the prestigious ILSVRC (ImageNet Large Scale Visual Recognition Challenge) Challenge in 2012. VGGnet was proposed by the Computer Vision Group at Oxford University and DeepMind in 2014 and won the first prize in the location

project at the ILSVRC Challenge that year [3]. DL technology has also been widely used in intelligent agriculture fields such as plant and fruit recognition, crop and weed detection and classification. Kamilaris et al. [4] summarized the development and application of DL technology in the field of agriculture in recent years, and believed that the performance of DL was better than traditional methods such as support vector machine and random forest classifier. Compared with traditional algorithms based on scale-invariant feature transform, histogram, statistics, texture, color and shape, the automatic feature extraction of DL model is more effective [5], [6].

In the field of crop leaf disease identification, various scholars have achieved much. For example, Mukhtar et al. [7] proposed a wheat disease recognition network based on one-shot learning. The model requires only a limited number of images for training and achieves more than 92% accuracy, 84% precision, and 85% recall. Brahimi et al. [8] used an improved AlexNet and GoogLeNet network model to identify nine kinds of tomato leaf diseases, and finally achieved an accuracy of 99.18%. DeChant et al. [9] used a convolutional neural network (CNN) to design the maize dead leaf disease identification system, which overcame the limitations of missing data and irregular crop images in the field, and finally achieved the maximum accuracy of 96.7% with better robustness. Gandhi et al. [10] first used a generative adversarial network to enhance the leaf disease dataset, trained two different pre-trained models, Inceptionv3 and MobileNet, on PlantVillage dataset, and embedded the trained model into the mobile application for leaf disease recognition. Finally, the classification accuracies of Inceptionv3 and MobileNet were 88.6% and 92%, respectively.

Despite the great progress in DL-based leaf disease identification, there are still numerous challenges and issues to be addressed. Due to the different sizes and scattered locations of leaf lesions, the traditional DL algorithms need to be improved to extract accurate disease feature information in mobile terminals. For example, Elhassouny and Smarandache [11] developed a method using an advanced CNN to recognize tomato leaf diseases, but the overall sample number of this work was small, with less than 1,000 images for each disease type, and the classification accuracy was less than 90% for each optimization method. Manso et al. [12] developed an algorithm to classify coffee leaf rust, and Piconet et al. [13] expanded Manso's work using an adapted deep residual neural network (ResNet)-based algorithm to deal with the detection of multiple plant diseases from infield acquired images. In this study, three diseases were identified by using ResNet from 8178 images, achieving 87% classification accuracy. However, there is still much room for improvement in classification accuracy. At the same time, smartphone applications have been developed to identify leaf diseases using the DL methods. Toseef and Khan [14] developed a fuzzy inference system for diagnosis of crop leaf diseases with smartphone. Mobile terminal applications of lightweight DL model are still a relatively new research
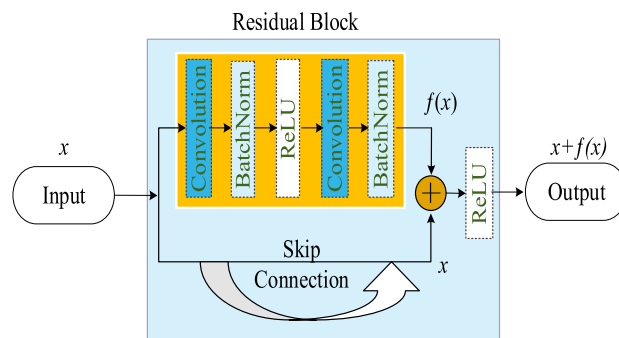

FIGURE 1. The basic architecture of residual block.

area in leaf disease identification. Loey et al. [15] proposed a hybrid deep-learning and image-classification-based model to detect leaf disease in 2020. However, the research work had several challenges. Geetha et al. [16] proposed a k-nearest neighbors algorithm to detect tomato leaf disease. But the accuracy of this method depended very much on leaf segmentation.

In recent years, more and more deep learning methods have been applied to leaf disease detection [17]. For example, in 2021, Chen et al. [18] proposed a hybrid transfer learning and image-processing-baded model to detect leaf disease. Chen improved leaf disease recognition and forecasting. However, data preparation in the method was very complex and not easy to generalize. In 2022, Harakannanavar et al. [19] proposed an improved deep-learning-based model to predict leaf disease in his research. This method adopted a variety of techniques to extract image features which were not easily transferred to mobile devices.

In this study, we aimed to overcome these shortcomings in two aspects: (1) in order to improve the expression ability of the network model, the deep variant residual network (VRNet) is proposed to replace the original ResNet; (2) in order to enhance the feature extraction ability of the network, the squeeze-and-excitation (SE) module integrates the attention mechanism to construct the deep SE variant residual network (SE-VRNet); (3) In order to be generalized in mobile devices, the method in this paper emphasizes lightweight.

The remainder of this manuscript is organized as follows. Section II reviews some fundamentals and related work. Section III describes the proposed method in detail. Section IV explains the extended dataset and the method of image preprocessing. Simulation results and evaluation are presented in Section V. Finally, the conclusions are drawn in Section VI.

II. FUNDAMENTALS AND RELATED WORK
Deep ResNet and SE module with attention mechanism are the important components of the proposed model in this study. A brief review of these frameworks is given below.
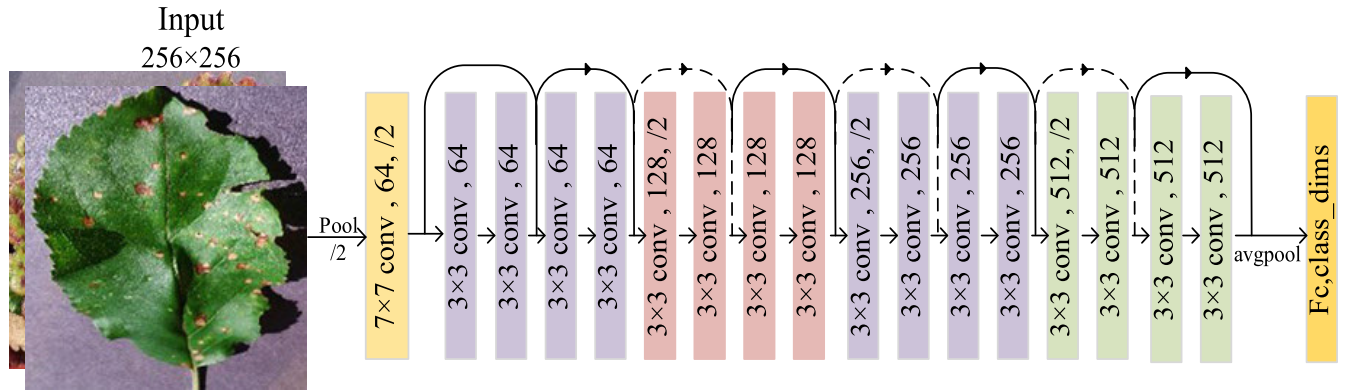
FIGURE 2. The specific structure of ResNet18.

