

Grape leaf disease detection based on attention mechanisms

Wenjuan Guo^{1,2}, Quan Feng^{1*}, Xiangzhou Li¹, Sen Yang¹, Junqi Yang¹

(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)

Abstract: Prevention and control of grape diseases is the key measure to ensure grape yield. In order to improve the precision of grape leaf disease detection, in this study, Squeeze-and-Excitation Networks (SE), Efficient Channel Attention (ECA), and Convolutional Block Attention Module (CBAM) attention mechanisms were introduced into Faster Region-based Convolutional Neural Networks (R-CNN), YOLOx, and single shot multibox detector (SSD), to enhance important features and weaken unrelated features and ensure the real-time performance of the model in improving its detection precision. The study showed that Faster R-CNN, YOLOx, and SSD models based on different attention mechanisms effectively enhanced the detection precision and operation speed of the models by slightly enhancing parameters. Optimal models among the three types of models were selected for comparison, and results showed that Faster R-CNN+SE had lower detection precision, YOLOx+ECA required the least parameters with the highest detection precision, and SSD+SE showed optimal real-time performance with relatively high detection precision. This study solved the problem of difficulty in grape leaf disease detection and provided a reference for the analysis of grape diseases and symptoms in automated agricultural production.

Keywords: disease detection, Faster R-CNN, YOLOx, SSD, attention mechanism

DOI: 10.25165/j.ijabe.20221505.7548

Citation: 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.

1 Introduction

Crop disease control is the best approach to the production of pollution-free vegetables, reducing loss and pesticide application in crop production, therefore, early prediction and prevention of disease are of key importance^[1]. Affected by the environment, the grape may suffer from powdery mildew, brown blotch, and anthracnose, which seriously affect its yield and quality. Traditional grape disease detection method totally relies on planters' experiences or experts' guidance, having the defects of low speed and efficiency and poor real-time performance. Since the disease-infected grape may have some spots on its leaves, image processing on grape leaves is generally used to identify and detect such diseases to provide guidance^[2].

With the rapid development of artificial intelligence technology, vision techniques are widely used in the field of image processing of crop diseases^[3], such as *K*-means clustering^[4], Bayesian classification^[5], support vector machines^[6], genetic algorithms^[7], radial basis functions^[8], ensemble learning^[9], and filter segmentation methods^[10], which are used in crop disease classification studies. However, crop disease classification and identification based on traditional methods require manual

selection of features and are subject to environmental factors. The development of deep learning, especially the upgrading of Convolutional Neural Network (CNN) has greatly improved the automatic detection and identification technology of crop diseases.

In recent years, an object detection algorithm based on CNN has presented its great advantage. Object detection algorithm is widely applied in face recognition^[11], navigation^[12], road obstacle detection^[13], pedestrian detection^[14], fruit detection^[15,16], and weed detection^[17]. CNN can extract the high-dimensional image features of an object, and make it possible to detect crop leaf diseases by object detection algorithm under complicated background. Hence, experts in China and abroad have conducted model researches on crop disease detection based on object detection algorithms. For example, Fuentes et al.^[18] completed object detection on the tomato disease dataset with the Faster Region-based Convolutional Neural networks (R-CNN), Region-based Fully Convolutional Networks (R-FCN), and Single Shot Multibox Detector (SSD) models, and concluded that the combined model of Faster R-CNN and VGG16 showed the highest disease detection rate. Qiao et al.^[19] detected the time serial images of grape leaves by Faster R-CNN, and realized dynamic detection of grape leaf diseases. Li et al.^[20] applied the improved Faster R-CNN in bitter melon leaf disease detection, and the model proved high robustness and efficiency. Ye et al.^[21] used the SSD model to realize crop disease detection based on self-built data set, achieving an average detection precision of 83.90%. Liu et al.^[22] proposed an improved model based on MobileNetv2 and YOLOv3, which conducted early detection of grey speck disease of tomato. The improved model has the advantages of small memory size, high detection precision, and fast identification.

The studies above demonstrate the feasibility of detecting grape leaf diseases through object detection technology, however, the operation speed of existing models is low with low detection

Received date: 2022-03-26 **Accepted date:** 2022-07-19

Biographies: Wenjuan Guo, PhD candidate, research interest: image processing and recognition, Email: 565105996@qq.com; Xiangzhou Li, PhD candidate, research interest: image processing and recognition, Email: 572395298@qq.com; Sen Yang, PhD candidate, research interest: image processing and recognition, Email: 1535361383@qq.com; Junqi Yang, Master candidate, research interest: image processing and recognition, Email: 3510945819@qq.com.

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

precision, which seriously restricts the application of current grape disease detection technologies. In order to further improve the efficiency and precision of grape disease detection, in this study, the attention mechanisms of Squeeze-and-Excitation Networks (SE), Efficient Channel Attention (ECA), and Convolutional Block Attention Module (CBAM) were integrated into the grape disease detection models of Faster R-CNN, YOLOx, and SSD, respectively, to strengthen concern for the diseases and improve the performance of the feature extraction network. Experiments were carried out on the self-built grape disease dataset, and results showed that the Faster R-CNN, YOLOx and SSD models based on different attention mechanisms effectively enhanced the detection precision and operation speed by slightly increasing parameters. The optimal models were selected among the three types of models for comparison, and results show that the Faster R-CNN+SE model presented low detection precision; the YOLOx+ECA model had the least parameters but achieved the highest detection precision; SSD+SE had the best real-time performance with high detection precision. This research can provide a reference for the scientific formulation of prevention strategies for grape diseases.

2 Materials and methods

2.1 Experimental data

The dataset in this study comes from the captured grape leaf disease images, and the samples in the dataset are the images of grape leaf diseases shot in the complicated background of real field environment. The self-built dataset includes 2300 images of grape leaf diseases, covering six kinds of grape diseases, i.e., powdery mildew, anthracnose, brown spot disease, gray mold, black pox, and downy mildew, as shown in Figure 1.

