Detection of maize leaf diseases using improved MobileNet V3-small

Ang Gao[1], Aijun Geng[1], Yuepeng Song[1,2,3*], Longlong Ren[1], Yue Zhang[1], Xiang Han[1]

(1. *College of Mechanical and Electrical Engineering, Shandong Agricultural University, Tai'an 271018, Shandong, China*;
2. *Shandong Provincial Engineering Laboratory of Agricultural Equipment Intelligence, Tai'an 271018, Shandong, China*;
3. *Shandong Provincial Key Laboratory of Horticultural Machineries and Equipment, Tai'an 271018, Shandong, China*)

Abstract: In order to realize the intelligent[2] identification of maize leaf diseases for accurate prevention and control, this study proposed a maize disease detection method based on improved MobileNet V3-small, using a UAV to collect maize disease images and establish a maize disease dataset in a complex context, and explored the effects of data expansion and migration learning on model recognition accuracy, recall rate, and F1-score instructive evaluative indexes, and the results show that the two approaches of data expansion and migration learning effectively improved the accuracy of the model. The structured compression of MobileNet V3-small bneck layer retains only 6 layers, the expansion multiplier of each layer was redesigned, 32-fold fast downsampling was used in the first layer, and the location of the SE module was optimized. The improved model had an average accuracy of 79.52% in the test set, a recall of 77.91%, an F1-score of 78.62%, a model size of 2.36 MB, and a single image detection speed of 9.02 ms. The detection accuracy and speed of the model can meet the requirements of mobile or embedded devices. This study provides technical support for realizing the intelligent detection of maize leaf diseases.

Keywords: maize leaf disease, image recognition, model compression, MobileNetV3-small

DOI: 10.25165/j.ijabe.20231603.7799

Citation: Gao A, Geng A J, Song Y P, Ren L L, Zhang Y, Han X. Detection of maize leaf diseases using improved MobileNet V3-small. Int J Agric & Biol Eng, 2023; 16(3): 225–232.

1 Introduction

Maize is a major food crop in the world for both grain and feed purposes[1]. Safeguarding the quality and increasing the yield of maize is of great practical importance for world food security. Prevention of maize pests and diseases is an important part of field management. Precise treatment and scientific decision-making have become important indicators in smart agriculture. In maize leaf diseases, traditional methods often rely on experienced experts and laborers on the spot to identify manually. This method is highly subjective and labor-intensive. Thus, the use of image processing technology for rapid and accurate diagnosis of maize diseases is of great importance for maize production. Therefore, the rapid and accurate diagnosis of maize diseases by image processing technology is of great significance to maize production.

With the rapid development of machine vision technology, scholars from domestic and international have done a lot of research in plant disease recognition by using deep learning and image processing techniques. Common deep learning classification models include Convolutional Neural Network (CNN)[2], Visual Geometry Group (VGG)[3], Residual Neural Network (RestNet)[4] and lightweight MobileNet V3[5], Inception V3[6], etc. In the neural network model, as the number of convolution layers increases, the recognition accuracy of the model will increase, but at the same time the volume of the model will also increase the running speed will decline, such as VGG16 compared to VGG19 volume is smaller, but the recognition accuracy has declined. The RestNet increases the residual structure, the network layer can be deeper without degradation, the network layer of RestNet152 was 152 layers, and the recognition accuracy has been greatly improved compared to RestNe50. However, the large size of the model is difficult to deploy. The lightweight model significantly reduces the model size, while the accuracy is slightly reduced to speed up the model operation so that it can run on mobile phones and embedded devices, which has become a hot spot for many scholars to study.

For the study of the MobileNet V3 deep learning model in agriculture, Ying et al.[7] detected various weeds in carrot fields, replacing the original YOLO-V4 backbone network with MobileNetV3-Small. Combining deep separable convolution, inverted residual structure, introducing lightweight attention mechanism, and reducing the memory required to process images, making the detection model more efficient. Tarek et al.[8] investigated various deep CNN techniques to detect and identify tomato leaf diseases. CNN models including ResNet50, InceptionV3, AlexNet, and three versions of MobileNet were trained on the Plant-Village dataset. Each model was trained using different optimization techniques such as Adam, Adagrad, RMSProp, and SGD. MobileNetV3 Large achieved a maximum accuracy of 99.81% using the Adagrad optimizer with a loss value of 0.0088. Tang et al.[9] dynamically collected images of leather eggs on a three-row egg conveyor, trained several deep learning classification models, and comprehensively compared the detection speed and detection effectiveness of the models. The comparison results showed that MobileNetV3-large was the best detection model with a detection time of only 4.267 s for 300 images and a detection accuracy of 96.3%.

Based on deep learning and image processing, the study of

Received date: 2022-07-21 **Accepted date:** 2023-04-14

Biographies: Ang Gao, PhD candidate, research interest: agricultural informatization and intelligent control technology, Email: 2019110103@sdau.edu.cn; **Aijun Geng**, PhD, Associate Professor, research interest: modern agricultural machinery design and theory, Email: gengaj@sdau.edu.cn; **Longlong Ren**, PhD, Lecturer, research interest: intelligent orchard agricultural machinery equipment, Email: renlonglong@sdau.edu.cn; **Yue Zhang**, PhD candidate, research interest: Cash crop machinery and equipment, Email: qq1964360182@163.com; **Xiang Han**, PhD candidate, research interest: intelligent agricultural equipment and technology, Email: 2021010102@sdau.edu.cn.