A. DEEP RESIDUAL NETWORK

The deep ResNet was proposed by He et al. [20] in 2015. They trained a deep ResNet with 18 to 152 layers by adding a residual cell structure to the neural network, and won the 2015 ILSVRC image classification task competition. ResNet managed to outperform the 2014 network models VGGNet and GoogleNet while having fewer parameters than VGGNet. The unique residual unit structure in ResNet can also effectively improve the training speed of the network, which plays a very important role in improving the performance of the network.

The ResNet algorithm makes excellent use of the idea of Highway Network [21] and adds multiple skip connection layers to the network, so that the original input information can complete nonlinear transformation and transfer downward from the original structure.

For the residual part, the network is linear in the forward propagation process, and the upper branch road (see Fig. 1) is used for learning the residual map ($f(x)$), and another road carries the input map (x) to the end for recovering the maps $(x + f(x))$. The basic architecture is illustrated in Fig. 1.

Compared to the convolution operation in a general neural network, the residual block network can not only avoid the vanishing and exploding gradients, but also accelerate the convergence of the network. In general, different ResNet network can be constructed with different numbers of residual structural units and different convolution structures. For example, ResNet18 and ResNet34 were designed respectively based on different residual structural units, and their convolutional structures were fine-tuned and improved. In this work, the ResNet18 network stacked by residual structural units is shown in Fig. 2.

B. ATTENTION MECHANISM

Attention plays an important role in selecting critical information. Therefore, the CNN integrated with the attention mechanism (AM) can focus on more abstract feature information. For example, Woo et al. [22] proposed an advanced attention module that derives attention maps to refine the features in the input feature map. Experimental results

showed a consistent improvement of classification performance. In 2021, Gan et al. [23] proposed a global attention mechanism (GAM) for brain tumor segmentation, and experimental results showed that the segmentation accuracy of the model based on GAM was considerably improved. In recent years, AM has made significant breakthroughs in areas such as target detection and image classification, and it has been proven to be beneficial in improving model performance in many applications. Therefore, in order to improve the feature extraction capability of the model, the AM was also introduced into the residual network model.

The attention mechanism introduced in this paper is mainly channel attention module, whose structure can be implemented in a special way. The specific implementation is shown in Fig.3. The attention block adopts the method of feeding the composite values of statistical data (mean value, maximum value, standard deviation and mean square error) of each channel to ReLU. Finally the four feature data were concatenated to the channel attention diagram with the size of $1 \times 1 \times 1$.

III. THE PROPOSED METHOD

In this paper, a new Squeeze-and-Excitation Variant Residual Network (SE-VRNet) was proposed by integrating the improved convolution module, SE and attention mechanism. The SE-VRNet is shown in Fig. 3.

Inspired by the channel separation in ShuffleNetV2 [24], we changed the convolution structure of residual block, so as to overcome the defect of ResNet model caused by the reduction of feature number in the network during down-sampling. We divided the original channel into two channels, and used the combination of 1×1 convolution and average pooling layers to replace the original 3×3 convolution, and finally obtained the new VRNet. The improved convolution structure in VRNet is shown in the convolution module of Fig. 3. In the convolution structure, the pooling layer can compress the output size and reduce the resolution of the feature map. The 1×1 convolution in each channel can make a linear combination of multiple feature maps, realize the information integration of this channel, reduce the dimension of the
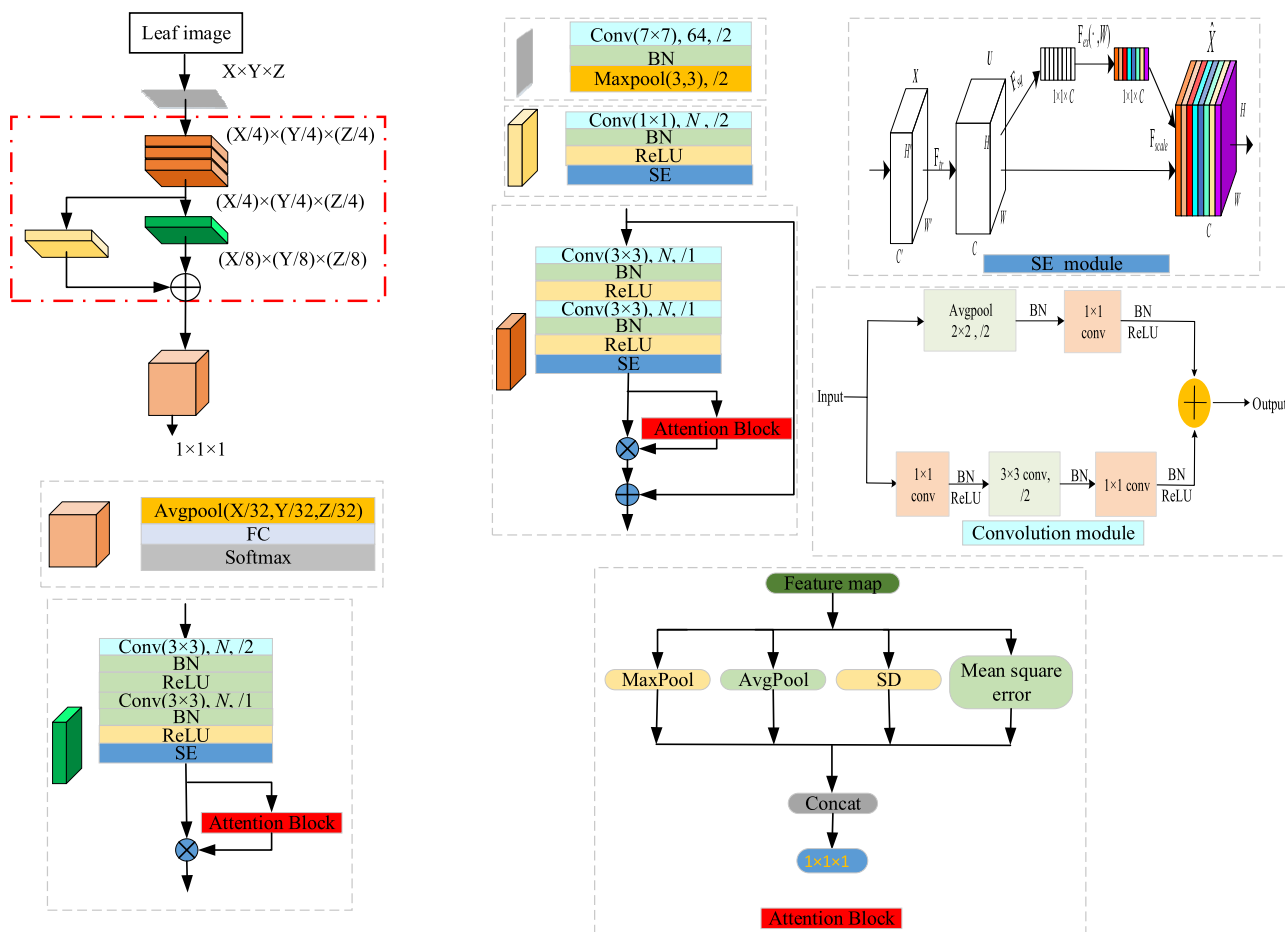
FIGURE 3. The structure of SE-VRNet.

