Cotton leaf disease detection method based on improved SSD

Wenjuan Guo[1,2], Shuo Feng[3], Quan Feng[1*], Xiangzhou Li[1], Xueze Gao[1]

(1. *College of Mechanical and Electrical Engineering, Gansu Agricultural University, Lanzhou 730070, China*;
2. *School of Cyber Security, Gansu University of Political Science and Law, Lanzhou 730070, China*;
3. *College of Mechanical and Electrical Engineering, Sichuan Agricultural University, Ya'an 625014, China*)

Abstract: In response to the problems of numerous model parameters and low detection accuracy in SSD-based cotton leaf disease detection methods, a cotton leaf disease detection method based on improved SSD was proposed by combining the characteristics of cotton leaf diseases. First, the lightweight network MobileNetV2 was introduced to improve the backbone feature extraction network, which provides more abundant semantic information and details while significantly reducing the amount of model parameters and computing complexity, and accelerates the detection speed to achieve real-time detection. Then, the SE attention mechanism, ECA attention mechanism, and CBAM attention mechanism were fused to filter out disease target features and effectively suppress the feature information of jamming targets, generating feature maps with strong semantics and precise location information. The test results on the self-built cotton leaf disease dataset show that the parameter quantity of the SSD_MobileNetV2 model with backbone network of MobileNetV2 was 50.9% of the SSD_VGG model taking VGG as the backbone. Compared with SSD_VGG model, the *P, R, F*1 values, and *mAP* of the MobileNetV2 model increased by 4.37%, 3.3%, 3.8%, and 8.79% respectively, while *FPS* increased by 22.5 frames/s. The SE, ECA, and CBAM attention mechanisms were introduced into the SSD_VGG model and SSD_MobileNetV2 model. Using gradient weighted class activation mapping algorithm to explain the model detection process and visually compare the detection results of each model. The results indicate that the *P, R, F*1 values, *mAP* and *FPS* of the SSD_MobileNetV2+ECA model were higher than other models that introduced the attention mechanisms. Moreover, this model has less parameter with faster running speed, and is more suitable for detecting cotton diseases in complex environments, showing the best detection effect. Therefore, the improved SSD_MobileNetV2+ECA model significantly enhanced the semantic information of the shallow feature map of the model, and has a good detection effect on cotton leaf diseases in complex environments. The research can provide a lightweight, real-time, and accurate solution for detecting of cotton diseases in complex environments.

Keywords: cotton disease detection, SSD, MobileNetV2, attention mechanism
DOI: 10.25165/j.ijabe.20241702.8574

Citation: Guo W J, Feng S, Feng Q, Li X Z, Gao X Z. Cotton leaf disease detection method based on improved SSD. Int J Agric & Biol Eng, 2024; 17(2): 211–220.

1 Introduction

Crop diseases constrain the sustainable development of agriculture and form one of the challenges that plague agricultural production[1]. Cotton is an important cash crop in China, and the output and consumption of cotton in China ranks the first in the world. However, cotton diseases can lead to a significant decrease in production and cause huge losses to the agricultural economy[2]. Therefore, early diagnosis and control of cotton diseases are important safeguard measures for high cotton yield.

Traditional cotton disease detection is mainly achieved through manual on-site diagnosis, which has the drawbacks of high workload and strong subjectivity, and cannot meet the needs of real-time monitoring of large-scale cotton diseases. The use of computer information technology for detecting and identifying cotton diseases is an advanced and effective method. Scholars have used classic machine learning methods to detect and recognize cotton diseases and have achieved high accuracy[3-6]. However, classic machine learning methods include three processes: disease image segmentation, disease feature extraction, and pattern recognition. The performance quality of these methods depends on whether useful disease features can be extracted. In addition, the process of generating features for cotton disease recognition is time-consuming and laborious, and the generalization performance of the methods is poor.

Compared with classical machine learning methods, deep learning adopts end-to-end learning, inputting raw data, outputting target tasks, and gradually abstracting the raw data into the required features for the target task through layer by layer extraction, which can avoid the impact of manually selected features on classification performance and effectively enhance the robustness of the model. In recent years, the emerging deep learning technology based on convolutional neural networks has been applied in the detection and recognition of crop diseases. For example, Nazki et al.[7] used GAN networks to identify tomato diseases and achieved high accuracy. Liu et al.[8] improved the SqueezeNet model to identify multiple types of leaf diseases. Wang et al.[9] proposed a bimodalNet crop disease identification model for disease identification of six crops. Li et al.[10] used an improved lightweight residual network to identify plant leaf diseases, which has a low error recognition rate. However,

Received date: 2023-10-10 **Accepted date:** 2024-03-06

Biographies: Wenjuan Guo, PhD candidate, research interests: image processing and recognition, Email: 565105996@qq.com; **Shuo Feng**, Undergraduate, research interests: image processing and recognition, Email: 191763445@qq.com; **Xiangzhou Li**, PhD candidate, research interests: Image processing and recognition, Email: 572395298@qq.com; **Xueze Gao**, PhD candidate, research interests: image processing and recognition, Email: gaoxueze@163.com.

***Corresponding author: Quan Feng**, PhD, Professor, research interests: image processing and recognition. College of Mechanical and Electrical Engineering, Gansu Agricultural University, Lanzhou 730070, China. Tel: +86-451-7632472, Email: fquan@sina.com.

there is relatively little research on the detection and recognition of cotton diseases based on convolutional neural networks in existing studies. For example, Zhang et al. used an improved VGG convolutional neural network to identify cotton diseases and achieved good classification results[11]. Wang et al.[12] used an adaptive discriminant deep confidence network for predicting cotton pests and diseases, with a prediction accuracy of 82.84%. Zhao et al.[13] recognized cotton leaf diseases and pests through transfer learning, which has a high classification accuracy. However, in the above research, the background of the dataset images is clear and simple, and the format of the images is standardized, which is not in line with the actual application environment. In the real agricultural production environment, crop images taken by farmers have complex and changeable backgrounds, and the location of diseases is generally not centered. Therefore, it is difficult for disease detection and recognition models in simple backgrounds to meet the needs of real agricultural production.

Given the above reasons, scholars have introduced object detection algorithms based on deep learning for crop disease detection in complex backgrounds[14]. Using object detection algorithms for detecting crop leaf diseases has high detection accuracy and fast detection speed[15]. For instance, Fuentes et al.[16] completed object detection on tomato disease dataset with the Faster R-CNN, R-FCN and SSD models, and the conclusions show that the combined model of Faster R-CNN and VGG16 showed the maximum disease detection rate. Li et al.[17] applied the improved Faster R-CNN in bitter gourd leave disease detection, and the improved model presented high robustness and efficiency. Ye et al.[18] used SSD model to realize crop disease detection based on self-built data set with complicated background and achieved an average detection precision of 83.90%. Liu et al.[19] proposed an improved model based on MobileNetv2 and YOLOv3, which conducted early detection on grey speck disease of tomato. The improved model has the advantages of small memory size, high detection precision and fast identification speed. Wen et al.[20] applied the improved YOLOV3 algorithm to detect the diseases of pseudo-ginseng leaves, achieving good detection results.

In order to further improve the detection performance of cotton leaf diseases in complex backgrounds, considering the characteristics of cotton leaf diseases, in this paper, an improved SSD model for cotton leaf disease detection was proposed, which addresses the failure of SSD model in fully utilizing shallow high-resolution feature maps and its inability in distinguishing feature weights. A lightweight network MobileNetV2 was applied to improve the backbone feature extraction network, which could provide more abundant semantic information and details by reducing the number of model parameters and the computation amount. Integrating different attention mechanisms help to screen out disease target features, effectively suppress feature information of jamming target, and balance the weight of feature information in the feature map.