***Corresponding author: Yuepeng Song**, PhD, Professor, research interest: intelligent agricultural machinery equipment, Shandong Agricultural University, Taian 271018, Shandong, China. Tel: +86-05388246822, Email: uptonsong@163.com.

maize leaf disease recognition has also been done by scholars. Priyadharshini et al.[10] proposed a deep CNN-based structure (modified LeNet) for maize leaf disease classification, and experiments were conducted using maize leaf images from the PlantVillage dataset for training and testing. The results showed that the final model achieved 97.89% accuracy of the final model. Mingjielv et al.[11] presented an image-enhanced and DMS robust Alexnet-based method for identifying maize leaf disease, which demonstrated the ability to recognize and distinguish healthy leaves from six different maize leaf diseases. First, the maize leaf disease features were enhanced. Then the DMS robust Alexnet was constructed for recognition and classification with 98.62% recognition precision. Zeng et al.[12] proposed a SKPSNet-50 convolutional neural network model. Firstly, the 3×3 convolutional kernels in the backbone network ResNet-50 were replaced with the Selective Kernel-Point-Swish_B (SKPS), and the ReLU activation function was replaced with the Swish_BB activation function. Compared with the original, the average recognition accuracy of the final model was 92.9%. The average recognition accuracy of the final model was 92.9%, which was 6% higher than that of the original model. It can be seen that scientists in maize leaf disease identification mostly use large networks, focusing on the accuracy of the network to achieve excellent detection accuracy. But lightweight networks of maize leaf disease rapid identification of research is relatively small.

The former obtained high accuracy rates in plant disease identification with the help of machine vision technology, which can provide technical support for intelligent disease judgment[13]. For the field maize crop, due to the complex field environment, the model generalization ability is poor and the accuracy is low. The model is not suitable for practical application if the disease images are studied in a single background. Therefore, this study proposed an improved MobileNet V3-small-based maize leaf disease detection method. Firstly, a drone was used to acquire maize disease images, classify the images acquired by the drone to create a maize leaf disease dataset, and expand the dataset. Data expansion and transfer learning techniques were trained to improve the recognition accuracy of the model in this study.

2 Experiments and methods

2.1 Image acquisition

The image data acquisition site for this experiment was the Science and Technology Innovation Experimental Field of Shandong Agricultural University, China in late August 2020, and the DJI Mavic 2 Zoom 12-megapixel UAV was used for image acquisition. As shown in Figure 1, the cruise video acquisition was conducted at a height of 0.5 m above the maizefield and the camera was adjusted for different angles, and the total duration of the final five video segments was 15 min. This method does not need to enter the maizefield for manual filming, which greatly improves the working environment of the filmmaker and reduces the work intensity.

1. UAV 2. Maizefield

Figure 1 Schematic diagram of the picture collection site

2.2 Data pre-processing

After the video acquisition, the open source computer vision library OpenCV-python in python 3.8 was used to take frames and save one image every 10 frames, and the saved images were cropped and classified to select four common maize diseases, maize small spot, maize large spot, maize rust, and healthy maize leaves and number each species, totaling 2034 images were selected. When the test sample was not rich enough, it would cause the recognition rate in the field environment to be lower than that in the laboratory condition by about 30%[14]. Therefore, in this study, the existing data set was expanded by rotating, flipping, and changing the brightness of the images. The effect of the image data before and after expansion was shown in Figure 2, and the total number of images after expansion was 8134, and the distribution was listed in Table 1. The expanded dataset was divided into 80% for model training, 10% for model validation, and 10% for model testing.

a. Original drawing b. Horizontal flip c. Enhanced contrast d. Gaussian noise e. Translation

Figure 2 Data enhanced image of maize leaf disease data

2.3 MobileNet deep learning network

The final application of the maize leaf disease recognition model was piggybacked on mobile and embedded devices, and this study used MobileNet V3, a lightweight neural network suitable for mobile and embedded devices. MobileNet V3 was a next-generation lightweight network proposed by Google after MobileNet V2 model using a combination of search technology and architectural design, which had a certain performance improvement compared with V2[15].

The core unit of MobileNet was depthwise separable

Table 1 Distribution of maize disease samples

Name of disease	No.	Number of original samples	Number of samples after expansion
Southern blight of maize	0	504	2515
Northern leaf blight of maize	1	505	2518
Puccinia polysora	2	520	2596
Normal blade	3	505	505
Totaling	--	2034	8134

convolution, which was an operation that decomposes the standard convolution into two parts: depthwise convolution, which applies each convolution kernel to all channels, and point-wise convolution, which turns the standard convolution into a 1×1 convolution kernel. Point-wise convolution was used to combine the output of all the convolution channels, which greatly reduces the number of parameters and the computational effort of the model[5].

MobileNet V3 introduces a lightweight attentional squeeze and excitation (SE) model[16] based on a compressed reward and punishment network based on V2. The SE model could assign its own weights on the feature map through the training process so that the network selectively amplifies valuable feature channels from global information and suppresses useless feature channels[17]. The H-Swish function was utilized in the model activation function, which mainly prunes the input x by ReLU6 (x) was a function that ensures that the output was in the range of 0 to 6 to reduce the responsibility of the additive calculations, which was given by

$$y = x\frac{\text{ReLU6}(x+3)}{6} \tag{1}$$

where, x is the input value; y is the output value; ReLU6 is the trimming function.

In this study, MobileNet V3-small was used as the basic network for network model design. As shown in Figure 3, the network structure of MobileNet V3-small consists of six parts, each of which includes "blocks". First, the sampled image passes through a 3×3 convolutional layer to generate partial features of sample images, and then further extract rich details in several bottleneck structures. Finally, a convolutional layer was applied to pool convolutional layer operations instead of a classifier to distinguish objects and foregrounds[18]. In these operations, the two bottleneck structures are combined by connecting the front and end of a few identical moving inversion bottleneck DIRC (MBCONV) outputs, adding Squeeze and Excitation (SE) and nonlinear H-shapes for optimization. Using a bottleneck structure in the network allows shallow neural networks to capture image features through a global receptive field and provide a description for the image.