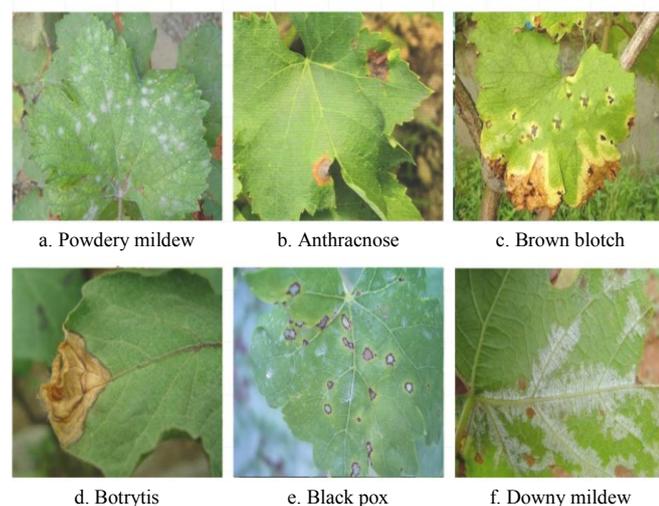


Figure 1 Samples of six types of grape disease

In this study, Gauss noise and random luminance methods are used to randomly process partial images to increase the richness of data samples in different sharpness and different light and shade conditions. In this study, 2300 images of grape leaf disease were precisely labeled by LabelImg software, and the data set of PASCAL VOC format was generated. In model training, the dataset was divided into a training set, a validation set, and a test set. In this study, based on the 9:1 proportion, 2070 images of the training set and validation set and 230 images of the test set were generated. Subsequently, based on the 9:1 proportion, 1863 images of the training set and 207 images of the test set were generated. The numbers of pictures of various diseases in the dataset are listed in Table 1.

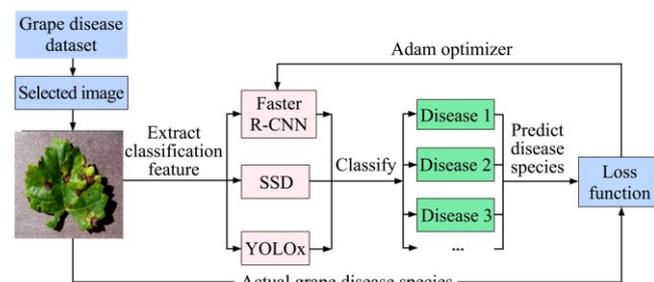
Table 1 The number of images of various diseases

Disease types	Graphics	Training set	Validation set	Test set
Powdery mildew	350	283	32	35
Anthracnose	385	311	35	39
Brown blotch	410	332	37	41
Botrytis	370	300	33	37
Black pox	405	329	36	40
Downy mildew	380	308	34	38

2.2 Models for grape leaf disease detection

There are mainly two types of object detection based on CNN, one is the object detection method based on the regional proposal, which first obtains the candidate region, then divides the candidate region, namely, two-stage object detection, such as R-CNN^[23], Fast R-CNN^[24] and Faster R-CNN^[25]. The other is the method without regional proposal, which is also called one-stage object detection. The method without regional proposal predicts the object's position and properties by CNN on the whole image, and its typical algorithms include SSD^[26] and YOLO series algorithm^[27-31].

In this study, the Faster R-CNN model, YOLOx model, and SSD model were taken as the models for grape leaf disease detection, and the training flow chart of the models for grape leaf disease detection is shown in Figure 2. Firstly, the selected grape leaf disease images were input. Secondly, the classification features were extracted. Finally, the Faster R-CNN model, YOLOx model, and SSD model were used to detect the disease, and the detection results were output. In the whole process, the loss function was calculated by predicting the difference between the disease species and the actual disease species, and the Adam optimizer was used to optimize the final output result.



Note: Faster R-CNN: Faster Region-based Convolutional Neural Network; SSD: Single Shot MultiBox Detector.

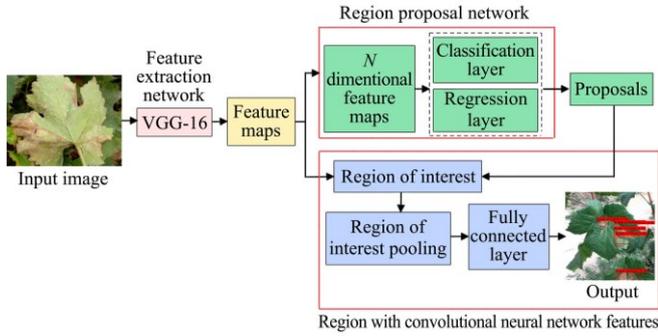
Figure 2 Training flow chart of grape leaf disease detection model

The Faster R-CNN model achieves end-to-end target detection, and the algorithm has high detection accuracy but low running speed; The YOLOx model runs fast, but it is not suitable for small target detection; The SSD model has higher detection accuracy and faster running speed, but algorithm training is excessively dependent on experience, and the performance of small target detection is still not as good as Faster R-CNN model. The three models are described as follows:

1) Faster R-CNN model: the model is composed of Feature Extraction Network, Region Proposal Network (RPN), and Region with Convolutional Neural Network Features (R-CNN), and its framework diagram is shown in Figure 3. Grape leaf disease detection based on Faster R-CNN mainly includes the following four aspects: generation of the candidate region of grape leaf diseases, extraction of disease characteristics, disease categorization, and bounding box regression.

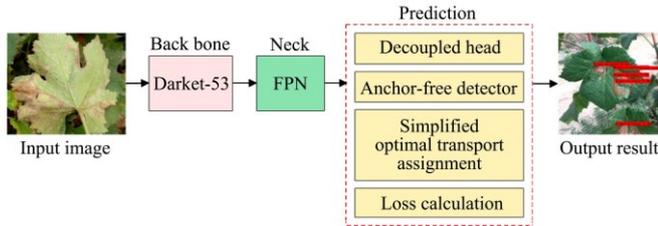
2) YOLOx model: this model has the advantage of high operation speed and flexibility, and the YOLOx-Darknet53 network was studied. The framework diagram of the

YOLOx-Darknet53 is shown in Figure 4, which includes input end, Backbone network, Neck, and Prediction. Compared with other YOLO series models, the YOLOx model updates YOLO Head into Decoupled Head in the part of Prediction, and updates Anchor Based method into Anchor Free method, at the same time, it adds the SimOTA method to do dynamic matching with the positive samples. The updates above help improve the detection precision and speed of the models and effectively reduce the parameters of the models.