2 Test data

The dataset in this study comes from the captured cotton leaf disease images, and the samples in the dataset are the images of cotton leaf diseases shot in the complicated background of real field environment. The self-built dataset collected 1666 cotton leaf images with different leaf sizes, disease types, and brightness levels. The dataset includes 6 types of cotton diseases, which are anthracnose, brown spot, verticillium wilt, fusarium wilt, ring rot, and white mold as shown in Figure 1.

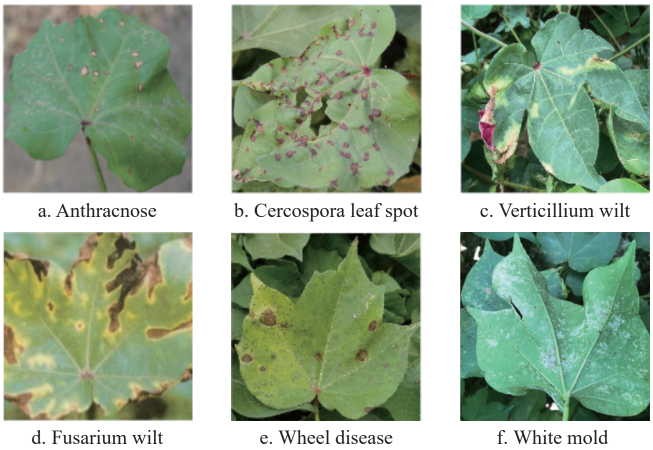

a. Anthracnose b. Cercospora leaf spot c. Verticillium wilt
d. Fusarium wilt e. Wheel disease f. White mold

Figure 1 Sample image of cotton disease

Then uniformly clip leaf images to 640×640 pixels in size, label the leaf images using the Labelimg tool. In model training, the dataset was divided into training set, validation set, and test set. In this study, based on the 9:1 proportion, 1500 images of training set and validation set and 166 images of test set were generated. Subsequently, based on the 9:1 proportion, 1350 images of training set and 150 images of test set were generated. The number of images of various diseases in the dataset is listed in Table 1.

Table 1 The number of images of various diseases

Disease types	Graphics	Training set	Validation set	Test set
Anthracnose	250	203	22	25
Cercospora leaf spot	285	231	26	28
Verticillium wilt	231	187	21	23
Fusarium wilt	270	219	24	27
Wheel disease	225	181	21	23
White mold	405	329	36	40

The construction process of disease data set is as follows: First, cotton disease data set is obtained after image preprocessing; Then, LabelImg software was used to annotate the images. Then, the data set is divided into training set, verification set and test set. Finally, the images were enhanced by brightness transformation, noise addition, scaling, rotation and mirroring, and finally cotton disease data set samples were obtained.

3 Model introduction

3.1 SSD model

There are mainly two types of object detection algorithms based on CNN: one is the object detection method based on regional proposal, namely, the two-stage object detection, represented by the Faster R-CNN algorithm[21]. This type of algorithm first generates candidate boxes based on heuristics or convolutional neural networks, and then classifies and regress the candidate boxes. It has the advantage of high precision; however, the algorithm has a longer training time and slower running speed. The second method is the one stage object detection with no-region proposal, in which typical algorithms are YOLO series algorithm and SSD algorithm[22-26]. This type of algorithm performs uniform and dense sampling at different scales at various positions of the image, followed by classification and regression by convolutional neural networks. It runs fast, but the accuracy of the algorithm is low and is not suitable for small target detection.

The SSD model combines the anchor mechanism of the Faster R-CNN algorithm and the regression idea of the YOLO algorithm to detect targets through multi-scale feature maps. The SSD model includes two parts: the prebase network and expanded network, and the network structure is shown in Figure 2. The prebase network is used to extract image features, and the VGG16 network was used as the backbone network for feature extraction[27]. The extended network consists of multi-scale feature maps for object classification and detection. The SSD model modified the last two full convolutional layers (FC6, FC7) of VGG16 to convolutional layers (Conv6, Conv7), while adding convolutional layers of Conv8, Conv9, Conv10 and Conv11. First, the SSD model generates target feature maps of different dimensions in cascaded convolution, which are 38×38, 19×19, 10×10, 5×5, 3×3 and 1×1. Then, multiple prior bounding boxes are set at each unit of the feature maps for localization and classification prediction. The prior bounding boxes are then decoded to obtain the prediction boxes, and a non maximum suppression algorithm is run to obtain the final detection results. Conv4_3 in SSD model and Conv7 belong to shallow feature layer, mainly used for detecting small targets. Conv8_ 2, Conv9_ 2, Conv10_ 2 and Conv11_ 2 belong to the deep feature layer and is mainly used for detecting medium to large targets.

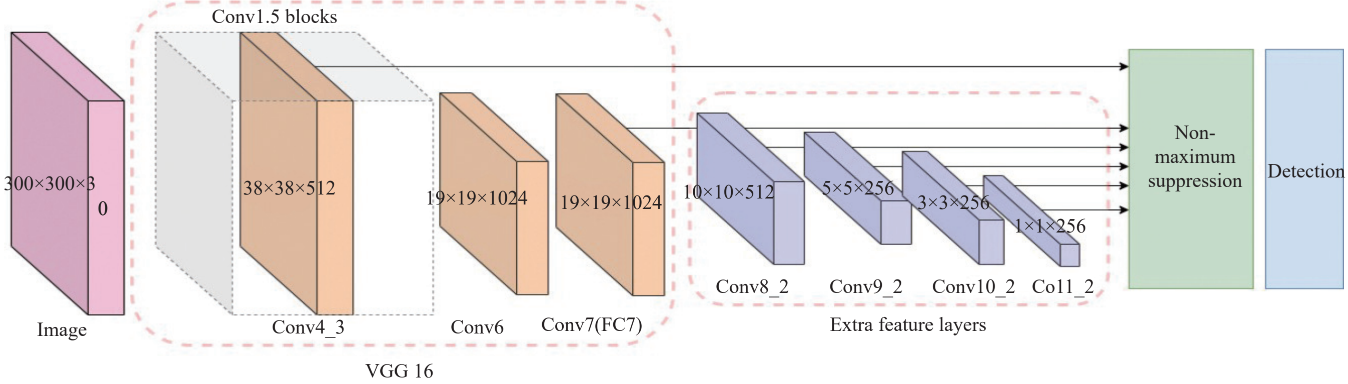


Figure 2 Framework diagram of SSD

The SSD uses a multi-scale prediction approach, with shallow networks detecting small targets and deep networks detecting large targets. Although shallow network contains rich geometric information and more accurate localization, it has small receptive field and weak semantic information representation ability. In contrast to shallow networks, deep networks have large receptive fields and rich semantic information, but their resolution is small and their ability to represent geometric information is weak. Therefore, SSD will have serious missing and false detection in target detection.

3.2 MobileNetV2 network

The MobileNetV2 network is updated based on the MobileNetV1 network, and it is a lightweight network [28-29]. Compared with MobileNetV1, MobileNetV2 improves network performance through reverse residual module and linear bottleneck layer.

The reverse residual module adopted by MobileNetV2 first realize dimension raising through 1×1 convolution kernel, and the activation function is ReLU6. Then extract the feature information in high-dimensional feature maps through 3×3 depthwise separable convolution; and finally realize dimension reduction by using 1×1 convolution kernel and reduce the number of channels, to ensure the contensistency of channel number at this time with that of input features. The structure of reverse residual module is shown in Figure 3.

In the dimension reduction convolution layer, MobileNetV2 replaces the nonlinear activation functions such as ReLU with a linear activation function to construct a linear bottleneck layer, which can avoid the loss of low dimensional feature information, and expand the network by setting coefficients and control the network size, where the step size is 1.

In MobileNetV2 network, deep separable convolution is introduced to replace ordinary convolution, and linear bottleneck and inverse residual structure are introduced to avoid information loss and improve accuracy, greatly reduce the number of model parameters and calculation amount, and thus improve the characterization ability of the network.