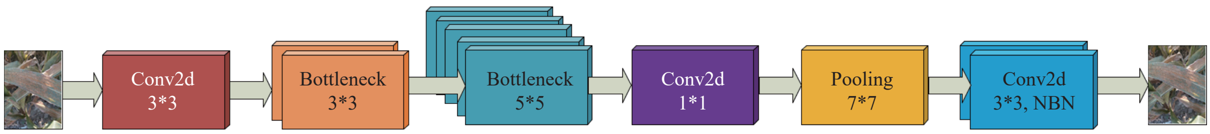


Figure 3 Network structure and process flow diagram of MobileNet V3-small

2.4 Transfer learning

As shown in Figure 4, transfer learning is a machine learning and deep learning approach that takes a convolutional neural network model on one task and makes it applicable to a new task with simple adjustments[19]. Unlike traditional machine learning, the source and target domains, source and target tasks of migration learning can be different[20], which can speed up the convergence speed of new network models when training to improve detection accuracy. The dataset studied in this study was relatively limited, and in order to improve the generalization ability of this model to speed up the training of the model and obtain good training results. The training weight parameters on the ImageNet dataset with about 1.2 million images were used to guide the training of the dataset for maize leaf diseases.

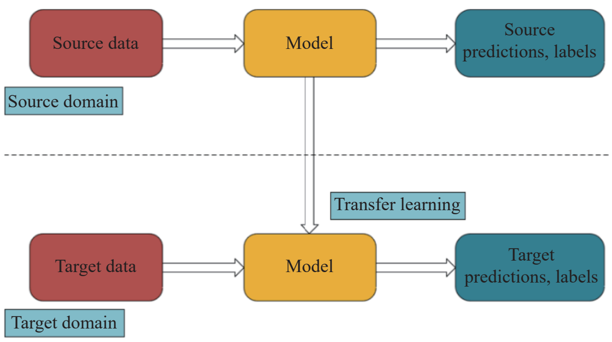


Figure 4 Schematic diagram of transfer learning

2.5 Improvements to MobileNet V3-small deep learning model

Although MobileNet deep learning network was suitable for embedded devices and mobile, there was still room for improvement in applying the model to maize disease identification. Therefore, the speed and accuracy of the MobileNet V3-based maize disease identification model should be improved to make the model obtain the best performance.

2.5.1 Network model structured compression

The classification accuracy of MobileNet V3-small tested on the public ImageNet dataset was about 65%. The ImageNet dataset was huge, with about 1.2 million images and 1000 classes of images[21]. The dataset in this study has only about 8000 images and 4 classifications, which is a small dataset in comparison. The original network model was bound to produce redundant parameters, so the model was structured to be compressed to reduce the model computation.

There were various ways to compress the model. The unstructured compression had a different operation for each layer of the convolutional kernel, making the convolutional kernel irregular, which made it extremely difficult to perform model transformation later to make it suitable for embedded device computation. Hence, this study adopted a structured compression approach to lightweight the MobileNet V3-small by changing the size of the convolutional kernel and removing the convolutional layers. Only 6 layers of the bneck layer of the model were retained, and four layers of the 5×5 convolutional kernel module were deleted. The expansion multiplier of each layer was redesigned, and the number of channels of the first convolutional 16 layer in the bneck was doubled to 32 channels in order to speed up the downsampling rate. The final compressed MobileNet V3-small network structure was listed in Table 2.

2.5.2 Squeeze and Excitation (SE) module appropriate location

The Squeeze and Excitation (SE) module mainly learns the correlation between channels and extracts channel-specific attention, which slightly increases the computational effort. But improves the model accuracy significantly was improved. The SE module was added after the internal and deep convolution of the bottleneck structure of the MobileNet V3-small network model. For embedded devices, extensive use of SE module will increase the running time of the model. Considering the balance between speed and accuracy of the model, the SE module was added to the fourth and last layer of the network bneck layer, which operation also achieved similar accuracy. The final improved MobileNet V3-small network structure was listed in Table 3.

Table 2 Structure of the compressed MobileNet V3-small model

Input (W×H×Number of channels)	Operator	exp size	Out	Stride
224×224×3	Conv2d, 3×3	--	32	2
112×112×32	Bneck, 3×3	32	32	2
56×56×32	Bneck, 3×3	64	32	2
28×28×32	Bneck, 3×3	120	40	1
28×28×40	Bneck, 5×5	240	48	2
14×14×48	Bneck, 5×5	288	96	2
7×7×96	Bneck, 5×5	288	112	1
7×7×112	Conv2d, 1×1	--	288	1
7×7×288	Pool, 7×7	--	--	1
1×1×288	Conv2d, 1×1.NBN	--	1024	1
1×1×1024	Conv2d, 1×1.NBN	--	*k*	1

Note: W: Width; H: Height.

Table 3 Structure of the improved MobileNet V3-small network model

Input (W×H×Number of channels)	Operator	exp size	Out	SE	Stride
224×224×3	Conv2d, 3×3	--	32	--	2
112×112×32	Bneck, 3×3	32	32	--	2
56×56×32	Bneck, 3×3	64	32	--	2
28×28×32	Bneck, 3×3	120	40	-	1
28×28×40	Bneck, 5×5	240	48	1	2
14×14×48	Bneck, 5×5	288	96	--	2
7×7×96	Bneck, 5×5	288	112	1	1
7×7×112	Conv2d, 1×1	--	288	--	1
7×7×288	Pool, 7×7	--	--	--	1
1×1×288	Conv2d, 1×1.NBN	--	1024	--	1
1×1×1024	Conv2d, 1×1.NBN	--	*k*	--	1

Note: SE: Squeeze and Excitation.