network, and add a nonlinear excitation through the activation layer to improve the expression ability of the network. Therefore, this approach can increase the number of channels in the network and compensate for the information loss caused by the low resolution of the feature map. We added the Batch Normalization (BN) layer [25] and Rectified Linear Unit (ReLU) layer after each 1×1 convolution layer in the VRNet. Among them, BN layer can keep the data distribution within a certain range to accelerate the convergence of the network. The fusion of BN and ReLU can effectively increase the sparsity of the network, reduce the interdependence between the parameters of the network model, improve the generalization performance of the network, and suppress the occurrence of over-fitting.

In order to solve the problem that it is difficult to accurately extract crop leaf disease features, the SE module was added into the VRNet to get the improved SE-VRNet.

The SE module can be embedded in neural network models to form a new network structure. The SE module is originally found in the SE-Net structure proposed by Momenta team [26]. This module focuses on the relationship between the feature channels and introduces AM

on the feature channels. Therefore, SE-Net can independently obtain the importance degree of each feature channel and further improve the feature expression based on the obtained importance degree, while suppressing the unimportant features. The key point is that the network model learns the feature weight values according to the loss function, which leads to better feature extraction ability and obtain better training results. In summary, SE-Net does all the work via the SE module. The SE module is shown in Fig. 3.

In the SE module, F_{tr} refers to the convolution operation of input X. Through the convolution layer, a three-dimensional $H \times W \times C$ matrix U was obtained. F_{sq} and Fex refer to squeeze and excitation operations, respectively. F_{scale} is the weighting operation of the original features of each channel.

The SE module realizes the feature extraction function through the following three operations.

1) The module compresses features along the dimension of space by extruding, so that all two-dimensional feature channels become a real number. That is, the $H \times W \times C$ input is converted into a simple $1 \times 1 \times C$ output, which is calculated using the formula shown in

TABLE 1. Types and quantities of samples in oridata (original plantVillage dataset).

Types of sample	Quantity	Types of sample	Quantity	Types of sample	Quantity
Healthy apple leaf	1645	Scab of apple	630	Rust of apple	275
Healthy blueberry leaf	1502	Black rot of apple	621	Powdery mildew of cherry	1052
Healthy cherry leaf	854	Gray spot of corn	513	Rust of corn	1192
Healthy maize leaf	1162	Leaf blight of corn	985	Black rot of grape	1180
Healthy soybean leaf	5090	Blackpox disease of grape	1383	Leaf blight of grape	1076
Healthy grape leaf	423	Spot disease of peach	2297	Yellow dragon disease of orange	5507
Healthy peach leaf	360	Spot disease of pepper	997	Late blight of potato	1000
Healthy pepper leaf	147	Early blight of potato	1000	Powdery mildew of pumpkin	1835
Healthy potato leaf	152	Tipburn of strawberry	1109	Spot disease of tomato	2127
Healthy strawberry leaf	456	Early blight of tomato	1000	Late blight of tomato	1909
Healthy tomato leaf	1591	Leaf mildew of tomato	952	Spot blight of tomato	1771
Healthy raspberry leaf	371	Leaf acariasis of tomato	1676	Ring spot of tomato	1404
Mosaic of tomato	373	Yellow smut of tomato	5357		

Equation 1.

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{W} \sum_{j=1}^{H} u_c(i, j) \qquad (1)$$

2) As shown in Equation 2, after excitation operation, W is used to generate weight values for each feature channel, where W is used to learn the correlation between feature channels.

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (2)$$

where z is the result from squeeze in the previous step, (z, W) represents a fully connected layer. δ and σ represent ReLU and sigmoid activation operations, respectively.

3) The last operation is the reweight calibration. The weight value of the output represents the importance of each feature channel. The value is weighted to the previous feature through the multiplicative channel to complete the re-calibration of the original feature in the dimension of the feature channel. The process is shown in Equation 3.

$$\widehat{X} = F_{scole}(u_c, s_c) = s_c \times u_c \qquad (3)$$

where u_c represents the cth channel feature map in U. s_c is the feature weight of the cth channel.

Combining the advantages of the above two modules and attention mechanism, a comprehensive block diagram of leaf disease recognition model SE-VRNet was proposed in this paper, as shown in Fig. 3. The early characteristics of leaf diseases in many kinds of crops are not obvious, the lesion area is small and the location is scattered. The learning ability of traditional machine learning models cannot accurately extract the characteristics of these leaf diseases.

With the improved convolution module, SE and attention mechanism, the proposed SE-VRNet can not only focus on the relationship between multiple channels of the image, but also obtain the importance of each channel feature. Therefore, it can resolve the difficulty of extracting the early characteristics of leaf disease. In this paper, a large dataset will be used to train the model, and then real field data will be used to verify its accuracy. Finally, the model will be optimized to have the best performance.

