IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 February 2024, accepted 11 March 2024, date of publication 18 March 2024, date of current version 29 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3377230

RESEARCH ARTICLE

A Plant Leaf Disease Image Classification Method Integrating Capsule Network and Residual Network

XIN ZHANG [1,2,3], **YUXIN MAO** [1], **QI YANG** [1], **AND XUYANG ZHANG** [1]

[1] School of Mechanical Engineering, Shenyang Ligong University, Shenyang 110159, China
[2] Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China
[3] School of Software, Northeastern University, Shenyang 110169, China

Corresponding authors: Xin Zhang (zx_sut@126.com) and Qi Yang (yangqi@sylu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 12202285, in part by the Project of the Education Department of Liaoning Province under Grant LJKZ0258, and in part by the Doctoral Research Foundation of Liaoning Science and Technology Department under Grant 2022-BS-187.

ABSTRACT In response to the challenge that traditional convolutional neural networks face in effectively capturing the posture and spatial relationships of plant disease lesions on leaves, leading to issues of low recognition accuracy and poor robustness, this paper proposes a plant leaf disease image classification method that integrates capsule networks and residual networks. Firstly, by optimizing and refining the traditional residual network (ResNet), the initial convolutional layer of ResNet is enhanced by replacing its kernel with a concatenation of 3×3 small convolutional kernels, aiming to more effectively extract features of plant leaf lesions. Subsequently, a channel attention mechanism is introduced into the residual block to heighten the model's focus on crucial features. Finally, the improved ResNet is effectively integrated with the capsule network (CapsNet). The initial pooling layer of ResNet is removed to reduce the loss of positional information. The output of the third residual module of ResNet is then connected with CapsNet, fully leveraging the strengths of both networks to enhance the model's robustness. Train and test the model on the PlantVillage, AI Challenger 2018, and Tomato Leaf Disease datasets and conduct comparative experiments with other typical classification models. The proposed SE-SK-CapResNet model has demonstrated a remarkable ability to accurately recognize diseased leaves, achieving an accuracy rate of 98.58%, 95.08%, and 97.19%, respectively. Furthermore, this model has exhibited superior performance in image rotation transformations and classification compared to traditional network models. These experimental results suggest that the SE-SK-CapResNet model is a promising solution for the detection of diseased leaves in the field of agriculture.

INDEX TERMS Deep learning, plant leaf disease, image recognition, capsule network, residual network.

I. INTRODUCTION

One of the frequent issues in agricultural production is plant disease, which has a negative effect on crop quality and productivity [1]. Traditional plant disease identification methods have some apparent limitations [2], usually relying on manual inspection and being limited by professional knowledge and experience, making it difficult to achieve fast and accurate disease identification.

In recent years, deep learning based on convolutional neural networks (CNNs) has been prominent in image recognition, and improved algorithms based on CNNs have now made significant progress in plant disease detection [3], [4], [5], [6], [7], [8]. For example, Gao et al. [9] enhanced the feature extraction ability of residual block by combining the improved two-branch attention module with residual block, and the accuracy rate of recognizing cucumber disease

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

reached 98.54%. Pandi et al. [10] proposed an improved CNN model, which effectively enlarged the range of the receptive field by replacing the traditional convolution with the dilated convolution, and the model could extract the image features more comprehensively. An accuracy of 96.5% was achieved in identifying rice leaf diseases. Liu et al. [11] proposed a multi-scale fusion network model based on CNN, which was improved based on EfficientNet, and by introducing the attention mechanism module, the recognition accuracy of cassava diseased leaves was significantly improved. Bezabih et al. [12] effectively enhanced the extraction of pepper leaf disease features by concatenating AlexNet and VGG16 feature extraction networks with an accuracy of 95.82%.

Despite the remarkable progress made by CNN in the field of plant disease detection, there are some limitations of traditional CNNs when dealing with plant leaf disease images [13], [14], [15], such as sensitivity to rotation, occlusion, and pose changes, which makes it challenging to capture pose and spatial relationships of diseased spots on plant diseased leaves. This means that the recognition performance of the network decreases when the placement angle, degree of twisting, or bending of the diseased leaf changes during the photographing process. Luo et al. [16] pointed out that the feature extraction method in CNNs suffers from the problem of information loss. To solve this problem, an improved multi-scale feature fusion ResNet model is proposed. The model adjusts the structure of the residual block and reduces the information loss by improving the downsampling method. The model's accuracy in recognizing apple diseases is 94.99% when facing a complex background. Pandian et al. [17] constructed a deep ResNet containing 197 layers. Using data enhancement techniques such as scaling, flipping, and random rotation, the model is better adapted to images with different scales, viewpoints, and poses, and the robustness and generalization ability of the model is improved. Alghamdi et al. [18] improved the VGG16 model by introducing spatial pyramid pooling blocks and densely connected blocks. This improvement enhanced the model's ability to perform multilevel and multiscale feature extraction and achieved 93.79% accuracy on the PlantVillage dataset.

The introduction of capsule networks addresses the limitations inherent in CNNs [19]. Capsule neural networks provide richer feature representations by introducing capsule units that can represent the feature vectors of an object, including pose information, scale information, and so on. Compared with traditional CNNs, capsule neural networks have a more vital ability to model spatial relationships and adaptability to changes [20], [21], [22]. However, capsule networks also have limitations [23], such as being relatively simple in the feature extraction network, making it challenging to extract deep features effectively. Samin et al. [24] proposed a CapsNet-based model that extracts features through multi-layer convolution and acquires the pose information of the plant in the image through the capsule layer, which achieves an accuracy of 93.07% in the test dataset. Xu et al. [25] enhanced the model's multi-scale feature extraction capability by introducing the Inception module instead of the convolution layer of CapsNet, which identifies the apple leaf disease with an accuracy of 93.11%.

To further improve the performance of the plant leaf disease recognition network, this paper proposes a model that fuses capsule and residual networks to enhance the network's performance in recognizing plant leaf diseases. To better adapt to the feature extraction of plant disease leaf spots, ResNet is optimized with a 3×3 small convolutional kernel (SK) in the first layer, and a channel attention mechanism is introduced to improve the model's attention to critical features. In terms of feature description, ResNet performs outstandingly. In contrast, CapsNet excels in extracting posture information. By fusing these two networks, the advantages of each are fully realized, and the model's robustness and classification performance are enhanced. The following are this paper's primary contributions:

(1) Our proposed model significantly improves plant leaf disease classification performance Compared with classical classification models such as AlexNet, VGG, GoogLeNet, etc.

(2) The improved CNN is combined with CapsNet to construct an integrated model. The model effectively captures plant disease leaf spot pose and spatial relationship information.

(3) The proposed model was trained using datasets collected in laboratory and field environments, both of which achieved high accuracy and verified the robustness and applicability of the model under different datasets.

II. FUNDAMENTALS AND RELATED WORK

Before introducing the proposed model, we will briefly review the key components, including the capsule network, the residual network, and the SE attention mechanism. These components are integrated to form a synergistic overall structure that gives the model unique properties and advantages.

A. CAPSULE NETWORK

Initially proposed by Hinton et al. [26], the capsule network (CapsNet) overcomes the problem of losing position information by the pooling layer in traditional CNNs by introducing capsule units. Traditional CNNs usually achieve the desired results when recognizing images similar to the training dataset. However, the recognition performance of traditional CNNs tends to suffer when there is some degree of rotation, distortion, or change in the relative positions of elements in the image. Traditional CNNs focus on local feature extraction and lack robustness to overall image changes. At the same time, pooling operations may further blur the positional information [27], compromising performance for images with significant spatial variations.