3 Results and discussion

3.1 Test environment

The hardware test environment in this study is a Lenovo laptop (y9000p) with an Intel Pentium i5-12700H processor at 3.5 GHz and GeForce GTX 30606 G GPU. The software test environment is Windows 10, the computer programming language is python 3.8, the machine learning library is pytorch 1.10.0, and the parallel computing architecture is CUDA10.2.

3.2 Evaluative indicators

In this study, in order to analyze and evaluate the performance of the trained model, recognition accuracy, recall, and F1-score were selected as valid indicators of the detection results, and the average recognition speed during testing was used to reflect the performance of the model together. The average recognition time is the ratio of the total time consumed for sample testing to the number of samples tested, and the average recognition accuracy, recall, and F1-score (F1) are calculated as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \times 100\% \tag{2}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \times 100\% \tag{3}$$

$$\mathrm{F1} = 2\times\frac{P\times R}{P+R} \times 100\% \tag{4}$$

where, *P* means precision, is the proportion of all positive predictions that are correct; *R* means Recall, is the proportion of all real positive observations that are correct; TP means True Positive, for example, where the true value of the model is 1 and the predicted value is 1; TN means True Negative, for example, where the true value of the model is 0 and the predicted value is 0; FP means False Positive, for example, where the true value of the model is 0 and the predicted value is 1; FN means False Negative, for example, where the true value of the model is 1 and the predicted value is 0; F1-score (F1) is the harmonic mean of precision and recall.

3.3 Experimental analysis

3.3.1 Analysis of migration training results

The training and validation datasets in the experiments were expanded maize disease image datasets, and seven network models, ResNet18[4], ResNet50[4], MobileNet V2[22], MobileNet V3-small[5], MobileNet V3-large[5], and Xception41[23] were used to compare the effects of migration training on the models. The experimental results in the test set were listed in Table 4.

Table 4 Different models were used to identify maize diseases

Model	Transfer training	Total test set accuracy/%	Total test set recall/%
ResNet 18	No	70.14	71.25
	Yes	73.61	73.68
ResNet 50	No	74.28	72.38
	Yes	76.05	75.83
MobileNet V2	No	64.02	63.87
	Yes	65.06	65.07
MobileNet V3-small	No	78.77	78.40
	Yes	80.66	79.83
MobileNet V3-large	No	80.36	80.12
	Yes	82.45	81.28
DenseNet121	No	65.48	63.14
	Yes	68.17	65.59
Xception 41	No	69.29	68.38
	Yes	72.25	72.20

From Table 4, it could be seen that whether the network model uses migration training or not had an effect on the precision and recall of the model, and there was more than a one percentage point improvement in precision and recall in all models. Accuracy improvement was most significant at 3.47% in model ResNet 18 after training with migration and at 1.04% in model MobileNet V2 after training with migration. In the MobileNet V3-small model, the accuracy improvement of the model after training with migration was 1.89%. The recall improvement was 1.43%, which proves that migration training was an effective way to improve the recognition accuracy of the model in this paper.

3.3.2 Data extension training results analysis

The purpose of data extension is to improve the grid model recognition accuracy and prevent overfitting[24]. In this subsection, the experimental results of training maize leaf disease recognition models for three models, MobileNet V3-large, MobileNet V3-small, and improved MobileNet V3-small, were shown in Figure 5.

From Figure 5, it could be seen that the data expansion had a significant impact on the precision, recall, and F1 metrics of the three models, and the evaluation metrics of all three models showed a 10 percentage points improvement after the data expansion. In the *P* metric MobileNet V3-small model showed the smallest improvement of 12.21% and the largest improvement was 13.24% for the improved MobileNet V3-small model. In the *R* metric, the MobileNet V3-large model has a minimum improvement of 11.73%, and the largest improvement was 12.82% for the improved MobileNet V3-small model. The smallest improvement in F1 was 11.73% for the MobileNet V3-large model, and the largest improvement was 14.53% for the improved MobileNet V3-small model. This indicates that data expansion is an effective way to improve model accuracy.

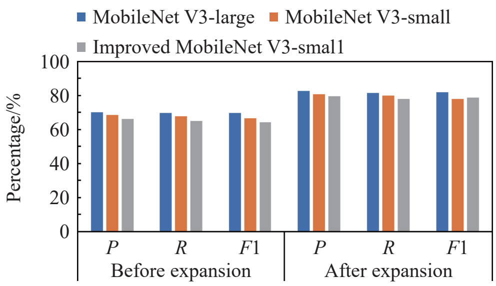


Figure 5 Comparison of model metrics before and after expansion

3.3.3 Analysis of the location results of structured compression and SE modules

The purpose of structured compression for the network model was to reduce redundant parameters and to find a balance between the recognition accuracy and computational effort of the network model[25]. In order to find a suitable model for the maize disease dataset in this study, the location and number of convolutional kernels and convolutional layers removed from the MobileNet V3-small network model were experimentally analyzed, as well as the appropriate location of the SE module in the model. Three schemes for the combination of structured compression and SE module proper location were designed and counted as Schemes A, B, and C, respectively. Scheme A retains only 8 layers for the bneck layer and removes three layers of the 5×5 convolutional kernel module in size while retaining the SE module in only the last two layers, and the network structure was listed in Table 5. Scheme B was the final model designed in this study, with the network structure listed in Table 3. Scheme C retains only 8 layers for the bneck layer, removes three layers of convolutional kernel modules of size 5×5, uses a channel number of 32 in the first layer of bneck, doubles the number of channels in this layer, increases the convolutional kernel in the last layer from 5×5 to 7×7, while retains the SE module only in the last two layers, with the network structure as listed in Table 6.

