Lightweight Convolutional Neural Networks for Pepper Diseases Detection and Classification

Xiangyu Zeng [1], Yuan Tang [2,*], Jun Li [3], Yang He [4], Youwan Tang [2]

[1] *College of Computer Science and Cyber Security Chengdu University of Technology Chengdu 610059, China.*

[2] *Chengdu Academy of Agricultural and Forestry Sciences Chengdu 611130, China.*

[*] *Corresponding author*

Abstract

In recent years, with the improvement of cultivation technology and the rapid development of logistics industry, pepper production has broken the regional and seasonal restrictions. According to data, the global pepper planting area in 2020 is about 1.999 million hectares, with a year-on-year growth of 3.3%, and the global pepper production is about 39.28 million tons. In the process of pepper cultivation and management, disease is an important factor restricting its quality improvement and yield growth, and the correct diagnosis of disease is a necessary prerequisite for effective prevention and control of diseases. At the same time, the traditional field diagnosis of diseases is done manually. However farmers lack professional knowledge in the identification and prevention of diseases and pests, and lack front-line guidance from expert. When diseases and pests occur, it is easy to cause significant economic losses. Therefore, the research of intelligent pest detection technology is very important to control the spread of diseases and pests.With the rise of artificial intelligence technology, automatic recognition and diagnosis of crop disease images using computer vision has become a hot research topic at home and abroad in recent years. In addition, the depth recognition model represented by convolutional neural network has made an important breakthrough. In this paper, the deep learning algorithm of Mobilenet-V2 convolutional neural network is used to extract the features 20% of the 10-000 healthy and diseased pepper leaf images taken, from the real field environment, so as to overcome the problems of complex image background and low contrast, and to achieve end-to-end image semantic segmentation. Finally, the detection and classification of pepper images of four types of single diseases and insect pests and multiple diseases and insect pests are achieved. Meanwhile, in order to compare with other neural network models, this paper uses the transfer learning method to apply the VGG16, AlexNet, GoogLeNet, ResNet as well as MobileNet-V2 neural network structures which have been trained in image recognition in advance to the recognition of pepper diseases, so as to improve the generalization performance of the model and effectively reduce the time and space complexity of the convolution layer.The results show that: the average accuracy of mobilenet-v2 model is 93.05%, compared with other comparative models, it has the characteristics of high stability, low computational complexity and low memory consumption. At the same time, the images obtained under the actual planting conditions play an important role in the development of automatic detection, diagnosis and classification of pepper diseases and pests. It can be applied to the early warning of pepper disease, to solve the problem of disease prevention and control in the absence of front-line expert.

Keywords*: Deep learning; Transfer learning; Image recognition; Mobilenet-V2; Pest detection.*

I. Introduction

As an excellent economic crop, pepper has a wide market and high economic benefits. It plays an important role in enriching people's vegetable basket and ensuring the balanced annual supply of vegetables. In recent years, with the improvement of cultivation technology and the rapid development of logistics industry, pepper production has broken the regional and seasonal restrictions. Due to the durable storage and transportation characteristics of pepper, fresh pepper is available in vegetable markets all over the world throughout the year, which promotes the

large-scale and commercialized development of pepper production. According to the data, in 2020, the global pepper planting area is about 1.999 million hectares, a year-on-year increase of 3.3%, and the global pepper production is about 39.28 million tons, a year-on-year increase of 0.4%, as shown in Figure 1. In 2019, China's annual output of pepper is 19007248 tons, ranking first in the world, while Mexico's annual output of pepper is 3238245 tons, ranking second in the world. Planting pepper is an ideal way to help farmers get rid of poverty and become rich, which has important economic and social significance for invigorating domestic trade circulation and earning foreign exchange through export.

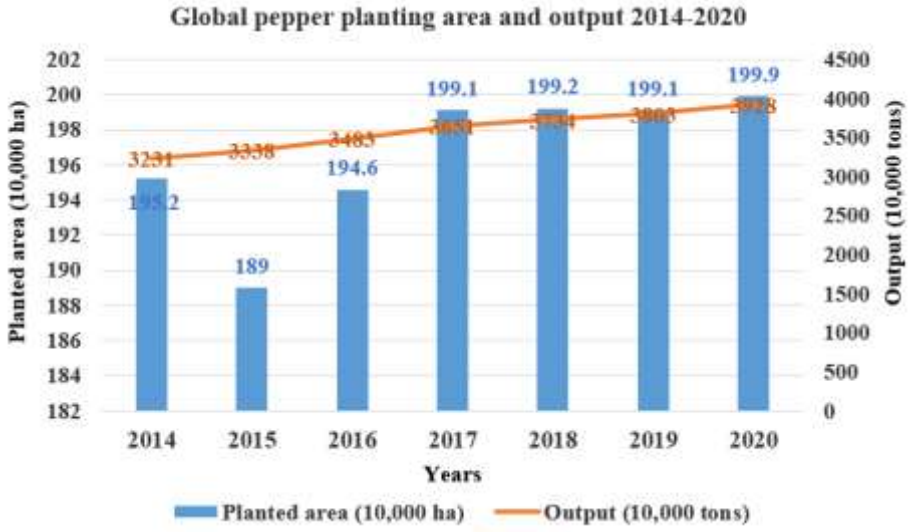

Figure 1: Global pepper planting and yield.

