

Article

Detection of Small Lesions on Grape Leaves Based on Improved YOLOv7

Mingji Yang¹, Xinbo Tong¹ and Haisong Chen^{2,*}

¹ Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application, Harbin University of Science and Technology, Harbin 150080, China; yangmingji@hrbust.edu.cn (M.Y.); 2120600050@stu.hrbust.edu.cn (X.T.)

² School of Undergraduate Education, Shenzhen Polytechnic University, Shenzhen 518055, China

* Correspondence: hschen@szpt.edu.cn

Abstract: The precise detection of small lesions on grape leaves is beneficial for early detection of diseases. In response to the high missed detection rate of small target diseases on grape leaves, this paper adds a new prediction branch and combines an improved channel attention mechanism and an improved E-ELAN (Extended-Efficient Long-range Attention Network) to propose an improved algorithm for the YOLOv7 (You Only Look Once version 7) model. Firstly, to address the issue of low resolution for small targets, a new detection head is added to detect smaller targets. Secondly, in order to increase the feature extraction ability of E-ELAN components in YOLOv7 for small targets, the asymmetric convolution is introduced into E-ELAN to replace the original 3×3 convolution in E-ELAN network to achieve multi-scale feature extraction. Then, to address the issue of insufficient extraction of information from small targets in YOLOv7, a channel attention mechanism was introduced and improved to enhance the network's sensitivity to small-scale targets. Finally, the CIoU (Complete Intersection over Union) in the original YOLOv7 network model was replaced with SIoU (Structured Intersection over Union) to optimize the loss function and enhance the network's localization ability. In order to verify the effectiveness of the improved YOLOv7 algorithm, three common grape leaf diseases were selected as detection objects to create a dataset for experiments. The results show that the average accuracy of the algorithm proposed in this paper is 2.7% higher than the original YOLOv7 algorithm, reaching 93.5%.

Keywords: YOLOv7; small target detection; E-ELAN network; attention mechanism



Citation: Yang, M.; Tong, X.; Chen, H. Detection of Small Lesions on Grape Leaves Based on Improved YOLOv7. *Electronics* **2024**, *13*, 464. <https://doi.org/10.3390/electronics13020464>

Academic Editors: KC Santosh and George A. Tsihrintzis

Received: 11 December 2023

Revised: 14 January 2024

Accepted: 19 January 2024

Published: 22 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The United States Department of Agriculture (USDA) released the “World Fresh apple, grape and pear market and Trade Report”, which shows that the 2022/2023 season global table grape production is expected to increase by 1.1 million tons to 27.3 million tons, achieving the fourth consecutive year of growth; China's grape production will reach 12.6 million tons, the world's largest grape producer [1]. At the same time, the output of the European Union and Turkey has increased significantly. However, Australia suffered heavy rains in the spring, and the attack of bacteria on grapes was extremely serious, affecting the yield of grapes. The wet conditions of South African grapes during the growing season have increased the pressure of fungal diseases, which have had a significant negative impact on the harvest, while the peak heat in December and January has brought sunburn to the grapes, and the lack of irrigation power during the high summer temperatures has resulted in crop losses [2].

When grapes grow, they are often eroded by diseases, which not only reduce the fruit and yield of grapes but even make the vineyard suffer a devastating blow in serious cases. Because grapes have important nutritional, medicinal and economic value, improving the efficiency of prevention and control of grape leaf diseases is the key to improving fruit

quality and yield [3]. In the process of grape growth, leaf diseases are mostly small targets. Affected by the growth time, grape leaf disease spots have problems such as random locations and different sizes of disease spots. Some disease spots have small pixels, which makes it easy to miss inspection and leads to difficulties in manual identification. Therefore, accurately and efficiently identifying the types of grape leaf diseases and providing corresponding prevention and control strategies is of great significance for the development of the grape industry [4].

At present, grape leaf diseases are still mainly observed manually. Relevant personnel need to observe the disease spots on the leaves with the naked eye and determine whether it is a specific disease based on their own experience. The accuracy of disease types depends on the experience of relevant personnel. In addition, some of the diseased areas on grape leaves are very small, which not only increases the difficulty of detection but also greatly affects the discrimination ability of personnel, leading to an increase in the rate of false detection rate [5,6].

In recent years, the research and application of artificial intelligence technology have been in full swing, especially in the field of modern agricultural precision planting, which has aroused strong interest from many researchers. The use of artificial intelligence technology for the intelligent identification of crop diseases has also become an important research hotspot. The crops involved are diverse, such as grapes, tomatoes, potatoes and corn. The types of methods involved are also diverse, mainly divided into two categories: traditional machine learning methods and deep learning methods [7].