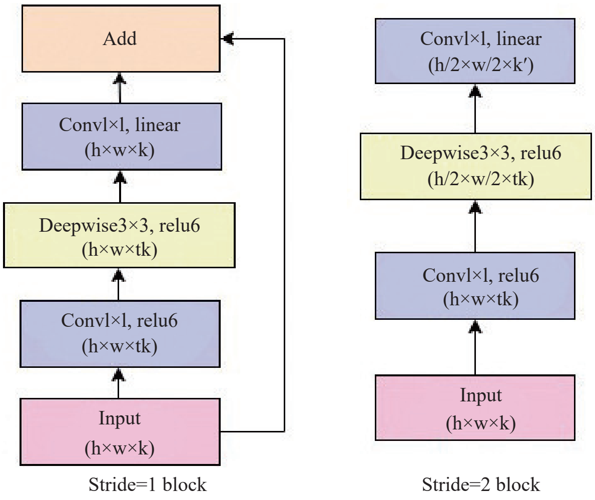


Figure 3 Structure of Inverted residual block

3.3 Attention mechanism

The invariance in translation of CNN lets the convolution kernel treat different regions and channels equally in extracting image features, resulting in failure of extracting useful features. In response to the above-mentioned shortcomings of convolutional neural networks, the introduction of attention mechanism can help disease detection models autonomously learn the weights of disease features, pay attention to important feature information while ignoring irrelevant information, reduce the complexity of detection tasks, and thus improve the detection efficiency of the models. In this paper, SE channel attention mechanism, CBAM spatial attention mechanism and ECA efficient channel attention mechanism are used. The complexity of SE attention mechanism model is low, and the new parameters and calculation amount are small. ECA attention mechanism is a lightweight channel attention module, which increases the complexity of the model less and improves the effect significantly. CBAM attention mechanism can improve network performance more effectively by connecting spatial domain and channel domain in series. Specifically:

3.3.1 SE channel attention mechanism

SE channel attention mechanism has an additional attention mechanism in the CNN direction [30], and it has the advantages of low complexity, small parameter quantity and little amount of calculation. SE includes two processes, compression and excitation, and tis network structure is shown in Figure 4.

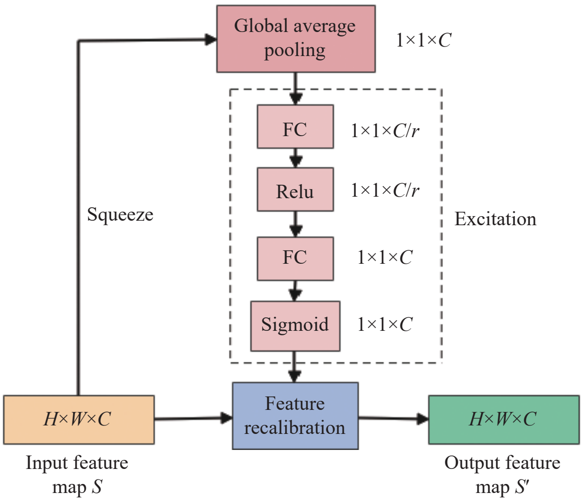


Figure 4 SE network structure diagram

The compression is made on feature maps based on spatial dimensions. During compression, global average pooling is used to compress the H×W×C feature maps into 1×1×C one-dimensional feature vector, which expands the receptive field of the feature maps, and then it turns to the activation process. The activation process includes two fully connected layers. The first one includes C/r neurons, and r is the dimension reduction scaling parameter, its input is 1×1×C and its output is 1×1×C/r. The second one includes C neurons, its input is 1×1×C/r and its output is 1×1×C. The using of two fully connected layers can better fit the complicated non-linear relationship of channels and reduce model complexity. After two fully connected layers, a one-dimensional vector is obtained through Sigmoid activation function, at last, the features are recalibrated and multiply the one-dimensional vector 1×1×C and the input feature map S according to channel weight, and then output the H×W×C feature map .

3.3.2 CBAM spatial attention mechanism

CBAM can enhance useful features in the input feature map while suppressing useless features, and is widely used in practical application[31]. The CBAM attention mechanism is a mixed domain attention mechanism, which is composed of the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). The network structure is shown in Figure 5.

Compared to the SE module, the CAM module has added a parallel maximum pooling layer. First, by averaging pooling and maximizing pooling, the H×W×C feature map is compressed into a 1×1×C one-dimensional feature vector. Next, the one-dimensional vector is sent into the multi-layer perception area MLP, which contains two fully connected layers. The first fully connected layer reduces the channel dimension from C to C/r, and the second connected layer increases the channel dimension from C/r to C. Then add the features according to the elements, and get a 1×1×C one-dimensional feature vector Mc through the Sigmoid activation function. Finally, Mc and the input feature map S are multiplied by elements to obtain H×W×C feature map, which is taken as input to the SAM module.

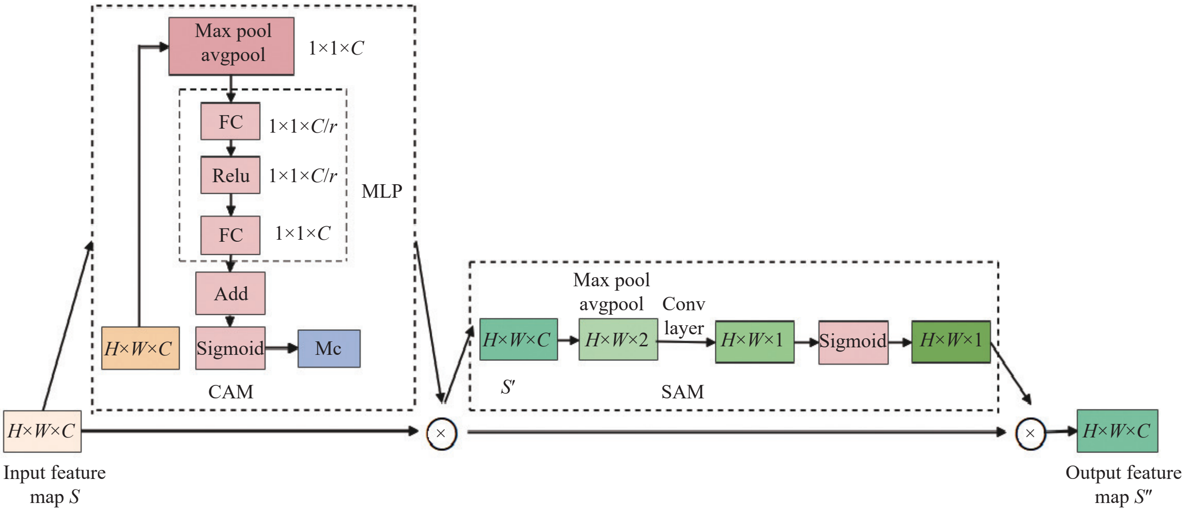


Figure 5 CBAM network structure diagram

The SAM module first applies the two feature maps H×W×1 obtained by maximum pooling and average pooling in the channel dimension, then splice the two feature maps into a H×W×2 feature map; secondly, utilize 3×3 convolutional layers to perform dimension reduction process on the spliced feature map and obtain a H×W×1 feature map. Then obtain an H×W×1 spatial attention feature map through the Sigmoid activation function. Finally, through multiplication of elements, based on the spatial attention feature map obtained from the previous step and the initially input feature map , the H×W×C output feature map can be obtained through calculation.

3.3.3 ECA efficient channel attention mechanism

ECA has some improvements based on SE; it is an extremely lightweight attention module[32]. ECA realized the local cross-channel interaction strategies without dimensionality reduction and the self-adaptive method of selecting the size of the one-dimensional convolutional kernel. It reduced the complexity of the module by improving the performance of the attention module. The network structure of the ECA is shown in Figure 6.