Note: VGG: Visual Geometry Group.

Figure 3 Framework diagram of Faster R-CNN



Note: FPN: Feature Pyramid Networks.

Figure 4 Framework diagram of YOLOx-Darknet53

3) SSD model: the model learns from the anchors mechanism of the Faster R-CNN model and the regression mechanism of YOLO model, with the help of a small convolution kernel and multi-dimensional feature prediction methods, it has fast detection speed and high detection precision. The framework diagram of SSD algorithm has two parts, as shown in Figure 5. The first part is the front-end deep learning network model, which is used to extract the initial characteristics of the disease object and helps to improve the model's ability to perceive the disease. The second part is the back-end multi-scale feature detection network, which uses Cascaded neural networks to classify features of different scales to obtain the category and location information of the disease, and then adds the features of the low-layer convolutional layer to improve the detection precision of the model, and finally uses non-maximum suppression (NMS) to obtain the final detection results.

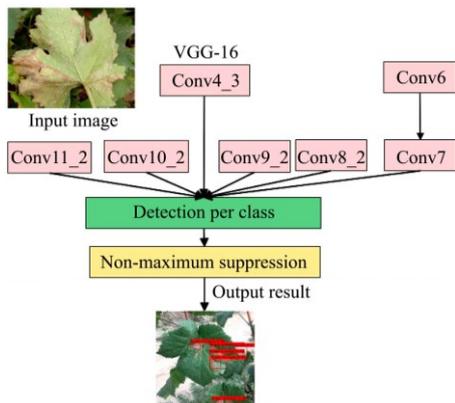
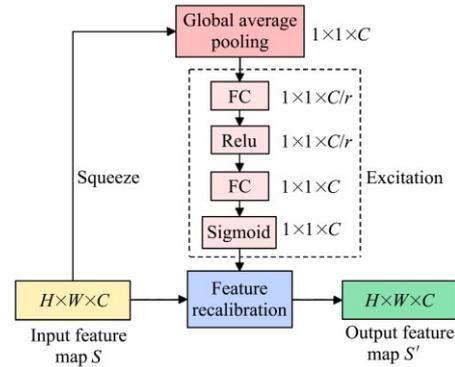


Figure 5 Framework diagram of SSD

2.3 Attention mechanism models

In this study, SE channel attention mechanism, ECA efficient channel attention mechanism, and CBAM spatial attention mechanism are adopted. The SE attentional mechanism model has the advantages of low complexity, fewer new parameters, and computation; The ECA attention mechanism is an improved version of SE attention mechanism, which is a lightweight channel attention module, the module adds little model complexity and has a significant improvement effect; The CBAM attention mechanism can improve network performance more effectively by connecting spatial domain and channel domain.

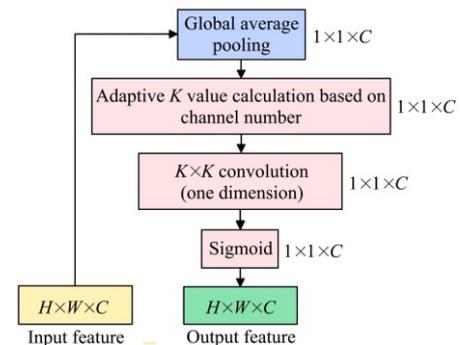
1) SE channel attention mechanism. SE channel attention mechanism extracts feature through the channel of CNN^[32]. Based on the method of feature recalibration, let the model do self-learning to capture the important information of each feature channel. SE includes two processes, squeeze, and excitation, and its network structure is shown in Figure 6. The squeeze process compresses the feature image based on spatial dimensions, while the excitation process is building a model on the correlation between feature channels after squeeze to obtain the importance of each channel, then excite the original feature images into corresponding channels.



Note: SE: Squeeze-and-Excitation; FC: Fully Connected; ReLU: Rectified Linear Unit; H represents the height of images; W represents the width of images; C represents the number of channels of images; r is the reduction ratio of the spatial dimensions of images.

Figure 6 SE network structure diagram

2) Efficient channel attention (ECA) mechanism. The efficient channel attention mechanism^[33] is an updated version of SE and it 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 efficient channel attention mechanism is shown in Figure 7.

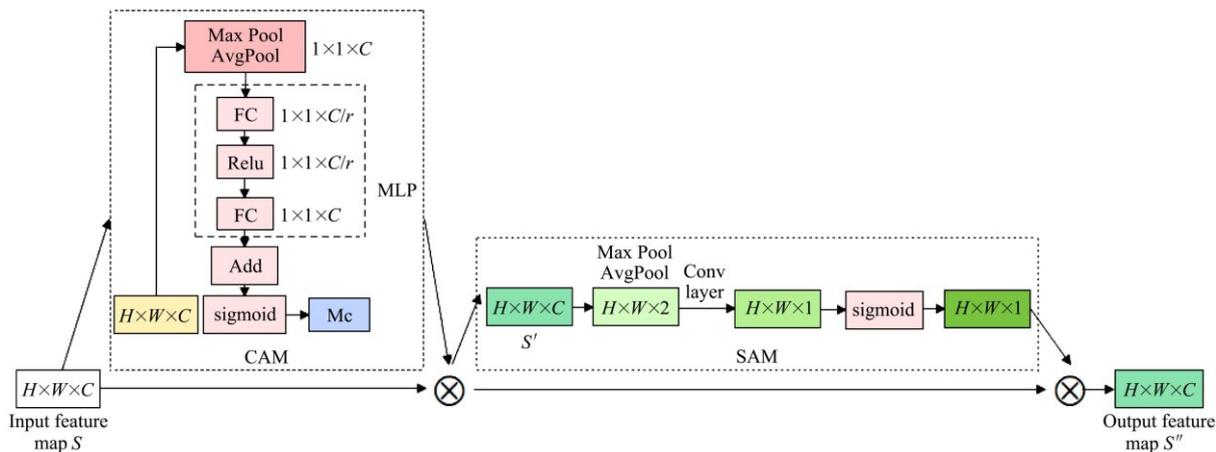


Note: ECA: Efficient Channel Attention.