However, there are the following common diseases in the cultivation of pepper: pepper white spot disease, capsicum scab disease, pepper bacterial leaf spot disease, pepper soft rot disease, which will directly affect the yield and quality of pepper [1]. The traditional field diagnosis of diseases is done manually. Due to the large planting area, the shortage of grass-roots personnel in plant protection stations in recent years, and the lack of relevant knowledge of disease identification among farmers, it is difficult to realize the timely and accurate positioning of disease outbreak center and the precise determination of disease monitoring, which increases the risk of stable and high yield of pepper. At the same time, it is time-consuming and labor-consuming to train farmers' professional knowledge and skills of disease identification on a large scale, and there are various limitations of artificial diagnosis of crop diseases in the field, such as easy fatigue and large subjective error. Therefore, there is an urgent need for a more convenient, fast and accurate method for diagnosing pepper diseases in order to prescribe the right medicine to prevent and control the diseases in time, and reduce the pressure on crop production management. In this context, intelligent recognition of pepper disease based on disease image has always been a challenging research topic in precision agriculture.

Image recognition [2] refers to the technology of using computers to process, analyze and understand images, in order to identify targets and objects of various modes, which is a practical application of deep learning algorithms. In recent years, convolutional neural network as a representative of the depth recognition model has made important breakthroughs, constantly refreshing the performance level of multiple tasks such as optical and SAR image scene classification, target detection, semantic segmentation and change detection. At the same time, with the rise of artificial intelligence technology, the use of computer vision for automatic recognition and diagnosis of crop disease images has become a major research hotspot in recent years [3,4]. Grinblat et al. [5] developed a relatively simple but powerful neural network to successfully recognize three different legumes according to the morphological pattern of leaf veins. By introducing convolution and sampling operator, Wang Xiping et al.

proposed the apple lesion image recognition method based on time-varying impulse learning, and the correct rate is as high as 97.45% [6].

Although the majority of researchers have made corresponding progress in image recognition, they mostly focus on the diagnosis and recognition of a single disease with a single neural network model. At the same time, part of the research is based on the common data set training, and the data quality reliability is not high, which leads to a large deviation in the practical application of the model. This paper proposes an improved method, mainly from the following two aspects: (1) Using convolutional neural network mobilenet-v2 deep learning algorithm, 20% of the 10000 healthy and diseased pepper leaf images from the real field environment are extracted for feature extraction. The algorithm overcomes the problems of complex background, low contrast and different object size, and realizes end-to-end image semantic segmentation. Then the model is trained to realize the detection and classification of four types of single pest, multi pest and health images of pepper planted in the real field environment. (2) In order to compare with the four current mainstream neural network models VGG-16, AlexNet, GoogLeNet as well as ResNet in the detection and classification ability, and speed up the convergence speed of the model in the training process, this study used the method of transfer learning [7] to apply the five mainstream neural network structures trained in image recognition to the recognition of pepper diseases. The generalization performance of the model is improved, and the time and space complexity of the convolution layer is effectively reduced.

The experimental results show that: compared with other comparative models, MobileNet-V2 model has higher stability, and the average recognition rate of four pepper diseases reaches 93.05%, which has real and better diagnosis effect compared with related research results. At the same time, TensorFlow Lite is used to lighten the model, and then the model is transplanted to Android App of mobile phone, which is applied to early warning of pepper disease, to solve the problem of pepper growers' disease prevention and control without the guidance of front-line experts.

II. Methods

2.1 Convolutional neural network model

Convolutional neural network is a kind of feedforward neural network which contains convolution calculation and has deep structure. As one of the representative algorithms of deep learning, it has the ability of representational learning. It can take the original image as input and form a more abstract high-level feature through the convolutional network structure, so as to obtain a deeper data feature representation and avoid the complexity of manual extraction of image features. The classic networks include LeNet-5, AlexNet, GoogLeNet, VGG-16, ResNet, MobileNet, etc.

2.1.1 Network structure

The convolutional neural network in deep learning is developed from traditional artificial neural network and widely used in the field of image recognition. Its network structure includes multiple convolution layer, pooling layer and full connection layer. The convolution layer uses the local correlation of the pixels in the image to extract features, and uses the convolution kernel as the intermediary to interact with the image to obtain the local information of the image. The image size after convolution will be as shown in Formula 1. Where S represents the step size, [N1, N1] represents the size of the original image, [N2, N2] represents the convolution kernel size.

$$\left[\frac{(N1-N2)}{S}+1,\frac{(N1-N2)}{S}+1\right] \tag{1}$$

The pooling layer is sandwiched between successive convolution layers to compress the amount of data and parameters and reduce over fitting. Pooling refers to the aggregation operation of computing the maximum or average value of a specific feature in a region of an image. Common pooling functions include maximum pooling and mean pooling.

Each neuron in the full connectivity layer is fully connected with all neurons in the previous layer to integrate the local information with class differentiation in the convolution layer or pooling layer. In general, the convolutional neural network uses gradient descent method and minimum loss function to adjust the weight parameters of the network layer by layer, and improves the accuracy of the network through multiple rounds of iterative training. The principle of convolution neural network is shown in Figure 2.