First, enter H×W×C feature map, compressed the feature map into 1×1×C one-dimensional feature vector by using global average pooling; next, the size of one-dimensional convolutional kernel K is adaptively selected, and the value of K is adaptively determined by the number of channels. Then cancel the latitude reduction operation to achieve local cross-channel connection, and use the K×K convolution layer to perform one-dimensional convolution operations on one-dimensional feature vectors. Finally, obtain H×W×C output feature map through Sigmoid activation function.

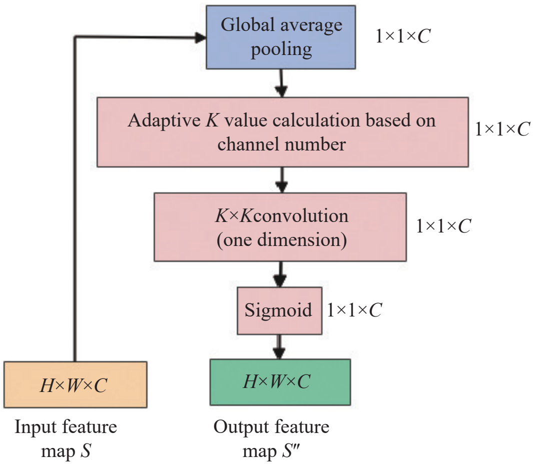


Figure 6 ECA network structure diagram

3.4 Improved SSD model based on different attention mechanisms

The SSD-based cotton leaf disease detection model has unsatisfactory detection performance in complex environments with irregular disease distribution and high timeliness requirements. Through analysis, the reasons are as follows:

(1) The shallow feature extraction network of SSD model lacks semantic information and seriously loses details in the process of feature extraction. The target feature information of cotton diseases in complex environment is less, so the semantic information and detail information of cotton diseases are particularly important for disease detection. Incomplete semantic information and detail information may affect the detection performance of the model.

(2) The feature layers used for detection in the SSD model have equal weights in both channel and spatial dimensions. Different channels and spatial dimensions represent different semantic information, so the feature layers with the same weight makes it difficult to distinguish the detection target and interference target of the disease.

In response to the above problems, in order to improve the detection performance of SSD detection models for cotton leaf diseases in complex environments, in this paper, the SSD model for cotton leaf disease detection was optimized by improving the backbone network and integrating attention mechanisms. The specific measures are as follows:

(1) Improving the backbone network

Due to the large parameter quantity of the VGG16 network, it takes much runtime in the feature extraction process. Nonlinear transformation in forward propagation will lead to the loss of key feature information. Therefore, enhancing the feature extraction capability of the backbone network is the key factor of improving the model detection precision and accelerating model detection speed. In this paper, MobileNetV2 lightweight network was used instead of VGG16 network for pruning processing. Through deeply separable convolution, the number of model parameters was significantly reduced to avoid the occurrence of gradient disappearance, weaken the dependency between parameters, and effectively alleviate the overfitting phenomenon. Through the shallow feature module, the receptive field of the feature map was expanded to provide rich semantic information and details of the cotton disease, so as to improve the feature extraction capability of the model and enhance the detection performance of small target diseases.

(2) Integration of different attention mechanisms

The introduction of SE attention mechanism into the SSD model is more concerned about the channel features with the largest amount of information by suppressing the unimportant channel features; the introduction of ECA into the SSD model achieves appropriate cross-channel interaction, significantly reducing the complexity of the model while maintaining good performance; the introduction of CBAM into the SSD model makes the disease detection model consider the importance of different pixels as well as the importance of pixels in different positions in the same channel. All the three of these attention mechanisms mentioned above can be seamlessly integrated into the SSD model, which helps the disease detection model learn the weight of disease features independently and realize end-to-end training. To solve this problem, 6 feature images with different sizes were extracted from the SSD model and input into the SE, ECA and CBAM attention modules to screen out disease object features, to enhance the feature images' representational ability in key feature information and improve the detection precision of SSD model on cotton disease objects.

The improved SSD model structure framework based on different attention mechanisms is shown in Figure 7, where the Attention Mechanism includes the SE attention module, ECA attention module, and CBAM attention module. After the cotton disease images were input, the lightweight network MobileNetV2 was used to extract the disease features. Then the multi-scale feature map is generated by the extended network. Then, 6 feature maps of different scales were input into different attention modules. Finally, the non-maximum suppression algorithm is run to get the final detection result.

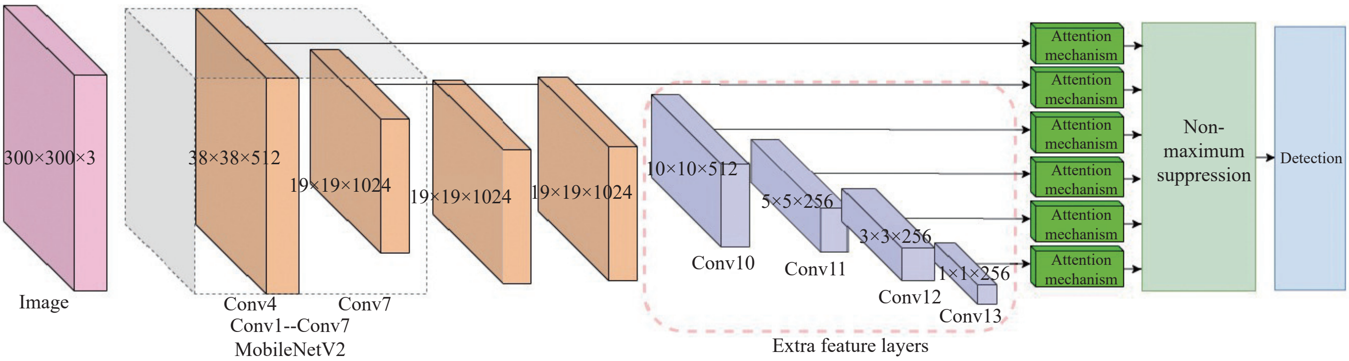


Figure 7 Framework diagram of improved SSD model

In summary, the detection of cotton leaf diseases in complex environments in this article is mainly based on improved SSD models with different attention mechanisms. First, in the data preprocessing stage, obtain the cotton disease dataset, complete the target detection labels of the disease, and divide the dataset proportionally. Second, in the model training stage, train the improved SSD model. When the model converges to a certain extent and the accuracy of the validation set no longer changes, the training is terminated to obtain the cotton disease detection model. Finally, in the cotton disease detection stage, a trained model is used to predict cotton diseases, and the performance of this model is evaluated based on the target detection results. The algorithm flow is shown in Figure 8.