Figure 7 ECA network structure diagram

3) CBAM Spatial Attention Mechanism. CBAM Spatial Attention Mechanism^[34] is composed of Channel Attention Module (CAM) and Spatial Attention Module (SAM). For the input feature map, CBAM infers the attention map on the channel and spatial dimensions and then multiplies the attention map and the

imported feature map to achieve optimization of self-adaptive features. The CBAM attention mechanism can enhance useful features in the input feature map while suppressing useless features and is widely used in practice. The network structure of CBAM is shown in Figure 8.



Note: CBAM: Convolutional Block Attention Module; CAM: Channel Attention Module; SAM: Spatial Attention Module; Conv: Convolution; MLP: Multilayer Perceptron. Mc is the channel attention map.

Figure 8 CBAM network structure diagram

2.4 Grape leaf disease detection model based on attention mechanisms

The introduction of SE attention mechanism in the disease detection model is more concerned about the channel features with the largest amount of information by suppressing the unimportant channel features; the introduction of ECA attention mechanism achieves appropriate cross-channel interaction, significantly reducing the complexity of the model while maintaining good performance; the introduction of CBAM attention mechanism makes the disease detection model consider the importance of different pixels and the importance of pixels in different positions in the same channel. All three of these attention mechanisms can be seamlessly integrated into Section 3.1 grape leaf disease detection model, enabling end-to-end training.

1) Faster R-CNN model based on different attention mechanisms

The Faster R-CNN model uses CNN to extract disease features to get the feature image. Due to the inherent locality of the convolution kernel, it can only retain local information of disease images rather than global information, causing information missing and reducing the detection precision of the Faster R-CNN model. To solve this problem and use the weight of transfer learning, under the premise of not changing feature extraction network structure by

backbone features, SE, ECA, and CBAM attention mechanisms were introduced by forwarding propagation after the last Identity block to improve the model, to obtain the feature information that has high contribution rate to disease object in the images. The framework diagram of the Faster R-CNN model based on different attention mechanisms is shown in Figure 9.

2) YOLOx model based on different attention mechanisms

Although YOLOx model has high detection speed with high detection precision, it has some disadvantages if it is directly applied to disease detection under complicated background. For example, the backbone network has insufficient ability in extracting features and cannot effectively integrate high-quality contextual feature information, thus reducing the detection precision of the model. Therefore, without changing the Darknet53 network structure of the YOLOx model, so that the weight of pre-training can be loaded directly in model training, in this study, based on the three output branches of the Backbone network Darknet53, SE, ECA, and CBAM attention mechanisms, the feature expression of each branch was enhanced to make the YOLOx model can selectively strengthen key features and effectively restrain useless features. The framework diagram of YOLOx model based on different attention mechanisms is shown in Figure 10.

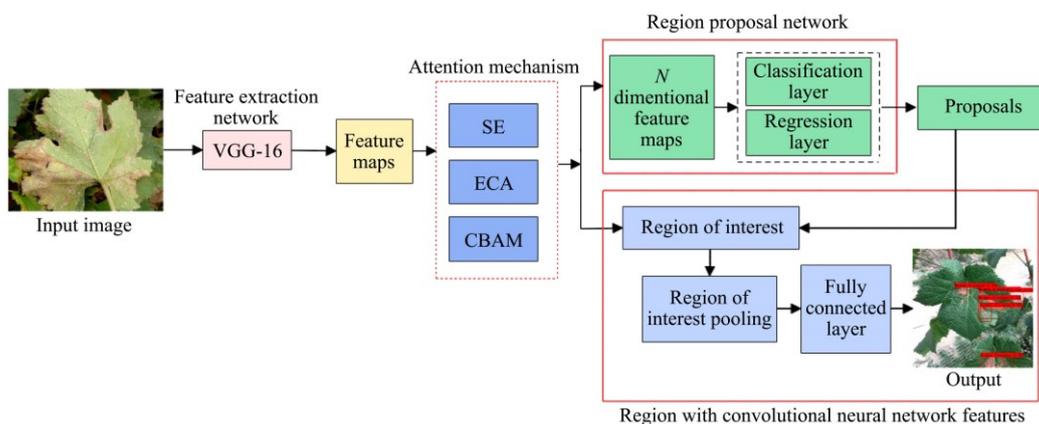


Figure 9 Framework diagram of Faster R-CNN model based on different attention mechanisms

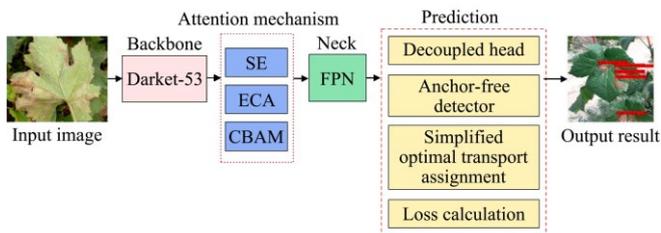


Figure 10 Framework diagram of YOLOx model based on different attention mechanisms

3) SSD model based on different attention mechanisms

It can be known from Section 3.1 that, the SSD model adopts multi-dimensional prediction method, and the front-end deep learning network detects small objects, while the rear-end multi-dimensional feature detection network detects big objects. The front-end deep learning network includes abundant geometric information and accurate positioning information, but the receptive field is small with the weak representational ability of semantic information. Relatively speaking, the rear-end multi-dimensional feature detection network has a relatively broad receptive field and semantic information with strong representational ability, however, the network resolution is low, and the representational ability in geometric information is low. Therefore, there might be information missing or errors in detecting diseases by SSD model. 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 disease objects. The framework diagram of SSD model based on different attention mechanisms is shown in Figure 11.