Table 5 Scheme A network model structure

Input (W×H×Number of channels)	Operator	exp size	Out	SE	Stride
224×224×3	Conv2d, 3×3	--	16	--	2
112×112×16	Bneck, 3×3	16	16	--	2
56×56×16	Bneck, 3×3	64	24	--	2
28×28×24	Bneck, 3×3	72	24	--	1
28×28×24	Bneck, 5×5	96	40	--	2
14×14×40	Bneck, 5×5	120	40	--	1
14×14×48	Bneck, 5×5	288	96	--	2
7×7×96	Bneck, 5×5	576	96	1	1
7×7×96	Bneck, 5×5	576	96	1	1
7×7×96	Conv2d, 1×1	--	576	--	1
7×7×576	Pool, 7×7	--	--	--	1
1×1×576	Conv2d, 1×1.NBN	--	1024	--	1
1×1×1024	Conv2d, 1×1.NBN	--	*k*	--	1

Table 6 Scheme B network model structure

Input (W×H×Number of channels)	Operator	exp size	Out	SE	Stride
224×224×3	Conv2d, 3×3	--	32	--	2
112×112×32	Bneck, 3×3	32	32	--	2
56×56×32	Bneck, 3×3	64	32	--	2
28×28×32	Bneck, 3×3	120	40	--	1
28×28×40	Bneck, 5×5	120	40	--	2
14×14×40	Bneck, 5×5	240	48	--	1
14×14×48	Bneck, 5×5	288	96	--	2
7×7×96	Bneck, 5×5	576	96	--	1
7×7×96	Bneck，5×5	672	112	1	1
7×7×112	Conv2d, 1×1	--	672	1	1
7×7×672	Pool，7×7	--	--	--	1
1×1×672	Conv2d, 1×1.NBN	--	1024	--	1
1×1×1024	Conv2d, 1×1.NBN	--	*k*	--	1

The recognition accuracy, recall, and F1 metrics were taken as evaluative metrics in the experiment. The results were listed in Table 7, while the model size (MB) and test set single test time were taken as evaluative metrics, as shown in Figure 5. As can be seen from Table 7, among the three metrics, the recognition accuracy of Scheme B was 79.52%, the recall rate was 77.90%, and the F1 metric was 78.62%. The recognition accuracy of Scheme A was the lowest, and the recognition accuracy of Scheme C was the highest. In Scheme C the recognition accuracy, recall rate, and F1 index of the model improved by more than five percentage points compared with the original model. It shows that different strengths of structured compression and different positions of SE modules had significant effects on the precision, recall, and F1 metrics of the model.

Table 7 Identification results of maize leaf diseases in different scenarios

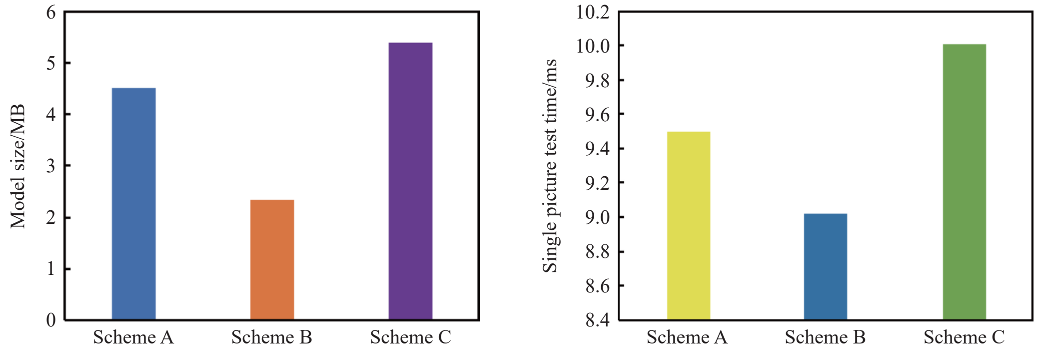
Scheme	*P*/%	*R*/%	F1-score/%
Scheme A	74.49	75.28	75.18
Scheme B	79.52	77.90	78.62
scheme C	85.67	85.02	85.36

As Figure 6 shows the maximum model of C among the three schemes was 5.41 MB, which exceeded the original model, and the test set single test time was 10.01 ms, while the model of Scheme B was the smallest in the comparison at 2.36 MB and the minimum single test time was 9.02 ms. The comparison found that the recognition accuracy of Scheme A decreased by 5.03% compared with Scheme B, but the model size was increased by 2.15 MB compared with Scheme B.

Considering that the detection speed of 10 ms/image can already meet the needs of real-time monitoring, the speed of recognition was guaranteed and the solution with high recognition accuracy was chosen as much as possible. Finally, it was considered that Scheme B could be the most suitable solution.