The development of capsule networks in the field of plant disease identification has demonstrated its unique

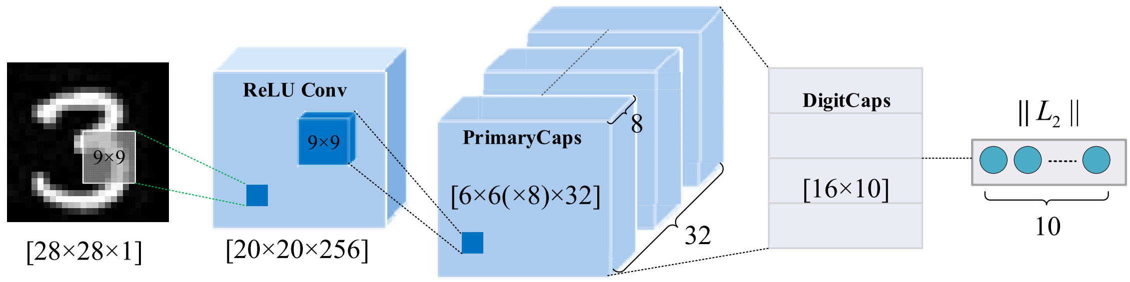

FIGURE 1. Capsule network structure diagram.

advantages [28]. Several studies have demonstrated the effectiveness of capsule networks in the accurate detection and classification of plant diseases [29], [30], [31], [32], [33], [34], [35]. Compared to convolutional neural networks, capsule networks can capture the pose and spatial relationship of plant leaf spots more efficiently, which improves the accuracy of disease detection. The hierarchical structure of capsule networks helps to learn more hierarchical feature representations and improves the ability to distinguish between different disease classes. Although applications in the field of plant diseases are still under research, capsule networks are expected to be an important tool for improving the performance and robustness of plant disease recognition.

Capsule networks have some limitations in a number of aspects, including the relative simplicity of the feature extraction network and limitations for large-size images [23]. First, the feature extraction network of capsule networks is relatively simple. Convolutional layers are often used as feature extractors in traditional capsule networks, and this structure performs well when dealing with simple image tasks. However, when dealing with more complex image tasks, the feature extraction network may be relatively shallow and difficult to capture deeper image features. Second, traditional capsule networks are usually designed to be trained and tested for fixed-size images. For example, the common MNIST dataset uses images of 28 $\times$ 28 pixels. This limitation means that capsule networks may struggle with images of larger sizes. Images with larger input sizes may need to be adapted and processed to fit the design of the capsule network.

The capsule network consists of a convolutional layer, a primary capsule layer, and a digital capsule layer. The capsule network structure is shown in Fig. 1. The convolutional layer is responsible for extracting local features. In contrast, the main capsule layer introduces capsule units to capture the image's spatial structure and pose information in a hierarchical relationship. The capsule units output the presence and attribute information of the features so that the network can retain the spatial information of the objects more accurately. The digital capsule layer then combines the outputs of different capsule units to form a comprehensive representation of the overall target. This structure allows CapsNet to exhibit greater robustness when dealing with images with significant spatial variations, especially in the presence of rotations, distortions, or changes in the relative positions of the elements, which improves recognition performance.

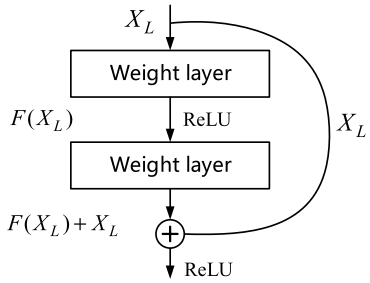

FIGURE 2. The structure of the residual block.

The dynamic routing mechanism is one of the critical features of CapsNet, which iteratively updates the weights to selectively assign the outputs of the capsules in the previous layer to the capsules in the next layer. This process is done automatically during training, gradually allowing the network to learn the correct routing method. Specifically, for the dynamic routing weight C_{ij}, a softmax function is used to make the sum of the weights 1. Then, using the weights obtained in the previous step, the outputs of the main capsule are weighted and summed to obtain the weighted sum S_j. Finally, a nonlinear activation function is used to obtain the output of the jth numerical capsule after the dynamic routing, v_j. The above three steps are looped and iterated until convergence. The corresponding equations are as follows:

$$C_{ij} = \frac{exp\left(b_{ij}\right)}{\sum_k exp\left(b_{ij}\right)} \tag{1}$$

$$S_j = \sum\nolimits_i C_{ij} \cdot u_{ij} \tag{2}$$

$$v_j = \frac{\left\|S_j\right\|^2}{1 + \left\|S_j\right\|^2} \cdot \frac{S_j}{\left\|S_j\right\|} \tag{3}$$

where b_{ij} is the initialized marginal probability, C_{ij} is the dynamic routing weight, u_{ij} is the output vector of the low-level capsule i, S_j is the input of the high-level capsule j, and v_j is the output of the high-level capsule j.

B. RESIDUAL NETWORK

The residual network (ResNet) model was proposed by He et al. [36], which significantly improves the performance

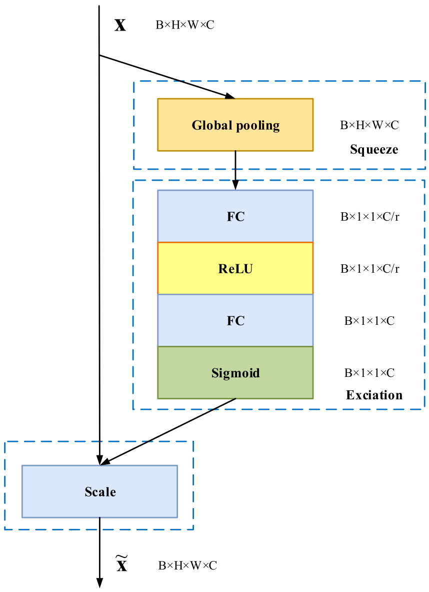

FIGURE 3. The structure of the SE module.

of deep networks with its unique residual connectivity compared to traditional CNNs. This mechanism enables the network to learn constant mappings easily, thus effectively solving the network degradation problem. The structure of the residual module of ResNet is shown in Fig. 2.

The residual block consists of the main path and the residual path. The main path learns the mapping of the input data, and the residual path learns the residuals of the input concerning that mapping $F(X_L, W_L)$. The residuals are added to the input to obtain the final output $H(X)$. This allows the network to learn the residual information directly, thus effectively solving the gradient vanishing problem. The process is shown in Equation 4.

$$H(X) = X_L + F(X_L, W_L) \tag{4}$$

where X_L is the L-layer feature map and W_L is the L-layer convolution parameter.

C. ATTENTION MECHANISM

Squeeze-and-excitation (SE) attention mechanism module: An essential part of improving neural network performance. The model's performance can be further enhanced by incorporating the SE attention mechanism in the residual block [37]. The structure of the SE module is shown in Fig. 3.

First, the features of each channel are compressed by a global average pooling operation. Compression operation compresses the size of the feature map from B×H×W×C (where H is the height, W is the width, and C is the number of channels) to B×1 × 1×C/r (r is the compression ratio) to obtain a global description vector. Next, the global description vector is mapped to a smaller dimension by two fully connected layers and processed by an activation function. Finally, the obtained weights are multiplied by the original feature map, which makes the network focus more on the essential features in the subsequent training process and realizes the dynamic scaling of features. Introducing the SE attention mechanism module can significantly improve the network's ability to perceive key features, thus improving the overall performance, especially in dealing with complex tasks and large-scale data with stronger performance and robustness.

III. THE PROPOSED METHOD

Residual networks and their family of derived networks are outstanding in image processing and feature description. They are easy to improve and optimize [38], especially since ResNet18 has a relatively small network depth, is suitable for medium-sized tasks, and has achieved remarkable results in many fields [39], [40], [41], [42]. By fusing the ResNet18 with CapsNet, the advantages of the two networks are thoroughly combined to improve further the performance in the plant leaf disease recognition task, thus achieving better recognition results. In this paper, we chose to improve the ResNet18 to make it a more suitable network for plant leaf disease recognition. The ResNet18 network structure is shown in Fig. 4.

For the recognition of plant leaf disease images, the first layer of the traditional ResNet18 network structure is a 7 × 7 convolutional kernel. As a larger convolutional kernel, its role is to achieve a larger range of feature extraction effects on the image [43]. However, for plant-diseased leaves, the leaf spot area is small, and all types of diseased leaves need a more fine-grained disease feature extraction in the spot area to enhance the model recognition effect on plant-diseased leaves.