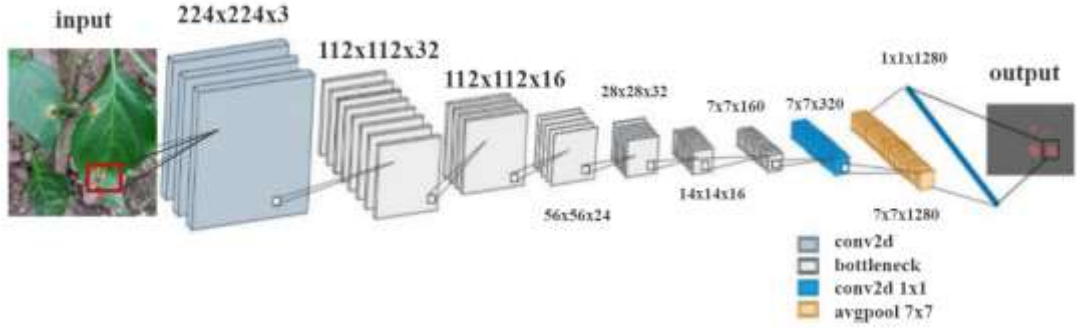

Figure 2: Basic principle of convolutional neural network.

2.1.2 End to end image semantic segmentation technology

In order to solve the problems of dense growth of pepper plants, disordered background formed by leaves and soil, and complex outdoor lighting in natural scenes, an end-to-end image segmentation method based on full convolution network is proposed to segment plants, leaves, fruits and other images, so as to reduce the interference of complex background on subsequent recognition. In the aspect of image enhancement, the lightweight convolutional neural network is used to construct the restoration network through the counter learning idea, and the clear and high separation rate image is directly generated. As shown in Figure 3.

Figure 3 : Image enhancement effect, left original image, right image enhanced image.

Macroscopically, image segmentation is the process of dividing the image into different regions and extracting the interested parts according to certain similarity criteria, as shown in Figure 4. Semantic segmentation is an operation to group pixels according to different semantic expressions in an image. The goal of segmentation is to label every pixel in the image and correspond each pixel to its category.

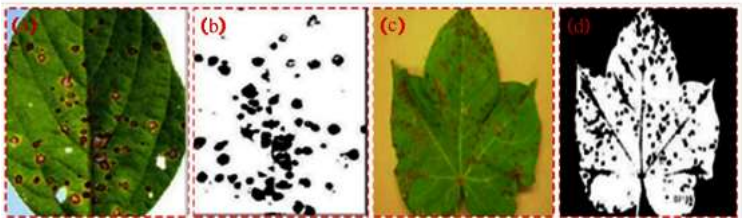

Figure 4: Schematic diagram of end-to-end semantic segmentation model.

The end-to-end semantic segmentation method used in this paper is to train a deep neural network classifier based on a large number of labeled images to classify each pixel in the image. Input the original image, take the true value of the image pixel as the supervision information, and output the semantic segmentation effect image corresponding to the original image, but the middle segmentation process is equivalent to a black box for users, which is the end-to-end output model.

2.2. Mobilenet model

2.2.1 MobileNet-V1

Mobilenet is a lightweight CNN neural network proposed by Google in 2017, which focuses on mobile devices and embedded devices, and has rapidly derived three versions of V1, V2 and V3 to solve the problem that complex models cannot be used on mobile terminals and embedded terminals [8]. Compared with other neural networks, the computational complexity and model memory of MobileNet are very small. The total network parameters are 1.32 million, and the memory is only 3% of vgg16. At the same time, the core of MobileNet model is depthwise separable convolution (DSC). The principle is shown in Figure 5. DSC decomposes standard convolution into depth convolution and point by point convolution.

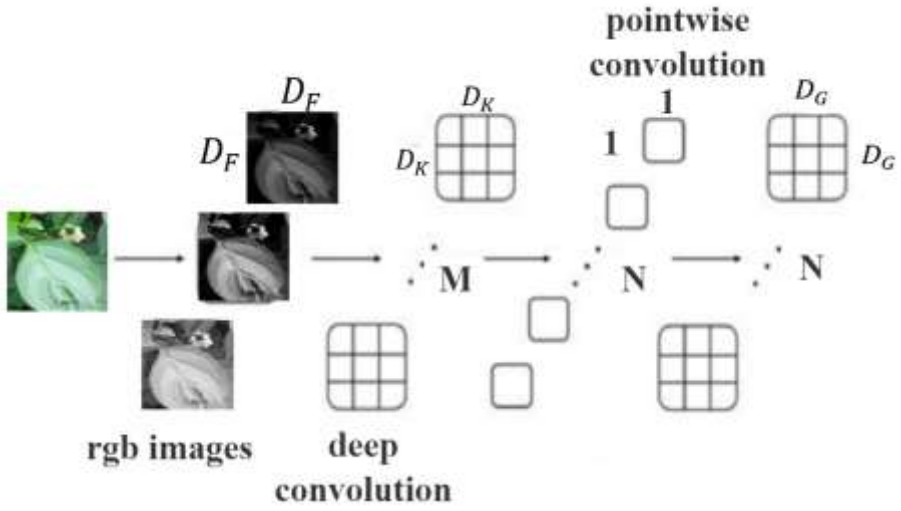

Figure 5 : Principle of deep separable convolution.