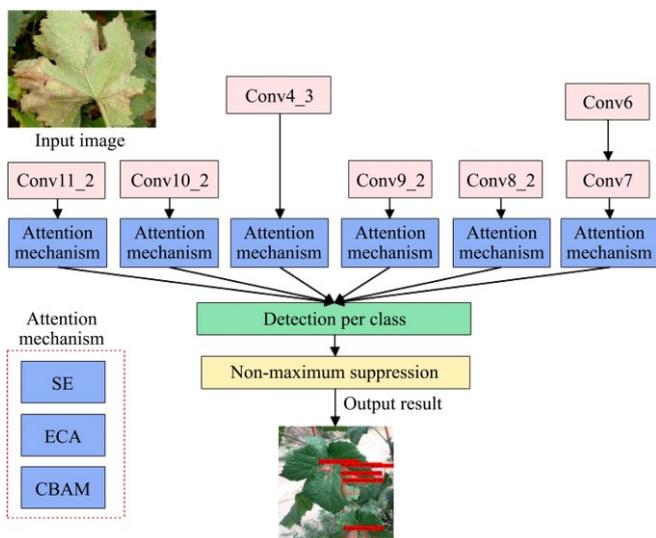


Figure 11 Framework diagram of SSD model based on different attention mechanisms

3 Experiment and analysis

3.1 Evaluation indexes

In this study, the commonly used target detection evaluation criteria are used to evaluate the detection results. Commonly used evaluation criteria include Precision (P), Recall (R), P-R curve, Average Precision (AP) for a single class of targets, and Mean Average Precision (mAP) for all classes. In this study, P , R , comprehensive evaluation index F1 value (F1 value is equivalent to the harmonic mean of the precision and recall), mAP, Frames Per

Second (FPS), and parameters were adopted to evaluate the detection results. rape leaf disease detection model.

The calculation equations of P , R , and F1 are Equations (1)-(3), respectively.

$$P = \frac{TP}{TP + FP} \times 100\% \quad (1)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

$$F1 = \frac{2 \times P \cdot R}{P + R} \quad (3)$$

where, P is the precision of the results, %; TP is the probability that positive samples are correctly detected; FP is the probability that negative samples are detected as positive samples; R is the recall of the results, %; FN is the probability that positive samples are detected as negative samples; F1 is the equivalent to the harmonic mean of the precision and recall, %. The higher TP value is, the more accurate the prediction and the better the performance of the model.

mAP (%) is the result 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).

$$AP = \int_0^1 PRdR \quad (4)$$

$$mAP = \frac{1}{N} \sum_{m=1}^N AP_m \quad (5)$$

where, N is the number of kinds of diseases, $N=6$; 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 identification speed of the algorithm.

3.2 Experiment platform and parameter setting

The experiment uses the Windows 10 operating system, the computer is equipped with 16 GB 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

3.3 Experiment results and analysis

Keeping the configuration information and training platform unchanged, the Faster R-CNN model, the YOLOx model, and the SSD model based on different attention mechanisms were compared with the classical models, Faster R-CNN, YOLOx, and SSD, all of which were trained and detected on the same grape disease dataset.

1) Experimental result analysis of the Faster R-CNN model based on different attention mechanisms

The Faster R-CNN model integrated with SE attention mechanism is simply called Faster R-CNN+SE, the Faster R-CNN model integrated with ECA attention mechanism is called Faster R-CNN+ECA, and the Faster R-CNN integrated with CBAM attention mechanism is called Faster R-CNN+CBAM. In the

same experiment environment, Faster R-CNN+SE, Faster R-CNN+ECA, Faster R-CNN+CBAM, and Faster R-CNN were adopted for disease detection on the grape disease dataset, the experiment results are listed in Table 3.

Table 3 Performance comparison of Faster R-CNN models based on different attention mechanisms

Model	P/%	R/%	F1/%	mAP/%	FPS	Size of parameters/MB
Faster R-CNN	75.06	74.42	74.74	79.12	13.16	522.91
Faster R-CNN+SE	79.80	84.23	81.96	85.39	13.22	523.04
Faster R-CNN+ECA	76.54	78.71	77.61	81.93	13.18	522.91
Faster R-CNN+CBAM	75.75	75.89	75.82	79.65	13.20	523.41

Note: *P* is the precision; *R* is the recall; F1 is the harmonic mean of the precision and recall; FPS is the frames per second. Faster R-CNN: Faster Region-based Convolutional Neural Network; SE: Squeeze-and-Excitation; ECA: Efficient Channel Attention; CBAM: Convolutional Block Attention Module.

It is shown from Table 3 that, compared with Faster R-CNN, the *P*, *R*, and F1 values of the Faster R-CNN+SE model were increased by 4.74%, 9.81%, and 7.22%, mAP increased by 6.27%, FPS increased by 0.06, the increase of attention modules on the network structure increased the number of parameters by 0.13 MB. Compared with Faster R-CNN, the *P*, *R*, and F1 values of the Faster R-CNN+ECA model increased by 1.48%, 4.29%, and 2.87%, respectively, mAP increased by 2.81%, FPS value increased by 0.02, and parameters kept unchanged. Compared with Faster R-CNN, the *P*, *R*, and F1 values of the Faster R-CNN+CBAM increased by 0.69%, 1.47%, and 1.08%, respectively, mAP increased by 0.53%, and FPS value increased by 0.04, and the added attention module increased parameters by 0.5 MB.

It can be seen from the above analysis that, although the parameters of Faster R-CNN+SE and Faster R-CNN+CBAM increased slightly, the performance of the three models after the introduction of attention mechanisms was better than that of the original Faster R-CNN. The reason is the introduction of attention mechanism can help obtain the feature information with high contribution rates to the disease object in the disease images, improve the detection precision, and accelerate the detection speed.