Therefore, to be more suitable for plant disease leaf feature extraction, the 7 × 7 convolution kernel in the first layer of the traditional ResNet is changed to a combination of 3 × 3 convolution kernels that can extract features from a smaller area to obtain SK-ResNet18 (Small Kernel ResNet18). In addition, using small 3 × 3 convolutional kernels in series significantly reduces the model's parameters and enhances the neural network's discriminative ability. This improvement helps to extract more detailed and rich information, which can better capture the subtle features in the disease images and improve the classification accuracy of plant leaf disease images.

By introducing the SE attention mechanism module to the ResNet18 model, SE-ResNet18 is obtained, making the model more focused on key features and reducing reliance on unimportant features. The structure of the residual module with the added SE attention mechanism is shown in Fig. 5.

To overcome the limitations of the traditional convolutional network in extracting plant leaf spot pose information and the capsule network in terms of simplicity of the feature extraction network and adapting to the large-size plant disease input image, we propose the SE-SK-CapResNet network model with the improved fusion of residual network and capsule network. The network structure of this model is

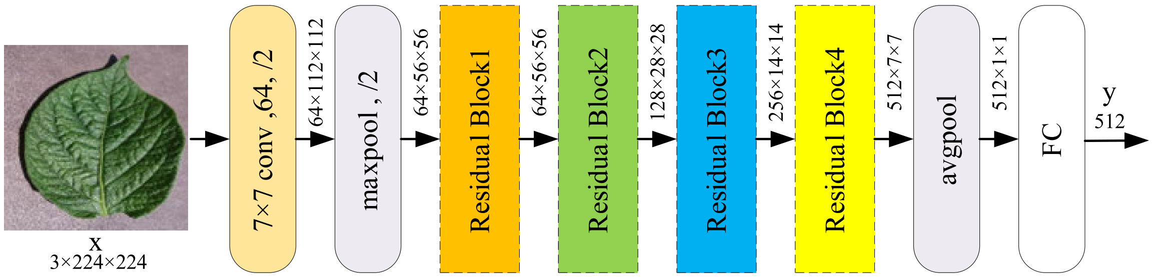

FIGURE 4. The structure of ResNet18.

shown in Fig. 6. By combining the two, the advantages of the improved residual network in feature extraction can be fully utilized, while the capsule network is used to extract the pose information of the object, thus improving the extraction ability of plant disease leaf features.

Conventional convolutional neural networks use pooling operations in the feature extraction process, which may lead to the loss of pose information of the target features. To solve this problem, we removed the SE-ResNet18 maximum pooling layer and the global average pooling layer to retain more low-level detail features, which is beneficial for the capsule network to extract the pose information of the object. The output of the third residual module of the improved residual network is connected to the capsule network, the parameters of the first two convolutional layers are adjusted, and the step spacing is set to 2 in both cases to ensure that the feature maps inputted to the first residual block are of the same size as those of the original network, so as to enable the model to adapt to the 224 × 224 plant disease images. Such improvement makes the SE-SK-CapResNet network model more advantageous and adaptable in plant disease recognition.

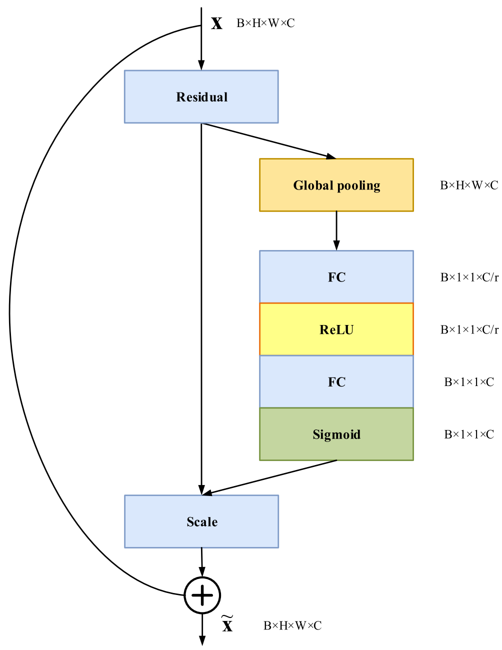

FIGURE 5. The structure of the SE residual block.

IV. DATASET AND PREPROCESSING

A. PLANTVILLAGE DATASET

This study utilizes a publicly available dataset from the PlantVillage project (www.plantvillage.org), containing a total of 54,303 images of healthy and diseased leaves of different plant species, among which images of diseased leaves of apple, potato, and corn were selected, including 630 images of apple scab, 622 images of apple black rot, 513 images of corn gray spot, 985 images of corn leaf blight, and 1,000 images of each of early and late blight of potato. A total of six disease leaf images of three species of plants selected from the dataset are shown in Fig. 7.

The dataset images are all 256 × 256 pixels, and the dataset is expanded by image enhancement techniques such as image flipping, gamma correction, and color enhancement to 1000 images for each category. Subsequently, the dataset was divided into training and test sets in the ratio of 4:1 to ensure the reliability of the experimental results. In addition, to maintain the consistency of the model inputs, we resized all the images to 224 × 224. We performed the normalization of the mean and standard deviation of each channel to enhance the stability and performance of the model training.

B. AI CHALLENGER 2018 DATASET

The dataset used in this study is derived from the AI Challenger 2018 competition, specifically the AI Challenger Plant Disease Recognition dataset [44]. The background of this dataset is set in a laboratory context and, similarly, it is a widely adopted public dataset. In comparison to the Plant Village dataset, it offers a more detailed categorization based on the severity of plant diseases. For this research, we selected grape leaf disease images from this dataset, including 336 images of healthy grape, 435 images of general grape black rot, 528 images of serious grape black rot, 577 images of general grape black measles, 478 images of serious grape black measles, and 720 images of grape leaf blight. Images with a part of disease severity not matching their labels were excluded from the dataset. Sample images of the six categories of grape leaf diseases in the dataset are shown in Fig. 8.

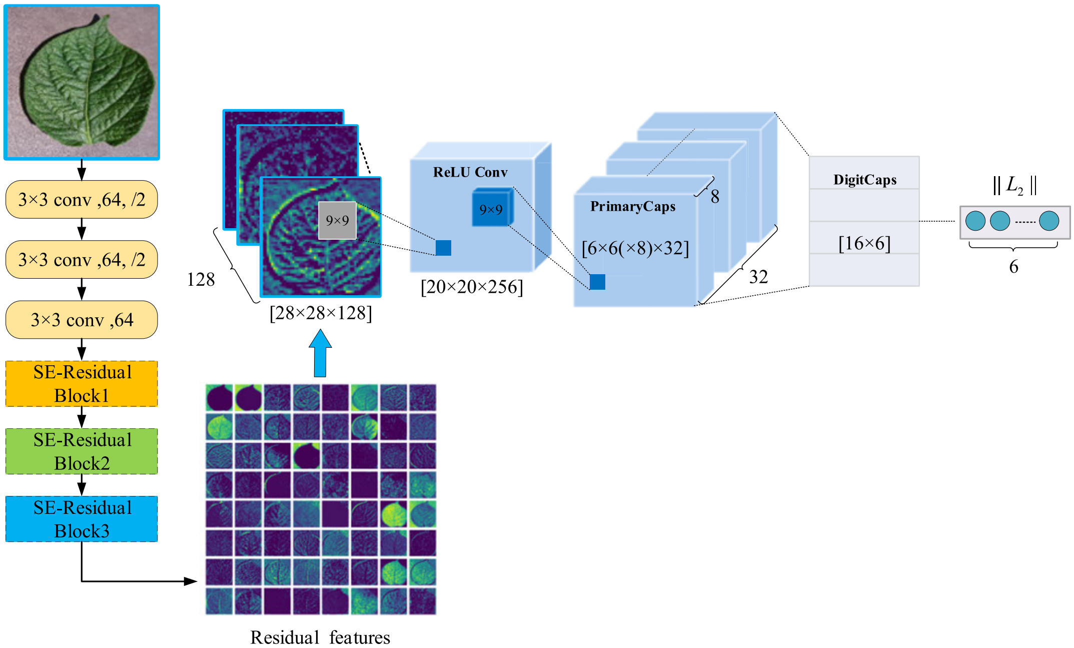

FIGURE 6. The structure of SE-SK-CapResNet.