Assuming that the dimension of the input image of the neural network is $DF \times DF \times M$, the size of the convolution kernel of the deep convolution is $DK \times DK \times 1$ and the number is M, the size of the convolution kernel of the point-by-point convolution is $1 \times 1 \times M$ and the number is N. After convolution, the output dimension is $DG \times DG \times N$. Then the calculation cost of the depth separable convolution and the ratio of the conventional convolution operation are shown in formula (2).

$$\frac{\text{DSC}}{\text{Std conv}} = \frac{1}{N} + \frac{1}{D_K^2} \tag{2}$$

DSC and Std conv represent the parameters of deep separable convolution and conventional convolution respectively; N is the number of convolution kernels; DK is the size of the deep convolution kernel. It can be seen from formula (1) that when the number of convolution kernels is less and the size is smaller, the depth separable convolution can reduce more parameters than the conventional convolution operation.

At the same time, MobileNet-V1 adopts two global super parameters: width multiplier and resolution multiplier, which can effectively balance the delay and accuracy, and select the appropriate model size according to the actual problem size of the modeler. Compared with other existing network models, MobileNet-V1 takes the optimization delay and model size into consideration, which effectively reduces the amount of computation.

2.2.2 MobileNet-V2

MobileNet-V2 is an improvement of MobileNet-V1. In essence, it is still a lightweight convolutional neural network, and its architecture is based on reverse residual structure. The common point is that both of them use depth convolution and point convolution to achieve feature extraction, and the amount of calculation is less than that of standard convolution. Mark Sandler et al. Pointed out in "mobilenetv2: inverted residuals and linear bottlenecks" [9] that the differences between mobilenet-v2 and mobilenet-v1 are mainly in two aspects: (1) there is a 1 * 1 expansion layer before the depthwise convolution, which aims to increase the number of channels and obtain more image features. Because DW itself cannot change the number of channels, the number of output channels depends on the number of upper transmission channels. If the number of upper channels is small, DW can not get better features in low dimension, adding PW before DW can extract features in high dimension. (2) MobileNet-V2 adopts linear instead of ReLU. Because the input features are mainly in low dimensional space, the integrity of input information is ensured by using ReLU. For low latitude space processing, ReLU is approximately a linear transformation. The purpose is to prevent ReLU from destroying features.

2.2.3 Reverse residual block

With the increase of network depth, training becomes more and more difficult, and the phenomenon of network degradation appears. This is mainly because in the network training process based on gradient descent algorithm, the multi-layer back propagation of error signal is easy to cause gradient dispersion, that is, the return training error signal is extremely weak when the ladder is too small, or the gradient explosion, that is, NaN phenomenon appears in the model due to the large gradient.

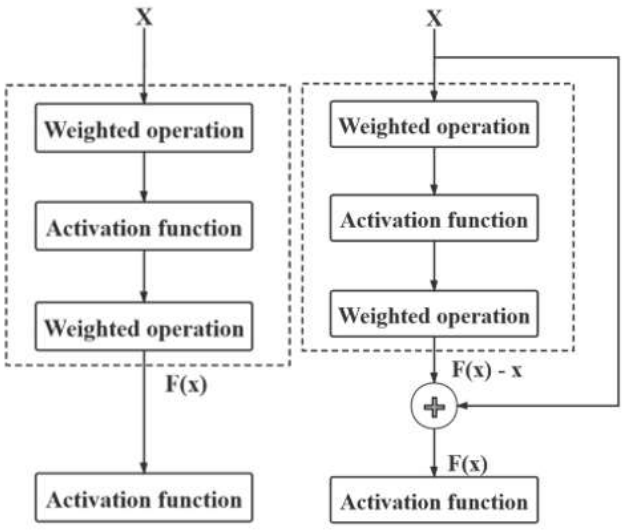

Figure 6: The structure diagram of residual learning, the left is no residual structure, the right is residual structure.

The structure of residual block and bottleneck network is very similar. Each residual block contains an input, then several bottlenecks are connected and expanded. The residual block is composed of the main path and the jump link. The output $F(x) - x$ is obtained after x is weighted by the first layer and the second layer by the nonlinear variation of relu. Compared with the network with or without residual structure, the deep residual network [10] divides a series of training into blocks to train, so as to minimize the error of each block and ultimately achieve the

goal of minimizing the overall error. Therefore, there will be no gradient dispersion phenomenon.

Although the input of the bottleneck layer contains all the necessary information, and the use of ReLU can increase the sparsity of the model, if the information is too sparse, it will be lost. In order to prevent the information loss, no ReLU activation function is added behind the leftmost layer in Figure 7 (b), which is the reason for the introduction of inversion. The main principle of reversing residual block is: the information is more abundant after the dimension is increased. At this time, the ReLU activation function is added, and then the dimension is reduced. In theory, all the necessary information is not lost, and this way of reversing memory efficiency will be much higher. Figure 7shows a comparison of the traditional residual block and the inverted residual block.