Among the three models introduced with attention mechanisms, Faster R-CNN+SE showed the optimal detection effect. Compared with Faster R-CNN+ECA, the *P*, *R*, and F1 values of the Faster R-CNN+SE model increased by 3.26%, 5.52%, and 4.35%, respectively, mAP increased by 3.46%, FPS value increased by 0.04, parameters increased by 0.13 MB. Compared with Faster R-CNN+CBAM, the *P*, *R*, and F1 values of the Faster R-CNN+SE increased by 4.05%, 8.34%, and 6.14%, respectively, the FPS value increased by 0.18, and parameters reduced by 0.37 MB.

Comprehensively considering the detection precision and operation speed of the models, the Faster R-CNN+SE model presented the optimal robustness by slightly increasing parameters, and it pays more attention to the channel features with the largest amount of information and suppresses unimportant channel features, showing ideal detection effect in grape disease dataset.

2) Experimental result analysis of the YOLOx model based on different attention mechanisms

The YOLOx model introduced with SE attention mechanism is YOLOx+SE for short; the YOLOx model introduced with ECA attention mechanism is YOLOx+ECA for short; the YOLOx model introduced with CBAM attention mechanism is YOLOx+CBAM for short. Under the same experimental environment, YOLOx+SE, YOLOx+ECA, YOLOx+CBAM, and YOLOx were

adopted to detect the diseases on the grape disease dataset, and the experimental results are shown in Table 4.

Table 4 Performance comparison of Faster YOLOx models based on different attention mechanisms

Model	P/%	R/%	F1/%	mAP/%	FPS	Size of parameters/MB
YOLOx	82.35	74.85	78.42	83.22	68.34	34.21
YOLOx+SE	82.46	82.21	82.33	84.02	68.88	34.38
YOLOx+ECA	87.77	86.07	86.91	88.66	73.68	34.87
YOLOx+CBAM	85.81	77.91	81.67	84.21	71.88	34.21

As can be seen from Table 4, compared with YOLOx, the *P*, *R*, and F1 values of the YOLOx+SE model increased by 0.11%, 7.36%, and 3.91%, respectively, the mAP increased by 0.8%, the FPS value increased by 0.54, and the increase of the attention module on the network structure increased the parameter amount by 0.17 MB; compared with YOLOx, the precision *P*, recall *R*, and F1 values of the YOLOx+ECA model increased by 5.42%, 11.22%, and 8.49%, respectively. mAP increased by 5.44%, FPS value increased by 5.34, and parameters increased by 0.66 MB. Compared with YOLOx, the *P*, *R*, and F1 values of the YOLOx+CBAM model increased by 3.46%, 3.06%, and 3.25%, respectively, mAP increased by 0.99%, FPS value increased by 3.54, and the parameters kept unchanged.

From the above analysis, it can be obtained that although the parameters of the YOLOx+SE model and the YOLOx+ECA model have increased slightly, the detection indexes of the three YOLOx models introduced with attention mechanisms were higher than those of the original YOLOx. The reason is that the introduction of attention mechanisms enabled the models to extract more comprehensive and rich features, and the model paid more attention to disease objects, thus increasing detection precision.

Among all models, the YOLOx+ECA model proved the optimal detection effect. Compared with YOLOx+SE, the *P*, *R*, and F1 values of the YOLOx+ECA model increased by 5.31%, 3.86%, and 4.58%, respectively; mAP increased by 4.64%, FPS value increased by 4.8, and parameters expanded by 0.49 MB. Compared with YOLOx+CBAM, the *P*, *R*, and F1 values of the YOLOx+ECA model increased by 1.96%, 8.16%, and 5.24%, respectively, mAP increased by 4.45%, FPS value increased by 1.8, parameters expanded by 0.66 MB.

In conclusion, compared with the other three models, although the YOLOx+ECA model had more parameters, it realized cross-channel interaction on the grape disease dataset to some extent and could achieve optimal detection results at a fast operation speed.

3) Experimental results of the SSD model based on different attention mechanisms

The SSD model introduced with SE attention mechanism is SSD+SE for short; the SSD model introduced with ECA attention mechanism is SSD+ECA for short; the SSD model introduced with CBAM attention mechanism is SSD+CBAM for short. Keeping the experimental conditions unchanged, the SSD+SE, SSD+ECA, SSD+CBAM, and SSD models were used to detect diseases on the grape disease dataset, and the experiment results are listed in Table 5.

Table 5 shows that, compared with SSD, the *P*, *R*, and F1 values of the SSD+SE model increased by 2.72%, 15.23%, and 9.45%, respectively, mAP increased by 10.73%, FPS value increased by 63.08, parameters expanded by 0.85 MB. Compared with SSD, the *P*, *R*, and F1 values of the SSD+ECA model increased by 1.35%, 8.77%, and 5.47%, respectively, mAP

increased by 6.67%, FPS went up by 0.55, and parameters kept unchanged. Compared with SSD, the P , R and $F1$ values of the SSD+CBAM increased by 0.94%, 3.61%, and 2.48%, respectively; mAP increased by 4.91%, FPS value increased by 8.47, parameters expanded by 3.38 MB.

Table 5 Performance comparison of SSD models based on different attention mechanisms

Model	P /%	R /%	$F1$ /%	mAP/%	FPS	Size of parameters/MB
SSD	80.74	68.87	74.33	76.23	67.69	99.76
SSD+SE	83.46	84.10	83.78	86.96	130.77	100.61
SSD+ECA	82.09	77.64	79.80	82.90	68.24	99.76
SSD+CBAM	81.68	72.48	76.81	81.14	76.16	103.14

Note: SSD: Single Shot MultiBox Detector.

Therefore, the increase of attention modules in a network structure expanded the parameters of the SSD+SE and SSD+CBAM models, and the three models could effectively position the interesting information in the feature images based on the importance of the features and restrain useless information. Thus, other detection indexes of the three models introduced with attention mechanisms were better than that of the SSD model.

