DBA_SSD A Novel End-to-End Object Detection Using Deep Attention Module for Helping Smart Device with Vegetable and Fruit Leaf Plant Disease Detection

Jun Wang
Guizhou University

Jing Yang (✉ jyang23@gzu.edu.cn)
Guizhou University   https://orcid.org/0000-0003-1915-9487

Liya Yu
Guizhou University

Hao Dong
Guizhou University

Kun Yun
Guizhou University

Yifan Wang
Guizhou University

Research

DBA_SSD：A Novel End-to-End Object Detection Using Deep Attention Module for Helping Smart Device with Vegetable and Fruit Leaf Plant Disease Detection

Jun Wang[1], Jing Yang[1,2,3,*], Liya Yu[1], Hao Dong[1], Yifan Wang[3]

[1]*School of Mechanical Engineering, Guizhou University, Guiyang 550025, China;*

[2]*Guizhou Provincial Key Laboratory of Public Big Data ,Guizhou University, Guiyang,550025,China;*

[3]*Key Laboratory of Advanced Manufacturing Technology of Ministry of Education, Guizhou University, Guiyang, 550025,China.*

Abstract: In response to the difficulty of detecting and classifying pests and vegetable and fruit leaves with pests and diseases, this study proposes a novel vegetable and fruit leaf pest detection method called deep block attention SSD (DBA_SSD) for the identification of pests and diseases and classification of the degree of pests and diseases of vegetable and fruit leaves. We propose three vegetable and fruit leaf pest detection methods, namely, squeeze-and excitation SSD (Se_SSD), DB_SSD, and DBA_SSD. Se_SSD fuses SSD feature extraction network and attention mechanism channel, DB_SSD improves VGG feature extraction network, and DBA_SSD fuses the improved VGG network and channel attention mechanism. To reduce the training time and accelerate the training process, the convolutional layers trained in the Image Net image dataset by the VGG model are migrated to this model, whereas the collected vegetable and fruit disease image dataset is randomly divided into training set, validation set, and test set in the ratio of 8:1:1. In addition, data enhancement methods, such as histogram equalization and horizontal flip were used to expand the image data. The performance of the three improved algorithms is compared and analyzed in the same environment and with the classical target detection algorithms YOLOv4, YOLOv3, Faster RCNN, and YOLOv4 tiny. Experiments show that DBA_SSD outperforms the two other improved algorithms, and its performance in comparative analysis is superior to other target detection algorithms.

Keywords: disease detection; degree classification of disease; data enhancement; target recognition; SSD

0 Introduction

Fruits and vegetables are susceptible to various diseases, thereby affecting their quality and yield seriously. The formulation of prevention and control plans as soon as possible before the outbreak of the disease can maximize the effect of prevention and control and reduce economic losses. Therefore, the identification of vegetable and fruit diseases and insect pests is an effective way to inhibit the rapid development of diseases and avoid their occurrence. Previously, people used to experience subjective judgments by category crop diseases. However, the ability of this process to distinguish various diseases is limited, and it is time consuming.

Agricultural detection based on artificial intelligence, such as crop yield prediction[1], weed identification processing[2], and pest and disease detection[3, 4]，is widely used with the development of artificial intelligence technology. Machine learning-based disease detection requires preprocessing the dataset, extracting the features of disease regions in the image using feature extraction algorithms, sending the obtained feature information to the classifier to obtain the model parameters, and obtaining the disease and pest categories and the degree of disease and pest of the object to be detected. However, the model generalization ability is weak because of the machine learning-based image recognition., When the number of categories is excessive, the features of each class cannot be distinguished effectively. Moreover, the categories can only be recognized in a specific image context. Thus, the needs of large-scale planting, based on which it is important to

45 research a fast end-to-end vegetable and fruit leaf pest detection method, cannot be met.

46 In recent years, deep learning algorithms based on convolutional neural networks (CNN)[5]
47 have been widely used, and CNN have a large number of adjustable parameters[6]. It is very effective
48 for real-life object detection, recognition and classification[7]. In the field of agricultural research,
49 deep learning technology in agriculture includes crop/weed recognition, fruit harvesting[8], and plant
50 recognition[9]. Similarly, recent studies have also focused on the identification of plant diseases.
51 Some of the latest network models have been applied to the classification of plant diseases. In
52 addition, some researchers have introduced deep learning algorithms to improve the performance
53 of plant species disease classification. Longsheng Fu[10] proposed an orchard kiwi fruit target
54 detection algorithm. According to the characteristics of kiwi fruit images, the 3*3 and 1*1
55 convolutions were introduced into the YOLOv3-tiny[11] model, DY3TNet model was proposed and
56 combined with R- CNN[12], YOLOv2[13] and YOLOv3-tiny are compared.The experimental results
57 show that the improved DY3TNet model is small in size and high in efficiency. Guoxu Liu et al.[14]
58 detected tomatoes based on the YOLOv3 model [15], combined dense structure for feature extraction,
59 replaced   traditional R-Bbox with C-Bbox, matched the shape of the tomato, reduced the number
60 of parameters, and compared YOLOv2 , and Faster RCNN[16]. Literature[17] proposed a tomato gray
61 spot recognition method based on Mobilenev2[18] and YOLOv3 lightweight network model. This
62 method improves the accuracy of tomato gray spot recognition by introducing GIOU regression loss
63 function, and uses a pre-training method that combines hybrid training and migration learning to
64 improve the generalization ability of the model. Literature[19] compared the performance of five
65 networks, namely, AlexNet[20], VGG-16[21], ResNet-101[22], DenseNet-161[23], and SqueezeNet[24]
66 for nutrient deficiency symptom identification based on the Deep Nutrient Deficiency for Sugar
67 Beet dataset and discussed their limitations.

68 Building a fast and high classification accuracy model is necessary to determine the detection
69 quality of plant and fruit leaf diseases and insect pests. The current mainstream target recognition
70 networks include YOLO series, Faster RCNN, SSD[25], and FPN[26]. The SSD target detection
71 network uses an end-to-end method to regress features and extracts different levels of image features,
72 which cover low-level and high-level semantic information. Previous studies have shown that the
73 SSD network is fast. However, the direct application of SSD methods to detect vegetable and fruit
74 pests and diseases cannot meet the high precision requirements in agricultural production. This
75 paper proposes a fusion residual network and 1*1 convolution feature extraction module. It
76 strengthens the feature extraction capability of SSD and improves the positioning and recognition
77 accuracy of SSD for detecting vegetable and fruit pests. We also use data-enhancement to perform
78 spatial transformation and pixel transformation on images, thereby not only improving the
79 abundance of algorithm features, detection accuracy, and detection efficiency but also reduces the
80 labor costs for agricultural fruit and vegetable pest detection.

81 This study focuses on proposing a novel end-to-end pest detection algorithm called Deep Block
82 Attention SSD (DBA_SSD) for vegetable and fruit leaves. Our main work and contributions are
83 presented as follows:

84 1.   We proposed a novel end-to-end detection algorithm for plant and fruit leaf diseases and
85 insect pests, DBA_SSD, by combining the attention mechanism and convolution kernel, which
86 combines the attributes of the plant and fruit leaf disease and insect pest pictures and pay more
87 attention to pest and disease details when testing fruit and vegetable leaves for pest and disease.

88 2.   We graded the health of the fruit and vegetable leaves. According to the research results

89   of the paper, different measures can be taken according to the severity of the diseases of the fruit
90   and vegetable leaves. Increasing the yield of fruits and vegetables is of great significance.

91   3.   We implemented the classic SSD, YOLOv4, YOLOv3, Faster RCNN, and YOLOv4 tiny
92   models and compared them with our proposed DBA_SSD. Our method is better than the classic
93   baseline method on the vegetable and fruit leaf data set.

94   The main structure of this article is presented as follows. The first chapter mainly introduces
95   the related work on the detection of leaf diseases and insect pests of fruits and vegetables and combs
96   the detection technology of leaf diseases and insect pests of fruits and vegetables. The second
97   chapter introduces the SSD model and related improvement modules and proposes two improved
98   methods for the SSD target detection algorithm. The third chapter introduces the environment of
99   algorithm experiment, data set structure, experiment procedure, and experiment evaluation standard.
100 The fourth chapter conducts a comparative analysis of the results of the two sets of experiments and
101 related ablation experiments on the proposed DBA_SSD. The other is a comparative analysis of the
102 results of SSD improved algorithms and other target detection algorithms. Finally, we summarize
103 and prospect the research in this article.

104 **1 Related work**

105 At present, the research methods on plant disease recognition mainly focuses on two aspects:
106 one is disease recognition based on machine learning, and the general steps include diseased leaf
107 image segmentation, feature extraction, and disease recognition; and the other is target recognition
108 technology based on deep learning, wherein terminal end-to-end target detection is favored by many
109 researchers because of its fast recognition speed and efficient feature extraction methods.

110 In the research on the identification of vegetable and fruit diseases based on machine learning,
111 Literature[27] proposed a DCNN-based apple tree leaf disease (ATLD) diagnosis method, and
112 established 5 common ATLDs and healthy leaf data sets. The DCNN model combines DenseNet
113 and Xception[28] models by using support vector machine to classify apple leaf diseases, the
114 experimental results show that the accuracy of the DCNN model better than and comparing
115 Inception-v3[29], MobileNet[30], VGG-16, DenseNet-201, Xception, VGG-INCEP. Shrivastava et
116 al.[31] proposed a rice disease image classification by only method using color features, and explored
117 the feature extraction methods of 14 different color channels. They obtained 172 different color
118 channel feature information and used 7 different classifiers. The performance is compared, and the
119 result shows that the classification accuracy of the support vector machine classifier is up to 94.65%.
120 Literature[32] introduced a hybrid method for detecting plant leaf diseases and insect pests. The first
121 stage corresponds to the image enhancement and image conversion scheme to overcome the
122 problems related to low illumination and noise. The second stage combines the feature extraction
123 technology of GLCM, complex Gabor filter, Curvelet, and image moments. The third stage uses the
124 extracted features to train the nerve fuzzy logic classifier, and the proposed combination of feature
125 extraction and image preprocessing can improve classification accuracy. Abdulridha[33] used
126 hyperspectral imaging and machine learning to develop a technique for detecting pumpkin powdery
127 mildew in the asymptomatic, early, middle, and late stages. This method uses a radial basis function
128 to treat the disease. Strains and healthy strains were distinguished, and the severity of diseased
129 strains was classified. Abdu[34]

130 proposed a method for identifying the surface of plant diseased leaves, extracting optimized
131 features from the diseased area, and identifying plant diseased leaves based on a feature-based

machine learning classifier. The diseased features are connected in series to form a pathological feature vector for disease recognition to improve detection accuracy.

In deep learning-based research on fruit and vegetable diseases, Salma Samiei[35] u used red clover and alfalfa as research objects and proposed CNN-LSTM models combined with denoising algorithms to classify the different growth stages of two different plant species. Based on high-resolution remote sensing data, Alin-Ionut, Ples, oianu et al.[36] and others proposed an integrated deep learning model for individual tree crown detection and species classification. Mohamed Kerkech et al.[37] proposed a new method of grape disease detection based on the SegNet[38] architecture for visible light and infrared image segmentation to identify shadows, ground, healthy and symptomatic vines, and finally merge the segmentation obtained from visible light and infrared images to generate the whole disease map of grapes. Literature[39], a U-Net method for pixel-level purple rapeseed segmentation was proposed to calculate the model parameters by adjusting the sample size. In the literature[40], a new thermal imaging method was proposed to calculate the color similarity problem between unripe citrus fruits and leaves, which were prone to temperature differences between fruit and leaf surfaces because of the varying rates of temperature change between the fruit and leaf surfaces caused by water mist and to build a deep learning model based on the thermal imaging system. Meanwhile, the disease detection algorithm is moving towards lightweight, thereby making deploy into embedded devices easy. Chongke Bi[41] proposed a lightweight method for apple leaf disease identification based on MobileNet model. This method was also compared with ResNet152 and InceptionV3. The method can provide stable recognition results and is easily deployed in mobile devices. Utpal Barman[42] compared MobileNet CNN and Self-Structured CNN (SSCNN) based on citrus disease dataset from smartphone images. The experiments show that SSCNN is more accurate in classifying citrus leaf diseases based on smartphone images and takes less computation time. After research, increasing number of scholars tend to detect plant diseases using deep learning-based target detection methods, especially YOLO, SSD, and other target detection algorithms represented by one-stage methods, which omit tedious machine learning steps, such as image preprocessing, segmentation, and feature extraction, in a one-step end-to-end method with high recognition accuracy. Therefore, this paper explores the effectiveness of target detection algorithms for vegetable and fruit leaf disease detection and grading by using SSD as baseline method.