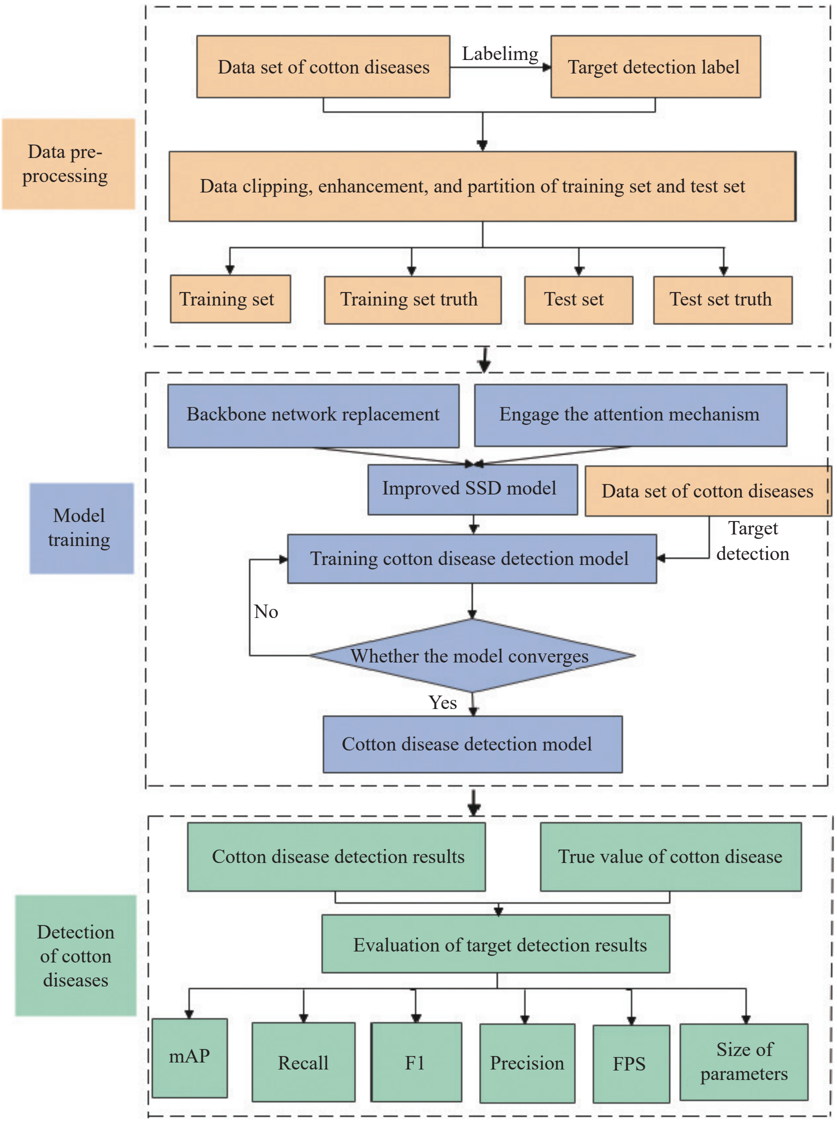

Data set of cotton diseases
Labelimg
Target detection label
Data pre-processing
Data clipping, enhancement, and partition of training set and test set
Training set
Training set truth
Test set
Test set truth
Backbone network replacement
Engage the attention mechanism
Improved SSD model
Data set of cotton diseases
Model training
Target detection
Training cotton disease detection model
No
Whether the model converges
Yes
Cotton disease detection model
Cotton disease detection results
True value of cotton disease
Detection of cotton diseases
Evaluation of target detection results
mAP
Recall
F1
Precision
FPS
Size of parameters

Figure 8 Algorithm flowchart

4 Test and result analysis

4.1 Evaluation indexes

In this study, P (Precision), R (Recall), comprehensive evaluation index $F1$ value, mAP (Mean Average Precision), FPS (Frames Per Second) and number of parameters were adopted to evaluate the detection results.

The Precision P represents the proportion of correct predictions being positive and all predictions being positive; The Recall R represents the proportion of correctly predicted positive and all positive samples. The calculation equations of P, R are Equations (1) and (2), respectively.

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}\times 100\% \tag{1}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}\times 100\% \tag{2}$$

where, TP represents the number of disease samples predicted to be positive in the disease dataset and actually positive; FP represents the number of disease samples predicted to be positive but actually negative; FN represents the number of disease samples predicted to be negative but actually positive.

$F1$ value is the harmonic mean of Precision P and Recall R. The calculation equation of $F1$ value is shown in Equation (3):

$$F1=\frac{2\times P\times R}{P+R} \tag{3}$$

mAP is the results of averaging the average precision AP of all diseases, it can measure a model's performance on all kinds of diseases. The definition of average precision AP is shown in Equation (4), and the definition of mAP is shown in Equation (5).

$$AP=\int_0^1 PR\mathrm{d}R \tag{4}$$

$$mAP=\frac{1}{N}\sum_{m=1}^{N}AP_m \tag{5}$$

where, N is the number of kinds of diseases; AP_m is the average precision of the m-th kind of disease.

FPS represents the number of images processed per second. The higher *FPS* is, the faster the detection speed of the algorithm.

4.2 Experiment platform and parameter setting

In this study, Windows 10 operating system was adopted. The computer is equipped with 16GB of memory, using Pytorch 1.10.1 as a deep learning framework, and the hardware configuration and model parameters related to the experiment are listed in Table 2.

Table 2 Test related hardware configuration and model parameters

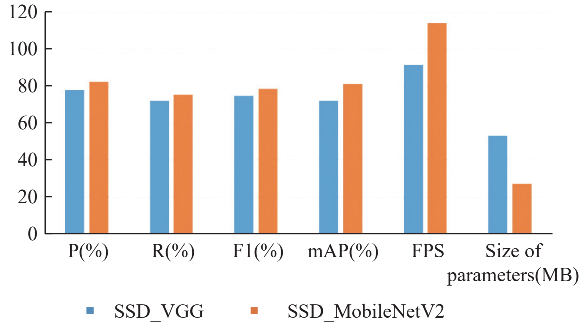
Name	Configuration	Name	Taking values
GPU	RTX3070Ti	Size of images	640×640
CPU	AMD Ryzen 7 5800X @ 3.8GHz	Learning rate	0.001
CUDA	11.3	Optimizer	Adam
CuDNN	8.2.1	Batch size	16

Note: CUDA: Compute Unified Device Architecture; CuDNN: NVIDIA CUDA® Deep Neural Network

4.3 Experiment results and analysis

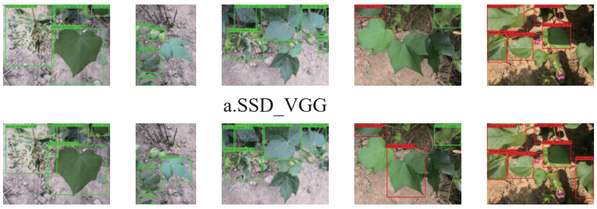
4.3.1 Detection effect of the SSD model using different backbone networks

To explore the impact of the SSD model based on different backbone networks on detection of cotton leaf diseases, *P*, *R*, *F*1 values, *mAP*, *FPS*, and number of parameters were compared under the same experimental conditions. The experimental results are shown in Figure 9, and the detection effect is shown in Figure 10.



Note: The SSD model with VGG16 as the backbone network is abbreviated as SSD_ VGG, the SSD model with MobileNetV2 as the backbone network is abbreviated as SSD_ MobileNetV2. The same below.

Figure 9 Detection results of cotton diseases based on different backbones



a.SSD_VGG

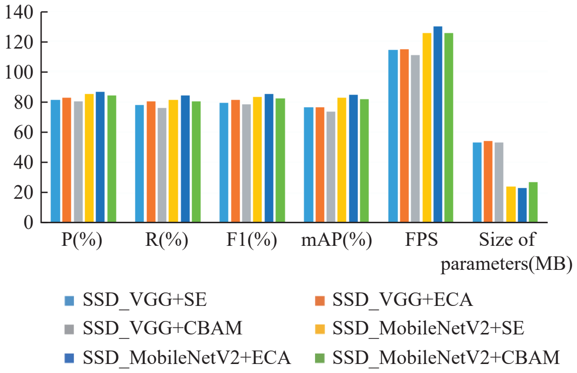
b.SSD_MobileNetV2

Figure 10 Detection effect of cotton diseases based on different backbones

As shown in Figure 9, the *P*, *R*, *F*1 values, *mAP*, and *FPS* of the SSD_MobileNetV2 are all higher than those of the SSD_ VGG model, and the parameter quantity of SSD_MobileNetV2 was significantly reduced. As shown in Figure 10, the detection performance of the SSD_MobileNetV2 model is significantly better than that of SSD_VGG model. The confidence level for most prediction boxes in the detection effect map is higher, and there is a significant improvement in leak detections. Therefore, MobileNetV2 was used to replace the backbone network VGG16 and significantly reduced the parameter quantity of the model, retained more location and marginal information and improved the detection precision rate.