Among the four models, the SSD+SE model proved the optimal detection effect, and the detection speed of this model was significantly faster than the other three models, showing optimal real-time performance. Compared with SSD+ECA, the P , R , and $F1$ values of SSD+SE increased by 1.37%, 6.46%, and 3.98%, respectively, mAP increased by 4.64%, FPS value increased by 63.63, and parameters expanded by 0.85 MB. Compared with SSD+CBAM, the P , R , and $F1$ values of SSD+SE increased by 1.78%, 11.62%, and 6.97%, respectively, mAP increased by 5.82%, FPS increased by 71.55, while parameters reduced by 2.53 MB.

The experimental results above showed that, since SE attention mechanism could optimize feature images, in terms of both detection precision and speed, SSD+SE was significantly better than the other three models, and it proved better comprehensive performance. Thus, it can be applied in the real-time detection of grape diseases.

4) Comparison analysis of the detection effect of the three optimal models after screening

The analysis above shows that Faster R-CNN+SE was the optimal model of Faster R-CNN based on different attention mechanisms. YOLOx+ECA was the optimal model of YOLOx based on different attention mechanisms. SSD+SE was the optimal model of SSD based on different attention mechanisms. In order to present the disease detection performance of each model, images were selected randomly in the experiment. By keeping the experimental environment unchanged, the three optimal disease detection models were screened for comparison, and the results are shown in Figure 12.

It can be found after observing the detection results of Image2, Image3, and Image4 that, Faster R-CNN+SE and SSD+SE had cases of information leakage when the two diseases had high contact ratios. For example, Faster R-CNN+SE in Image2 missed the disease information on the upper left; Faster R-CNN+SE in Image3 and SSD+SE missed the disease information in the middle part; Faster R-CNN+SE in Image4 missed the disease information in small spots in the middle; Faster R-CNN+SE and SSD+SE in Image3 took brown blotch as anthracnose of grape. Compared with the other two models, YOLOx+ECA successfully detected small disease spots in Image3 and Image4, without any information missing or errors.

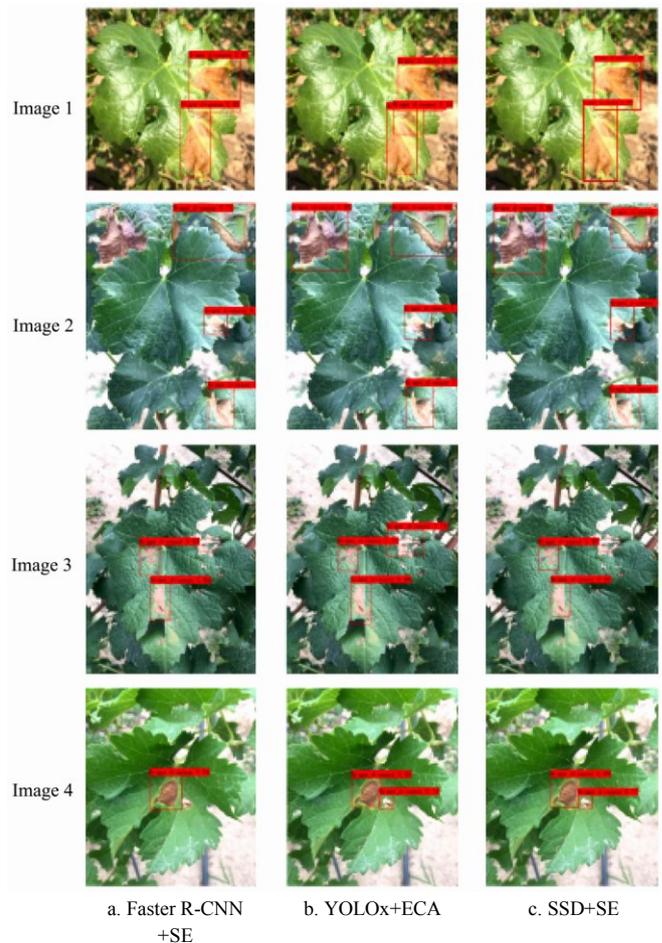


Figure 12 Comparison of different model results

In general, among the three models of Faster R-CNN+SE, YOLOx+ ECA, and SSD+SE, the detection precision of Faster R-CNN+SE was the lowest at a low operation speed with the most parameters; SSD+SE had the fastest operation speed with high precision, thus it can be applied in real-time detection of field grape diseases; YOLOx+ECA had the highest detection precision with the least parameters, and it effectively enhanced the detection rate of small objects and objects under occlusion, showing strong robustness.

4 Conclusions

1) In view of the low detection precision of the Faster R-CNN model, the three attention mechanisms of SE, ECA, and CBAM were introduced in this study on the basis of the original model. Experimental results showed that the P , R , and $F1$ values of Faster R-CNN+SE, Faster R-CNN+ECA, and Faster R-CNN+CBAM were all higher than that of Faster R-CNN. With the increase of the attention module, the parameters increased a little. Among the models above, Faster R-CNN+SE proved the optimal detection effect in the grape disease dataset.

2) In order to overcome the defect of low precision under different environments, three attention mechanisms of SE, ECA, and CBAM were introduced to the YOLOx model. Experimental results showed that the P , R , $F1$, mAP, and FPS values of YOLOx+SE, YOLOx+ECA, and YOLOx+CBAM after introduced with attention mechanisms were all higher than that of YOLOx, and the parameters increased slightly. Among the four models, YOLOx+ECA demonstrated the fastest speed and optimal performance.

3) In order to avoid information missing or errors in disease

detection, the attention mechanisms of SE, ECA, and CBAM were introduced to the SSD model. Experimental results showed that the P , R , $F1$, mAP , and FPS values of SSD+SE, SSD+ECA, and SSD+CBAM were all higher than that of SSD, and the parameters expanded slightly. Among the four models, the detection speed of SSD+SE was significantly faster than that of the other three models, and its detection performance was the best.

4) Comparison analysis was carried out on the three optimal models of Faster R-CNN+SE, YOLOx+ECA, and SSD+SE, and results showed that Faster R-CNN+SE had lower detection precision with more parameters; YOLOx+ECA had the least parameters but the highest detection precision; SSD+SE showed optimal real-time performance with relatively high detection precision.