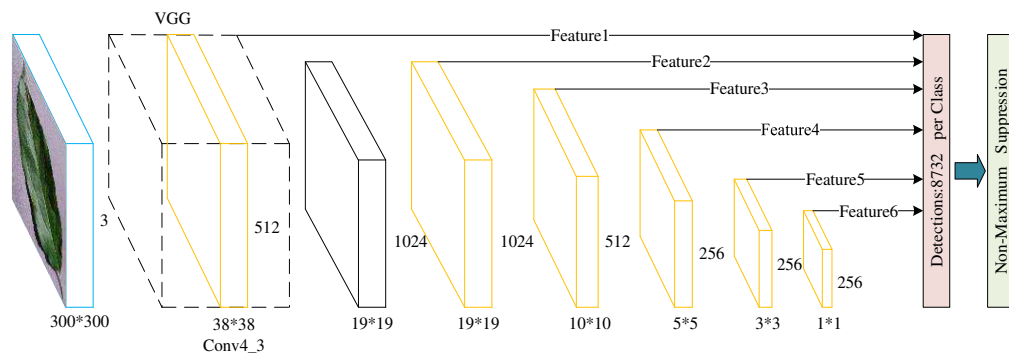## 2 Novel end-to-end method for leaf disease detection of fruits and vegetables

2.1 SSD Network

The SSD algorithm model is a one-stage real-time target detection model proposed simultaneously with YOLO series. SSD combines the one-stage regression prediction idea of the YOLO series and the Anchor Box mechanism of the Faster RCNN by using VGG as the base feature extraction network and extracting six different size feature layers from the bottom to the top layer as the regression prediction features. The advantage of SSD is that it improves the operation speed of the algorithm greatly while maintaining the detection accuracy. Moreover, the detection of small targets and large objects are considered. Figure 1 shows the SSD backbone network structure.

The loss function of SSD contains log loss for classification and smooth L1 for regression, and controls the proportion of positive and negative samples, which can improve the speed of optimization and the stability of training results. The total loss function is the sum of the errors of classification and regression. α is used to adjust the weight between the confidence loss and location loss, default =1, and N denotes the total number of default boxes that eventually match with Ground Truth boxes. Confidence loss is a typical softmax loss, and location loss is a typical smooth L1 loss.

Total loss:
$$L(x,c,l,g)=\frac{1}{N}(L_{conf}(x,c)+\alpha L_{loc}(x,l,g)) \tag{1}$$

Classified losses:
$$L_{conf}(x,c)=-\sum_{i\in Neg}^{N}x_{ij}^{p}\log(\hat{c}_{i}^{p})-\sum_{i\in Neg}\log(\hat{c}_{i}^{0}) \quad \hat{c}_{i}^{p}=\frac{\exp(c_{i}^{p})}{\sum_{p}\exp(c_{i}^{p})} \tag{2}$$

Of which:
$$x_{ij}^{p}=\{1,0\} \tag{3}$$

represents whether the i-th regression box matches the j-th GroundTruth box of type P.

Regression of losses:
$$L_{loc}(x,l,g)=\sum_{i\in Pos}^{N}\sum_{m\in\{cx,cy,w,h\}}x_{ij}^{k}smooth_{L1}(l_{i}^{m}-\hat{g}_{j}^{m}) \tag{4}$$

Although SSD adopts the direct regression prediction of the full-roller machine and no longer generates candidate frames, which greatly improves the detection speed of the SSD network, the SSD algorithm will miss and mis-detect when facing similar surface features and leaf occlusion situations, which often occur in actual leaf disease detection. For this reason, SSD needs to be improved to enhance feature recognition.

2.2 Squeeze-and Excitation SSD (Se_SSD) Network

Se_Block[43] mainly focuses on the relationship between channels and can explicitly model the interdependencies between feature channels with the structural unit "Squeeze-and Excitation (SE)" module, which adaptively adjusts the feature response values of each channel and internal dependencies between channels. The Se_Block module works as shown in Fig.2, First, feature compression is performed along the spatial dimension of the feature map, and each two-dimensional feature channel is turned into a real number, that has a global perceptual field to a certain extent. The output has the same number of dimensions as the input feature channels. Then, based on the correlation between the feature channels, a weight is generated for each feature channel to represent the importance of the feature channels. Finally, the original features are re-calibrated in the channel dimension by multiplying the channel-by-channel weights onto the previous features.
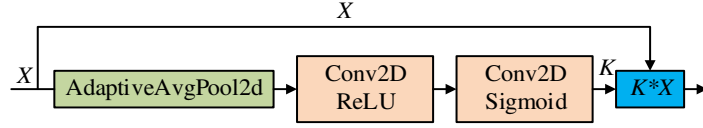
Fig. 2 Se_Block Attention Module

To increase the feature extraction capability of SSD feature extraction model and focus more on the feature layers with higher importance, this paper adds Se_Block attention mechanism module in front of the last six effective feature layers used for regression prediction on the basis of SSD model. The feature layers are rescaled by channel dimension. The structure of Se_SSD network is shown in Fig.3.
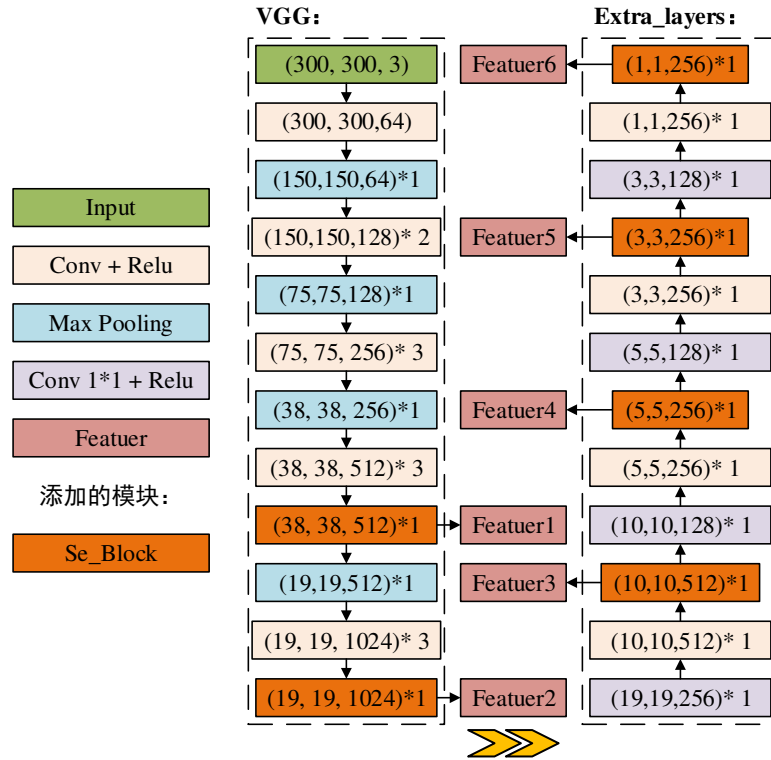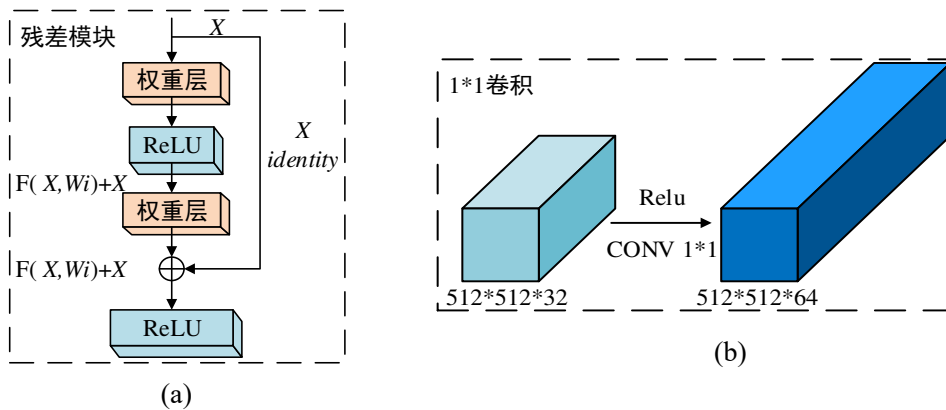

Fig.3 Se_SSD network structure

2.3 DBA_SSD Network

Fig. 4 (a) Residual network module and (b) 1*1 convolution

The residual network module, which is a module with good application in the last two years,

210 is shown in Figure 4(a). X is the input feature map, Wi is the weight of the ith layer network, F(X,
211 Wi) + X is the feature output, and F(X, Wi) + X is how the data are computed in the module. The
212 residual network is superior to the traditional convolutional network. The residual network module
213 implements an ultra-deep network and avoids the bottleneck problem of saturating the neural
214 network with correctness due to continuous deepening. In addition, by directly connecting the input
215 and output to achieve the goal of simplifying the learning objective and difficulty. 1*1 convolution
216 [46] is shown in Figure 4(b), and 1*1 convolution is usually followed by a nonlinear layer of Relu
217 for nonlinearization to learn more features. In addition to this 1*1 convolution's can change the
218 dimensionality of the image and transform the original image by 1*1 convolution to improve the
219 generalization ability to reduce overfitting, and at the same time reduce the computational effort by
220 boosting and reducing the number of channels to achieve cross-channel information interaction and
221 feature integration in the process.

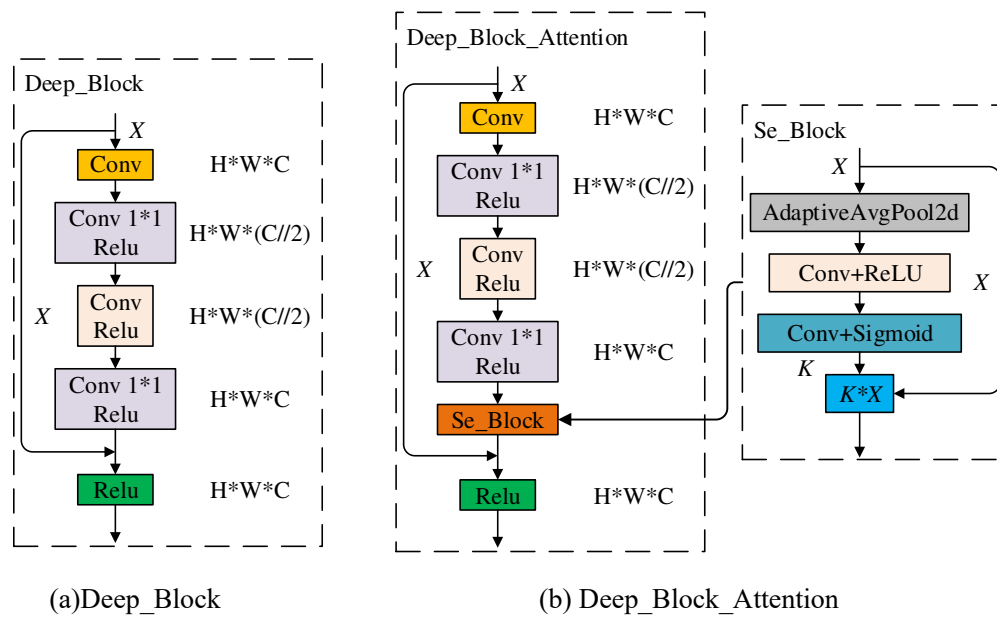(a)Deep_Block   (b) Deep_Block_Attention

222 Figure 5 Enriched feature extraction module ((a) Deep_Block, a feature extraction module
223 combining residual network and 1*1 convolution; (b) Deep_Block_Attention, a feature extraction
224 module adding an attention mechanism to (a))
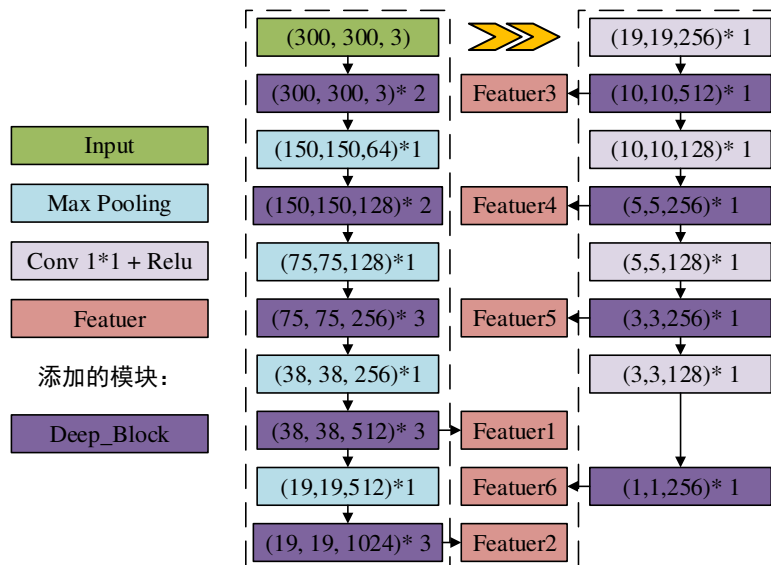

225

226   Fig.6 DBA_SSD network structure

227   As shown in Figure 5, two kinds of rich feature extraction modules are designed in this paper,
228 as shown in Figure 5(a), Deep_Block is used to enhance the network feature extraction capability
229 by using 1*1 convolution to reduce the number of channels after convolution, fusing multi-channel
230 information, while introducing a residual structure to prevent the loss of feature layer information.
231 Deep_Block_Attention adds a channel attention mechanism at the end of the Deep_Block structure
232 for fine-tuning at the channel level. As shown in Figure 5(b), the feature extraction network of SSD
233 is reconstructed with the rich feature extraction module as the basic feature extraction unit, as shown
234 in Figure 6, to deepen the feature extraction of each layer and increase the richness of feature
235 learning by the rich feature extraction module.