The grape leaf disease dataset was augmented through image flipping, with 600 images selected for each category. Additionally, the dataset was divided into training and testing sets in a 4:1 ratio, and all images were resized to 224 × 224 pixels to ensure consistency across the dataset.

C. TOMATO LEAF DISEASE DATASET

The dataset is derived from the Kaggle website platform (https://www.kaggle.com/datasets/responsibleailab/crop -disease-ghana/data), and the original dataset is a series of annotated crop images containing images of leaf diseases on three different species of plants: corn, pepper, and tomato. The photos in this dataset were taken in a field setting, which makes the background more realistic and complex and more in line with the production environment of natural agriculture than the dataset collected in a laboratory setting. The tomato category was selected from the dataset. Then, the images were cropped to retain only the portion containing tomato leaf diseases, resulting in a new dataset that includes five different categories of diseased and healthy tomato leaves. The five disease classes include Early Blight, Late Blight, Leaf Curl, Mosaic, and Septoria Leaf Spot. Sample images of the six categories of leaves in the dataset are shown in Fig. 9.

The new dataset was augmented with the same data to have about 600 images per category. Similarly, the dataset is divided into training and testing sets in the ratio of 4:1, and all the images are resized to 224 × 224.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The version of the toolkit used for training in this study is PyTorch1.7, CUDA11.2, Python3.7, the operating system is Windows10 Professional, the hardware environment is

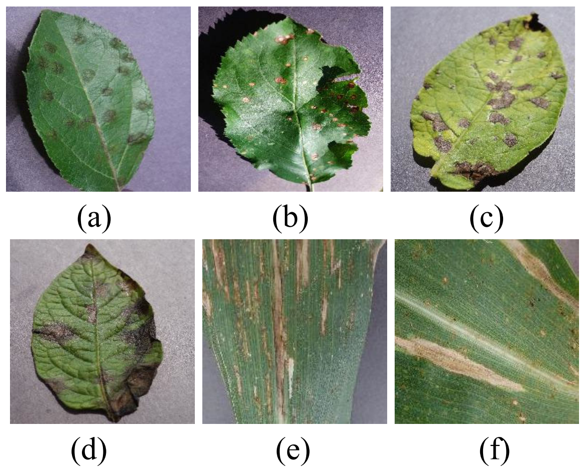

FIGURE 7. Image samples of PlantVillage Dataset: (a) Apple scab, (b) Apple Black rot,(c) Potato Early blight, (d) Potato Late blight, (e) Corn gray Leaf spot, (f) Corn Leaf blight.

NVIDIA GTX1060 for the GPU, and i7-8750 for the CPU. The number of capsules in the main capsule layer of the capsule network is set to 8, and the number of capsules in the digital capsule layer is set to 6. The number of dynamic routing iterations in the main capsule layer is set to 3, and the number of output channels in the main capsule layer and in the digital capsule layer, are set to 32 and 16, respectively. The training model epoch is 60 times, the bit size is 32, the initial learning rate is 0.0001, the cross-entropy loss function, and the Adam optimizer optimized the model.

In order to comprehensively evaluate the performance of the proposed model in this paper in the task of plant leaf

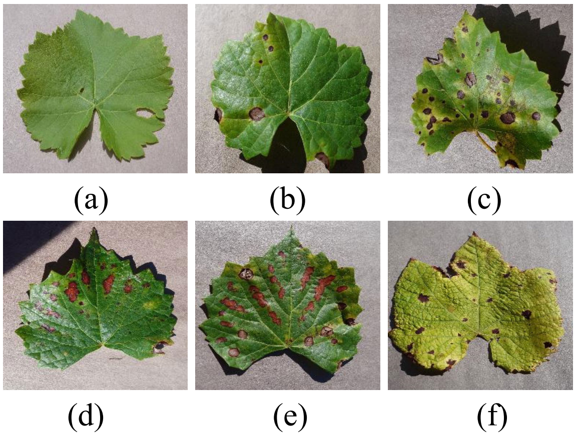

FIGURE 8. Image samples of grape leaf disease dataset: (a) Health leaf, (b) General grape black rot,(c) Serious grape black rot, (d) General grape black measles, (e) Serious grape black measles, (f) Grape leaf blight.

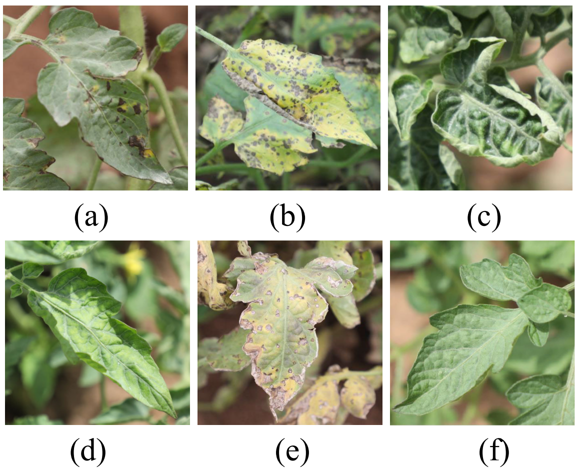

FIGURE 9. Image samples of tomato leaf disease dataset: (a) Early blight, (b) Late blight, (c) Leaf curl, (d) Mosaic, (e) Septoria leaf spot, (f) Health leaf.

disease image recognition, a series of key evaluation metrics are selected, including precision, recall, accuracy, F_1. They are defined as:

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

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

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{8}$$

where TP is the number of true positive instances that the model correctly predicted; FP is the number of false positive instances, which the model incorrectly predicts as positive; FN is the number of false negative instances, which the model incorrectly predicts as positive; and TN is the number of true negative instances that the model correctly predicts.

TABLE 1. Comparison of evaluation criteria for different improvement strategies.

Model	Precision	Recall	Accuracy	F1
ResNet18	93.83%	94.22%	93.83%	93.78%
SK-ResNet18	95.10%	95.08%	95.08%	95.07%
SE-ResNet18	97.05%	97.05%	97.05%	97.05%
CapResNet18	98.07%	98.07%	98.07%	98.07%
SE-SK-CapResNet18	98.58%	98.58%	98.58%	98.58%

In addition to focusing on the traditional classification performance metrics, this paper also integrates the model's inference time, parameter size, and memory footprint.

A. ABLATION EXPERIMENT

For different improvement strategies, including replacing the 7×7 convolutional kernel in the first layer of the traditional ResNet with a combination of 3×3 convolutional kernels to form the SK-ResNet18 model; introducing the SE attention mechanism into the ResNet18 model to improve the model's focus on the key features to form the SE-ResNet18 model; fusing ResNet with CapsNet to form the CapResNet model; and combining the improved residual network with CapsNet to obtain the SE-SK-CapResNet model. Ablation tests on the model in the PlantVillage dataset were carried out to confirm and assess the model's performance in the test set under various improvement techniques. The outcomes are displayed in Table 1.

By analyzing the results of the ablation experiments, we can observe the positive impact of each improvement strategy on the model performance. Changing the 7×7 convolutional kernel in the first layer to three small 3×3 convolutional kernels in series (SK-ResNet18) and introducing the SE attention mechanism (SE-ResNet18) improved the accuracy to 95.08% and 97.05%, respectively. After fusing the residual network with the capsule network (CapResNet18), the model performance is again significantly improved with an accuracy of 98.07%. Finally, the SE-SK-CapResNet18 model performed the best, achieving an accuracy of 98.58%, which verifies the performance advantage of the model improvement strategy proposed in this paper in plant leaf disease image recognition.

B. COMPARISON OF DIFFERENT ALGORITHMIC

The preprocessed dataset image dataset was input into the improved network model for training, and the SE-SK-CapResNet model with the highest accuracy of 98.58% was obtained after 60 training rounds. To verify the performance of the SE-SK-CapResNet model proposed in this paper in plant leaf disease image recognition, a comparative analysis is performed with other classical classification models (including AlexNet, GoogLeNet, VGG16, ResNet18, and EfficientNet) under the same experimental conditions. The model's performance in recognizing plant disease leaves is

TABLE 2. Comparison of evaluation criteria for different network models (PlantVillage dataset).