This paper mainly studied the classification of various leaf diseases, so softmax activation function and cross entropy loss function were used. Cross entropy describes the difference between two probability distributions. However, the output of the model network in this paper was a vector, not a probability distribution, so this paper used softmax activation function to normalize the vector into a probability distribution, and then used cross entropy loss function to calculate the loss. In addition, the mean square error loss function was also used in the study. This loss function is simple and effective, which is suitable for mobile terminal devices.

IV. DATASET AND PREPROCESSING
The experimental data used in this study are respectively the original dataset OriData, the augmented dataset New-Data from PlantVillage, and the self-built leaf disease dataset SelfData. The image acquisition methods of OriData and SelfData are various, and some images are deformed, out of focus, motion blurred, distorted and non-standard due to the influence of different factors.

To mitigate the negative impact of the lack of standardization of the dataset on the recognition model, the statistical analysis and corresponding pre-processing of the images in the dataset is performed before the model training.

The data preprocessing includes image enhancement, image denoising and normalization to enable better generalization performance of the model.

A. PLANTVILLAGE DATASET
Crop leaf disease images in PlantVillage dataset includes 38 disease types of 14 crop categories, with a total of 54,305 leaf images. The types and quantities of the dataset are shown in Table 1.
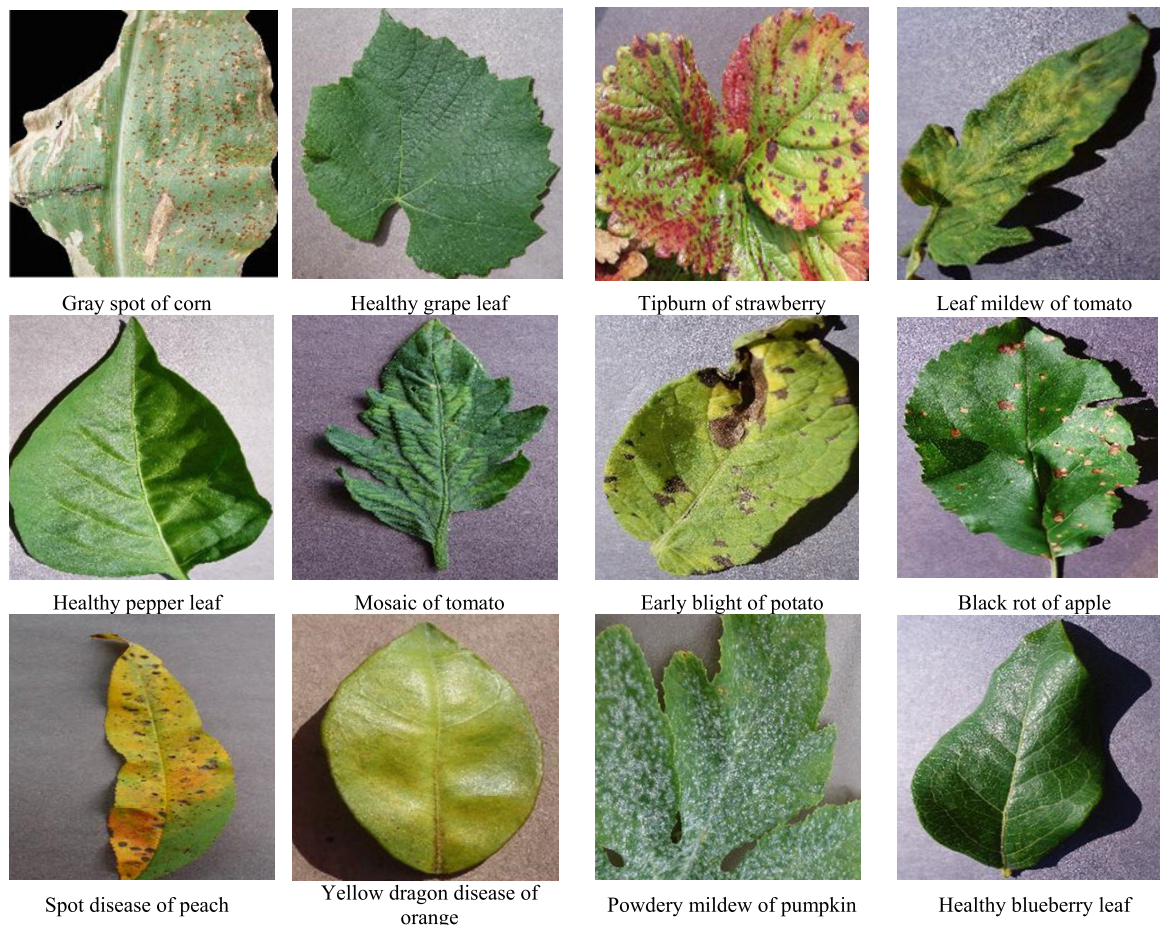
FIGURE 4. The partial leaf types in the OriData.

We find from Table 1 that the following problems mainly exist in the original PlantVillage dataset.

1) The distribution of disease categories in the dataset is particularly uneven. Tomato has the largest number of images, containing 15,569 images for 9 diseases, while potato healthy leaf and apple rust have the smallest number of images in the dataset, with only 152 and 275 images, respectively. In this case, the accuracy of the trained model tends to be somewhat biased, with higher accuracy for certain disease types with more image data.

2) The images dataset is of low quality and the image's background is not uniform. The shooting method, lighting conditions, and background are different for all images, and the difference between the leaf disease image and the background image is evident.

To address the issue of imbalanced data distribution in Ori-Data, we first clean them, and then use data enhancement method to obtain the augmented dataset NewData. After data cleaning and enhancement, the NewData dataset has a total of 87,867 image samples, with an average of about 2,300 disease images per category. Compared with OriData, NewData has a higher balanced distribution of disease categories.

TABLE 2. Types and quantities of samples in SelfData dataset.

Types of sample	Quantity	Types of sample	Quantity
Gray spot of corn	5193	Rice blast disease	5089
Healthy corn leaf	5161	Spot disease of apple	5340
Leaf blight of corn	5206	Brown spot of apple	5653
Rust of corn	5077	Gray spot of apple	4809
Brown spot of rice	5053	Mosaic of apple	4874
Healthy rice leaf	5124	Rust of apple	5692
Dicladispaarmigera	5131		

The images of partial leaf diseases are shown in Fig. 4.