236 **3 Experimental environment and experimental design**

237 **3.1 Experimental environment**

238   This experiment is a deep learning model built under the Pytorch deep learning framework,
239 using a dataset of 3000 vegetable and fruit leaves, and the final output prediction frame identifies
240 the leaf species and determines the severity of leaf disease. The experimental environment uses
241 AMD Ryzen 7 4800H processor, NVIDIA GeForce RTX 2060 graphics card, 32G RAM, and
242 Pytorch deep learning framework.

243 **3.2 Dataset**

244   Benefiting from the convenience and simplicity of Labelimg, this experiment uses Labelimg
245 software to label the dataset and obtain data in VOC format for training, with label files as .xml files
246 and pictures as .jpg files. The dataset of the experiment has 3000 images, which are divided into 5
247 major categories: Apple, Tomato, Potatoes, Strawberry and Chili; each major category is divided
248 into 3 subcategories according to the severity of leaf disease: healthy, general, and severe. In total,
249 15 subcategories are noted, and the image resolution is around 255*470*3 pixels. The ratio of test,
250 train, and val in the total data set is 1:8:1. Figure 7 shows the composition of the data set.
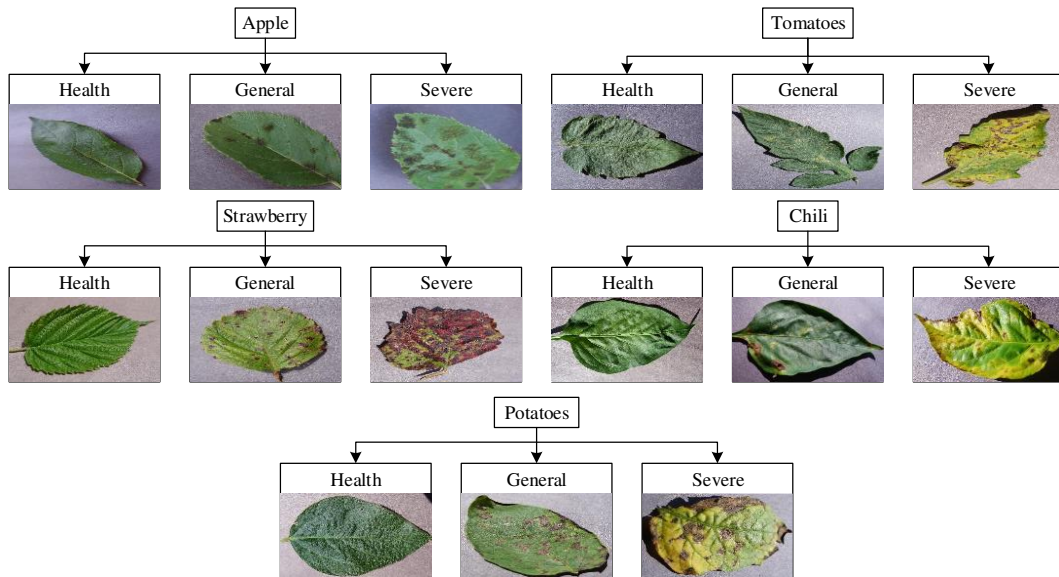

251
252   Fig.7 Data set composition structure

3.3 Experimental design

To ensure the equalization of the dataset and to increase the richness and quality of the dataset, data enhancement and image preprocessing were performed on the images before the experimental tests[44]. The means of enhancement are Histogram Equalization, Horizontal Flip + Hue Saturation Value, Vertical Flip + Channel Shuffle, Horizontal Flip + Vertical Flip+ Channel Shuffle. The enhanced images are shown in Figure 8, with each of the 15 classes expanded to 1,000 images, and the number of data sets expanded from 3,000 to 15,000, with the training, validation, and testing ratios randomly assigned according to 1:8:1.

Among them, Histogram Equalization is a means of image pixel processing that serves to bring out the objects with insignificant contrast. It is essentially a grayscale transformation that spreads concentrated grayscale intervals over the entire grayscale interval. It is a nonlinear transformation. A digital image histogram with a gray level in the range [0, L-1] is a discrete function of.

$$s_k = \sum_{j=0}^{k} \frac{n_j}{n} \quad k = 0, 1, 2, \dots L - 1 \tag{5}$$

where n is the sum of the pixels in the image, is the number of pixels in the current gray level, and L is the total number of possible gray levels in the image. Color histogram equalization is the fusion of each of the three channels of the image after equalization. Histogram equalization is performed on the dataset considering the presence of blur and insignificance in the dataset.

(a) Positive sample (healthy)

| Original Image | Histogram Equalization | Horizontal Flip and Hue Saturation Value | Vertical Flip andChannel Shuffle | Horizontal Flip ＆Vertical Flip& Channel Shuffle |
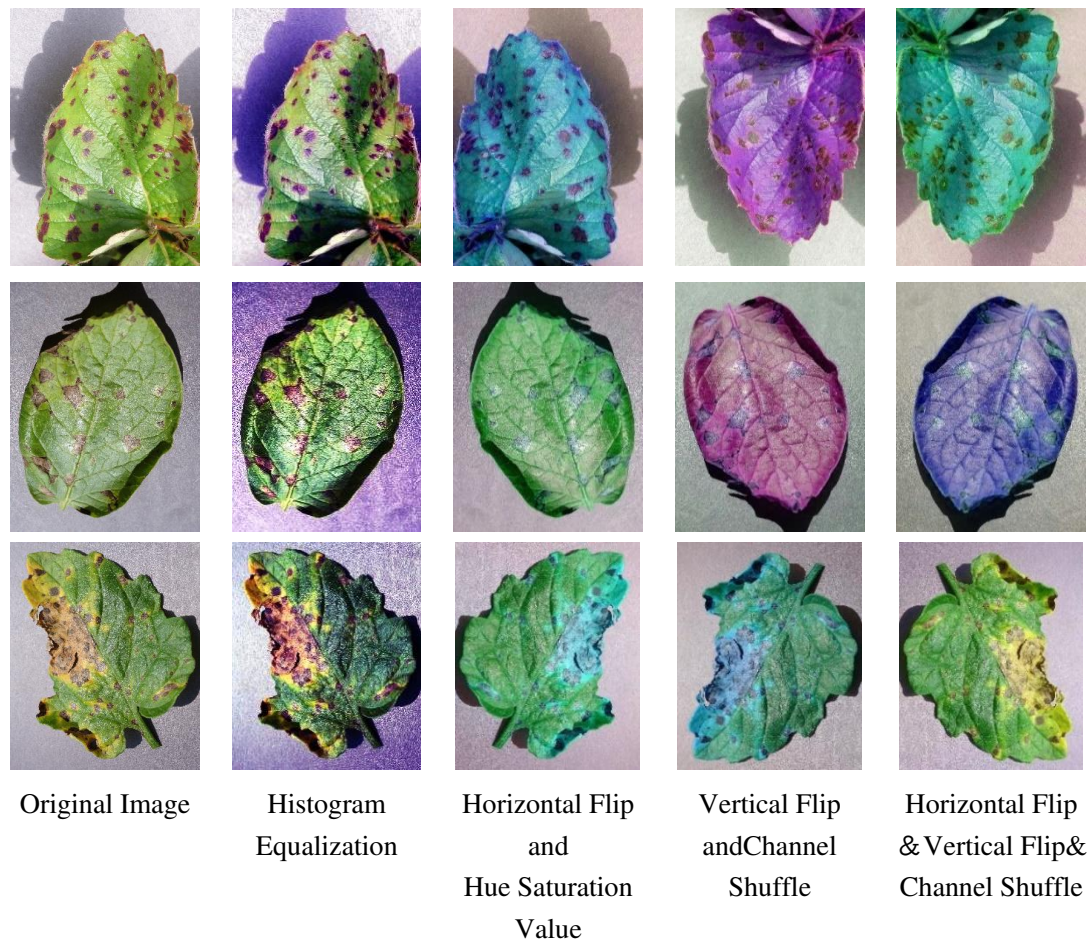
(b) Negative sample (disease)

Fig. 8 Data Enhancement

To better test the performance of the improved algorithm, four experiments were designed. Se_SSD with channel attention mechanism added at the end of the feature extraction network, DB_SSD (Deep Block SSD) with improved VGG feature extraction network, DBA_SSD with fusion of the improved VGG network and channel attention mechanism, and SSD of the original network are compared, and the VGG model trained on Image Net image dataset is trained by migrated convolutional layers to this model.

Experiment 1: The Se_SSD network with the Se_Block channel attention mechanism added is trained and the average accuracy of this network for the detection of fruits and vegetables leaves is tested.

Experiment 2: The DB_SSD network with the Deep_Block module added is trained in the environment and hardware conditions of Experiment 1, where the Deep_Block module does not contain the attention mechanism, and the effect of the network with the added attention mechanism on the detection of fruits and vegetables leaves is tested.

Experiment 3. The DBA_SSD network with the Deep_Block_Attention module added, and the Deep_Block_Attention module containing the attention mechanism, i.e., the SSD network with both improved structure and added attention mechanism, is trained and tested under the environment and hardware conditions of Experiment 1.

Experiment 4. The original SSD network is trained and tested under the environment and hardware conditions of Experiment 1.

290   All the four experiments were trained on the basis of 15,000 vegetable and fruit pest leaf
291 datasets and tested 1500 randomly selected images. The experiments followed the experimental
292 flow in Fig.9, the experiment-comparison-optimization-experiment pattern, to obtain the average
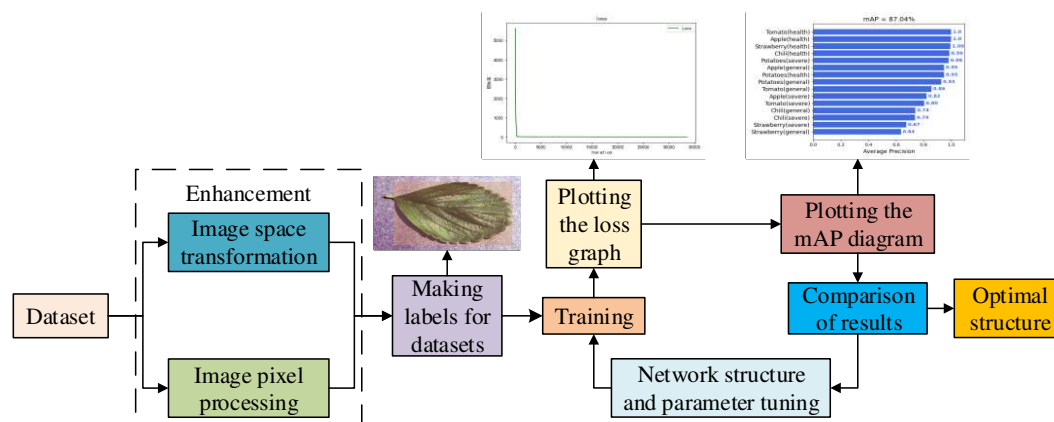293 accuracy mAP under this model and to compare the mAP values of different models.

294

295   Fig.9 Experimental flow

296 **3.4 Performance Evaluation Metrics**

297   Precision is a measure of the accuracy of a model's prediction, and its value is equal to the
298 number of correctly predicted positive samples over the total number of positively predicted samples.
299 Recall (Recall) is a measure of the model's ability to identify positive samples, and its value is the
300 number of correctly predicted positive samples over the total number of positively predicted samples.
301 The prediction results of the model are shown in Table 1 for   TP, FP, FN, and TN.