4.3.2 SSD model detection effect by adopting different attention mechanisms

To explore the impact of the SSD model based on different attention mechanisms on the detection of cotton leaf diseases, *P*, *R*, *F*1 values, *mAP*, *FPS*, and number of parameters were compared under the same experimental conditions. The experimental results are shown in Figure 11, and the detection effect is shown in Figure 12.



Note: The SSD_VGG model after introducing the SE attention mechanism is called SSD_VGG+SE for short; the SSD_VGG after introducing the ECA is called SSD_VGG+ECA for short; the SSD_VGG model after introducing CBAM is called SSD_VGG+CBAM for short; the SSD_MobileNetV2 model after introducing SE is called SSD_MobileNetV2+SE for short; the SSD_MobileNetV2 model after introducing ECA is called SSD_MobileNetV2 +ECA for short; the SSD_MobileNetV2 model after introducing CBAM is called SSD_MobileNetV2+CBAM for short. The same below.

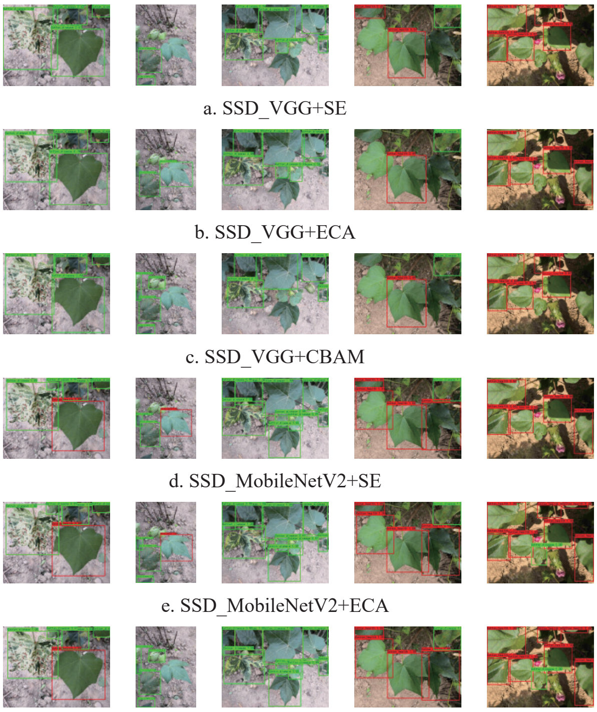
Figure 11 Detection results of cotton diseases based on different attention mechanism

It can be obtained from Figure 11 that:

(1) After introducing SE, ECA and CBAM into the SSD_VGG model, SSD_VGG+ECA showed optimal detection effect, followed by SSD_VGG+SE model and finally the SSD_VGG+CBAM model. The *P*, *R*, *F*1 values, *mAP* and *FPS* of SSD_VGG+ECA were all higher than that of SSD_VGG+SE model and SSD_VGG+CBAM model, while the parameter quantity of the three models were basically equal.

(2) After introducing SE, ECA and CBAM into the SSD_MobileNetV2 model, the SSD_MobileNetV2+ECA model showed the best detection performance, followed by SSD_ MobileNetV2+SE model, and finally the SSD_MobileNetV2 + CBAM model. The *P, R, F*1 values, *mAP*, and *FPS* of the SSD_MobileNetV2+ECA were all higher than those of the SSD_MobileNetV2+SE model, and the SSD_MobileNetV2+ CBAM, and the parameter quantity of the three models was basically equal.

(3) After introducing SE, ECA and CBAM into the SSD_MobileNetV2 model, the detection performance was better than the SSD_VGG after introducing the three attention mechanisms above, and the parameter quantity of the SSD_ MobileNetV2+SE model, the SSD_MobileNetV2+ECA model and the SSD_Mobile NetV2+CBAM model were less than half that of the SSD_VGG+SE model, the SSD_VGG+ECA model and the SSD_VGG+CBAM model.



a. SSD_VGG+SE

b. SSD_VGG+ECA

c. SSD_VGG+CBAM

d. SSD_MobileNetV2+SE

e. SSD_MobileNetV2+ECA

f. SSD_MobileNetV2+CBAM

Figure 12 Detection effect of cotton diseases based on different attention mechanism

It can be obtained from Figure 12 that:

(1) Higher detection precision rate was obtained by introducing ECA into the SSD_VGG model and the SSD_MobileNetV2 model, thus better detection performance was obtained. It shows that, the ECA module utilized the global information of cotton disease images, and selectively enhanced the channel weight of the cotton disease by effectively suppressing the unimportant feature information in the environment, thus significantly improved the precision of disease detection boundary.

(2) Some circumstances of leak detection occurred after introducing SE into the SSD_VGG model and the SSD_ MobileNetV2 model, showing that the SE module only selected some interactive coverage, and ignored some channel information of the cotton disease.

(3) Some circumstances of leak detection and false detection occurred after introducing CBAM into the SSD_VGG model and the SSD_MobileNetV2 model, showing that, although the CBAM was introduced into the spatial attention mechanism, its discrimination of cotton diseases and the spatial and location information of the environment was poorer than that of channel attention mechanisms.

4.3.3 Detection performance by using different target detection models

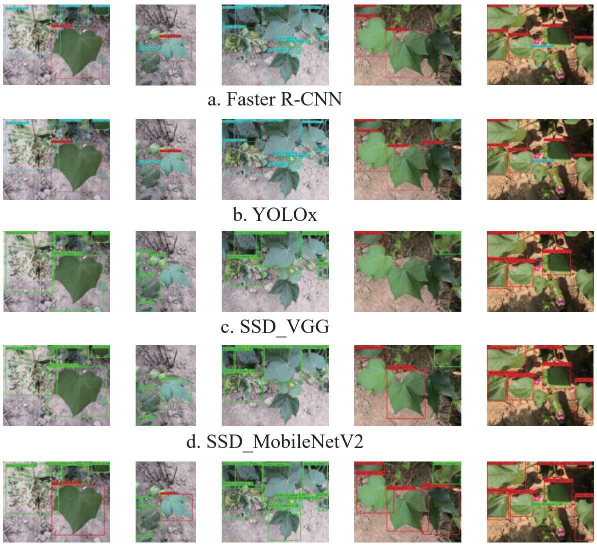
To explore the impact of the different detection models on the detection of cotton leaf diseases, *P*, *R*, *F*1 values, *mAP*, *FPS*, and number of parameters were compared under the same experimental conditions. The experimental results are listed in Table 3, and the detection performance is shown in Figure 13.

It can be obtained from Table 3 that:

(1) *P*, *R*, *F*1 value and *mAP* of Faster R-CNN model were 6.64, 9.65, 8.22 and 4.42 percentage points lower than SSD_ MobileNetV2+ECA model, respectively, and *FPS* value was 12.22% of that of SSD_MobileNetV2+ECA model. The number of parameters in this model is the highest among all models, which is 24.58 times that of SSD_MobileNetV2+ECA model. The Faster R-CNN model uses the feature information extracted from the last layer of the backbone network as the input part of the region generation network and feature prediction head during the detection process, resulting in the loss of a large amount of cotton disease information and unsatisfactory detection results.