Model	Precision	Recall	Accuracy	F1	Parameter size	Inference time	Memory footprint
AlexNet	96.28%	96.25%	96.25%	96.24%	217.55 MB	2.10 ms	226.49 MB
GoogLeNet	98.10%	98.08%	98.08%	98.07%	39.37 MB	15.6 ms	134.28 MB
VGG16	96.00%	96.13%	96.00%	95.99%	512.26 MB	2.13 ms	731.41 MB
ResNet18	93.83%	94.22%	93.83%	93.78%	42.65 MB	3.02 ms	106.01 MB
EfficientNet	97.97%	97.92%	97.92%	97.91%	15.32 MB	9.89 ms	196.72 MB
SE-SK-CapResNet	98.58%	98.58%	98.58%	98.58%	33.26 MB	28.08 ms	85.29 MB

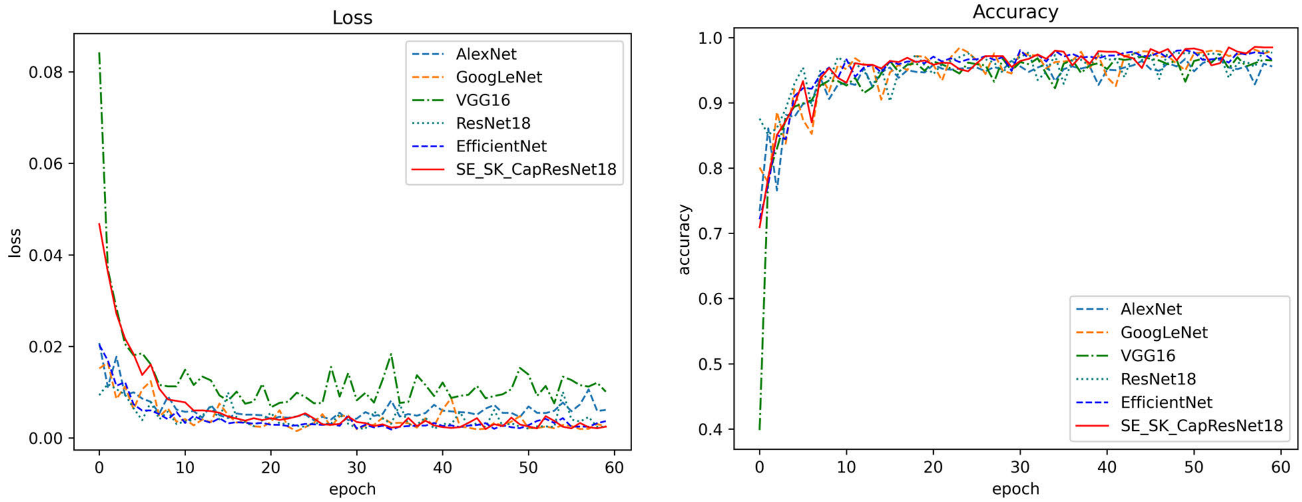

FIGURE 10. The change diagram of the loss curve and accuracy curve.

TABLE 3. Comparison of evaluation criteria for different network models (AI Challenger 2018 dataset).

Model	Precision	Recall	Accuracy	F1
AlexNet	86.28%	84.44%	84.44%	84.41%
GoogLeNet	91.52%	90.56%	90.56%	90.59%
VGG16	88.66%	88.33%	88.33%	88.40%
ResNet18	87.51%	83.47%	83.47%	83.91%
EfficientNet	88.65%	87.36%	87.36%	87.36%
SE-SK-CapResNet	95.26%	95.08%	95.08%	95.11%

TABLE 4. Comparison of evaluation criteria for different network models(Tomato Leaf Disease dataset).

Model	Precision	Recall	Accuracy	F1
AlexNet	89.33%	87.65%	87.81%	87.78%
GoogLeNet	92.72%	92.18%	92.18%	92.30%
VGG16	94.54%	94.52%	94.38%	94.41%
ResNet18	92.07%	91.88%	91.72%	91.72%
EfficientNet	95.42%	95.39%	95.31%	95.38%
SE-SK-CapResNet	97.32%	97.20%	97.19%	97.24%

observed on the test set, and the loss change curve and accuracy change curve are shown in Fig. 10.

As can be observed from Fig. 10, the SE-SK-CapResNet model performs well in terms of loss function and accuracy, and after 15 rounds of training, its loss function curve starts to converge gradually, and the final loss value converges to only 0.0022, while the accuracy curve continues to rise, and finally reaches an accuracy rate of 98.58%. Comparatively speaking, AlexNet, GoogLeNet, VGG16, ResNet18, and EfficientNet models perform relatively weakly in terms of loss function and accuracy, especially showing large fluctuations during the training process, showing the superiority of the classification performance of this SE-SK-CapResNet model on the dataset.

Table 2 compares the experimental results of different network models in the plant leaf disease image recognition task. In terms of precision, recall, accuracy, and F1 value, the SE-SK-CapResNet model performs well, reaching 98.58%, which is higher than other classical models such as AlexNet (96.28%), GoogLeNet (98.10%), VGG16 (96.00%) and ResNet18 (93.83%). It is worth mentioning that the SE-SK-CapResNet model is also relatively low in terms of the number of parameters and memory usage, which are 33.26MB and 85.29MB, respectively, showing a high model efficiency, which can satisfy the needs of embedded device deployment. In addition, the SE-SK-CapResNet model can meet real-time demand. Although the inference time for a single image is slightly higher, it remains within a reasonable range.

TABLE 5. A comparison between the proposed model vs. other works.

Author(s), Ref. No.	Dataset	Algorithm(classification)	Precision	Recall	Accuracy	F1
Z. Xiao, Y. Shi, G. Zhu, J. Xiong, and J. Wu.[8]	PlantVillage	Deep variant residual network with SE module	97.20%	97.19%	97.19%	97.18%
R. Gao, R. Wang, L. Feng, Q. Li, and H. Wu.[9]	AI Challenger 2018	Enhance ResNet with DECA module	84.15%	83.42%	86.35%	83.60%
H. Alghamdi and T. Turki.[18]	PlantVillage	A CNN-based Plant Disease Diagnosis Framework	92.06%	92.71%	93.79%	92.36%
N. Xie and X. Wan.[23]	PlantVillage	CapPlant: a capsule network based framework	93.07%	93.07%	93.07%	93.07%
M. A. Jasim and J. M. Al-Tuwaijari.[45]	PlantVillage	CNN	98.27%	98.01%	98.03%	98.14%
J. Chen, W. Wang, D. Zhang, A. Zeb, and Y. A.[46]	PlantVillage	Mobile - DANet	96.20%	95.60%	95.89%	95.83%
J. Chen, D. Zhang, A. Zeb, and Y. A. Nanehkaran.[47]	PlantVillage	pre-trained MobileNet-V2	97.49%	95.83%	96.68%	96.64%
Y. Ai, C. Sun, J. Tie, and X. Cai.[48]	AI Challenger 2018	Inception-ResNet	-	-	86.10%	-
Y. Zhong and M. Zhao[49]	AI Challenger 2018	DenseNet-121 with multi-label classification	93.55%	93.31%	93.31%	93.39%
Y. Liu, G. Gao, and Z. Zhang.[50]	AI Challenger 2018	SqueezeNext incorporates multi-scale convolution kernels	89.23%	88.83%	91.94%	-
Our proposed model	PlantVillage	SE-SK-CapResNet	98.58%	98.58%	98.58%	98.58%
Our proposed model	AI Challenger 2018	SE-SK-CapResNet	95.26%	95.08%	95.08%	95.11%

To verify the performance of the proposed model on different datasets, the proposed model is experimentally validated on the AI Challenger 2018 dataset and the tomato leaf disease dataset. The results are shown in Table 3 and Table 4. Unlike the images collected in the laboratory environment in the PlantVillage dataset, the tomato leaf disease dataset is derived from natural field environments where the background is more realistic and complex. Therefore, this experiment simultaneously evaluates the applicability and performance of the model in real agricultural scenarios. Table 5 compares the proposed model with other works using the same dataset.