302   Table 1 Confusion matrix

	True class	
	TP	FP
	True Positive	False Positive
Predict class	FN	TN
	False Negative	True Negative

303   True Positives (TP): indicates the number of correctly identified positive samples; True
304 Negatives (TN): indicates the number of correctly identified negative samples; False Positives (FP):
305 indicates the number of incorrectly identified negative samples; False Negatives (FN): indicates the
306 number of incorrectly identified positive samples.

$$precision = \frac{TP}{TP + FP} \qquad (6)$$

$$recall = \frac{TP}{TP + FN} \qquad (7)$$

307   The PR curve is a graph drawn with Recall as the horizontal axis and Precision as the vertical
308 axis; Precision is negatively correlated with Recall, and the recall rate decreases as precision
309 increases. AP (Average Precision) as a single category indicator is the integration of PR curve.

$$AP = \int_0^1 p(r)\, d(r) \qquad (8)$$

310   The value of mAP(mean average precision), as one of the important metrics for the evaluation

311   of the whole model, is the average of the summation of all the category APs.

$$mAP = \frac{\sum_{n=1}^{N} AP(n)}{N} \tag{9}$$

312   where n is the category and N is the total number of categories.

4. Analysis of experimental results

313

4.1 DBA _SSD model experimental comparison analysis

314

315   The first 50 Epochs were trained by freezing some of the network layer weights, and each batch
316   was trained with 8 images. For the last 50 Epochs, the frozen layers were unfrozen and the full
317   network was trained. The learning rate started at 5e-4, and after unfrozen the learning rate was 1e-
318   4. Fine tuning of the model parameters was performed. As shown in Figure 12, the horizontal
319   coordinate is the number of Epochs trained, and the vertical coordinate is the loss value at the end
320   of training for each Epoch. different line shapes indicate different improvement algorithms. As
321   shown in Figure 11, the loss value of the model decreases as the number of iterations increases. The
322   loss values in the training log gradually stop changing around 90 - 100 Epoch. The red thin solid
323   line in the figure indicates the loss value of DBA_SSD, whose value is lower compared with the
324   loss of SSD, Se_SSD, and DB_SSD algorithms.
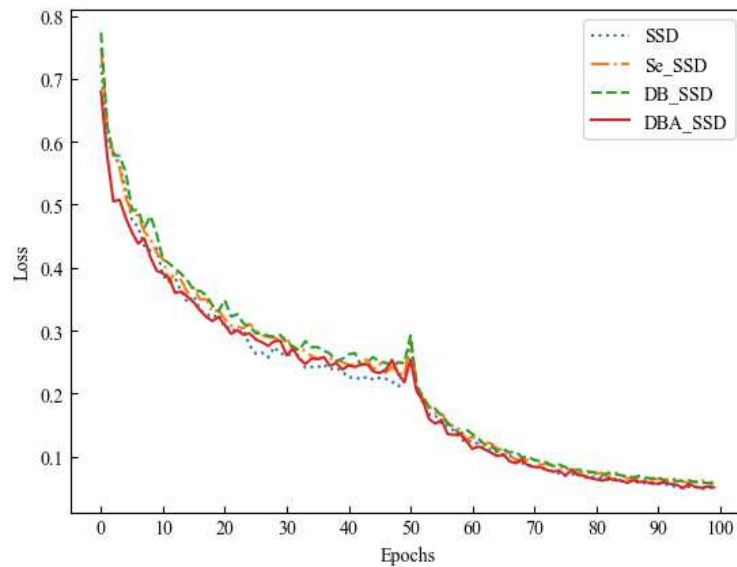
325
326   Fig.11 SSD and its improved algorithm loss variation graph
327   The test results between SSD and its improved algorithm are shown in Table 2. DBA_SSD has
328   the highest accuracy because Deep Block strengthens the network's feature extraction ability on the
329   one hand, and it incorporates the channel attention mechanism to accelerate the network learning
330   on the other hand, so that the network focuses on the channels with high information content for
331   feature learning.The prediction accuracy between its SSD and its improved algorithm for predicting
332   different species of fruit and vegetable diseases is shown in Figure 12. The prediction accuracy of
333   DBA_SSD is relatively high among most of the categories, and the mAP value of DBA_SSD is
334   92.20%, while the mAP values of SSD, Se_SSD, and DB_SSD are 9.96%, 90.77%, and 89.93%,
335   respectively.
336   Table 2 Comparison of accuracy of improved SSD algorithm

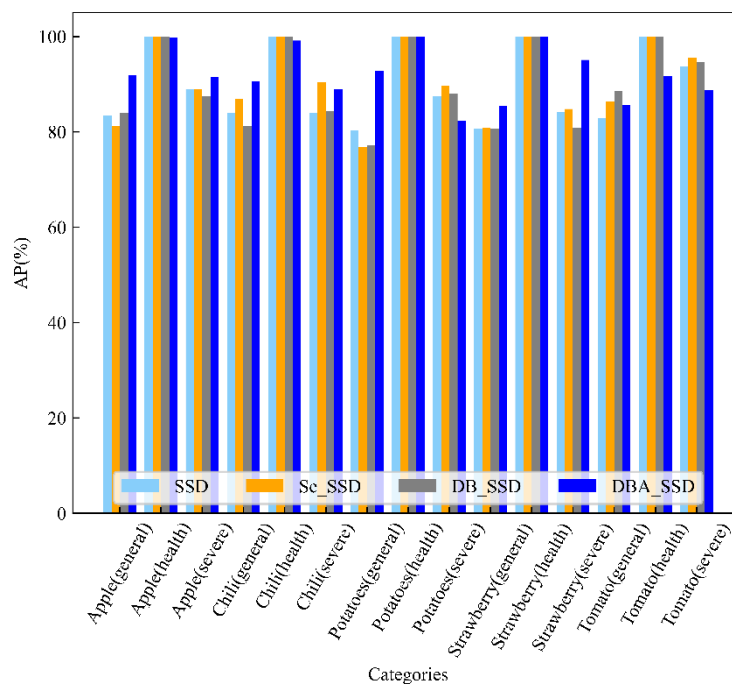| Target identification methods	Inserted modules	mAP
SSD	\	89.96%
Se_SSD	Se_Block	90.77%
DB_SSD	Deep_Block	89.93%
DBA_SSD	Deep_Block_Attention	**92.20%**

Fig.12 AP diagram of SSD and its improved algorithm for the detection of different kinds of diseases

Further observe the data distribution of the experimental results in Figure 13. The horizontal coordinates indicate the improved algorithm types, the vertical coordinates are the distribution of predicted AP values for the 15 types, the points of the triangle indicate the mean, and the thin solid line in the middle of the rectangle indicates the median. From Figure 13, we can see that among the four algorithms SSD, Se_SSD, DB_SSD, and DBA_SSD, DBA_SSD prediction accuracy is more concentrated. Moreover, the median and mean are the highest. DBA_SSD algorithm has better performance compared with other improved algorithms.

347
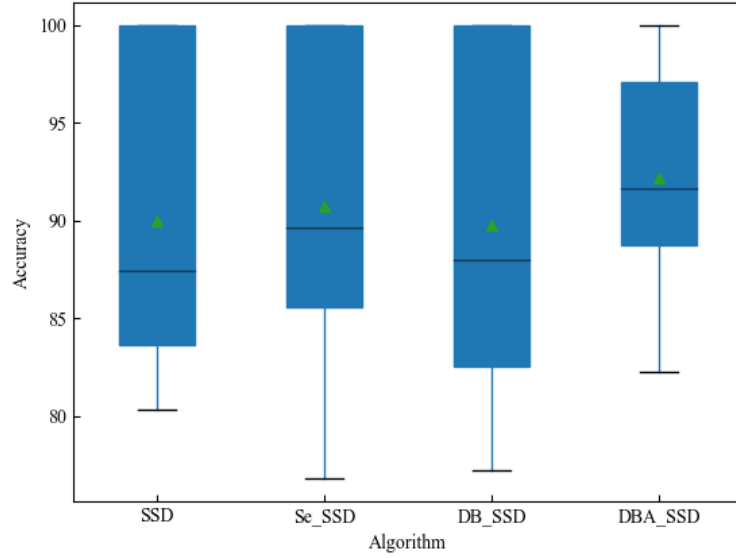
348   Fig.13 Box diagram of SSD and its improvement algorithm

349   **4.2 Comparative analysis with classical target detection algorithms**

350   This experiment compares and analyzes the test results of the classical target detection

351   algorithms YOLOv4[45], YOLOv4 tiny[46], Faster RCNN, and YOLOv3. This experiment is

352   conducted with the same dataset in the same experimental environment, and its Loss variation of

353   each algorithm is shown in Figure 14.

354
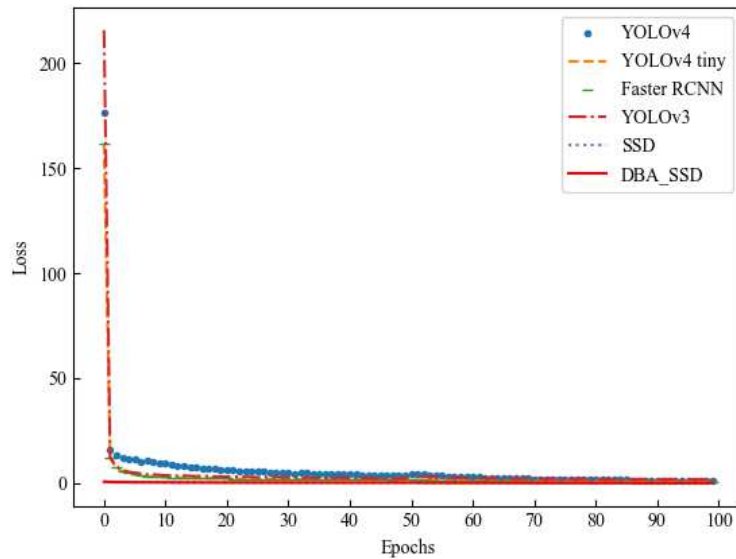
355   Fig.14 Target detection algorithm loss diagram

356   Each vegetable and fruit plant leaves in this paper can be classified into three categories

357   according to the degree of disease, which are healthy, general and severe (Table 3). Figure 15 then

358   averages the detection accuracy of the same leaves on the basis of Table 3. The prediction accuracy

359   of this category is the average of the sum of the prediction accuracy of the three degrees of leaves.

360   Therefore, its horizontal coordinates indicate different target detection algorithms, and its vertical

361   coordinates indicate the average prediction accuracy and the total average prediction accuracy (mAP)

362   of different kinds of fruits and vegetables leaves.

363   Compared with DBA_SSD, YOLOv4 has lower prediction accuracy for Strawberry and Chili,

YOLOv4 tiny has weaker prediction ability for Tomato, and YOLOv3 has lower prediction accuracy for Strawberry. This is   the learning difference caused by different algorithms of feature extraction networks focusing on different information of the learned images, and DBA_SSD solves this deficiency by covering all levels of semantic information. The rightmost column indicates the average detection accuracy of the DBA_SSD algorithm in different categories, with the highest classification accuracy of 100% and the lowest of 82.24%.