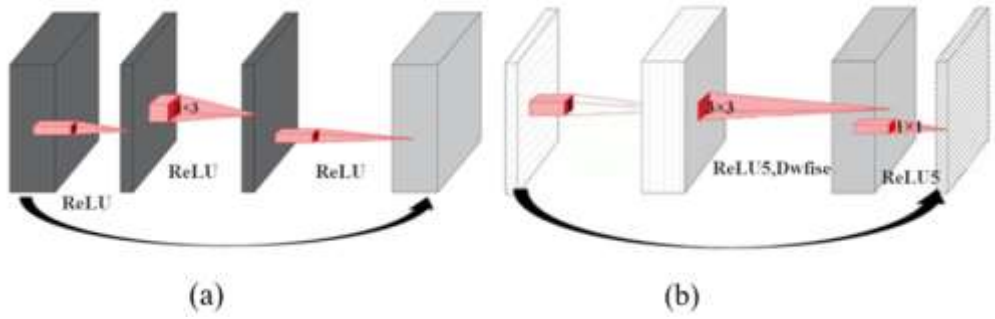

Figure 7: (a): traditional residual block (b): reverse residual block.

2.3 Transfer learning

Transfer learning solves the problem of how to transfer learned knowledge from one scenario to another. It is necessary to combine deep learning with transfer learning because the combination of the two can save resources while improving the accuracy of the model. When we lack enough data to complete the training, we can realize the generalization ability of the model itself through transfer learning.

The transfer learning first keeps the structure of the model convolutional layer unchanged, and then loads the trained weights and parameters into the convolutional layer. Then, a full-connection layer suitable for the new task is designed, the original full-connection layer is replaced with the newly designed full-connection layer, and a new convolutional network model is formed with the previous convolutional layer. Finally, the new model is trained with the new data set images. Comparison between traditional machine learning and transfer learning is shown in Figure 8.

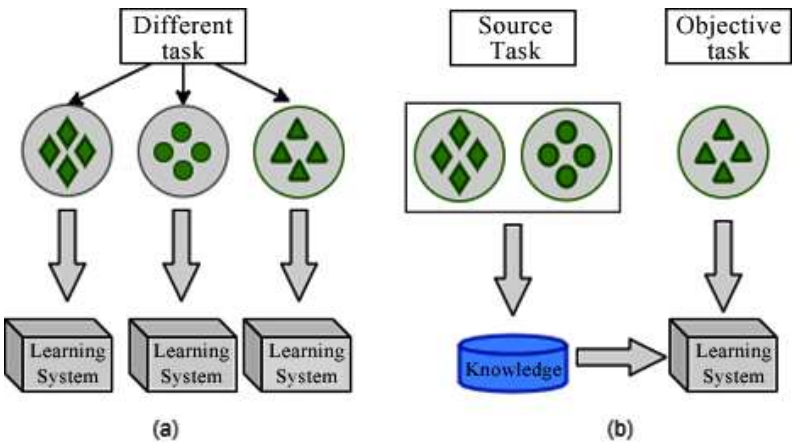

Figure 8: (a) traditional machine learning principle, (b) transfer learning principle.

III. Materials and Experiments

3.1 Data set preprocessing

As shown in Table 2, the data of this study come from: (1) the open image database of pepper diseases accounts for 30%. (2) The pepper disease database of Chengdu Academy of agriculture and Forestry Sciences (taken in Xichang pepper planting base from April to may 2021) accounts for 50%. (3) 20% of the field photos were taken by researchers. The data set contains 10000 images of pepper leaves and fruits, including 2000 images of pepper scab, pepper white spot, pepper soft rot, pepper bacterial leaf spot and healthy pepper. It should be noted that soil, multi leaf overlap, other plants, light as well as other factors will have an impact on the detection and classification of pepper diseases and pests under the real field conditions, so the image complexity increases, making the experiment closer to the actual situation.

Table 1: Description of experimental sample data set

diseases / *sources*	**bacterial spot**	**white spot**	**soft rot**	**bacterial leaf spot**	**healthy**
Network(30%)	550	600	600	600	650
research institute(50%)	1000	1000	1000	1000	1000
field shooting(20%)	450	400	400	400	350

4 common diseases of pepper are illustrated in Figure 8. Pepper scab is one of the main diseases in pepper production. The incidence rate of common plots is 20-30%, and the serious disease region can reach 100%, which causes premature leaves of the damaged plant leaves, seriously affecting the yield and quality of capsicum, and reducing the yield of capsicum. The cotyledons first appeared silver white spots, and then gradually turned into light black sunken spots with rough surface like scab; In the late stage of pepper white spot disease, the middle part of leaf spot is perforated, which will seriously affect the quality and yield of pepper. The soft rot of pepper mainly damages the fruits. The diseased fruits first appear dark green spots, then turn brown and soft rot, with a foul smell; The inner flesh rotted and the peel turned white; The leaf of pepper bacterial leaf spot first appeared spot disease, water immersion, irregular shape, chlorosis, and then developed into brown to rust color, and the diseased part was membranous. When the weather is dry, the disease spot is rust color, the disease spot is crisp, and some perforation

Figure 9: Images of four common pepper diseases

In this experiment, the above data set is randomly divided into training set and verification set according to the ratio of 8:2. The training set is used for model training, the verification set is used for model parameter tuning, and the test set is used to test the generalization ability of the model. In order to reduce the data size and speed up the model training process, all images are compressed to 640 × 480 size. Some preprocessed images are shown in Figure 10.