Table 3 Detection results of cotton diseases based on different target detection model

Model	*P*/%	*R*/%	*F*1/%	*mAP*/%	*FPS*	Size of Parameter/MB
Faster R-CNN	79.80	74.42	77.02	80.38	15.87	569.37
YOLOx	81.01	65.72	72.57	74.79	76.91	34.38
SSD_VGG	78.05	72.26	75.04	72.43	91.68	53.16
SSD_MobileNetV2	82.42	75.56	78.84	81.22	114.18	27.06
SSD_MobileNetV2+ECA	**86.44**	**84.07**	**85.24**	**84.80**	**129.92**	**23.16**



a. Faster R-CNN

b. YOLOx

c. SSD_VGG

d. SSD_MobileNetV2

e. SSD_MobileNetV2+ECA

Figure 13 Detection effect of cotton diseases based on different target detection models

(2) Using YOLOx model, the *P*, *R*, *F*1 values and *mAP* were 5.43%, 18.35%, 12.67%, and 10.01% lower than the SSD_ MobileNetV2+ECA model, the *FPS* value was 53.01 frames/s, which is lower than that of the SSD_MobileNetV2+ECA model, model parameter quantity was 11.22 MB higher than that of the SSD_ MobileNetV2+ECA model. It shows that the YOLOx model did not fully utilize the features of cotton diseases.

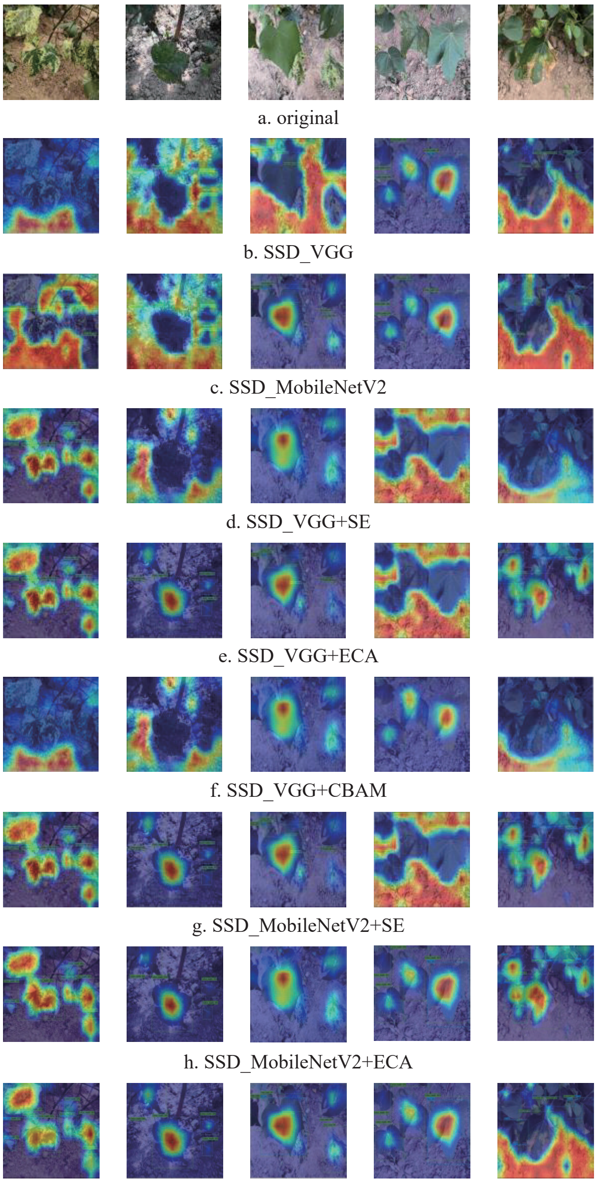
(3) Using the SSD_VGG model, *P*, *R*, *F*1 values and *mAP* were 8.39%, 11.81%, 10.2%, and 12.37% lower than that of the SSD_ MobileNetV2+ECA model, respectively, and the *FPS* value was 38.24 frames/s lower than the SSD_MobileNetV2+ECA model, the model parameter quantity was 30 MB higher than that of the SSD_ MobileNetV2+ECA model. It shows that the anchor frame size used in the SSD_VGG model was not suitable for cotton diseases, since its precision in disease feature extraction was not high.

(4) In using the SSD_MobileNetV2 model, the *P*, *R*, *F*1 values and *mAP* were 4.02%, 8.51%, 6.4%, and 3.58% lower than that of the SSD_MobileNetV2+ECA model, the *FPS* value was 15.74 frames/s lower than that of the SSD_MobileNetV2+ECA model, model parameter quantity was 3.9 MB higher than that of the SSD_ MobileNetV2+ECA model. It shows that the SSD_ MobileNetV2 chose larger feature dimensions in prediction, thus it is not suitable for detection cotton diseases.

As can be seen from Figure 13, when using the Faster R-CNN model, YOLOx model, SSD_VGG model and SSD_MobileNetV2 model to detect cotton diseases, there were certain missing and false detection in all images. SSD_MobileNetV2+ECA model correctly detected all cotton diseases and had the best detection effect.

4.3.4 Model visualization

In a convolutional neural network, the features of the fully connected layer are difficult to understand, but the last layer of convolutional units contains the most comprehensive semantic information, and each channel can detect different activation regions of the target. Therefore, the characteristic information of the last layer of convolutional unit is fully used to explain the network model, and the internal characteristics of the neural network are understood by visualization technology to realize the interpretation of the model decision. In order to better reflect the advantages of attention mechanisms, the gradient weighted class activation map Grad-CAM was used as a visualization tool[33]. Grad-CAM obtains the weight values of each channel by calculating the gradient information of the last convolutional layer, and maps the weighted feature map to the original image in the form of a heat map. The pixel values in the heat map represent the importance of the pixel area to the model's detection results, and the redder the color, the more attention the model pays to the area. The visualization results of different models are shown in Figure 14.


a. original

b. SSD_VGG

c. SSD_MobileNetV2

d. SSD_VGG+SE

e. SSD_VGG+ECA

f. SSD_VGG+CBAM

g. SSD_MobileNetV2+SE

h. SSD_MobileNetV2+ECA

i. SSD_MobileNetV2+CBAM

Figure 14 Grad-CAM diagram of different models

As can be seen from Figure 14, compared with the original figure, the color parts of the SSD_VGG model and SSD_MobileNetV2 model are distributed around the leaves, and after the introduction of the attention module, the color parts of the model are mostly concentrated on the leaves, indicating that the model focuses on the cotton leaves. Among them, the color part of SSD_MobileNetV2+ECA model concentrated on the disease parts of the leaves, and paid the most attention to the disease spots. Therefore, the addition of attention module makes the model strengthen the ability to extract important disease features, restrain the extraction of background features, and extract key disease features significantly stronger than the unimproved model, and improves the classification accuracy. In summary, the improved model can well solve the problems of low accuracy and low generalization of cotton leaf diseases.

5 Conclusions

An improved SSD model for detecting cotton leaf diseases in complex backgrounds was proposed in this paper. The improved SSD model has good timeliness and robustness, and could achieve more comprehensive and precise cotton disease detection in complex field environments. It can inspire ideas for the detection of cotton leaf diseases in practical applications. The main conclusions are as follows:

(1) Adopting the lightweight network MobileNetV2 to replace the original backbone network VGG16 of SSD effectively reduced the parameter and computational quantity of the model. The parameter quantity of the SSD_MobileNetV2 was just 50.9% that of the SSD_VGG. At the same time, the SSD_MobileNetV2 model effectively integrated the rich detail information in the shallow layer with the rich semantic information in the deep layer, significantly improving the extraction capability of cotton disease features, and accelerating the running speed of the algorithm. Compared to SSD_VGG, the *P, R, F*1 values, and *mAP* of SSD_MobileNetV2 model increased by 4.37%, 3.3%, 3.8%, and 8.79%, respectively, while *FPS* increased by 22.5 frames/s.