Table 3 Comparison of the accuracy of the improved SSD model and other target detection algorithms for the detection of different kinds of diseases

Algorithm / Category	YOLOv4	YOLOv4 tiny	YOLOv3	SSD	Faster RCNN	DBA_SSD
Apple(general)	94.79%	78.32%	88.87%	83.45%	74.85%	91.83%
Apple(health)	100.00%	99.78%	94.83%	100.00%	100.00%	99.73%
Apple(severe)	82.27%	88.01%	90.71%	88.93%	88.20%	**91.56%**
Chili(general)	73.99%	92.32%	81.60%	83.89%	90.54%	90.65%
Chili(health)	98.75%	100.00%	100.00%	100.00%	100.00%	99.12%
Chili(severe)	73.74%	72.70%	91.55%	83.94%	100.00%	88.86%
Potatoes(general)	92.71%	89.41%	88.37%	80.35%	88.65%	**92.72%**
Potatoes(health)	94.74%	100.00%	100.00%	100.00%	99.80%	**100.00%**
Potatoes(severe)	98.08%	91.18%	89.32%	87.46%	82.88%	82.24%
Strawberry(general)	63.52%	80.89%	59.83%	80.69%	73.48%	**85.37%**
Strawberry(health)	99.52%	100.00%	100.00%	100.00%	100.00%	**100.00%**
Strawberry(severe)	67.46%	76.44%	78.95%	84.11%	92.64%	**95.07%**
Tomato(general)	85.70%	85.64%	95.81%	82.85%	85.02%	85.54%
Tomato(health)	100.00%	94.58%	100.00%	100.00%	89.14%	91.67%
Tomato(severe)	80.35%	64.31%	78.24%	93.69%	86.33%	88.65%
mAP	87.04%	87.57%	89.21%	89.96%	90.10%	**92.20%**
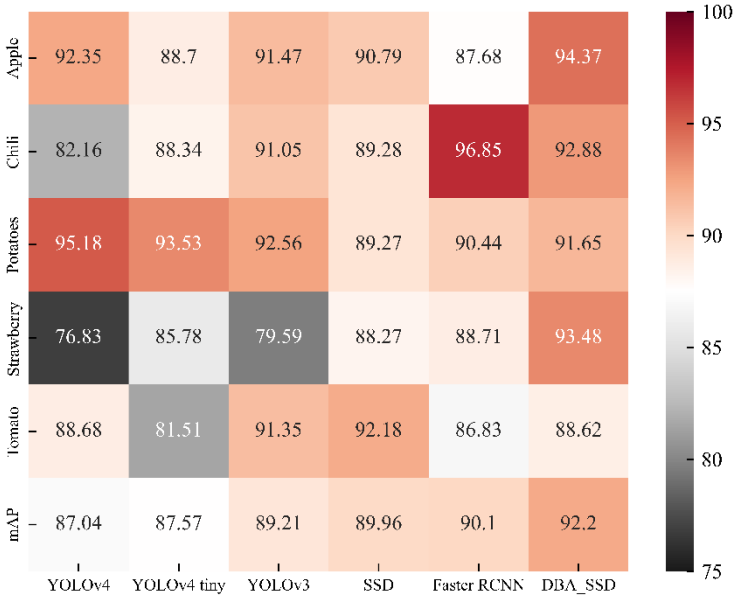
Fig.15 Heat map of correlation between different target detection algorithms and vegetable and

374   fruit leaf types

375   Figure 16 shows that YOLOv4 corresponds to the largest rectangular box area, and its upper

376   quartile edge is close to 100%, indicating the existence of a certain number of prediction accuracies

377   higher than 95%. However, its predicted category accuracy is more discrete. YOLOv3 has a smaller

378   rectangular area, but its distance at the top of the rectangle is not as far as DBA_SSD, indicating

379   that the number of its higher accuracy is not as high as DBA_SSD. Although the upper quartile line

380   of SSD is in contact with the 100% line, its rectangle area is larger, indicating that the prediction

381   accuracy varies widely and is unstable. The rectangle box area of DBA_SSD is the smallest among

382   other algorithms, indicating that the prediction accuracy is more concentrated and is closer to the

383   100% line, suggesting that a large part of the prediction accuracy is high and the prediction of each

384   kind is more stable. The experiment shows that the DBA_SSD model has a high accuracy rate for

385   the recognition of fruit and vegetable leaves, and the SSD is a one-stage target recognition algorithm

386   with the advantage of fast recognition speed. The comprehensive performance of DBA_SSD has

387   been improved compared with the previous SSD, and the performance is also higher compared with

388   other target detection algorithms. The detection effect is shown in Fig.17.

389
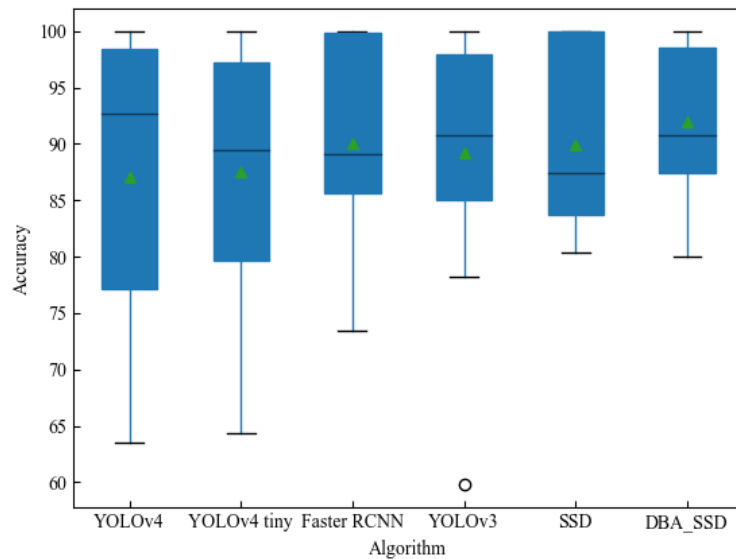
390   Fig.16 Box plot of AP statistics under target detection algorithm

391

392
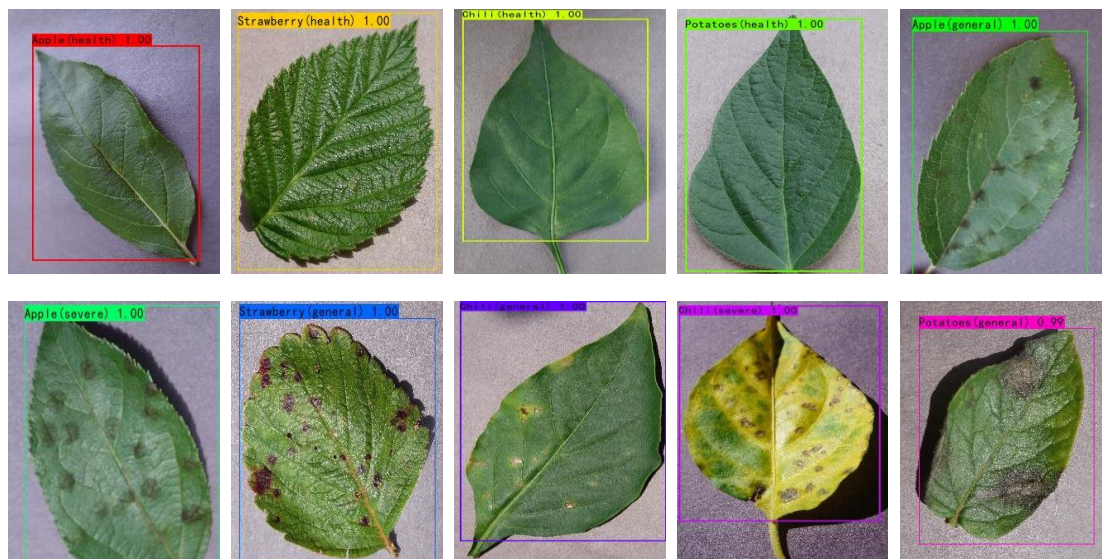
Fig.17 DBA_SSD recognition effect

5 Summary and Outlook

This paper discusses the work related to vegetable and fruit pest leaf detection, which augments the dataset with spatial transformation, as well as pixel processing on top of the original one. To address the problem wherein the recognition rate of the SSD model is not high and the detail information is not paid attention to, which resulted in high accuracy and incorporated 1×1 convolution, residual network and attention mechanism into the SSD algorithm, the DBA_SSD network model for vegetable and fruit leaf pest detection is proposed to compare and analyze the experimental effect of multiple sets of classical target detection algorithms in vegetable and fruit leaf health detection. It makes the SSD algorithm improve to 92.20% on the original basis with high robustness and speed. In this paper, we mainly performed research on vegetable and fruit pest identification algorithms, but a gap for the application of target detection algorithms in actual production, and future work will mainly focus on the reduction and optimization of algorithms to implant embedded devices for the application of real-time monitoring of agricultural plant diseases.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Reza M N, Na I S, Baek S W, et al. Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images[J]. BIOSYSTEMS ENGINEERING. Biosystems Engineering, 2019, 177: 109-121. doi:http://doi.org/10.1016/j.biosystemseng.2018.09.014

[2] Quan L, Feng H, Li Y, et al. Maize seedling detection under different growth stages and complex field environments based on an improved Faster R-CNN[J]. BIOSYSTEMS ENGINEERING. 2019, 184: 1-23. doi:http://doi.org/10.1016/j.biosystemseng.2019.05.002

[3] Sun Y, Liu X, Yuan M, et al. Automatic in-trap pest detection using deep learning for pheromone-based Dendroctonus valens monitoring[J]. BIOSYSTEMS ENGINEERING. 2018, 176: 140-150. doi:http://doi.org/10.1016/j.biosystemseng.2018.10.012

[4] Arnal Barbedo J G. Plant disease identification from individual lesions and spots using deep learning[J]. BIOSYSTEMS ENGINEERING. 2019, 180: 96-107. doi:http://doi.org/10.1016/j.biosystemseng.2019.02.002

[5] Li G, Zhang M, Li J, et al. Efficient densely connected convolutional neural networks[J]. PATTERN RECOGNITION. 2021, 109(107610). doi:http://doi.org/10.1016/j.patcog.2020.107610

[6] Hasan R I, Yusuf S M, Alzubaidi L. Review of the State of the Art of Deep Learning for Plant Diseases: A Broad Analysis and Discussion[J]. Plants, 2020, 9(10): 1302. doi:http://doi.org/10.3390/plants9101302

[7] Yang J, Li S, Wang Z, et al. Using Deep Learning to Detect Defects in Manufacturing: A Comprehensive Survey and Current Challenges.[J]. Materials, 2020, 13(24): 5755. doi:http://doi.org/10.3390/ma13245755

[8] Fu L, Majeed Y, Zhang X, et al. Faster R-CNN-based apple detection in dense-foliage fruiting-wall trees using RGB and depth features for robotic harvesting[J]. BIOSYSTEMS ENGINEERING. 2020, 197: 245-256. doi:http://doi.org/10.1016/j.biosystemseng.2020.07.007

[9] Liu Z, Wang J, Tian Y, et al. Deep learning for image-based large-flowered chrysanthemum cultivar recognition[J]. PLANT METHODS. 2019, 15(1): 1-11. doi:http://doi.org/10.1186/s13007-019-0532-7

[10] Fu L, Feng Y, Wu J, et al. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model[J]. PRECISION AGRICULTURE. 2020. doi:http://doi.org/10.1007/s11119-020-09754-y

[11] Zhao H, Zhou Y, Zhang L, et al. Mixed YOLOv3-LITE: A Lightweight Real-Time Object Detection Method[J]. Sensors, 2020, 20(7): 1861. doi: http://doi.org/10.3390/s20071861

[12] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[13] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[J]. arXiv e-prints. 2016: 1612-8242.

[14] Liu G, Nouaze J C, Mbouembe P L T, et al. YOLO-Tomato: A Robust Algorithm for Tomato Detection Based on YOLOv3[J]. Sensors, 2020, 20(7): 2145. doi:http://doi.org/10.3390/s20072145

[15] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. arXiv e-prints. 2018: 1804-2767.

[16] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. 2017, 39(6): 1137-1149. doi:http://doi.org/10.1109/TPAMI.2016.2577031

[17] Liu J, Wang X. Early recognition of tomato gray leaf spot disease based on MobileNetv2-YOLOv3 model[J]. PLANT METHODS. 2020, 16(1): 1-16.doi:http://doi.org/10.1186/s13007-020-00624-2

[18] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C]. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).2018.

[19] Yi J, Krusenbaum L, Unger P, et al. Deep Learning for Non-Invasive Diagnosis of Nutrient Deficiencies in Sugar Beet Using RGB Images[J]. Sensors, 2020, 20(20): 5893. doi:http://doi.org/10.3390/s20205893

[20] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90. doi:http://doi.org/10.1145/3065386

[21] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. arXiv 1409.1556. 2014.

[22] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. arXiv e-prints. 2015: 1512-3385.

[23] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[24] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size[J]. arXiv preprint arXiv:1602.07360, 2016.

[25] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[J]. arXiv e-prints. 2015: 1512-2325.

[26] Lin T, Dollár P, Girshick R, et al. Feature Pyramid Networks for Object Detection[C]. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).Honolulu, HI, USA: IEEE, 2017. 936-944.

[27] Chao X, Sun G, Zhao H, et al. Identification of Apple Tree Leaf Diseases Based on Deep Learning Models[J]. Symmetry, 2020, 12(7): 1065. doi:http://doi.org/10.3390/sym12071065

[28] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions[J]. arXiv e-prints. 2016.

[29] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.

[30] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.