Acknowledgements

The authors acknowledge that this work was financially supported by the National Natural Science Foundation of China (Grant No. 31971792), 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), School Level Youth Project of Gansu University of Political Science and Law (Grant No. GZF2019XQNLW08), Gansu Education Department Innovation Fund project (Grant No. 2022B-144).

References

- [1] Food and Agriculture Organization of the United Nations. Plant Diseases and pests. Available: <http://www.fao.org/emergencies/emergency-types/plant-pests-and-diseases/en/>. Accessed on [2020-02-05].
- [2] Carlson G A. A decision theoretic approach to crop disease prediction and control. *American Journal of Agricultural Economics*, 1970; 52(2): 216–223.
- [3] Wang X Y, Wen H J, Li X X, Fu Z T, Lyu X J, Zhang L X. Research progress analysis of mainly agricultural diseases detection and early warning technologies. *Transactions of the CSAM*, 2016; 47(9): 266–277. (in Chinese)
- [4] Wang Z B, Wang K Y, Pan S H, Han Y Y. Segmentation of crop disease images with an improved k -means clustering algorithm. *Applied Engineering in Agriculture*, 2018; 34(2): 277–289.
- [5] Bi C G, Chen G F. Bayesian networks modeling for crop diseases. In: 2010 International Conference on Computer and Computing Technologies in Agriculture, Computer and Computing Technologies in Agriculture IV (CCTA 2010), Springer, 2011; pp.312–320.
- [6] Zhang D Y, Chen G, Zhang H H, Jin N, Gu C Y, Weng S Z, et al. Integration of spectroscopy and image for identifying fusarium damage in wheat. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2020; 236: 118344. doi: 10.1016/j.saa.2020.118344.
- [7] Zhang C L, Zhang S W, Yang J C. Apple leaf disease identification using genetic algorithm and correlation based feature selection method. *Int J Agric & Biol Eng*, 2017; 10(2): 74–83.
- [8] Reza G, Jullada L, Daciana I. Plant pest and disease diagnosis. *Journal of Plant Diseases and Protection*, 2012; 119(5): 200–207.
- [9] Ferentinos, Konstantinos P. Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 2018; 145: 311–318.
- [10] Radhika K, Mamatha B, Srikanth P. Crop and weed discrimination using Laws' texture masks. *Int J Agric & Biol Eng*, 2020; 13(1): 191–197.
- [11] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New York: IEEE, 2015; pp.815–823. doi: 10.1109/CVPR.2015.7298682.
- [12] Yang W L, Fan S S, Xu S X, King P, Kang B, Kim E. Autonomous underwater vehicle navigation using sonar image matching based on convolutional neural network. *IFAC Papers On Line*, 2019; 52(21): 156–162.
- [13] Sayanan S, Mohan M T. Active learning for on-road vehicle detection: A comparative study. *Machine Vision and Applications*, 2014; 25(3): 599–611.
- [14] Zhang L L, Lin L, Liang X D, He K M. Is faster R-CNN doing well for pedestrian detection? *European Conference on Computer Vision*. Springer, 2016; pp.443–457.
- [15] Liu F, Liu Y K, Lin S, Guo W Z, Xu F, Zhang B. Fast recognition method for tomatoes under complex environments based on improved YOLO. *Transactions of the CSAM*, 2020; 51(6): 229–237. (in Chinese)
- [16] Long J H, Zhao C J, Lin S, Guo W Z, Wen C W, Zhang Y. Segmentation method of the tomato fruits with different maturities under greenhouse environment based on improved Mask R-CNN. *Transactions of the CSAE*, 2021; 37(18): 100–108. (in Chinese)
- [17] Quan L Z, Xia F L, Jiang W, Li H L, Li H D, Lou Z X, et al. Research on recognition of maize seedlings and weeds in maize field based on YOLO v4 convolutional neural network. *Journal of Northeast Agricultural University*, 2021; 52(7): 89–98. (in Chinese)
- [18] 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.
- [19] Qiao H, Feng Q, Zhang R, Liu T Y. Dynamic monitoring of grape leaf disease based on sequential images tracking. *Transactions of the CSAE*, 2018; 34(17): 167–175. (in Chinese)
- [20] 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)
- [21] Ye Z H, Zhao M X, Jia L. Image recognition of crop diseases in complex background. *Transactions of the CSAM*, 2021; 52(S1): 118–124,147. (in Chinese)
- [22] Liu J, Wang X. Early recognition of tomato gray leaf spot disease based on MobileNetv2-YOLOv3 model. *Plant Methods*, 2020; 16: 83. doi: 10.1186/s13007-020-00624-2.
- [23] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus: IEEE, 2014; pp.580–587. doi: 10.1109/CVPR.2014.81.
- [24] Girshick R. Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision, Santiago: IEEE, 2015; pp.1440–1448. doi: 10.1109/ICCV.2015.169.
- [25] Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 2015; pp.91–99.
- [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, 2016; pp.21–37.
- [27] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, real-time object detection. In: 2016 Proceedings of the IEEE conference on computer vision and pattern recognition, 2016; pp.779–788.
- [28] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. In: 2017 IEEE Conference on Computer Vision & Pattern Recognition (CVPR), Honolulu: IEEE, 2017; pp.6517–6525. doi: 10.1109/CVPR.2017.690.
- [29] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. arXiv: 1804.02767v1.
- [30] Bochkovskiy A, Wang C Y, Liao H. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv: 2004.10934.
- [31] Zheng G, Liu S T, Wang F, Li Z M, Sun J. YOLOx: Exceeding Yolo series in 2021. arXiv:2107.08430.
- [32] Hu J, Shen L, Albanie S, Sun G, Wu E H. Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020; 42(8): 2011–2023.
- [33] Wang Q L, Wu B G, Zhu P F, Li P H, Zuo W M, Hu Q H. Supplementary material for “ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks”. In: 2020 Proceeding of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Piscataway: IEEE, 2020; pp.11531–11539.
- [34] Woo S, Park J, Lee J Y, Kweon I S. CBAM: Convolutional Block Attention Module. *Springer European Conference on Computer Vision*. Munich: Springer, 2018; pp.3–19.