3.3.4 Analysis of results before and after model improvement

The test indexes before and after model improvement were listed in Table 8. By comparing the test indexes before and after improvement, it was found that the size of the improved MobileNet V3-small network model was reduced by 2.53 MB, which was 51.74% compared to the model size before improvement. It greatly reduced the amount of computation for later porting to embedded devices. The running time of each test image was reduced by 0.87 ms, while the average recognition accuracy of the model was reduced by only 1.14%, the average recall was reduced by 1.92%,



a. Size comparison diagram of model structures of different schemes

b. Comparison diagram of test time of single picture of models in different schemes

Figure 6 Model size, single image test time comparison chart

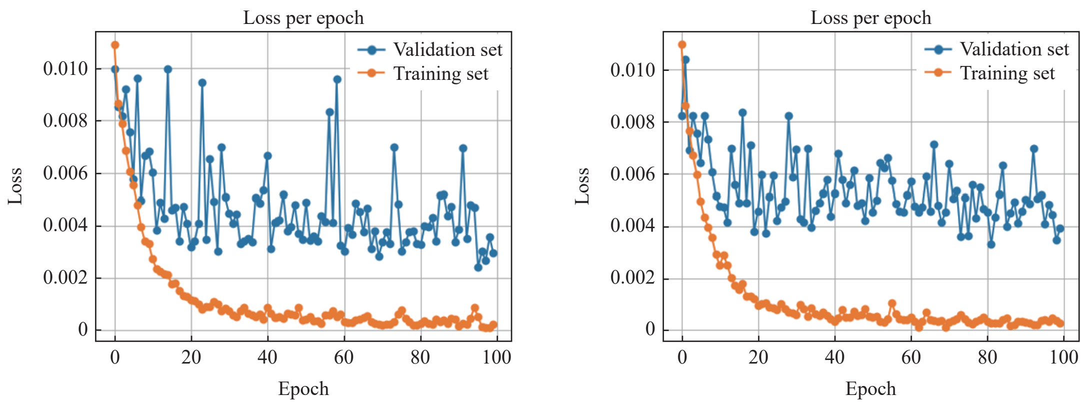
Table 8 Comparison test results before and after model improvement

Model	Model size/MB	Speed of single image recognition/ms	Types of leaf diseases	*P*/%	*R*/%	F1-score/%
MobileNet V3-small	4.89	9.89	0	79.13	88.91	83.73
			1	78.02	75.59	76.79
			2	78.21	75.97	81.23
			3	87.26	78.85	69.49
			Average	80.66	79.83	77.81
Improved MobileNet V3-small	2.36	9.02	0	80.85	85.5	83.11
			1	74.81	77.15	75.96
			2	85.74	77.82	81.59
			3	76.67	71.15	73.81
			Average	79.52	77.91	78.62

and the average F1 index was increased by 0.81%. Among them, the recognition accuracy of different maize leaf disease types had improved and decreased, such as the accuracy of the improved model to recognize leaf disease type 0 increased by 1.73%, and the F1-score of the improved model to recognize leaf disease type 1 decreased by 0.83%.

3.4 Loss and accuracy of model training and validation sets

The loss curve consisting of the loss function values of each training round could measure the goodness of the model prediction. In this study, before and after improving the model, the loss curves of the training and validation datasets were compared as shown in Figure 7. The loss values of the model training set decreased rapidly in the first 20 rounds of iteration, and the loss values of the model decreased slowly and approached 0 value gradually and smoothly after 20 rounds. It was found from the loss values of the validation set before and after the model improvement that the jump of the improved loss value was significantly smaller than that of the model before the improvement indicating that the improved model was more robust, which was proved by the change of the loss curve that the improved model training was effective.



a. MobileNet V3-small loss

b. Improved MobileNet V3-small loss

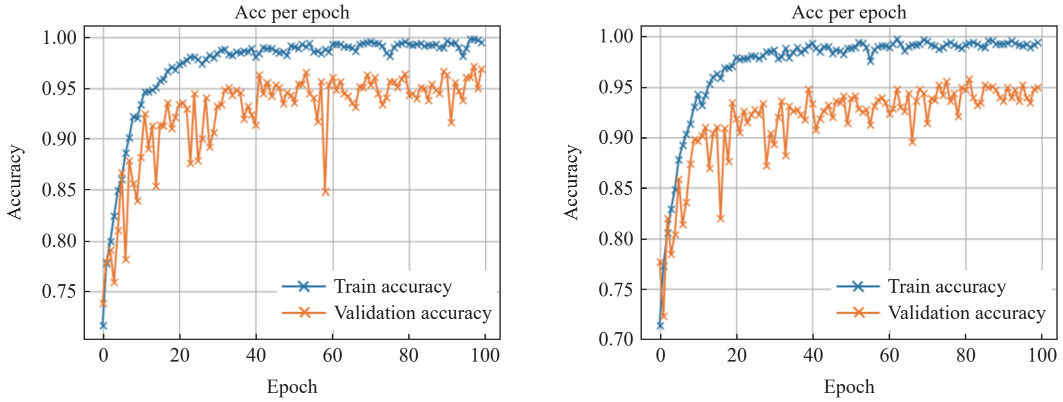
Note: val: Validation set; train: Training set.

Figure 7 Comparison of model loss curves before and after improvement

After each round of training, the recognition accuracy was calculated for both the training and validation sets to measure the goodness of the model prediction. As shown in Figure 8, this study compared the recognition accuracy curves of the training and validation datasets of the model before and after the improvement. The recognition accuracy of the training and validation sets of the model increases rapidly in the first 20 rounds of iterations, and the recognition accuracy of the training and validation sets of the model increases slowly and tends to level off in the second 20 rounds. The accuracy of the training set gradually approaches 1, and the accuracy of the validation set gradually approaches 0.95. The comparison of the accuracy of the model before and after the improvement in different rounds proves that the recognition accuracy of the improved model meets the requirements.