[31] Shrivastava V K, Pradhan M K. Rice plant disease classification using color features: a machine learning paradigm[J]. JOURNAL OF PLANT PATHOLOGY. 2020. doi:http://doi.org/10.1007/s42161-020-00683-3

[32] Rao A, Kulkarni S B. A Hybrid Approach for Plant Leaf Disease Detection and Classification Using Digital Image Processing Methods[J]. The International Journal of Electrical Engineering & Education, 2020: 0020720920953126. doi:http://doi.org/10.1177/0020720920953126

[33] Abdulridha J, Ampatzidis Y, Roberts P, et al. Detecting powdery mildew disease in squash at different stages using UAV-based hyperspectral imaging and artificial intelligence[J]. BIOSYSTEMS ENGINEERING. 2020, 197: 135-148. doi:http://doi.org/10.1016/j.biosystemseng.2020.07.001

[34] Abdu A M, Mokji M M, Sheikh U U. Automatic vegetable disease identification approach using individual lesion features[J]. COMPUTERS AND ELECTRONICS IN AGRICULTURE. 2020, 176(105660). doi:http://doi.org/10.1016/j.compag.2020.105660

[35] Samiei S, Rasti P, Ly Vu J, et al. Deep learning-based detection of seedling development[J]. PLANT METHODS. 2020, 16(1031). doi: http://doi.org/10.1186/s13007-020-00647-9

504 [36] Plesoianu A, Stupariu M, Sandric I, et al. Individual Tree-Crown Detection and Species
505 Classification in Very High-Resolution Remote Sensing Imagery Using a Deep Learning Ensemble
506 Model[J]. REMOTE SENSING. 2020, 12(242615). doi: http://doi.org/10.3390/rs12152426

507 [37] Kerkech M, Hafiane A, Canals R. Vine disease detection in UAV multispectral images using
508 optimized image registration and deep learning segmentation approach[J]. COMPUTERS AND
509 ELECTRONICS IN AGRICULTURE. 2020, 176: 105660.
510 doi:http://doi.org/10.1016/j.compag.2020.105446

511 [38] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder
512 architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence,
513 2017, 39(12): 2481-2495.

514 [39] Zhang J, Xie T, Yang C, et al. Segmenting Purple Rapeseed Leaves in the Field from UAV RGB
515 Imagery Using Deep Learning as an Auxiliary Means for Nitrogen Stress Detection[J]. Remote Sensing,
516 2020, 12(9): 1403. doi:http://doi.org/10.3390/rs12091403

517 [40] Gan H, Lee W S, Alchanatis V, et al. Active thermal imaging for immature citrus fruit detection[J].
518 Biosystems Engineering. 2020, 198: 291-303. doi:http://doi.org/10.1016/j.biosystemseng.2020.08.015

519 [41] Bi C, Wang J, Duan Y, et al. MobileNet Based Apple Leaf Diseases Identification[J]. MOBILE
520 NETWORKS & APPLICATIONS. 2020. 1-9doi:http://doi.org/10.1007/s11036-020-01640-1

521 [42] Barman U, Choudhury R D, Sahu D, et al. Comparison of convolution neural networks for
522 smartphone image based real time classification of citrus leaf disease[J]. Computers and Electronics in
523 Agriculture, 2020, 177: 105661. doi:http://doi.org/10.1016/j.compag.2020.105661

524 [43] Hu J, Shen L, Albanie S, et al. Squeeze-and-Excitation Networks[J]. IEEE TRANSACTIONS ON
525 PATTERN ANALYSIS AND MACHINE INTELLIGENCE. 2020, 42(8): 2011-2023.
526 doi:http://doi.org/10.1109/TPAMI.2019.2913372

527 [44] Buslaev A, Iglovikov V I, Khvedchenya E, et al. Albumentations: fast and flexible image
528 augmentations[J]. Information, 2020, 11(2): 125.doi: https://doi.org/10.3390/info11020125 .

529 [45] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal Speed and Accuracy of Object
530 Detection[J]. arXiv preprint arXiv:2004.10934, 2020.

531 [46] Jiang Z, Zhao L, Li S, et al. Real-time object detection method based on improved YOLOv4-tiny[J].
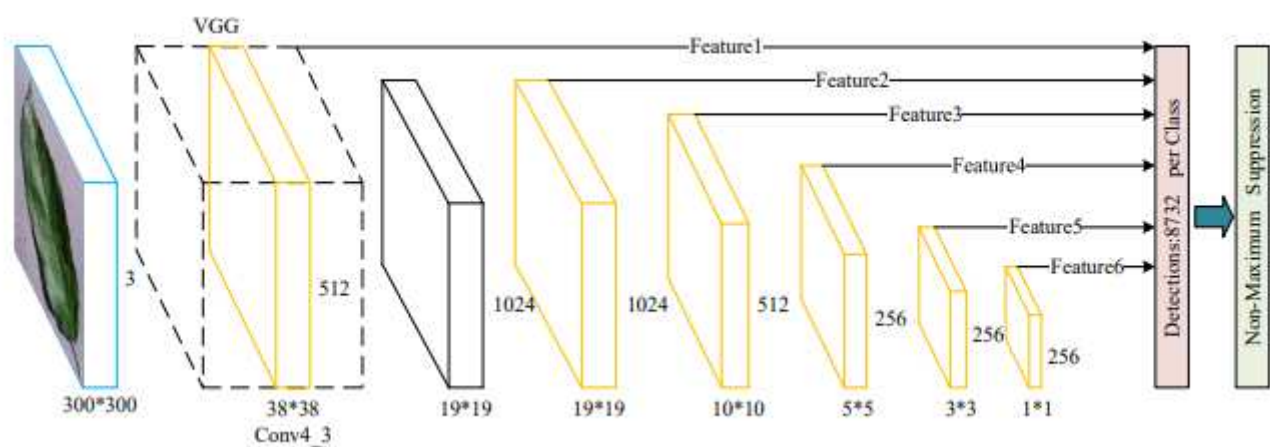532 arXiv preprint arXiv:2011.04244, 2020.

533

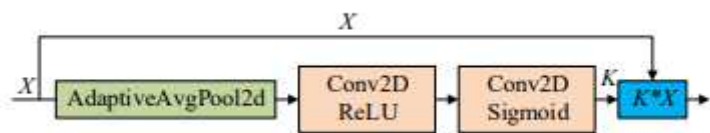# Figures

Figure 1

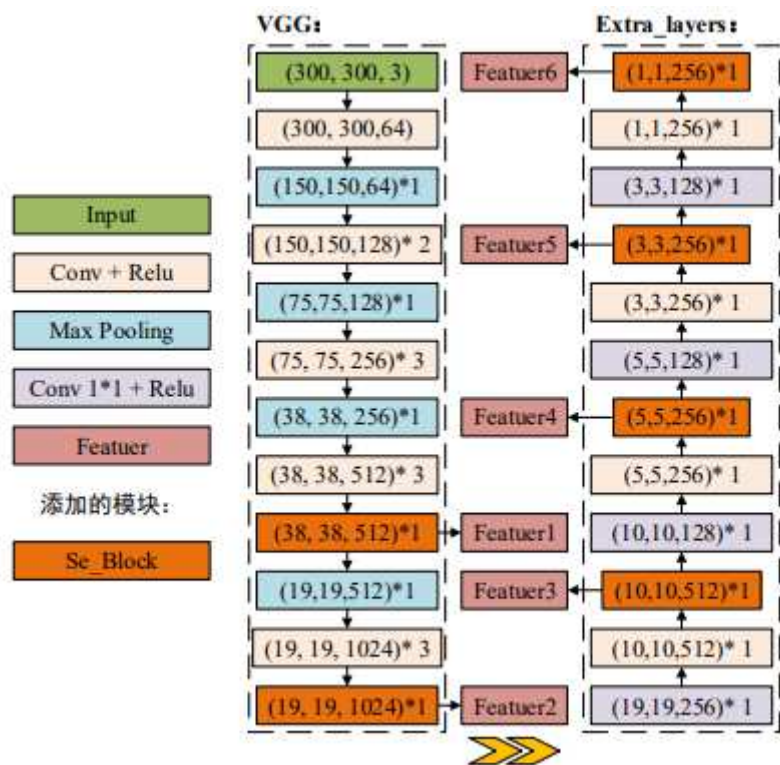SSD backbone network structure

Figure 2

Se_Block Attention Module

Figure 3

Se_SSD network structure

(a)

(b)
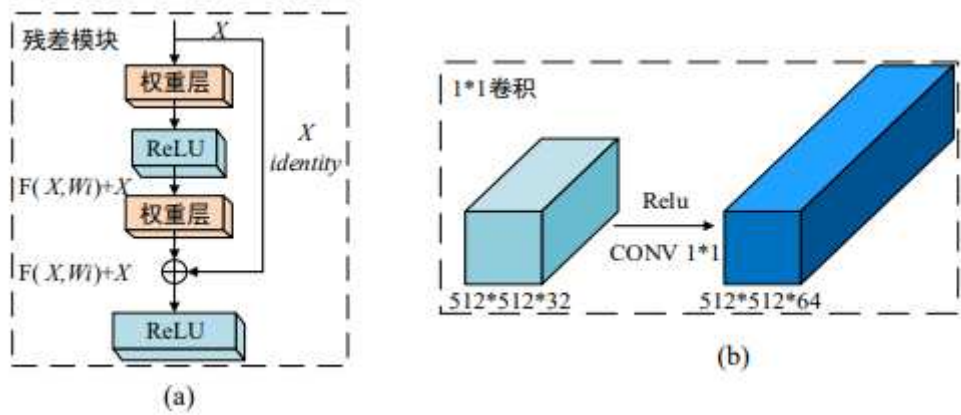
Figure 4

(a) Residual network module and (b) 1*1 convolution
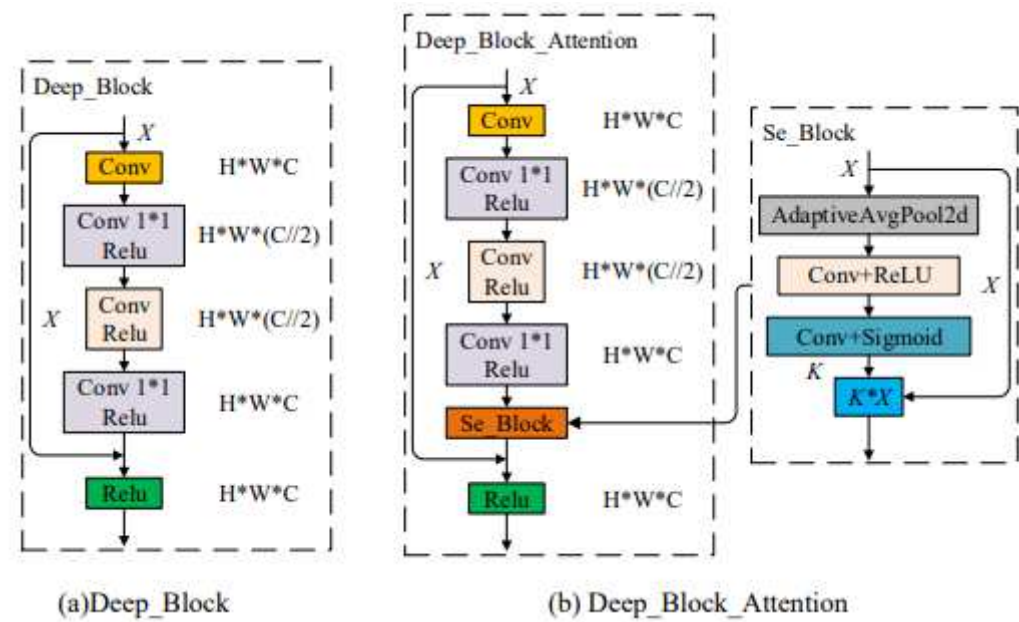