The application of machine learning theory and technology in crop pest identification began in the 1980s and 1990s. In 2005, Tian Youwen [8] proposed a method for identifying grape diseases by applying SVM and chromaticity moment analysis, which had good classification and generalization ability for small sample datasets, and its recognition rate of black spot, powdery mildew and downy mildew is more than 90%. In 2007, Youwen [8] used SVM to identify grape leaf diseases with an average recognition rate of 87.5%. In 2016, Mohan [9] used the Scale Invariant Feature Transform (SIFT) algorithm to extract color features of rice leaf blights, brown spots and bacterial wilt. Finally, SVM and K-nearest neighbor algorithm (KNN) were used for classification, and a rice disease recognition method with an accuracy of 91.10% was proposed. Padol [10] used SVM classification technology to detect and classify grape leaf diseases. The method first found the lesion area through K-means clustering segmentation, then extracted the color and texture features, and finally used SVM to classify the leaf disease. Waghmare [11] proposed a grape leaf disease recognition system based on multi-class support vector machines, and the classification accuracy of this method for grape leaf disease is 96.66%. In 2019, Jaisakthi [12] used the global threshold method to extract the diseased parts of grape leaves and then classified the grape leaf images through the support vector machine, obtaining an accuracy of 93.04% on the test images.

The above methods are all crop disease recognition methods based on traditional machine learning technology. Compared to early manual recognition methods, their recognition accuracy and efficiency have been significantly improved. However, these methods have problems such as low efficiency, weak anti-interference ability and poor universality, which seriously limits their promotion in actual agricultural production. With the development of deep learning technology, convolutional neural network (CNN) replaces the traditional artificial feature extraction method, which has a stronger feature expression ability and significantly improves the accuracy and efficiency of plant disease recognition [5].

CNN is the most widely used method in deep learning image classification technology and is favored by many plant disease recognition researchers; it can automatically extract image features to classify and recognize leaf disease types in images. In 2016, Mohanty [13] proposed a model based on the integration of AlexNet and GoogleNet to detect 26 kinds of diseases in the Plant Village dataset, with a detection accuracy of 99.35%. In 2017, Ramcharan [14] proposed an automated cassava disease identification method based on the

Inceptionv3 network, which used transfer learning to train CNN, and the classification accuracy reached 93%. Bin Liu [15] designed a new deep convolutional neural network model based on AlexNet, which used apple leaf disease images for pre-processing to generate new datasets, and the accuracy rate reached 97.62% in the datasets of four common apple leaf diseases. In 2018, Liu [16] proposed a CNN-based recognition model for strawberry leaf powdery mildew disease. The CNN structure with nine different network depths was designed by this method, and the nine models were trained and tested with different proportions of training sets and test sets. The results showed that the average accuracy reached 98.61%. In 2019, Too, E. C. [17] used VGGNet, Inception V4, ResNet and DenseNet to detect the apple disease dataset and found that DenseNet performed best, with a test accuracy of 99.75%. In 2020, Gong An [18] proposed using multiple CNNs to identify diseases. The experimental results showed that the recognition accuracy of the fused model is 87.19%. Long Yin [19] used ResNet50 and other networks as the basic framework to train and identify crop disease data. Experimental results showed that the improved model has the highest recognition accuracy of 90.2%.