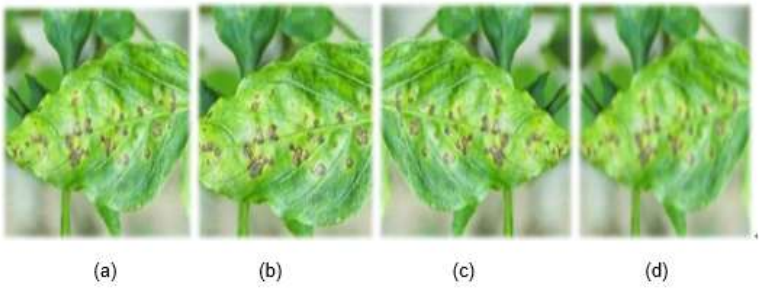

Figure 10: Data preprocessing results: (a) original image, (b) clipping, (c) flipping, (d) smoothing

3.2 model training

3.2.1 five classification function

In order to recognize common pepper diseases efficiently and accurately, this paper improves the classifier to 5-objective softmax classifier. Softmax classifier is an algorithm that divides a target variable into several categories. It can accept the characteristic matrix of full connection layer input and output different probability values of each category corresponding to the input target. For example, there are N input targets $\{x_i \cdot y_i\}_{i=1}^{k}$, and each target is labeled $y_i = \{1,2,3 \ldots\}$, k is the number of types of model output categories ($k \geq 2$). In this paper, five classifiers were made for scab, white spot, soft rot, bacterial leaf spot and healthy pepper and the value of k is 5. For a given input x_i using hypothesis function $f_\theta(x_i)$ estimate the corresponding class probability j value $P(y_i = j|x_i)$. The function is shown in formula (3).

$$f_\theta(x_i) = \begin{bmatrix} P(y_i = 1 \mid x_i; \theta) \\ P(y_i = 2 \mid x_i; \theta) \\ P(y_i = 3 \mid x_i; \theta) \\ P(y_i = 4 \mid x_i; \theta) \\ P(y_i = 5 \mid x_i; \theta \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x_i}} \begin{bmatrix} e^{\theta_1^T x_i} \\ e^{\theta_2^T x_i} \\ e^{\theta_3^T x_i} \\ e^{\theta_4^T x_i} \\ e^{\theta_5^T x_i} \end{bmatrix} \tag{3}$$

θ is the parameter of softmax classifier. In order to ensure that the sum of probabilities is 1, formula 3 is used for normalization. The loss function of softmax classifier is shown in formula 4, where $(y_i = j)$ is an indicative function whose value is consistent with the true value in brackets, that is, when $y_i = j$ is true, the function value is 1, otherwise it is 0.

$$\frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x_i}} \tag{4}$$

$$J(x, y, \theta) = -\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{j=1}^{k} 1\{y_i = j\} \log_2 \left(\frac{e^{\theta_j^T x_i}}{\sum_{j=1}^{k} e^{\theta_j^T x_i}}\right)\right] \tag{5}$$

3.2.2 Mobilenet-V2 model

The process of model training and testing is shown in Figure 11. In this paper, several common data enhancement methods such as normalization, rotation, flipping, displacement, scaling and smoothing are used for the original

image, so as to improve the invariance of translation and scaling of the model and effectively prevent the model from over fitting. At the same time, in order to accelerate the convergence speed of the model in the training process, this study uses the transfer learning method to transfer the pre-trained weights of different network structures on the data set to the corresponding different network structures, and on this basis to train different neural network models.

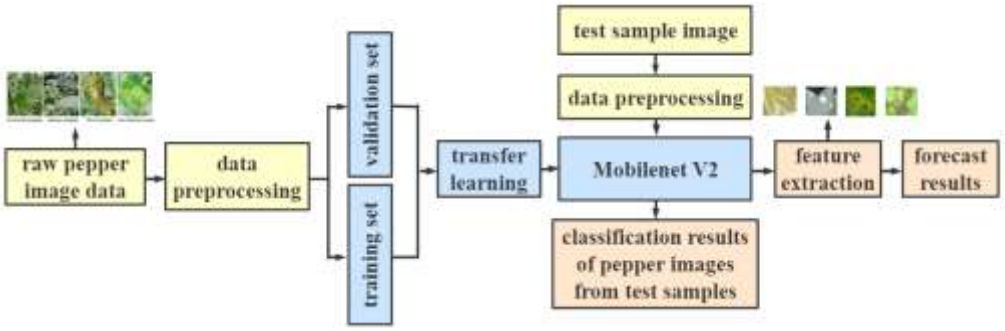

Figure 11: Process of model training and testing

At the same time, TF-Hub is used to load Mobilenet-V2 model, and linear classifier is integrated into feature extractor layer and hub models. In order to prevent over fitting and improve the generalization ability of the model, two groups of dropout layers are added in the full connection layer, with dropout rate of 0.5. When the activation function Leaky ReLU is added to the dropout of each layer, it is a better activation function than ReLU. In addition to inheriting the advantages of relu, such as reducing the amount of computation, solving the problem of gradient vanishing [11] and alleviating over fitting, Leaky ReLU gives a non-zero slope to all negative values of input, which solves the problem that neurons do not learn when the ReLU function enters the negative range. Moreover, the value of the parameter alpha controls the gradient of the negative part of the linear function.