As seen in Table 3, the proposed SE-SK-CapResNet model performs well relative to other classical models (AlexNet, GoogLeNet, VGG16, ResNet18, EfficientNet). Its high precision, recall, accuracy, and F1 value indicate that the model maintains its excellent performance in the face of more finely divided datasets.

As seen in Table 4, the model performance on the Tomato Leaf Disease dataset is slightly underperforming relative to the PlantVillage dataset. This mainly stems from the increased background complexity. Images in field environments contain more interfering factors, making feature extraction and classification of diseased leaves more complex. The SE-SK-CapResNet model performs well on the tomato leaf disease dataset, with precision, recall, accuracy, and F1 values exceeding 97%. In comparison, the other models are slightly inadequate. The experimental results again demonstrate the robustness and applicability of the proposed model on different datasets and natural environments, showing its significant advantages in the plant leaf disease image recognition task.

C. CONFUSION MATRIX

A confusion matrix is an essential tool for evaluating classification model performance because it visualizes model performance in each category and serves as a reference for evaluation. To have a more comprehensive understanding of the recognition performance of each model on different plant disease categories, we conducted a comparative analysis of the confusion matrix of AlexNet, GoogleNet, VGG16, ResNet18, EfficientNet, and SE-SK-CapResNet proposed in this paper under the same experimental conditions, and the confusion matrices of different models are shown in Fig. 11 shows.

By comparing the confusion matrices, we can observe that AlexNet, GoogLeNet, VGG16, and EfficientNet perform differently in recognizing different disease categories. These models perform well under most categories, but some perform poorly on specific diseases. For example, GoogLeNet and VGG16 showed relatively poor performance in the classification of corn leaf blight, while AlexNet and ResNet18 were also relatively weak in the classification of corn leaf blight as well as apple blackstar disease, and EfficientNet presented relatively poor results in the classification of corn gray spot. In contrast, SE-SK-CapResNet demonstrated a high level of performance under all disease categories. In identifying apple black star disease and apple black rot, SE-SK-CapResNet completely and correctly classified the samples into the corresponding classes, showing apparent advantages.

D. IMAGE ROTATION ROBUSTNESS EXPERIMENT

To evaluate the robustness of the improved capsule residual network under the change of image rotation, the test set images are rotated with the rotation angle ranging from 0 to

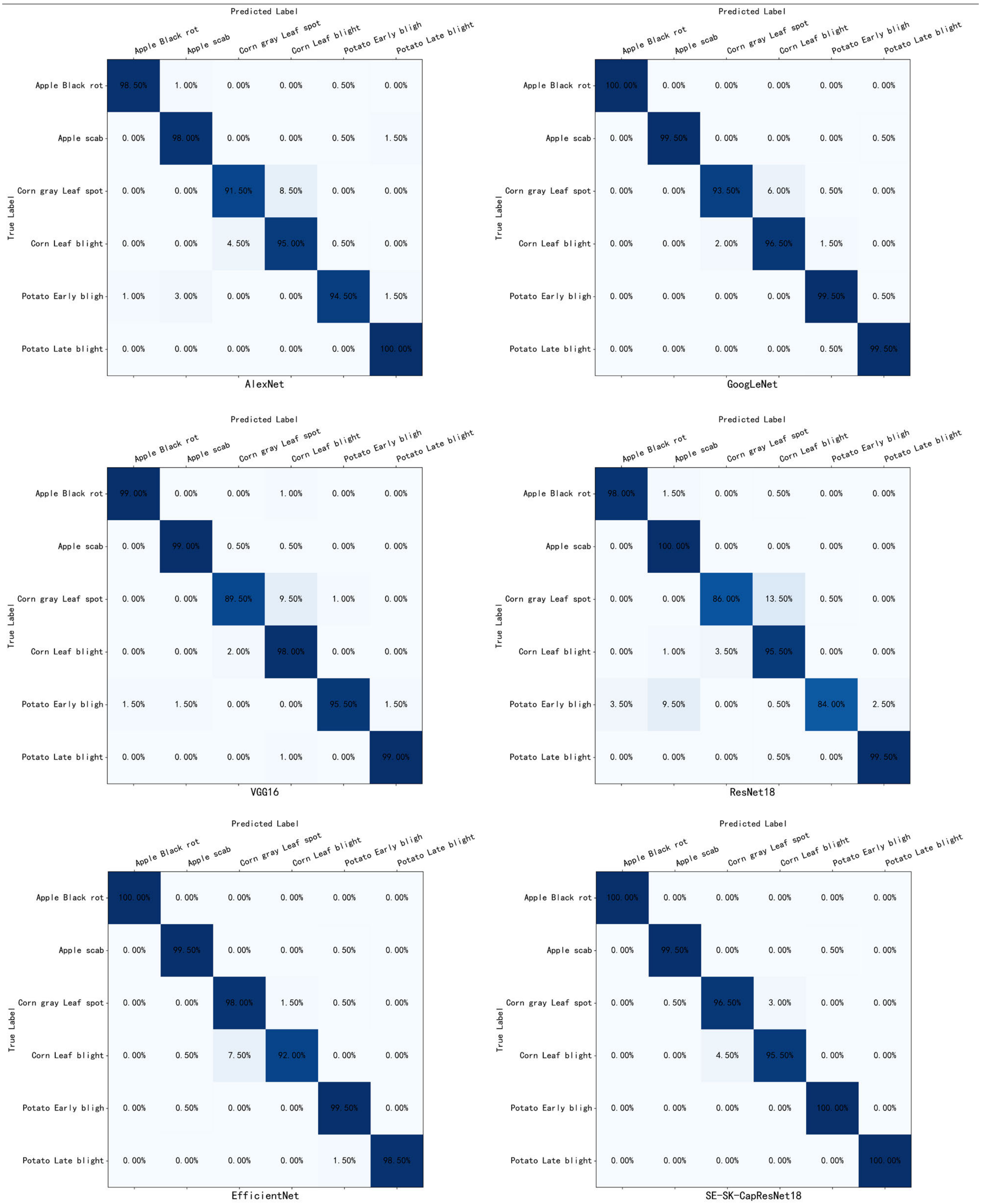

FIGURE 11. The confusion matrix of different network models.

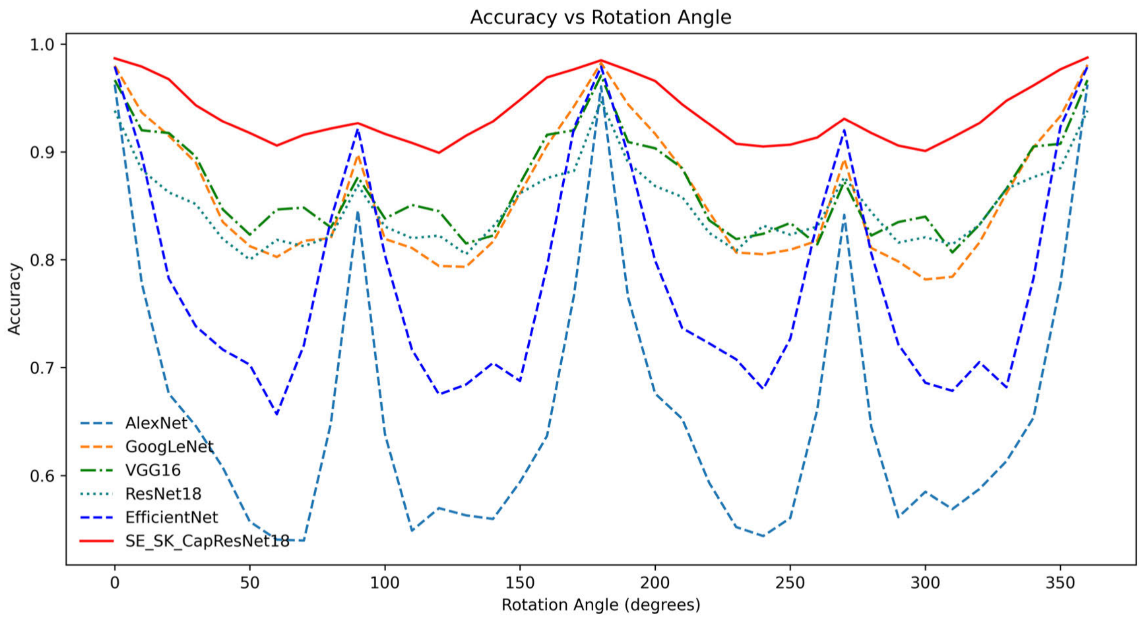

FIGURE 12. Comparison of experimental results on image rotation robustness.