With the rise of CNN-based target detection methods, these methods can not only identify the type of leaf disease but also locate the location of the disease. In 2017, Fuentes [20] proposed a method based on deep learning target detection to identify and locate tomato diseases. In this method, Faster R-CNN, R-FCN and SSD were used to detect tomato diseases and insect pests. The results show that R-FCN has the highest mAP (mean average precision) of 85.98%. In 2018, Adhikari [21] proposed a tomato disease detection method based on YOLO, which used Raspberry Pi for disease image acquisition and recognition; the deployment of the model was realized, and its mAP reached 76%. Liu Tanyu [22] first removed the background in the grape disease image to reduce the interference of the disease region and adopted fast-RCNN technology to extract leaf disease spots, which made grape leaves adaptable for detection. For common grape diseases, the average mAP of this method was 75.52%. Xin [23] proposed a real-time grape detection model using the YOLOv3 model as the basic framework. The model replaced the Darknet-53 backbone network in YOLOv3 with the EfficientNet network to effectively balance the image resolution and the depth and width of the training network, and the final recognition rate reached 97.29%. In 2019, Jiang [24] expanded the apple leaf disease dataset through data enhancement and combined Inception structure in GoogleNet and Rainbow concatenation with SSD network, and the experiment showed that the mAP reached 78.80%. Its detection speed is 23.13 fps. Zhao [25] adopted the YOLO network to identify tomato diseases, and the experimental results show that the mAP of this method is 97.24%. In 2020, He Xin [26] proposed a method for identifying grape leaf diseases. The method first improved the pooling and mask branch in mask R-CNN. Then, the SENet attention module was integrated into the ResNet50 classification network to form a new multi-feature fusion classification network, and the recognition accuracy was 90.83%. Zhu Xu [27] proposed a blueberry canopy fruit detection and recognition method based on Faster R-CNN. By constructing their own dataset and using improved Faster R-CNN for training, the average recognition accuracy of the obtained model could reach 94%. In 2021, He Zifen [28] used the SE attention mechanism and asymmetric mixing convolutional module in Mask R-CNN to detect three diseases of apple and obtained an average segmentation accuracy of 94.7%, reducing the training memory. In 2022, Wang Zhuo [29] proposed a real-time apple detection model, YOLOv4-CA. In this model, MobileNet v3 was used as the backbone feature extraction network, and deep separable convolution was introduced into the feature fusion network with an average accuracy of 92.23%. Huang Tongbin [30] artificially realized the recognition of citrus fruits in the natural environment by introducing the attentional mechanism module (CBAM) to improve the feature extraction capability of the network and proposed a citrus recognition method based on the improved YOLOv5 model. The test results show that the average accuracy of the model reaches 91.3%.

In 2023, YOLOv7 rose strongly, and many researchers used YOLOv7 to detect items and achieved good results. Zhao Zhenbing [31] proposed an anti-vibration hammer corro-

sion detection algorithm based on improved YOLOv7. The method first carried out color space transformation through the HSV color model to highlight corrosion features, then introduced multi-scale feature fusion, and made full use of feature maps under different scales to obtain weighted feature fusion. It can accelerate the inference speed and improve the accuracy of the model. The test results show that the detection accuracy of the improved algorithm is increased by 1.6%, and the inference speed is increased by 17%. Jialiang [32] proposed the detection of dragon fruit based on YOLOv7, and the accuracy, recall rate and average accuracy were 84.4%, 92.4% and 93.2%, respectively. LiangLiang Z [33] proposed a multi-scale UAV aerial image target detection model MS-YOLOv7 based on YOLOv7, aiming at the common problems of a large number of targets and a high proportion of small targets in UAV aerial images. The new network uses multiple detection heads and CBAM convolutional attention module to extract features of different scales, and the average detection accuracy of MS-YOLOv7 is increased by 4.9% compared with the YOLOv7 object detection algorithm.

With the progress of computing technology, as well as the remarkable innovation and breakthrough of image detection of crop leaf diseases and related technologies, many scholars at home and abroad have used Faster R-CNN, SSD and other algorithms to detect common disease spots in crop leaf diseases images, and the average accuracy rate has reached over 90%, which provides new possibilities for effective management of crop leaf diseases. However, the existing detection methods still have some shortcomings, such as low accuracy and slow detection speed for small target diseases of plant leaves. Therefore, it is worth exploring and studying to detect small target lesions on grape leaves quickly and accurately by using the advantages of deep convolution neural networks. This paper improved on the original YOLOv7 model and added an improved channel attention mechanism. After training, a grape disease detection model is obtained. The main contributions of the proposed method are summarized as follows:

1. In response to the problem of low accuracy in identifying small targets in the original YOLOv7 model, a specialized output prediction branch was added to detect small defect targets and perform feature fusion.
2. In order to increase the feature extraction ability of E-ELAN, asymmetric convolutional ACNet was introduced into E-ELAN to enhance its ability to extract small target features.
3. In response to the insufficient information extraction for small target lesions on grape leaves using the existing YOLOv7 algorithm, an improved CA attention mechanism, which can effectively identify small lesions, was added.
4. The SIoU loss function was introduced to enhance the localization ability of the network, thereby improving the detection accuracy of the network and reducing false or missed detections of small targets during the detection process.