In addition, in order to control the complexity of the model and reduce the over fitting, the weights are regularized. The improved network structure parameter configuration is shown in Table 2. Through these parameters, you can see the composition of each layer of the model. After the data passes through each layer, the dimensions of the data and the number of parameters of each layer are output. For CNN model, the calculation method of Param is :(length of convolution kernel, number of channels of width of convolution kernel +1) * number of convolution kernel.

Table 2: Configuration of improved network structure parameters

Layer	parameter	value
Dropout	rate	0.5
Leaky-ReLU	alpha	0.05
Dense	units	512
	Activation	relu
Dropout	rate	0.5
Leaky-ReLU	alpha	0.05
Dense	units	train_generator.num_classes
	activation	softmax
Note: rate is dropout rate, rate = 0.5 means 50% of random nodes will be hidden in each training, units is the number of neurons, activation is the activation function		

3.3 Super parameter setting

After many experiments and comparisons, the main super parameters of the model are set as shown in Table 3. The model optimizer is Adam. At the same time, the loss function is set as the classification cross entropy function, and the network evaluation index is labeled as accuracy.

Table 3: Super parameter configuration

parameter	value	explain
epochs	140	training times
learning_rate	0.0001	learning rate
batch_size	128	sample batch
momentum	0.9	quantity of motion
weight decay	0.0005	weight decay

IV. Results & Discussion

4.1 Model evaluation criteria

The accuracy of the model is the proportion of the number of correctly classified samples to the total number of samples, which can reflect the prediction ability of a model as a whole. And its definition is shown in formula 6.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{6}$$

TP (true positive) refers to the number of positive classes identified by the model, TN (true negative) refers to the number of negative classes identified by the model, FP (false positive) refers to the number of false positives of the model, that is, the number of negative class images predicted as positive classes, FN (false negative) refers to the number of missed reports of the model, namely the number of positive classes that have not been recognized.

4.2 Result analysis

In order to determine the best model for pepper disease detection and classification, this study compared MobileNet-V2 with VGG16, AlexNet, GoogLeNet as well as ResNet in terms of accuracy, loss function and parameters. Figure 11 shows the change of verification set accuracy and loss function during iteration. It can be seen from Figure 12 that different models show different characteristics in the change of verification set accuracy during the iteration process. Specifically, in addition to GoogleNet, the accuracy of the five models in the iterative process jitters. For mobilenet-v2 model, although the accuracy of verification set is low in the initial iteration, with the deepening of training, the accuracy has been on the rise, and after 75 rounds, it surpasses the accuracy of the other five models. The average accuracy is 93.05%, which is slightly inferior to the image recognition accuracy only under laboratory conditions, but closer to the actual situation of field detection.

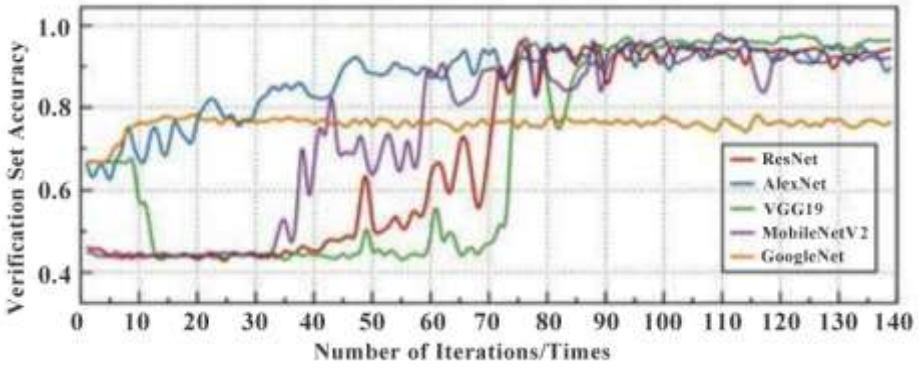

Figure 12: change of accuracy of verification set.

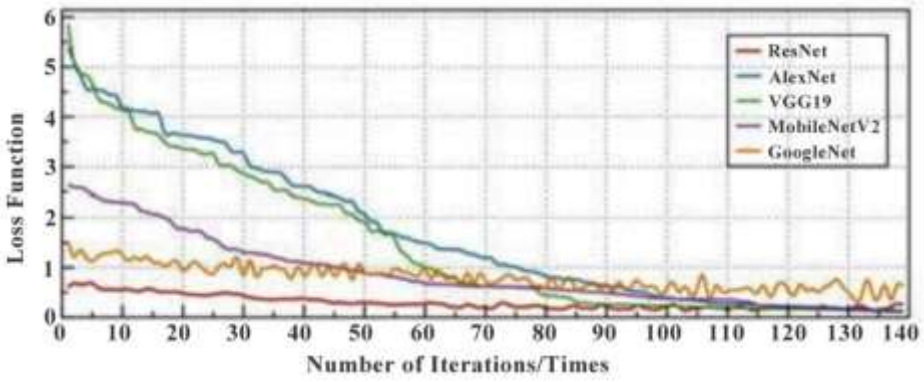

Figure 13: Change of loss function of verification set.

Table 4shows the comparison of the accuracy of the final verification set of the five models. It can be seen from the table that although VGG16 and AlexNet have high accuracy, the memory occupancy of the model with 128×128 3-channel input reaches 1055.4 and 510.3MB, respectively, which will inevitably reduce the running speed of some devices such as smart phones and take up a large amount of memory space.