3.5 Confusion matrix for the test set

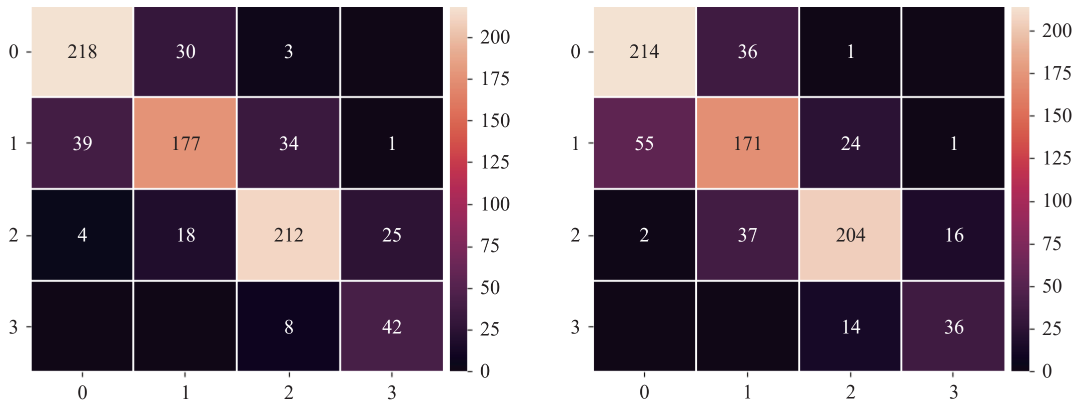
A confusion matrix is a specific matrix used to visualize the performance of an algorithm for supervised learning, indicating the presence or absence of confusion in multiple categories in an N×N square matrix, where each column represents the predicted value and each row represents the actual category. In this study, MobileNet V3 was used with modified MobileNet V3 for the test



a. MobileNet V3-small accuracy b. Improved MobileNet V3-small accuracy

Figure 8 Accuracy diagram of the model before and after improvement

set to identify maize leaf diseases and produce a confusion matrix with numbers 0, 1, 2, and 3 corresponding to the four maize small spot diseases, maize large spot disease, maize rust, and healthy maize leaf labels, respectively. As shown in Figure 9, it was observed that the MobileNet V3 model made the most errors in identifying Label 1 for maize small spot disease and the least errors in identifying Label 3 for healthy maize leaves. The number of errors for all classifications identified by the improved MobileNet V3 model increased slightly compared to the original model, indicating a slight decrease in the accuracy of classification identification. The same improved model has the most errors for Label 1 and the least errors for Label 3, indicating that the model is most difficult to identify the small spot disease in the maize leaf disease, while the model is least difficult to identify the healthy maize leaf with obvious characteristics. The number of confusion matrix error tiles in the test set before and after the overall improvement met the requirements and effectively proved the effectiveness of the model in this study.



a. MobileNet V3-small b. Improved MobileNet V3-small

Figure 9 Confusion matrix of the model before and after improvement

3.6 Analysis of public data set test results

The Plant-village dataset includes more than 50 000 healthy and diseased images divided into 38 categories, and this subsection extracts the dataset of maize leaf diseases in a single context for the validation of the improved before-and-after model. The extracted dataset was classified in the same way as the dataset in this study, with 433 images of maize small spot disease, 354 images of maize large spot disease, 187 images of maize rust disease, and 432 images of healthy maize leaves, totaling 1406 images, and the classification labels were indicated by the numbers 0, 1, 2, and 3. This dataset was partitioned using the same way of making the dataset of this study, and the model was trained and tested in the same software environment. MobileNet V3-large, lMobileNet V3-small, and modified MobileNet V3-small were used to test this dataset without any other optimization methods, and the results were listed in Table 9.

From Table 9, it could be seen that training and testing the model on the public dataset achieved excellent results, and the improved MobileNet V3-small model *P*, *R*, and F1 metrics generally outperformed the unimproved MobileNet V3-small model, with P improving by 2.75 percentage points. Compared with the MobileNet V3-large model, the improved model reduces *P* by 2.12 percentage points, *R* by 0.15 percentage points, and F1 by 1.14 percentage points, but the size of the model was only 15.24% of that of the MobileNet V3-large model. The superiority of the improved model is fully demonstrated.

Table 9 Experimental results of the public data set model

Model	Model size/MB	*P*/%	*R*/%	F1/%
MobileNet V3-large	15.3	95.12	94.50	94.81
MobileNet V3-small	4.89	90.25	91.16	90.70
Improved MobileNet V3-small	2.36	93.00	94.35	93.67

4 Conclusions

In this study, maize disease image acquisition of maize fields by UAV was used to establish a maize disease dataset in a complex context, and the lightweight convolutional neural network MobileNetV3 was improved by keeping only 6 layers in the bneck layer of the model, redesigning the expansion multiplier of each layer, using a 32-fold fast downsampling method in the first layer, optimizing the position of the SE module in the structure, and using only two SE module operations. The improved model was subjected to migration learning on the expanded dataset to obtain the maize disease classification model, and the recognition accuracy, recall, F1 index, operation speed, and model size of the model in this study were compared with other models for experimental analysis, and the following conclusions were obtained.

In comparing the model accuracies of data expansion and migration learning for the self-built dataset, the detection accuracy of the model with data expansion improved by 10 percentage points on average, while the best accuracy of the model trained with migration learning improved by 3 percentage points, proving that both treatments can improve the recognition accuracy of the model.

This model had an average accuracy of 79.52% in the test set of the self-built dataset, which was 1.14 percentage points less than the average accuracy of the unimproved model, a single test time of 9.02 ms, which was 0.87 ms less than the original model test time, as well as a model size of 2.36 MB which was 51.74% less than the original model, greatly improving the overall performance of the model. The results demonstrate that the detection accuracy and detection speed of the convolutional neural network-based maize disease identification model could meet the usage requirements.