2. Analysis of the Characteristics of Grape Leaves with Diseases

The main diseases of grape leaves are blight, black rot and black measles, as shown in Figure 1. Their lesions mainly have the following characteristics:

1. The grape leaf blight affects the normal metabolism of the leaves and causes the grape leaves to fall off, leading to a reduction in grape yield. The grape leaf blight disease mainly harms the grape leaves. According to the area of the disease spot, it is generally divided into two kinds: big blight and small blight. The edge of the big blight is dark brown, and the shape is generally polygonal or quasi-circular. The small blight first produces small yellow-green spots on the grape leaves, and then the spots gradually increase. The shape is circular, and the color changes from dark brown to light brown as the lesion increases, as shown in Figure 1a.
2. The black rot disease spot of grape leaves shows a black circular outline in the leaves. With the expansion of the disease spot, the center of the disease spot becomes gray-white, and then the edge becomes black, as shown in Figure 1b.

3. At the initial stage of grape leaf black measles, red-brown irregular spots appear first and then expand into round spots with a diameter of about 2 cm. The front of the lesion has obvious concentric rings of the same depth and color, while the back of the lesion is covered with a light brown layer of mold during humid weather, as shown in Figure 1c.
4. Even under a single condition, the size of the lesions is diverse, and the same lesion may even show different sizes due to factors such as angle and image size.