B. SELFDATA DATASET
To enrich the diversity of the data and verify the applicability of the model in real-world scenarios, the SelfData dataset was constructed in this study according to the relevant data standards. This dataset is a combination of open datasets for apple, maize and rice, and contains images of
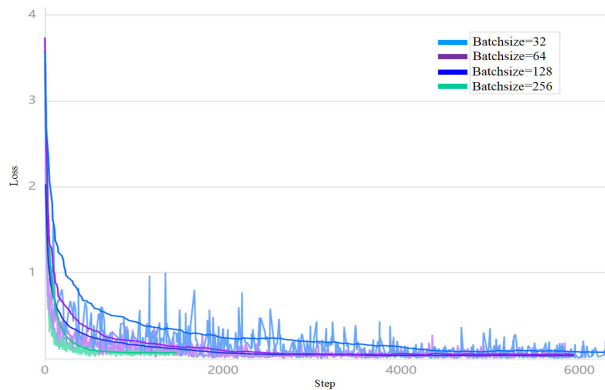
FIGURE 5. The change diagram of Loss curve.

13 diseases in three crop groups. The apple dataset was produced by Northwest A&F University and collected at Baishui apple orchard, Luochuan apple orchard and Qingcheng apple orchard of Northwest A&F University. The datasets of corn and rice were taken from the corn data in PlantVillage and the Kaggle data platform, respectively. Due to the small number of corn and rice samples in the original dataset, this study carried out data amplification for the original images of rice and maize by means of data enhancement. The total sample size of the SelfData dataset reaches 67,642, and the average number of images for each category is about 5,000. The information of the SelfData dataset is summarized in Table 2.

The Table 2 shows that the distribution of disease categories in SelfData is relatively balanced. The total number of apple image samples was the largest, accounting for 26, 368 of 5 diseases, while corn and rice images both contain 4 diseases, with 20,637 and 20,397 samples, respectively.

V. RESULTS AND DISCUSSION
A. DEVELOPMENT ENVIRONMENT AND EXPERIMENTAL SETTINGS
The proposed SE-VRNet was implemented under Windows 10 with an I7-8700 CPU, 32 GB of RAM, and two GeForce RTX 2080Ti GPUs. Pytorch 1.3.0 was chosen as the deep learning library.

The image input size was normalized to 256×256 and the batch size(BatchSize) was 128. The optimizer uses a phased learning descent method with an initial learning rate set to 0.025 and a weight decay set to 0.1. The learning rate decay strategy adopted the fixed decay strategy, which is one tenth of the previous stage after every 15 training epochs. The total number of training epochs was 60.

The datasets used in the work were the original PlantVillage (OriData), NewData and SelfData. PlantVillage is an open dataset commonly employed in the area of crop pest and disease identification. NewData was established by cleaning and expanding the PlantVillage database by the Author. SelfData contains leaf images taken from three apple orchards in order to improve the applicability of the model in the real-world scenarios.
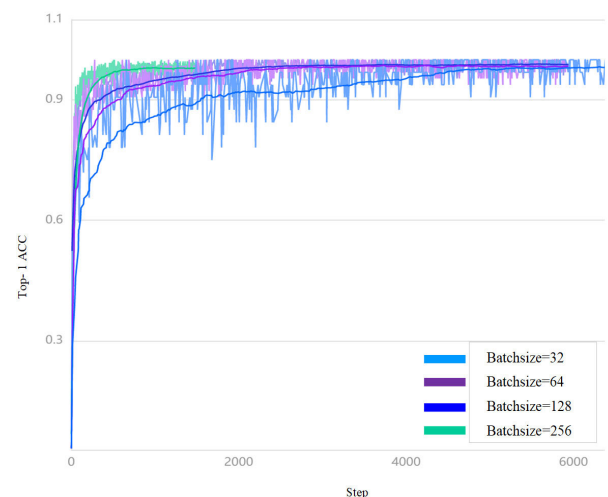

FIGURE 6. The change diagram of Top-1 accuracy curve.

B. THE INFLUENCE OF BATCH SIZE ON MODEL TRAINING
During the training of DL models, the batch size was an essential parameter that affects model optimization and training speed, which referred to the number of samples read each time during network training [27]. Therefore, in order to explore the impact of BatchSize on model training, this study first built a ResNet18 model based on the residual structure, and then set the batch size to 32, 64, 128, and 256, respectively. The loss and Top-1 accuracy curves as they changing during training are shown in Fig. 5 and Fig. 6, respectively.

They show that in the training process, the increase of the BatchSize can effectively improve the training speed and accelerate the convergence of the network, but this will also consume additional computing resources, and when the network training reaches a certain level, the overall performance of the network tends to be stable. Therefore, BatchSize is set to 128 in a compromise that not only efficiently saves the computational cost, but also speeds up model training.

C. CROSS VALIDATION
In this study, permutation tests were conducted for each group test dataset, and then the accuracy of all leaf diseases was computed by using cross validation (CV). Result of permutation tests showed that the order of test dataset almost did not affect the accuracy of the detection. According to the results of previous studies and the amount of dataset, this study selected a 10 fold CV.

In the 10 fold CV, firstly, the datasets (such as OriData, NewData and SelfData) are randomly partitioned into 10 groups. Of the 10 groups, 9 groups are chosen as the training set, and the remaining one is used as the test data. The validation is repeated 10 times such that each group is used exactly once as test data. The final estimation is produced by averaging the 10 results during the validation. The classifier predicted the label of the subjects that were left out, and the

TABLE 3. Comparative results of different training methods on different datasets.

Method	Dataset	Model	P	R	F_1	Top-1	Top-3	Model size
No pre-training	OriData	ResNet18	97.20%	97.19%	97.18%	97.19%	99.67%	64.2 MB
		ResNet34	97.23%	97.21%	97.20%	97.21%	99.73%	122.0 MB
	NewData	ResNet18	98.52%	98.50%	98.49%	98.50%	99.88%	64.2 MB
		ResNet34	98.48%	98.46%	98.46%	98.46%	99.90%	122.0 MB
Pre-training	OriData	ResNet18	98.02%	98.00%	98.00%	98.00%	99.85%	64.2 MB
		ResNet34	98.03%	98.02%	98.02%	98.02%	99.87%	122.0 MB
	NewData	ResNet18	98.84%	98.82%	98.82%	98.82%	99.95%	64.2 MB
		ResNet34	98.50%	98.49%	98.49%	98.51%	99.93%	122.0 MB

TABLE 4. Comparison of classification results for different models.