Table 4: Performance of different network structures on datasets.

network model	verification set accuracy/%	memory usage/MB
VGG-16	96.22	1055.4
AlexNet	90.12	510.3
GoogLeNet	75.64	98.7
ResNet	94.46	178.6
MobileNet-V2	93.05	24.5

The results show that: the average accuracy of mobilenet-v2 model is 93.05%, although its accuracy is slightly lower than VGG16 and ResNet, it has the advantages of high stability, low computational complexity, minimum network parameters and low memory consumption compared with other comparative models. At the same time, the image recognition under the actual cultivation conditions is more difficult and complex than that under the laboratory conditions. Adding the images obtained under the actual cultivation conditions in the field plays an important role in the development of practical application of automatic detection, diagnosis and classification of pepper diseases.

V. Conclusions

As an excellent cash crop, pepper has a wide market and high economic benefits. It plays an important role in enriching people's vegetable basket and ensuring the balanced annual supply of vegetables. According to the data, in 2020, the global pepper planting areas were about 1.999 million hectares, with a year-on-year growth of 3.3%, and the global pepper output was about 39.28 million tons. In the process of cultivation and management of pepper, disease is an important factor that restricts its quality improvement and yield growth. Due to the lack of professional knowledge and expert guidance in pest identification and control, it is easy to cause significant economic losses when pests occur.

With the development of artificial intelligence technology, automatic recognition and diagnosis of crop disease images by computer vision has become a hot research topic at home and abroad in recent years. In this paper, the convolutional neural network MobileNet-V2 deep learning algorithm is used to extract the features of 100000 healthy and diseased pepper leaf images, which are 20% from the real field environment. The problems of complex

background and low contrast are overcome, and the end-to-end image semantic segmentation is realized. As a result, we can detect and classify four types of single and multiple diseases and pests as well as healthy pepper images. At the same time, in order to compare with other neural network models, this paper uses the transfer learning method to apply the VGG16, AlexNet, GoogLeNet, ResNet and MobileNet-V2 neural network structures which have been trained in image recognition in advance to the recognition of pepper diseases, so as to improve the generalization performance of the model and effectively reduce the time and space complexity of the convolution layer.

The results show that the average accuracy of MobileNet-V2 model is 93.05%, although its accuracy is slightly lower than VGG16 and ResNet, it has the characteristics of high stability, low computational complexity, minimum network parameters and low memory consumption compared with other comparison models. At the same time, the images obtained under the actual cultivation conditions in the field play an important role in the development of practical application of automatic detection, diagnosis and classification of pepper diseases and pests. It can be used in the early warning of pepper disease, solve the problem of pepper growers in the absence of front-line expert guidance for disease control, and provide a contribution to the realization of rural revitalization.

Another problem that needs to be improved is that the real pictures of pepper diseases in this paper are only from Sichuan Province, and we plan to obtain more image data from different geographical locations and cultivation conditions in the future. At the same time, more crops and diseases are included to expand the existing data set, so as to further train and improve the mobile net-v2 convolutional neural network model. Of course, this process is challenging and will take more time. In fact, deep learning is not only convolutional neural network, but it is the most widely used and the most in-depth research in crop disease recognition. We hoped that researchers can make more attempts on other types of deep learning networks in the future. Despite the limitations of artificial neural network, the computer does not have the pre-cognitive knowledge reserve of things, can’t reach the complexity of human brain neural network, and its learning ability can’t be compared with human beings, but there is no denying the fact that deep learning still is currently the most potential crop disease image recognition method of modernization.

Acknowledgments

This work was sponsored by Science and Technology Program of Sichuan Province, China (Grant No. 2020YFS0090, 2021YFN0117).

References

[1] Milan Šević, Katarina Gašić, Maja Ignjatov, Mirjana Mijatović, Anđelka Prokić, Aleksa Obradović. Integration of biological and conventional treatments in control of pepper bacterial spot[J]. Crop Protection,2019,119.

[2] Wang J H. Research on image recognition technology based on convolution neural network[C]，Proceedings of 4th International Workshop on Materials Engineering and Computer Science，2019: 160-164.

[3] KAMILARIS A, PRENAFETA-BOLDÚ F X. A review of the use of convolutional neural networks in agriculture[J]. The Journal of Agricultural Science, 2018, 156(3): 312-322.

[4] Rahman C R, Arko P S, Ali M E, et al. Identification and recognition of rice diseases and pests using convolutional neural network[J]. Biosystems Engineering, 2019, 194: 112-120.

[5] Guillermo L. Grinblat, Lucas C. Uzal, Mónica G. Larese, Pablo M. Granitto. Deep learning for plant identification using vein morphological patterns[J]. Computers and Electronics in Agriculture,2016,127.

[6] WANG X P, HUANG T, TAN W X, et al. Apple lesion image recognition method based on convolutional network [J]. Computer engineering, 2015, 41(12): 293-298.

[7] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data

Engineering, 2009, 22(10): 1345-1359.

[8] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [EB/OL]. 2017. arXiv: 1704.04861 [cs.CV].

[9] MobileNetV2: Inverted Residuals and Linear Bottlenecks, Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. arXiv preprint. arXiv:1801.04381, 2018.

[10] He K，Zhang X，Ren S，et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016: 770-778.

[11] Bengio Y，Simard P，Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157-166.

ISSN: 0010-8189