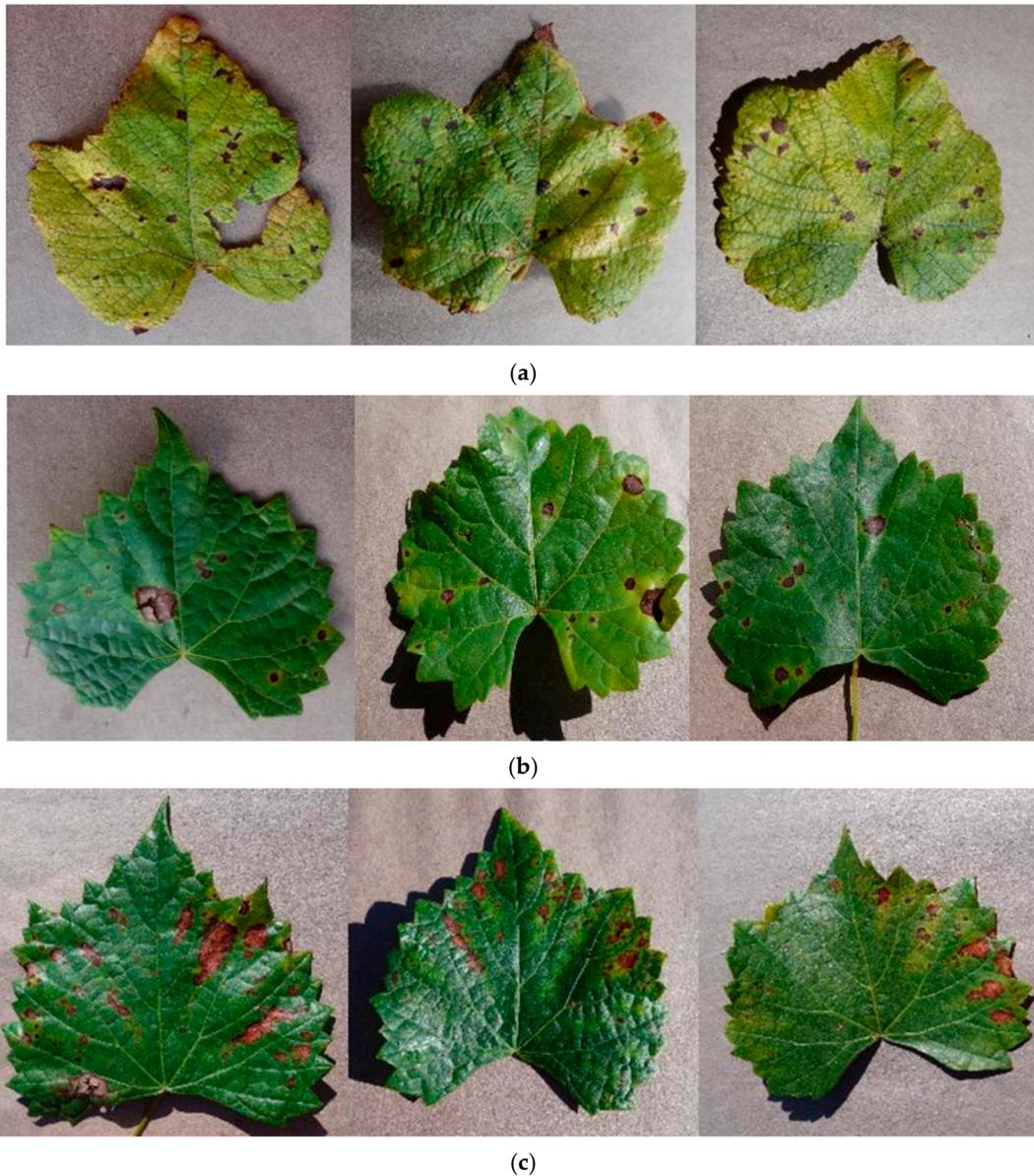


Figure 1. The examples of the grape leaves with diseases:(a) Grape leaves with blight disease; (b) Grape leaves with black rot disease; (c) Grape leaves with black measles disease.

According to the characteristics of the above-mentioned grape leaf disease images, the target shape and scale of grape leaf disease are changeable, which makes it very difficult for YOLOv7 to detect small disease spots because YOLOv7 mainly relies on convolution operation in feature extraction. For small-scale disease spots, due to their limited pixel area, many details may be lost after convolution processing, which makes it difficult to extract effective features. And because the positioning accuracy of YOLOv7 depends on the size of its receptive field, for small-scale targets, the receptive field may not be enough to cover the whole target, resulting in inaccurate positioning.

3. Improvement of YOLOv7 Model

YOLO series algorithms are known for their fast speed, small model size and simple structure, so they are widely used in industrial production and life. From the performance comparison of the YOLOv1 to YOLOv4 algorithm, the comprehensive performance of the algorithm is improved with each new version. Although the detection accuracy of YOLOv4 is lower than that of YOLOv3, the speed has been greatly improved, which is very suitable for real-time target detection. YOLOv5 integrates some characteristics of YOLOv3 and YOLOv4, and the speed is faster. By improving the accuracy of the detection algorithm, YOLOv7 makes the detected target more accurate and faster than the previous version and can process a large number of images in a short time. This paper chooses to optimize the YOLOv7 model to improve the performance of the algorithm further.

As shown in Figure 2, YOLOv7 adopts the backbone network composed of BConv module, E-ELAN module and MP module for downsampling and feature extraction, and adopts SPPCSPC module to expand the receptive field. Then, three feature maps of different scales are input into the feature fusion module, which is composed of the CatELAN module, UpSample module and MP module. After the feature fusion module, Cat is used to fuse the context information, the output is re-parameterized through the REPCONV layer, and the output tensor is finally obtained. YOLOv7 introduces model reparameterization into the network architecture to reduce the calculation amount of the model and introduces label allocation strategy and efficient E-ELAN network architecture to form a model feature extraction module. Finally, YOLOv7 puts forward the training method of the auxiliary head, which increases the training cost and improves the accuracy without affecting the reasoning time.

The YOLOv7 algorithm uses a top-down and bottom-up bidirectional fusion network to aggregate features at different levels. It uses shallow features to distinguish simple targets and deep features to distinguish complex targets, and it enhances the semantic information of the entire pyramid by transferring high-level strong semantic features. However, the detection effect of YOLOv7 on small blade targets is poor, mainly because the resolution of small targets is low, and they are susceptible to interference from noise, complex background and other factors, resulting in more difficult detection of small targets. In view of the above shortcomings, the detection performance of YOLOv7 is mainly improved from the following aspects:

- (1) To solve the problem of YOLOv7's insufficient accuracy in identifying minor lesions, this paper adds a detection layer dedicated to detecting small targets and involves the new feature layer in the feature fusion process. Through this scheme, the network can detect more small targets with lower resolution.
- (2) Aiming at the problems of low resolution and insufficient information for small targets, asymmetric convolution ACNet is introduced into E-ELAN to further enhance the ability to extract features.
- (3) To solve the problem of image noise, background and other factors interfering with the accuracy of grape leaf lesions, this paper adopts an improved CA attention mechanism to strengthen key features and suppress the interference of noise, which can improve the model's attention to small targets.
- (4) Aiming at the problem of poor positioning ability of CIoU loss function, SIoU is introduced in the original YOLOv7 network model to optimize the loss function, reduce the degree of freedom of the loss function and enhance the positioning ability of the network, thus improving the network detection accuracy and reducing the missed detection of small targets in the detection process.

The improved YOLOv7 model is shown in Figure 2. In the figure, the red part represents the Input section of YOLOv7; The yellow section represents the Backbone section of YOLOv7; The purple part is the Neck part of YOLOv7; The light blue part is the Head part of YOLOv7; The dark blue section represents the specific components of each module of YOLOv7; The green section is an added module based on YOLOv7.