Models	NewData		SelfData		Model size	Execution time	Running memory
	Top-1	Top-3	Top-1	Top-3			
GoogLeNet	91.19%	91.83%	84.26%	94.49%	050.3 MB	73.54 ms	603 MB
AlexNet	95.64%	99.66%	89.74%	99.44%	333.7 MB	350.21 ms	2880 MB
VGG16	98.74%	99.92%	95.61%	99.82%	790.5 MB	700.78 ms	6000 MB
Xception39	92.12%	93.10%	90.26%	96.23%	039.0MB	49.62 ms	480 MB
Xception145	98.68%	99.56%	94.32%	99.50%	145.0MB	185.12 ms	1740 MB
ResNet18	98.02%	99.87%	94.17%	99.65%	064.2 MB	78.34 ms	769 MB
VRNet18	99.10%	99.95%	95.07%	99.87%	064.6 MB	82.11 ms	780 MB
SE-VRNet18	99.38%	99.98%	95.20%	99.88%	065.0 MB	86.89 ms	798 MB
ResNet34	98.17%	99.93%	93.21%	99.45%	122.0 MB	126.17 ms	1464 MB
VRNet34	99.49%	99.97%	94.29%	99.66%	125.3 MB	129.25 ms	1512 MB
SE-VRNet34	99.73%	99.98%	95.71%	99.89%	119.8 MB	113.38 ms	1016 MB

accuracy of detection was assessed. Finally, total accuracy was computed for each leaf disease category.

D. COMPARATIVE RESULTS
In this study, the ResNet18 and ResNet34 models were used to perform comparative experiments on OriData and NewData. The performance of ResNet18 and ResNet34 for different datasets were evaluated by Precision (P), Recall (R, Sensitivity), F_1 and accuracy. They are defined as:

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

$$F_1 = \frac{2PR}{P + R} \tag{6}$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{7}$$

where TP (True positive) is the number of successfully detected diseased leaf, TN (True negative) is the number of successfully detected healthy leaf, FP (False positive) is the number of healthy leaf that were mis-detected as diseased leaf, and FN (False negative) is the number of diseased leaf that were mis-detected as healthy leaf. Top-1 accuracy is the accuracy of the top predicted category consistent with the

actual category, and Top-3 accuracy is the accuracy of the top three predicted categories that contain the actual category. P is the ratio of true positives over all predicted positives and R is the ratio of true positives over all positives. F_1 takes both P and R into account and is usually a better measure to use when one of P and R is high and the other is low, as it balances both parameters.

To verify the performance of the transfer learning [28] from pre-trained model, extensive tests were conducted using OriData and NewData, and the comparative results were presented in Table 3.

The pre-training in Table 3 was the transfer learning of the pre-trained model on ImageNet using ResNet.

By analyzing the results in Table 3, it can be seen that the model accuracy was improved to a certain extent by using the pre-training for network migration learning. The performance of the model was further effectively improved by using the augmented dataset NewData and adopting a more balanced sample distribution.

Therefore, in the subsequent comparative study for different network algorithms, this paper directly used the augmented dataset NewData and applied the pre-training method to train the proposed network. We proposed the modified model VRNet and SE-VRNet on the base of ResNet.
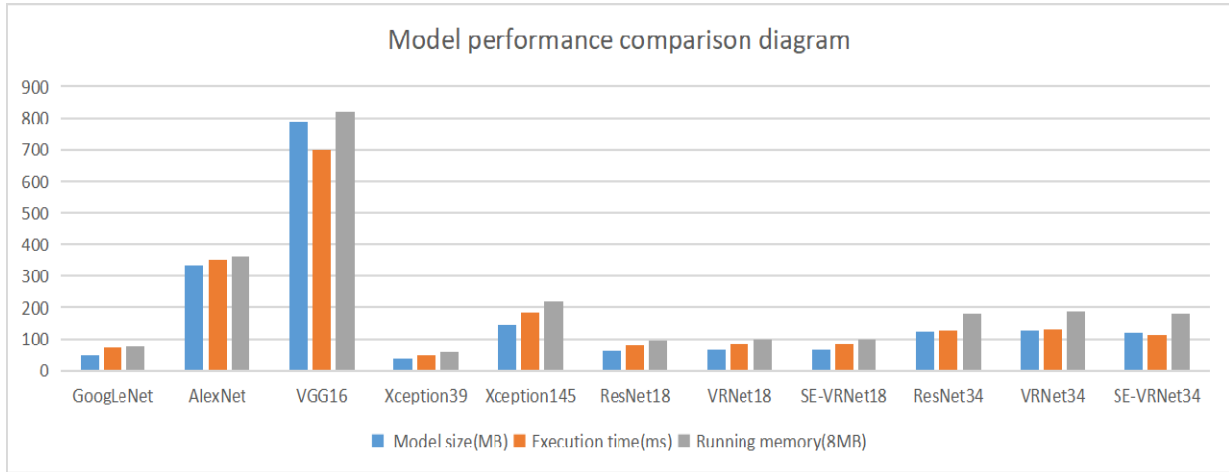
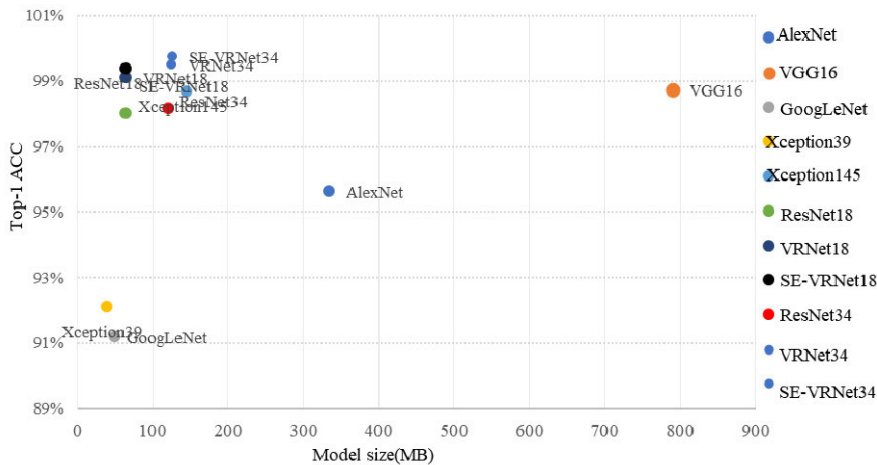**FIGURE 7.** The model performance comparison diagram.