(a)Deep_Block

(b) Deep_Block_Attention

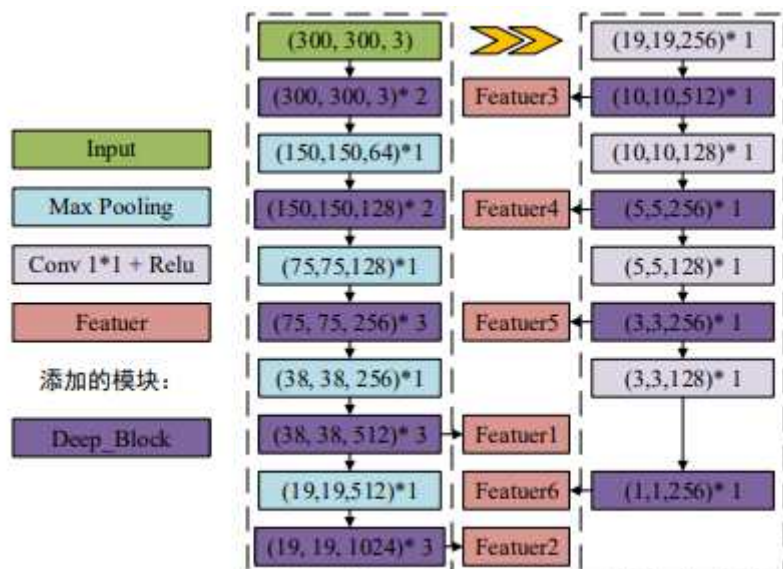## Figure 5

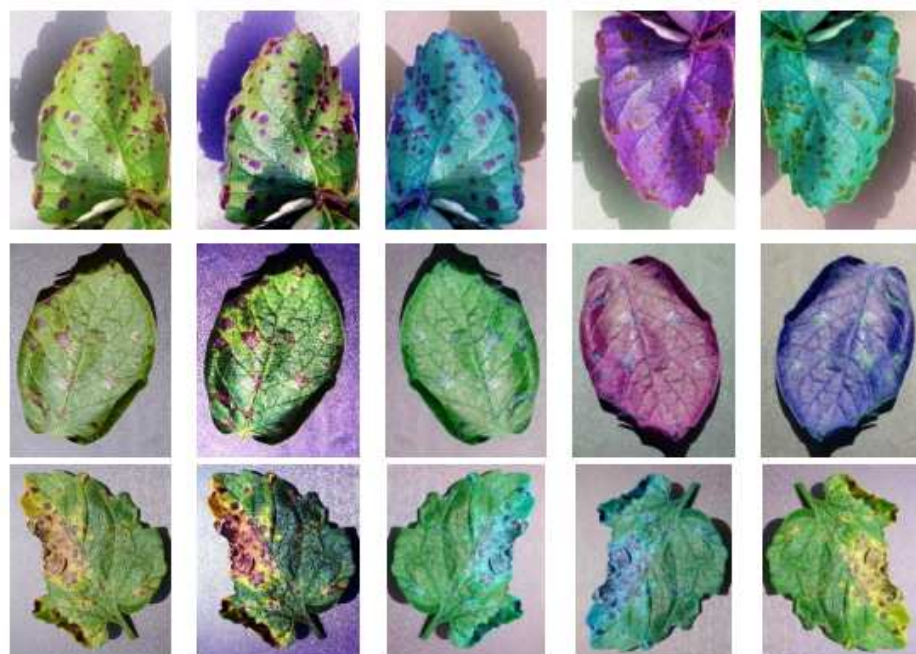Enriched feature extraction module

Figure 6

DBA_SSD network structure
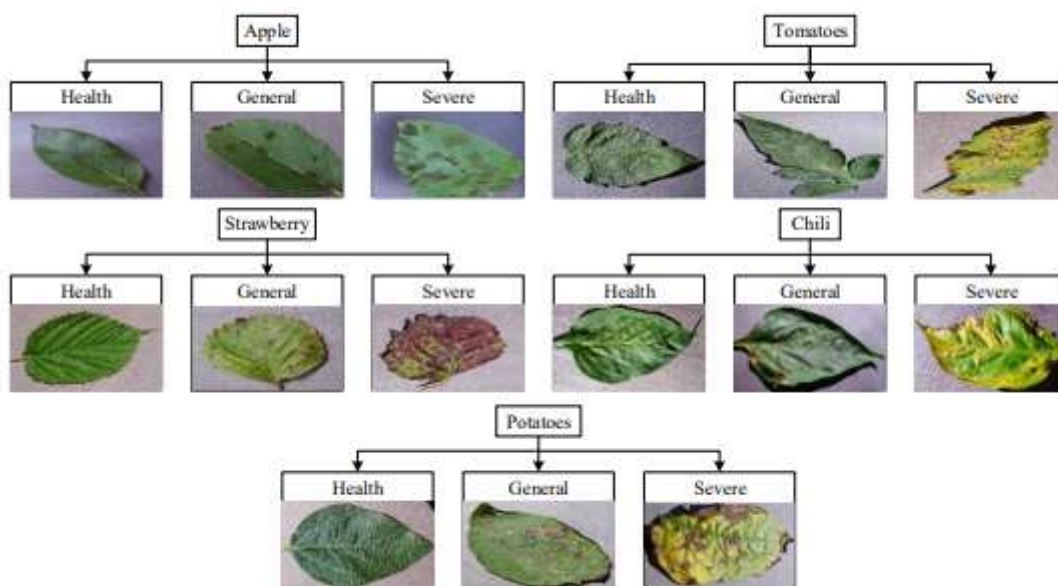
Figure 7

Data set composition structure

(a) Positive sample (healthy)

| Original Image | Histogram Equalization | Horizontal Flip and Hue Saturation Value | Vertical Flip andChannel Shuffle | Horizontal Flip & Vertical Flip& Channel Shuffle |

(b) Negative sample (disease)

Figure 8

Data Enhancement
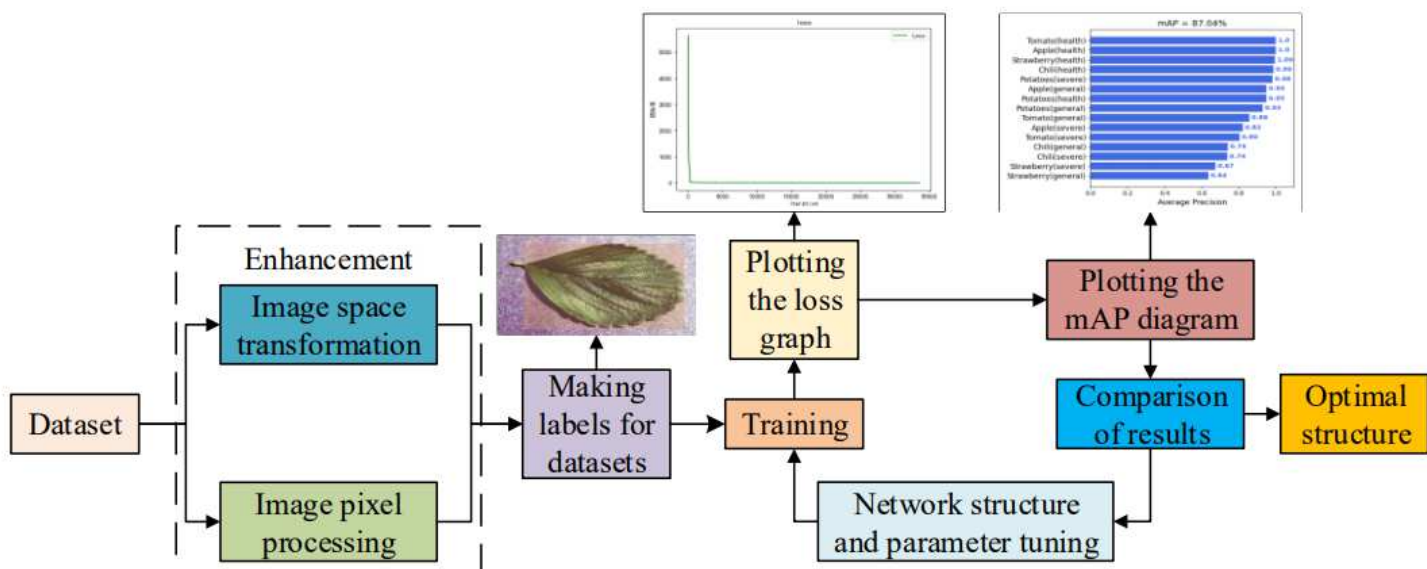
Figure 9

Experimental flow
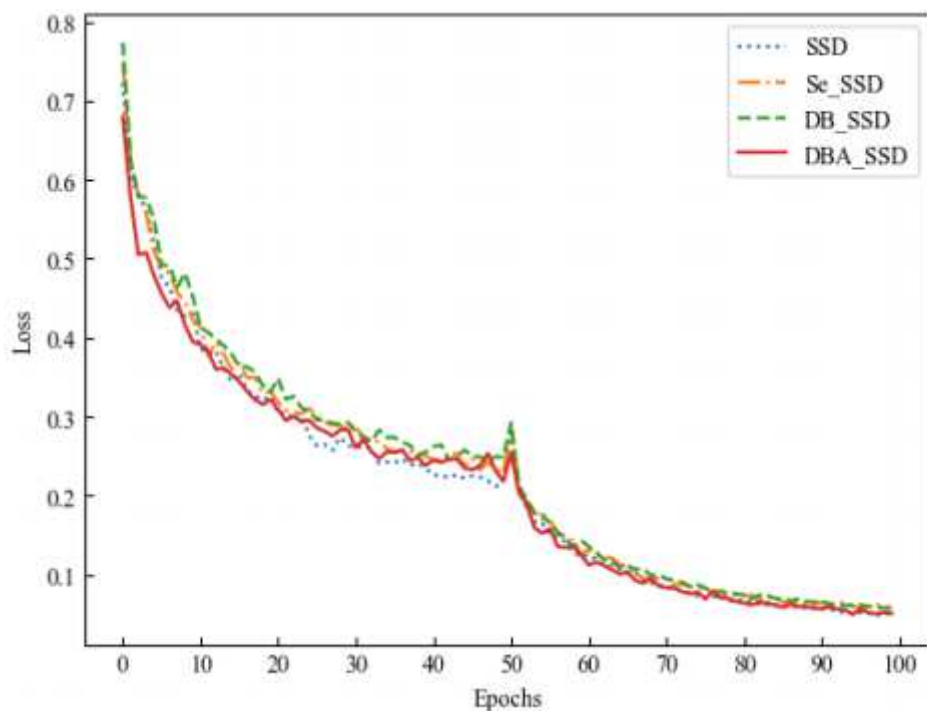
Figure 10

SSD and its improved algorithm loss variation graph
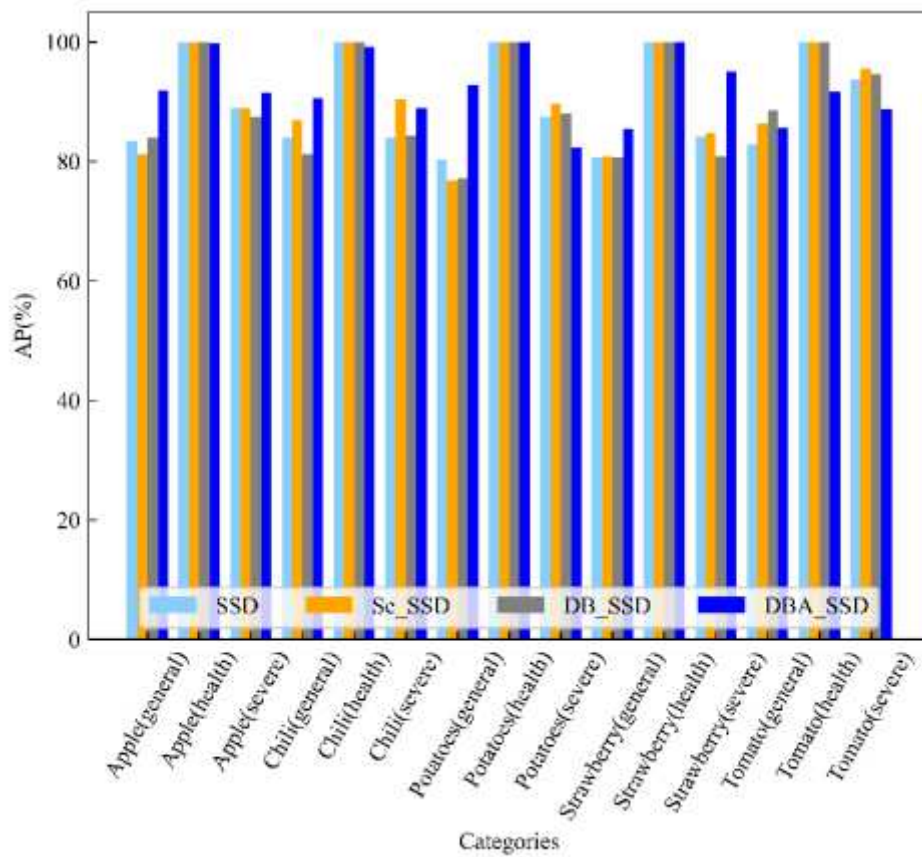
Figure 11