(2) After introducing SE, ECA and CBAM into the SSD_VGG model, the SSD_VGG+ECA model presented the optimal detection performance, followed by the SSD_VGG+SE model, and finally the SSD_VGG+CBAM model. After introducing SE, ECA and CBAM into the SSD_MobileNetV2 model, the SSD_MobileNetV2 +ECA model presented the optimal detection performance, followed by the SSD_MobileNetV2+SE model and finally the SSD_MobileNetV2+CBA model. Visualize the focus areas of each model using Grad-CAM. After introducing the three attention mechanisms into the SSD_MobileNetV2 model, the detection performance of the model was better than that of the SSD_VGG model after also introducing the three attention mechanisms. Therefore, the introduction of the ECA attention mechanism improved the feature utilization rate of cotton diseases, enhanced the equalization of integration of disease features, and effectively promoted the detection performance of cotton leaf diseases. The SSD_MobileNetV2+ECA model proved optimal detection performance, and its *P*, *R*, *F*1 values, *mAP* and *FPS* were all higher than that of other models that were introduced with attention mechanisms, moreover, for its smaller parameter quantity and higher running speed, it is more suitable for cotton disease detection in complex environments.

(3) The *P*, *R*, *F*1 values, *mAP*, and *FPS* of the SSD_MobileNetV2+ECA model were higher than that of the other four commonly used object detection models, and the model has significant advantages in parameter quantity to achieve better

detection results.

Acknowledgements

The authors acknowledge that this work was financially supported by the National Natural Science Foundation of China (Grant No. 32160421), the Industrialization Support Project from the Education Department of Gansu Province (Grant No. 2021CYZC-57), Youth Science and Technology Foundation of Gansu Province (Grant No. 21JR7RA572), Gansu Education Department Innovation Fund project (Grant No. 2022B-144).

[References]

[1] Guo W J, Feng Q, Li X Z, Yang S, Yang J Q. Grape leaf disease detection based on attention mechanisms. Int J Agric & Biol Eng, 2022; 15(5): 205–212.

[2] Wang J W, Zhao L H, Shi Y Q, Wei F, Feng H J, Zhu H Q, et al. Preliminary report on the whole process control techniques of cotton diseases. China Cotton, 2020; 47(5): 20–22, 46. (in Chinese)

[3] Zhang J H, Ji R H, Yuan X, Li H, Qi L J. Recognition of pest damage for cotton leaf based on RBF-SVM algorithm. Transactions of the CSAM, 2011; 42(8): 178–183. (in Chinese)

[4] Zhang J H, Qi L J, Ji R H, Wang H, Huang S K, Wang P. Cotton diseases identification based on rough sets and BP neural network. Transactions of the CSAE, 2012; 28(7): 161–167. (in Chinese)

[5] Zhang J H, Han S Q, Zhai Z F, Kong F T, Feng X, Wu J Z. Improved adaptive watershed method for segmentation of cotton leaf adhesion lesions. Transactions of the CSAE, 2018; 34(24): 165–174. (in Chinese)

[6] Zhai Z F, Xu Z, Zhou X Q, Wang L L, Zhang J H. Recognition of hazard grade for cotton blind stinkbug based on Naive Bayesian classifier. Transactions of the CSAE, 2015; 31(1): 204–211. (in Chinese)

[7] Nazki H, Yoon S, Fuentes A, Dong S P. Unsupervised image translation using adversarial networks for improved plant disease recognition. Computers and Electronics in Agriculture, 2020; 168: 105–117.

[8] Liu Y, Gao G Q. Identification of multiple leaf diseases using improved SqueezeNet model. Transactions of the CSAE, 2021; 37(2): 187–195. (in Chinese)

[9] Wang C S, Zhao C J, Wu H R, Zhou J, Li J X, Zhu H J. Recognizing crop diseases using bimodal joint representation learning. Transactions of the CSAE, 2021; 37(11): 180–188. (in Chinese)

[10] Li S Q, Chen C, Zhu T, Liu B. Plant leaf disease identification based on lightweight residual network. Transactions of the CSAM, 2022; 53(3): 243–250. (in Chinese)

[11] Zhang J H, Kong F T, Wu J Z, Zhai Z F, Han S Q, Cao S S. Cotton disease recognition model based on improved VGG convolutional neural network. Journal of China Agricultural University, 2018; 23(11): 167–177. (in Chinese)

[12] Wang X F, Zhang C L, Zhang S W, Zhu Y H. Forecasting of cotton diseases and pests based on adaptive discriminant deep belief network. Transactions of the CSAE, 2018; 34(14): 157–164. (in Chinese)

[13] Zhao L X, Hou F D, Lyu Z C, Zhu H C, Ding X L. Image recognition of cotton leaf diseases and pests based on transfer learning. Transactions of the CSAE, 2020; 36(7): 184–191. (in Chinese)

[14] Chen K Q, Zhu Z L, Deng X M, Ma C X, Wang H A. Deep learning for multi-scale object detection: A survey. Journal of Software, 2021; 32(4): 1201–1227. (in Chinese)

[15] Zhou H R, Wu B M. Advances in research on deep learning for crop disease image recognition. Journal of Agricultural Science and Technology, 2021; 23(5): 61–68. (in Chinese)

[16] Fuentes A, Yoon S, Kim S C, Park D S. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors 2017; 17(9): 2022. doi: 10.3390/s17092022.

[17] Li J H, Lin L J, Tian K, Alaa A A. Detection of leaf diseases of balsam pear in the field based on improved Faster R-CNN. Transactions of the CSAE, 2020; 36(12): 179–185. (in Chinese)

[18] Ye Z H, Zhao M X, Jia L. Image recognition of crop diseases in complex background. Transactions of the CSAM, 2021; 52(S0): 118–124,147. (in Chinese)

[19] Liu J, Wang X. Early recognition of tomato gray leaf spot disease based on MobileNetv2-YOLOv3 model. Plant Methods, 2020; 16: 83.

[20] Wen B, Cao R X, Yang Q L, Zhang J, Zhu H, Li Z C. Detecting leaf disease for Panax notoginseng using an improved YOLOv3 algorithm. Transactions of the CSAE, 2022; 38(3): 164–172. (in Chinese)

[21] Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017; 39(6): 1137–1149.

[22] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. IEEE Conference on Computer Vision and Pattern Recognition, 2016; pp.779–788.

[23] Redmon J, Farhadi A. YOLOv3: An incremental improvement. (2018-04-08).https://arxiv.org/pdf/1804. 02767. pdf.

[24] Bochkovskiy A, Wang C Y, Liao H. YOLOv4: Optimal speed and accuracy of object detection. (2020-04-23). https://arxiv.or/abs/2004. 10934.

[25] Zheng G, Liu S T, Wang F, Li Z M, Sun J. YOLOx: Exceeding YOLO series in 2021. arXiv preprint, 2107; 08430.

[26] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y, et al. SSD: Single shot multibox detector. European Conference on Computer Vision, Springer, Cham, 2016; pp.21–37.

[27] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint, 2015; pp.1409–1556.

[28] Sandler M, Howard A, Zhu M L, Zhmoginov A, Chen L. MobileNetV2: inverted residuals and linear bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018; pp.4510-4520.

[29] Dai F, Li X Z, Shi R J, Zhang F W, Zhao W Y, Guo W J. Film identification method based on improved deeplabv3+ for full-film double-ditch corn seedbed. Int J Agric & Biol Eng, 2023; 16(5): 165–172.

[30] Hu J, Li S, Sun G. Squeeze-and-excitation networks. The IEEE conference on computer vision and pattern recognition. Piscataway, New York, USA: IEEE, 2018; pp.2011–2023.

[31] Woo S, Park J, Lee J Y, Kweon I S. CBAM: Convolutional block attention module. European Conference on Computer Vision, Munish: Springer, 2018; pp.3–19.

[32] Wang Q L, Wu B G, Zhu P F, Li P H, Zuo W M, Hu Q H. ECA-Net: Efficient channel attention for deep convolutional neural networks. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Piscataway, NJ: IEEE Press, 2020.

[33] Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. 2017 IEEE International Conference on Computer Vision (ICCV), Piscataway, NJ, USA, IEEE, 2017: pp.618–626.