FIGURE 8. The scatter plot of model performance.

To validate their performance and lightweight class [29], we compared them with AlexNet, VGG, Xception and GoogLeNet models.

Table 4 shows their comparisons of different models on the NewData and SelfData. The results such as execution time and running memory are obtained on the basis that the image input size of leaf disease after normalization processing is 256×256 and the experimental hardware. The above results will vary with different input size and hardware platform.

The model performance comparison diagram of size, execution time and running memory is shown in Fig. 7. In order to show the comparison between the models' sizes and model accuracy more vividly, we use the scatter diagram for visualization, as shown in Fig. 8

As can be seen from Table 4, the GoogLeNet model has the worst performance, followed by AlexNet. The performance of VGG16 on the NewData is better than that of ResNet, and the performance of VGG16 on the SelfData is the best, but its size is 790.5 MB, and its running memory is the largest. The good news is that the SE-VRNet34 has the best overall

performance. Compared to ResNet34, SE-VRNet34 shows significant improvement on both crop disease datasets. The best accuracy of the NewData was 99.98% and the Top-1 accuracy of the SelfData increased by 2.5 percentage points. Meanwhile, the size of the SE-VRNet34 model is 119.8MB. Compared to other models, such as VGG16 and Xception145, SE-VRNet34 is lightweight and can be easily transplanted to mobile terminals. There are a few promising lightweight design methods [30], [31] that have been proposed. Some improvement works of such methods are left for our future research.

For the tests of NewData and SelfData, the accuracy of SE-VRNet proposed in this manuscript is superior to the ResNet model in Top-1 accuracy and Top-3 accuracy, and the advantage of SE-VRNet model is outstanding.

VI. CONCLUSION

Compared with articles [32], [33], [34], our experiments demonstrated the effectiveness and versatility of the crop leaf disease recognition model SE-VRNet. This model proposed

in the paper improves the recognition accuracy on crop leaf disease with the NewData and SelfData datasets. Due to the addition of a large number of orchard infield leaf images to the training data, the model also improves the accuracy of infield leaf disease detection. In addition, the model size is only 119.8MB, so it is lightweight and is more suitable for mobile terminals and convenient for crop leaf disease detection. The method proposed in this paper not only improves the recognition accuracy, but also realizes the lightweight of the model. Therefore, compared with the traditional model, this model has the significance of popularization.

REFERENCES

[1] Y. Xu, S. Kong, Z. Gao, Q. Chen, Y. Jiao, and C. Li, "HLNet model and application in crop leaf diseases identification," *Sustainability*, vol. 14, no. 14, p. 8915, Jul. 2022, doi: 10.3390/su14148915.

[2] A. Pandey and K. Jain, "A robust deep attention dense convolutional neural network for plant leaf disease identification and classification from smart phone captured real world images," *Ecol. Informat.*, vol. 70, Sep. 2022, Art. no. 101725, doi: 10.1016/j.ecoinf.2022.101725.

[3] C. Wang, Y. Tang, X. Zou, W. SiTu, and W. Feng, "A robust fruit image segmentation algorithm against varying illumination for vision system of fruit harvesting robot," *Optik*, vol. 131, pp. 626–631, Feb. 2017.

[4] A. Kamilaris, F. Gao, F. X. Prenafeta-Boldu, and M. I. Ali, "Agri-IoT: A semantic framework for Internet of Things-enabled smart farming applications," in *Proc. IEEE 3rd World Forum Internet Things*, Reston, VA, USA, Dec. 2016, pp. 442–447.

[5] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018.

[6] N.-R. Zhou, T.-F. Zhang, X.-W. Xie, and J.-Y. Wu, "Hybrid quantum–classical generative adversarial networks for image generation via learning discrete distribution," *Signal Process., Image Commun.*, vol. 110, Jan. 2023, Art. no. 116891, doi: 10.1016/j.image.2022.116891.

[7] H. Mukhtar, M. Z. Khan, M. U. G. Khan, and H. Younis, "Wheat disease recognition through one-shot learning using fields images," in *Proc. Int. Conf. Artif. Intell. (ICAI)*, Apr. 2021, pp. 229–233.

[8] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep learning for tomato diseases: Classification and symptoms visualization," *Appl. Artif. Intell.*, vol. 31, no. 4, pp. 299–315, 2017.

[9] C. DeChant, T. Wiesner-Hanks, S. Chen, E. L. Stewart, J. Yosinski, M. A. Gore, R. J. Nelson, and H. Lipson, "Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning," *Phytopathology*, vol. 107, no. 11, pp. 1426–1432, Nov. 2017.

[10] R. Gandhi, S. Nimbalkar, N. Yelamanchili, and S. Ponkshe, "Plant disease detection using CNNs and GANs as an augmentative approach," in *Proc. IEEE Int. Conf. Innov. Res. Develop. (ICIRD)*, Bangkok, Thailand, May 2018, pp. 1–5.

[11] A. Elhassouny and F. Smarandache, "Smart mobile application to recognize tomato leaf diseases using convolutional neural networks," in *Proc. Int. Conf. Comput. Sci. Renew. Energies (ICCSRE)*, Jul. 2019, pp. 1–4.

[12] G. L. Manso, H. Knidel, R. A. Krohling, and J. A. Ventura, "A smartphone application to detection and classification of coffee leaf miner and coffee leaf rust," 2019, *arXiv:1904.00742*.

[13] A. Picon, A. Alvarez-Gila, M. Seitz, A. Ortiz-Barredo, J. Echazarra, and A. Johannes, "Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild," *Comput. Electron. Agricult.*, vol. 161, pp. 280–290, Jun. 2019.

[14] M. Toseef and M. J. Khan, "An intelligent mobile application for diagnosis of crop diseases in Pakistan using fuzzy inference system," *Comput. Electron. Agricult.*, vol. 153, pp. 1–11, Oct. 2018.

[15] M. Loey, A. Elsawy, and M. Afify, "Deep learning in plant diseases detection for agricultural crops: A survey," *Int. J. Service Sci., Manage., Eng., Technol.*, vol. 11, no. 2, pp. 41–44, 2020.

[16] G. Geetha, S. Samundeswari, G. Saranya, K. Meenakshi, and M. Nithya, "Plant leaf disease classification and detection system using machine learning," *J. Phys., Conf. Ser.*, vol. 1712, no. 1, Dec. 2020, Art. no. 012012.