In this study, the model was trained and tested on a maize leaf disease dataset in a single context in the public Plant-village dataset, and the improved model improved P by 2.75 percentage points, R by 3.19 percentage points, and F1 by 2.97 percentage points compared with the unimproved model, and the size of the model was only 15.24% of the MobileNet V3-large model. The generalized and superiority of the improved model were fully demonstrated.

In this study, an improved MobileNetV3 maize leaf disease recognition method was proposed based on the improved MobileNetV3 through training and testing, which can quickly and accurately classify maize leaf diseases, and can further study the use of recognition models to guide maize leaf disease control on this basis. In the future, the multimodal recognition method is planted to further use to detect the disease at the early stage of maize leaf disease, so as to make disease judgments and control earlier.

Acknowledgements

This study was supported by the Fruit Industry Innovation Team Project of the Modern Agricultural Industry Technology System of Shandong Province (SDAIT-06-12) and the "Double First-class" Award and subsidy fund project of Shandong Agricultural University (SYL2017X).

[References]

[1] Han T F, Li Y Z, Qu X L, Ma C B, Wang H Y, Huang J, et al. Spatio-temporal evolutions and driving factors of wheat and maize yields in China. Transactions of the CSAE, 2022; 38(1): 100–108. (in Chinese)

[2] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2012; 60(6): 84–90.

[3] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint, 2014. arXiv: 1409.1556.

[4] He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas: IEEE, 2016; pp.770–778. doi: 10.1109/CVPR.2016.90.

[5] Howard A, Sandler M, Chen B, Wang W J, Chen L-C, Tan M X. Searching for MobileNetV3. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019; pp.1314–1324. doi: 10.1109/ICCV.2019.00140.

[6] Christian S, Vincent V, Sergey I, Jonathon S, Zbigniew W. Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas: IEEE, 2015; pp.2818–2826. doi: 10.1109/CVPR.2016.308.

[7] Ying B Y, Xu Y C, Zhang S, Shi Y G, Liu L. Weed detection in images of carrot fields based on improved YOLO v4. Traitement du Signal, 2021; 38(2): 341–348.

[8] Tarek H, Aly H, Eisa S, Abul-Soud M. Optimized deep learning algorithms for tomato leaf disease detection with hardware deployment. Electronics, 2022; 11(1): 140.

[9] Tang W Q, Hu J C, Wang Q H. High-throughput online visual detection method of cracked preserved eggs based on deep learning. Applied Sciences, 2022; 12(3): 952.

[10] Priyadharshini R A, Arivazhagan S, Arun M, Mirnalini A. Maize leaf disease classification using deep convolutional neural networks. Neural Computing and Applications, 2019; 31: 8887–8895.

[11] Lyu M J, Zhou G X, He M F, Chen A B, Zhang W Z, Hu Y H. Maize leaf disease identification based on feature enhancement and DMS-robust alexnet. IEEE Access, 2020; 8: 57952–57966.

[12] Zeng W H, Li H D, Hu G S, Liang D. Identification of maize leaf diseases by using the SKPSNet-50 convolutional neural network model. Sustainable Computing:Informatics and Systems, 2022; 35: 100695.

[13] Cen H Y, Zhu Y M, Sun D W, Zhai L, Wan L A, Ma Z H, et al. Current status and future perspective of the application of deep learning in plant phenotype research. Transactions of the CSAE, 2020; 36(9): 1–16. (in Chinese)

[14] Fan X P, Zhou J P, Xu Y, Peng X. Corn disease recognition under complicated background based on improved convolutional neural network. Transactions of the CSAM, 2021; 52(3): 210–217. (in Chinese)

[15] Sun H, Li S, Li M Z, Liu H J, Qiao L, Zhang Y. Research progress of image sensing and deep learning in agriculture. Transactions of the CSAM, 2020; 51(5): 1–17. (in Chinese)

[16] Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020; 42(8): 2011–2023.

[17] Xu Z N, Wang J K, Hu Y J, Zhang Z C, Zhao X D, Yang L L, et al. Potato disease recognition based on MobileNetV3. Jiangsu Agricultural Sciences, 2022; 50(10): 176–182. (in Chinese)

[18] Fang J L, Hu Y H, Dai B S, Wu Z D. Detection of group-housed pigs based on improved CenterNet model. Transactions of the CSAE, 2021; 37(16): 136–144. (in Chinese)

[19] Xu J H, Shao M Y, Wang Y C, Han W T. Recognition of corn leaf spot and rust based on transfer learning with convolutional neural network. Transactions of the CSAM, 2020; 51(2): 230–236, 253. (in Chinese)

[20] Wu Q L, Zheng J, Wu K M. Research and application of crop pest monitoring and early warning technology in China. Frontiers of Agricultural Science and Engineering, 2022; 9(1): 19–36.

[21] Li Y H, Wang N N, Zhu M R, Yang X, Gao X B. Recent advances in neural architecture search: A survey. Journal of Software, 2022; 33(1): 129–149.

[22] Sandler M, Howard A, Zhu M. MobileNetV2: Inverted residuals and linear bottlenecks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City: IEEE, 2018; pp.4510–4520. doi: 10.1109/CVPR.2018.00474.

[23] Chollet F. Xception: Deep learning with depthwise separable convolutions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu: IEEE, 2017; pp.1800–1807. doi: 10.1109/CVPR.2017.195.

[24] Ferrag M A, Shuo L, Friha O, Yang X. Cyber security intrusion detection for agriculture 4.0: Machine learning-based solutions, datasets, and future directions. IEEE/CAA Journal of Automatica Sinica, 2022; 9(3): 407–436.

[25] Zhang Z K, Pang W G, Xie W J, Liu M S, Wang Y. Deep learning for real-time applications: A survey. Journal of Software, 2020; 31(9): 2654–2677. (in Chinese)