AP diagram of SSD and its improved algorithm for the detection of different kinds of diseases
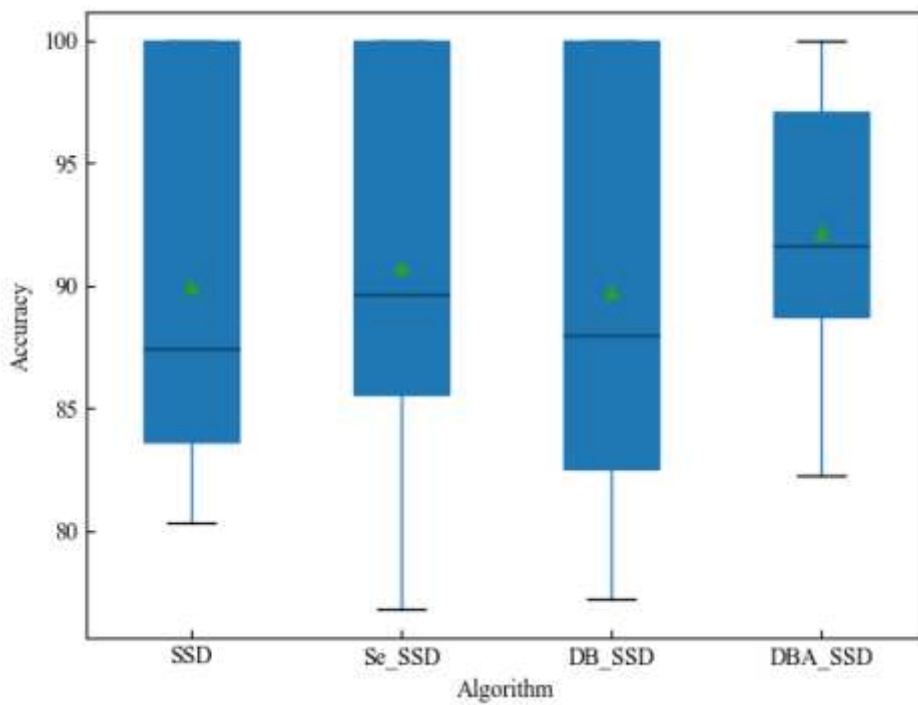
Figure 12

Box diagram of SSD and its improvement algorithm
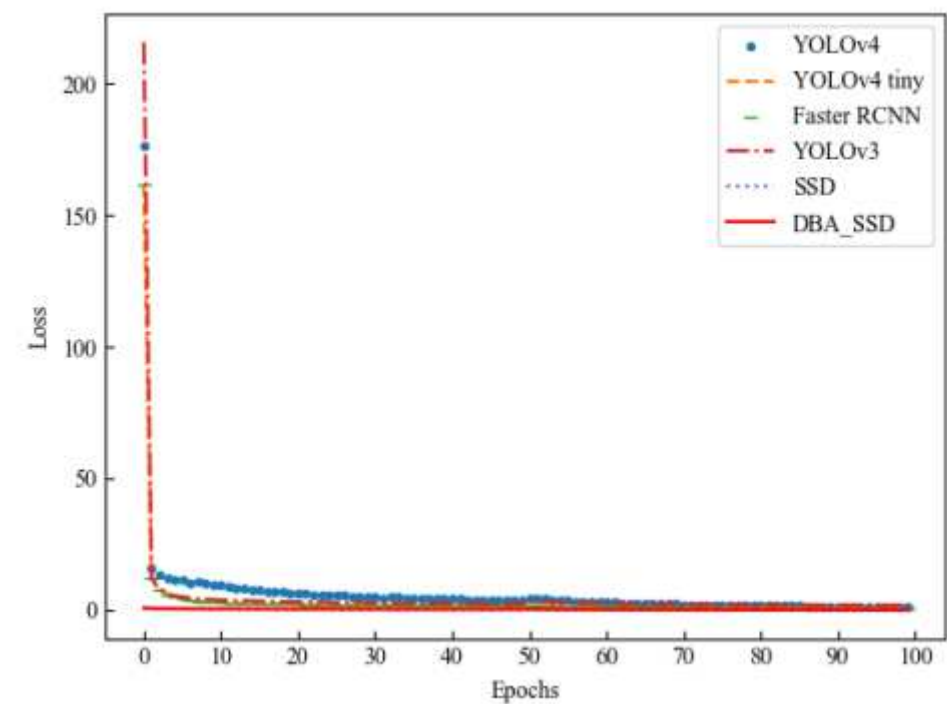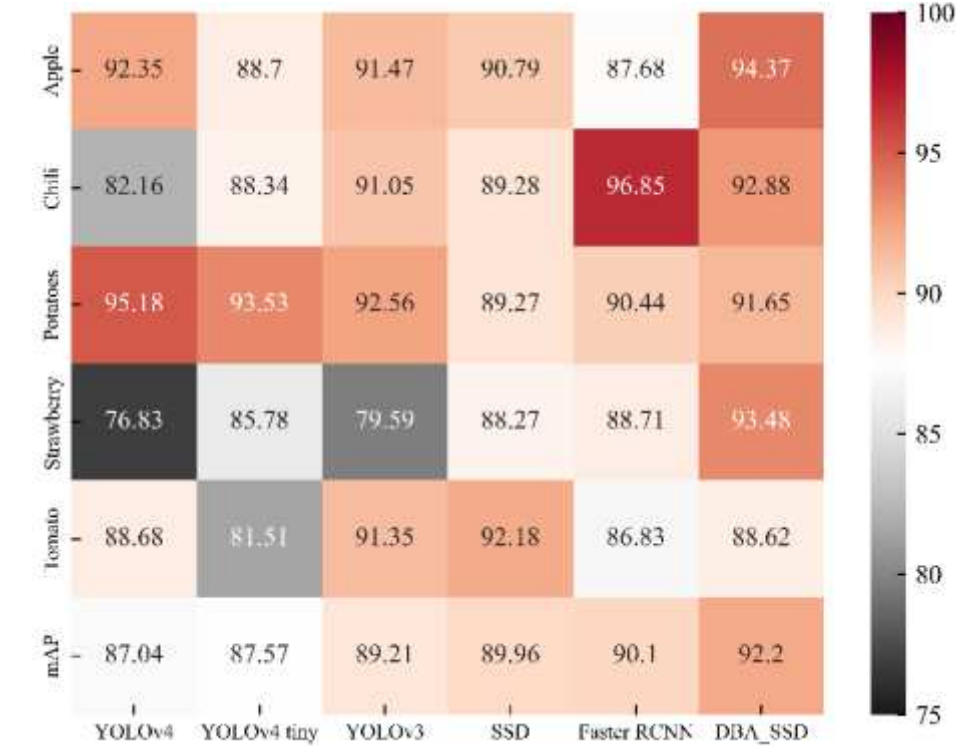

Figure 13

Target detection algorithm loss diagram

Figure 14

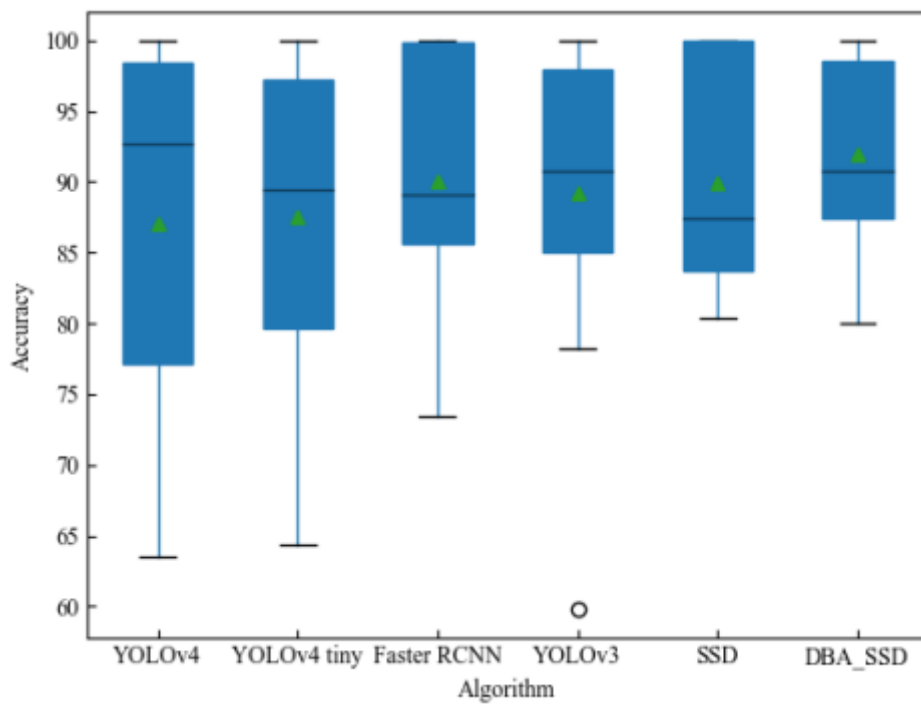Heat map of correlation between different target detection algorithms and vegetable and fruit leaf types

Figure 15

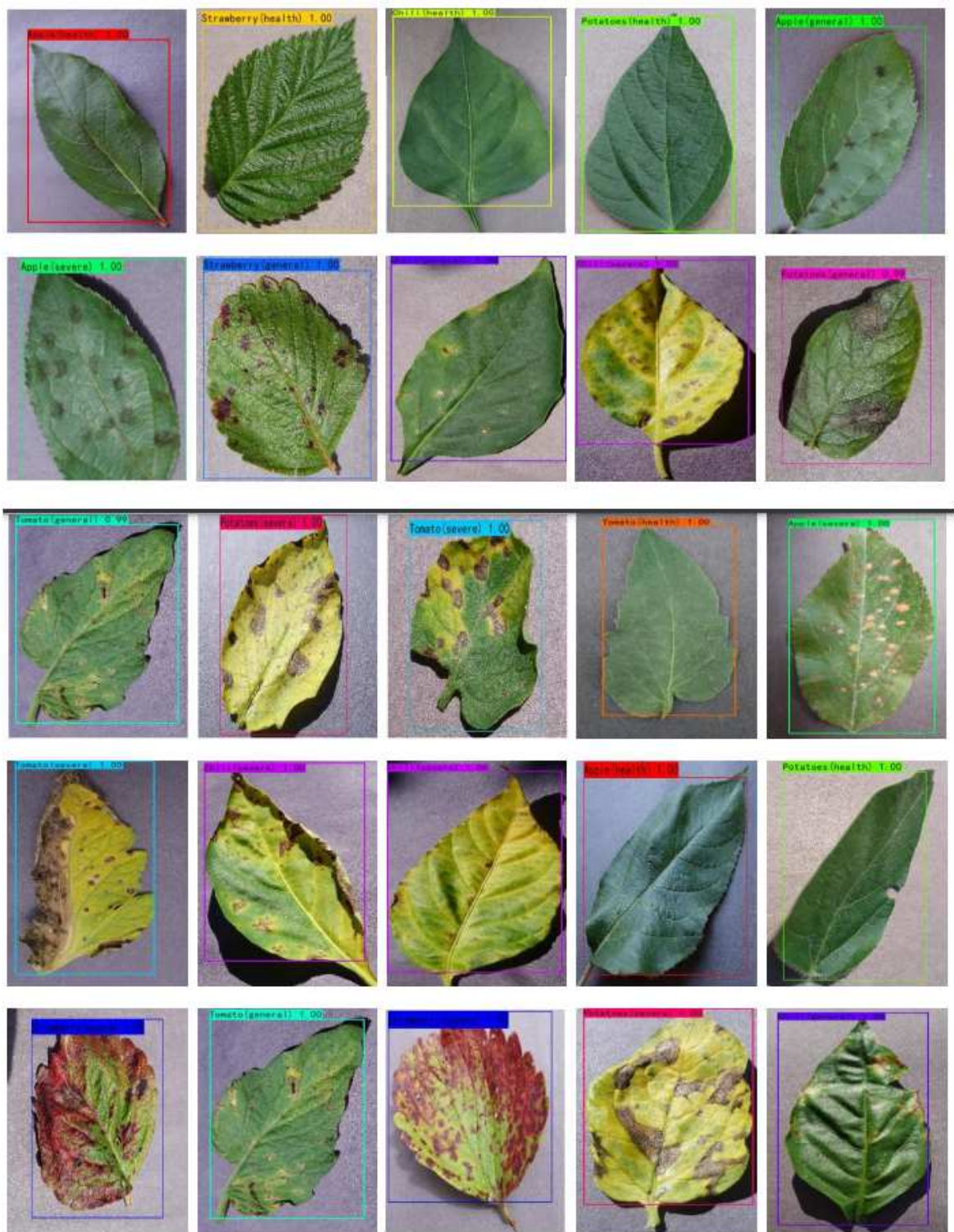Box plot of AP statistics under target detection algorithm

Figure 16

DBA_SSD recognition effect