[17] M. V. Applalanaidu and G. Kumaravelan, "A review of machine learning approaches in plant leaf disease detection and classification," in *Proc. 3rd Int. Conf. Intell. Commun. Technol. Virtual Mobile Netw. (ICICV)*, Feb. 2021, pp. 716–724.

[18] C.-C. Chen, J. Y. Ba, T. J. Li, C. C. K. Chan, K. C. Wang, and Z. Liu, "EfficientNet: A low-bandwidth IoT image sensor framework for cassava leaf disease classification," *Sensors Mater., Int. J. Sensor Technol.*, vol. 33, no. 11, pp. 4031–4044, 2021.

[19] S. S. Harakannanavar, J. M. Rudagi, V. I. Puranikmath, A. Siddiqua, and R. Pramodhini, "Plant leaf disease detection using computer vision and machine learning algorithms," *Global Transitions Proc.*, vol. 3, no. 1, pp. 305–310, Jun. 2022.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[21] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," 2015, *arXiv:1507.06228*.

[22] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, German, Sep. 2018, pp. 8–14.

[23] X. Gan, L. Wang, Q. Chen, Y. Ge, and S. Duan, "GAU-Net: U-Net based on global attention mechanism for brain tumor segmentation," *J. Phys., Conf. Ser.*, vol. 1861, no. 1, Mar. 2021, Art. no. 012041.

[24] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2018, pp. 116–131.

[25] L. Zhang, G. Zhou, C. Lu, A. Chen, Y. Wang, L. Li, and W. Cai, "MMDGAN: A fusion data augmentation method for tomato-leaf disease identification," *Appl. Soft Comput.*, vol. 123, Jul. 2022, Art. no. 108969.

[26] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141.

[27] U. K. Lilhore, A. L. Imoize, C.-C. Lee, S. Simaiya, S. K. Pani, N. Goyal, A. Kumar, and C.-T. Li, "Enhanced convolutional neural network model for cassava leaf disease identification and classification," *Mathematics*, vol. 10, no. 4, p. 580, Feb. 2022.

[28] X. Fan, P. Luo, Y. Mu, R. Zhou, T. Tjahjadi, and Y. Ren, "Leaf image based plant disease identification using transfer learning and feature fusion," *Comput. Electron. Agricult.*, vol. 196, May 2022, Art. no. 106892.

[29] Y. Li, D. Zhang, and D.-J. Lee, "IIRNet: A lightweight deep neural network using intensely inverted residuals for image recognition," *Image Vis. Comput.*, vol. 92, Dec. 2019, Art. no. 103819.

[30] S. Qin and S. Liu, "Efficient and unified license plate recognition via lightweight deep neural network," *IET Image Process.*, vol. 14, no. 16, pp. 4102–4109, Dec. 2020.

[31] L. Zhao and L. Wang, "A new lightweight network based on MobileNetV3," *KSII Trans. Internet Inf. Syst.*, vol. 16, no. 1, pp. 1–15, 2022, doi: 10.3837/tiis.2022.01.001.

[32] Y. A. Nanehkaran, D. Zhang, J. Chen, Y. Tian, and N. Al-Nabhan, "Recognition of plant leaf diseases based on computer vision," *J. Ambient Intell. Hum. Comput.*, no. 9, pp. 1–18, Sep. 2020, doi: 10.1007/s12652-020-02505-x.

[33] J. Chen, J. Chen, D. Zhang, Y. A. Nanehkaran, and Y. Sun, "A cognitive vision method for the detection of plant disease images," *Mach. Vis. Appl.*, vol. 32, no. 1, Jan. 2021, Art. no. 31, doi: 10.1007/s00138-020-01150-w.

[34] J. Chen, W. Chen, A. Zeb, D. Zhang, and Y. A. Nanehkaran, "Crop pest recognition using attention-embedded lightweight network under field conditions," *Appl. Entomol. Zool.*, vol. 56, no. 4, pp. 427–442, Nov. 2021, doi: 10.1007/s13355-021-00732-y.

ZHIYONG XIAO received the master's degree in control theory and engineering from the East China University of Science and Technology, Shanghai, China, in 2008, and the Ph.D. degree in mechatronic engineering from the School of Mechatronic Engineering, Nanchang University. He was a Research Assistant with Jiangxi Agriculture University, from 2009 to 2020. He is currently an Assistant Professor with the Jiangxi Institute of Economic Administrators. His research interests include image processing, computer vision, and pattern recognition.

YONGGE SHI received the bachelor's degree in radio technology from Jiangxi Polytechnic University, Nanchang, China. From 1982 to 2018, he was a Lecturer, an Assistant Professor, and a Professor with Nanchang University, Nanchang. He is currently a Professor with the Gongqing Institute of Science and Technology, Jiujiang, China, where he is also the Head of the Department of Information Engineering. His research interests include communications and computer networks.

JIANPING XIONG received the bachelor's degree in electronics and information engineering and the master's degree in communications and information systems from Nanchang University, Nanchang, China, in 2002 and 2005, respectively. He is currently an Associate Professor with the Industrial Center, Shenzhen Polytechnic, Shenzhen, China. His research interests include image and signal processing, electronic technology, and deep learning.

GAILIN ZHU received the bachelor's degree in communication engineering and the master's degree in electronic communications and engineering from Nanchang University, Nanchang, China, in 2016 and 2020, respectively. From 2016 to 2017, he was employed with the Guangdong Telecom Planning and Design Institute, Guangzhou, China. He is currently with Zhuhai Unitech Power Technology Company Ltd., Zhuhai, China. His research interests include signal processing, deep learning, and pattern recognition.

JIANHUA WU received the bachelor's degree in information engineering from the Harbin Institute of Technology, Harbin, China, in 1982, the M.Sc. degree in communication and electronic systems from the South China University of Technology, Guangzhou, China, in 1985, and the Ph.D. degree in image and signal processing from the University of Poitiers, Poitiers, France, in 2005. He is currently a Professor with the Department of Electronic Information Engineering, Nanchang University, Nanchang, China. He has published more than 20 articles in academic journals, such as *Applied Soft Computing*, *Optics Communications*, *Optics and Laser Technology*. His research interests include image and signal processing, power quality signal analysis, medical imaging, and pattern recognition.

● ● ●