360 degrees, and the step size is set to 10 degrees. The network is compared with a classical image classification model under the same experimental conditions. The comparison of the accuracy of different network models under image rotation is shown in Fig. 12.

As can be observed from Fig. 12, in the image rotation robustness experiments, with the gradual increase of the rotation angle, the accuracy of all models generally experienced a process of decreasing and then increasing. Specifically, the accuracy of each model peaked when the rotation angles were 90 degrees, 180 degrees, 270 degrees, and 360 degrees. This phenomenon suggests that the models are more likely to recognize images correctly at specific rotation angles, while performance degradation may occur at other angles. This phenomenon may be because the transformations, such as flipping and rotating the images during data enhancement, increased the model's ability to recognize plant disease images at specific angles. It is interesting to note that the SE-SK-CapResNet18 model shows higher robustness relative to other classical models throughout the rotation experiments. Especially at larger rotation angles, SE-SK-CapResNet18 maintains a relatively stable high accuracy, further highlighting its superior performance in processing rotated images. This result validates the excellent performance of the SE-SK-CapResNet18 model proposed in coping with rotational transformations of plant disease images.

VI. CONCLUSION

(1) Two key improvements addressed the need for leaf spot feature extraction for plant diseases. First, by replacing the 7×7 large convolutional kernel in the first layer of ResNet with three series-connected 3×3 small convolutional kernels, the ability to extract the details of the lesions was effectively enhanced. Second, the channel attention mechanism was introduced to strengthen the attention to the key features of plant disease leaves.

(2) Aiming at the limitations of traditional CNNs in terms of rotational invariance, a network model fusing capsule and residual networks is proposed. The improved ResNet is effectively connected with CapsNet, the initial pooling layer of ResNet is removed to reduce the loss of positional information, and the output of the third residual module of ResNet is connected with CapsNet, which sufficiently combines the advantages of the two networks, so that the model still maintains stable performance when the image rotation angle changes significantly, and enhances the robustness of the model.

(3) Comparative experiments were conducted on the publicly available dataset PlantVillage, AI Challenger 2018, and the tomato leaf disease dataset, and the results show that the proposed SE-SK-CapResNet model performs well in plant disease image recognition, with classification accuracies of 98.58%, 95.08%, and 97.19%, respectively, which are higher than those of other classical models. It is worth mentioning that the SE-SK-CapResNet18 model can maintain high accuracy relatively stably when dealing with plant disease images with different rotational angles, showing stronger robustness and superior classification performance compared with the traditional network models.

REFERENCES

[1] V. S. Dhaka, S. V. Meena, G. Rani, D. Sinwar, K. Kavita, M. F. Ijaz, and M. Woźniak, "A survey of deep convolutional neural networks applied for prediction of plant leaf diseases," *Sensors*, vol. 21, no. 14, p. 4749, Jul. 2021.

[2] Ramanjot, U. Mittal, A. Wadhawan, J. Singla, N. Z. Jhanjhi, R. M. Ghoniem, S. K. Ray, and A. Abdelmaboud, ''Plant disease detection and classification: A systematic literature review,'' *Sensors*, vol. 23, no. 10, p. 4769, May 2023.

[3] 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.

[4] L. Li, S. Zhang, and B. Wang, ''Plant disease detection and classification by deep learning—A review,'' *IEEE Access*, vol. 9, pp. 56683–56698, 2021.

[5] J.-R. Xiao, P.-C. Chung, H.-Y. Wu, Q.-H. Phan, J.-L.-A. Yeh, and M. T.-K. Hou, ''Detection of strawberry diseases using a convolutional neural network,'' *Plants*, vol. 10, no. 1, p. 31, Dec. 2020.

[6] M. Turkoglu, M. Aslan, A. Arı, Z. M. Alçin, and D. Hanbay, ''A multi-division convolutional neural network-based plant identification system,'' *PeerJ Comput. Sci.*, vol. 7, p. e572, May 2021.

[7] E. Elfatimi, R. Eryigit, and L. Elfatimi, ''Beans leaf diseases classification using MobileNet models,'' *IEEE Access*, vol. 10, pp. 9471–9482, 2022.

[8] Z. Xiao, Y. Shi, G. Zhu, J. Xiong, and J. Wu, ''Leaf disease detection based on lightweight deep residual network and attention mechanism,'' *IEEE Access*, vol. 11, pp. 48248–48258, 2023.

[9] R. Gao, R. Wang, L. Feng, Q. Li, and H. Wu, ''Dual-branch, efficient, channel attention-based crop disease identification,'' *Comput. Electron. Agricult.*, vol. 190, Nov. 2021, Art. no. 106410.

[10] S. S. Pandi, A. Senthilselvi, J. Gitanjali, K. ArivuSelvan, J. Gopal, and J. Vellingiri, ''Rice plant disease classification using dilated convolutional neural network with global average pooling,'' *Ecolog. Model.*, vol. 474, Dec. 2022, Art. no. 110166.

[11] M. Liu, H. Liang, and M. Hou, ''Research on cassava disease classification using the multi-scale fusion model based on EfficientNet and attention mechanism,'' *Frontiers Plant Sci.*, vol. 13, Dec. 2022, Art. no. 1088531.

[12] Y. A. Bezabh, A. O. Salau, B. M. Abuhayi, A. A. Mussa, and A. M. Ayalew, ''CPD-CCNN: Classification of pepper disease using a concatenation of convolutional neural network models,'' *Sci. Rep.*, vol. 13, no. 1, p. 15581, Sep. 2023.

[13] V. Maeda-Gutiérrez, C. E. Galván-Tejada, L. A. Zanella-Calzada, J. M. Celaya-Padilla, J. I. Galván-Tejada, H. Gamboa-Rosales, H. Luna-García, R. Magallanes-Quintanar, C. A. Guerrero Méndez, and C. A. Olvera-Olvera, ''Comparison of convolutional neural network architectures for classification of tomato plant diseases,'' *Appl. Sci.*, vol. 10, no. 4, p. 1245, Feb. 2020.

[14] S. Mousavi and G. Farahani, ''A novel enhanced VGG16 model to tackle grapevine leaves diseases with automatic method,'' *IEEE Access*, vol. 10, pp. 111564–111578, 2022.

[15] C. P. Lee, K. M. Lim, Y. X. Song, and A. Alqahtani, ''Plant-CNN-ViT: Plant classification with ensemble of convolutional neural networks and vision transformer,'' *Plants*, vol. 12, no. 14, p. 2642, Jul. 2023.

[16] Y. Luo, J. Sun, J. Shen, X. Wu, L. Wang, and W. Zhu, ''Apple leaf disease recognition and sub-class categorization based on improved multi-scale feature fusion network,'' *IEEE Access*, vol. 9, pp. 95517–95527, 2021.

[17] A. J. Pandian, K. Kanchanadevi, N. R. Rajalakshmi, and G. Arulkumaran, ''An improved deep residual convolutional neural network for plant leaf disease detection,'' *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, Sep. 2022.

[18] H. Alghamdi and T. Turki, ''PDD-Net: Plant disease diagnoses using multilevel and multiscale convolutional neural network features,'' *Agriculture*, vol. 13, no. 5, p. 1072, May 2023.

[19] M. K. Patrick, A. F. Adekoya, A. A. Mighty, and B. Y. Edward, ''Capsule networks—A survey,'' *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 1, pp. 1295–1310, Jan. 2022.

[20] T. Zeng, H. K.-H. So, and E. Y. Lam, ''RedCap: Residual encoder–decoder capsule network for holographic image reconstruction,'' *Opt. Exp.*, vol. 28, no. 4, p. 4876, Feb. 2020.

[21] X. Wen, Z. Han, X. Liu, and Y.-S. Liu, ''Point2SpatialCapsule: Aggregating features and spatial relationships of local regions on point clouds using spatial-aware capsules,'' *IEEE Trans. Image Process.*, vol. 29, pp. 8855–8869, 2020.

[22] R. Lei, C. Zhang, X. Zhang, J. Huang, Z. Li, W. Liu, and H. Cui, ''Multiscale feature aggregation capsule neural network for hyperspectral remote sensing image classification,'' *Remote Sens.*, vol. 14, no. 7, p. 1652, Mar. 2022.

[23] N. Xie and X. Wan, ''Residual vector capsule: Improving capsule by pose attention,'' *IEEE Access*, vol. 9, pp. 129626–129634, 2021.

[24] O. B. Samin, M. Omar, and M. Mansoor, ''CapPlant: A capsule network based framework for plant disease classification,'' *PeerJ Comput. Sci.*, vol. 7, p. e752, Nov. 2021.

[25] C. Xu, X. Wang, and S. Zhang, ''Dilated convolution capsule network for apple leaf disease identification,'' *Frontiers Plant Sci.*, vol. 13, Nov. 2022, Art. no. 1002312.

[26] S. Sabour, N. Frosst, and G. E. Hinton, ''Dynamic routing between capsules,'' in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 3859–3869.

[27] X. Shi, F. Xing, K. Xu, P. Chen, Y. Liang, Z. Lu, and Z. Guo, ''Loss-based attention for interpreting image-level prediction of convolutional neural networks,'' *IEEE Trans. Image Process.*, vol. 30, pp. 1662–1675, 2021.

[28] P. K. Mensah and M. A. Ayidzoe, ''Overview of CapsNet performance evaluation methods for image classification using a dual input capsule network as a case study,'' *Int. J. Comput. Digit. Syst.*, vol. 12, no. 1, pp. 29–43, Jul. 2022.

[29] S. Verma, A. Chug, and A. P. Singh, ''Exploring capsule networks for disease classification in plants,'' *J. Statist. Manage. Syst.*, vol. 23, no. 2, pp. 307–315, Feb. 2020.

[30] G. Altan, ''Performance evaluation of capsule networks for classification of plant leaf diseases,'' *Int. J. Appl. Math. Electron. Comput.*, vol. 8, no. 3, pp. 57–63, Oct. 2020.

[31] N. Vasudevan and T. Karthick, ''A hybrid approach for plant disease detection using E-GAN and CapsNet,'' *Comput. Syst. Sci. Eng.*, vol. 46, no. 1, pp. 337–356, 2023.

[32] S. Sapna, S. Sandhya, V. Acharya, and V. Ravi, ''Apple foliar leaf disease detection through improved capsule neural network architecture,'' *Multimedia Tools Appl.*, early access, Nov. 2023.

[33] V. H. Kondekar and S. K. Bodhe, ''Automation in plant pathology: Optimized attentional Capsule_BiLSTM optimized with chaotic sparrow algorithm for colour feature-based plant disease detection,'' *Multimedia Tools Appl.*, early access, Oct. 2023.

[34] Y. Li, M. Qian, P. Liu, Q. Cai, X. Li, J. Guo, H. Yan, F. Yu, K. Yuan, J. Yu, L. Qin, H. Liu, W. Wu, P. Xiao, and Z. Zhou, ''The recognition of Rice images by UAV based on capsule network,'' *Cluster Comput.*, vol. 22, no. S4, pp. 9515–9524, Jul. 2019.

[35] L. M. Abouelmagd, M. Y. Shams, H. S. Marie, and A. E. Hassanien, ''An optimized capsule neural networks for tomato leaf disease classification,'' *EURASIP J. Image Video Process.*, vol. 2024, no. 1, p. 2, Jan. 2024.

[36] 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.

[37] Y. Qu, X. Li, Z. Qin, and Q. Lu, ''Acoustic scene classification based on three-dimensional multi-channel feature-correlated deep learning networks,'' *Sci. Rep.*, vol. 12, no. 1, p. 13730, Aug. 2022.

[38] M. Jampour, S. Abbaasi, and M. Javidi, ''CapsNet regularization and its conjugation with ResNet for signature identification,'' *Pattern Recognit.*, vol. 120, Dec. 2021, Art. no. 107851.

[39] H. Wu, S. Zhao, L. Li, C. Lu, and W. Chen, ''Self-attention network with joint loss for remote sensing image scene classification,'' *IEEE Access*, vol. 8, pp. 210347–210359, 2020.

[40] G. Abosamra and H. Oqaibi, ''Using residual networks and cosine distance-based K-NN algorithm to recognize on-line signatures,'' *IEEE Access*, vol. 9, pp. 54962–54977, 2021.

[41] Z. Tong, J. Zhong, J. Li, J. Wu, and Z. Li, ''A power quality disturbances classification method based on multi-modal parallel feature extraction,'' *Sci. Rep.*, vol. 13, no. 1, p. 17655, Oct. 2023.

[42] H. J. Al Dawasari, M. Bilal, M. Moinuddin, K. Arshad, and K. Assaleh, ''DeepVision: Enhanced drone detection and recognition in visible imagery through deep learning networks,'' *Sensors*, vol. 23, no. 21, p. 8711, Oct. 2023.

[43] X. Du, K. Li, Y. Lv, and S. Qiu, ''Motor imaging EEG signal recognition of ResNet18 network based on deformable convolution,'' *Electronics*, vol. 11, no. 22, p. 3674, Nov. 2022.

[44] AI Challenger. *AI Challenger 2018 Datasets*. Accessed: Feb. 20, 2024. [Online]. Available: https://github.com/AIChallenger/AI_Challenger_2018

[45] M. A. Jasim and J. M. AL-Tuwaijari, ''Plant leaf diseases detection and classification using image processing and deep learning techniques,'' in *Proc. Int. Conf. Comput. Sci. Softw. Eng. (CSASE)*, Apr. 2020, pp. 259–265.

[46] J. Chen, W. Wang, D. Zhang, A. Zeb, and Y. A. Nanehkaran, "Attention embedded lightweight network for maize disease recognition," *Plant Pathol.*, vol. 70, no. 3, pp. 630–642, Apr. 2021.
[47] J. Chen, D. Zhang, A. Zeb, and Y. A. Nanehkaran, "Identification of Rice plant diseases using lightweight attention networks," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114514.
[48] Y. Ai, C. Sun, J. Tie, and X. Cai, "Research on recognition model of crop diseases and insect pests based on deep learning in harsh environments," *IEEE Access*, vol. 8, pp. 171686–171693, 2020.
[49] Y. Zhong and M. Zhao, "Research on deep learning in apple leaf disease recognition," *Comput. Electron. Agricult.*, vol. 168, Jan. 2020, Art. no. 105146.
[50] Y. Liu, G. Gao, and Z. Zhang, "Crop disease recognition based on modified light-weight CNN with attention mechanism," *IEEE Access*, vol. 10, pp. 112066–112075, 2022.

XIN ZHANG received the B.S. degree in mathematics and applied mathematics, the M.S. degree in applied mathematics, and the Ph.D. degree in mechanical engineering from Shenyang University of Technology, China, in 2010, 2013, and 2020, respectively. She is currently an Associate Professor with Shenyang Ligong University. Her research interests include deep learning, intelligent robots, and slam.

YUXIN MAO received the B.S. degree in mechanical electronic engineering from Shenyang Ligong University, China, in 2021, where he is currently pursuing the M.S. degree in mechanical electronic engineering. His research interests include deep learning and image processing.

QI YANG received the M.S. and Ph.D. degrees in pattern recognition and intelligent systems from Northeastern University, China, in 2007 and 2012, respectively. From 2015 to 2016, he was a Visiting Scholar with the University of California at Merced, Merced. He is currently a Professor with Shenyang Ligong University. His research interests include image processing, pattern recognition, robot vision, and intelligent control research.

XUYANG ZHANG received the B.S. degree in mechanical engineering from Dalian University of Science and Technology, China, in 2019. He is currently pursuing the M.S. degree in mechanical electronic engineering with Shenyang Ligong University, China. His research interests include image processing and pattern recognition.