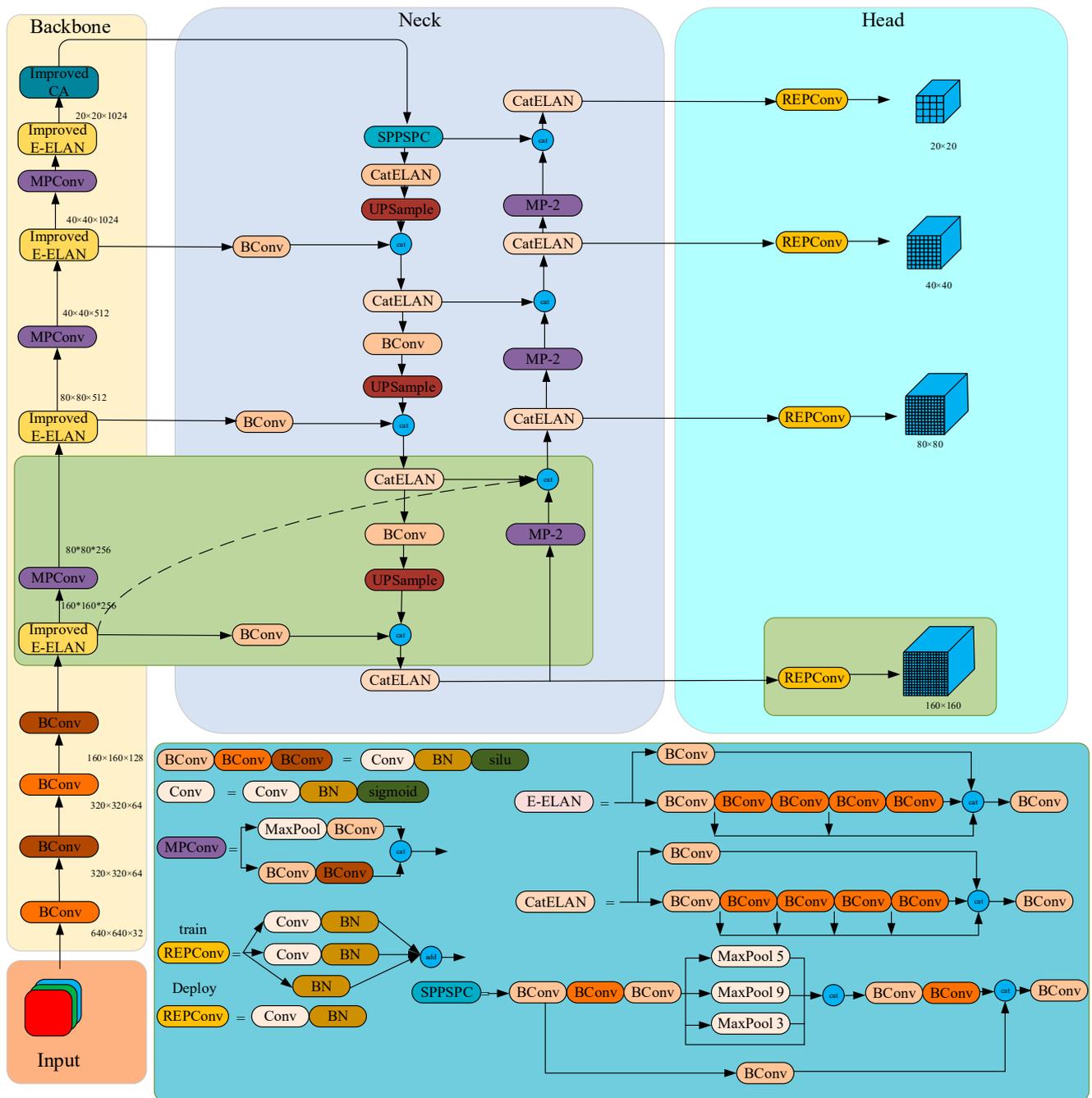


Figure 2. Schematic diagram of improved algorithm of YOLOv7.

3.1. Improvement of Feature Fusion Network

When using a convolutional neural network to extract image features, with the deepening of convolution layers, the resolution of the feature map decreases gradually, and the range of the local receptive field to perceive the original image expands continuously. Moreover, the feature map closer to the top tends to pay attention to the global information of the image, so the deep features extracted by the deep neural network are very unfavorable for small target detection. Although the shallow feature map has a small local receptive field and contains more position and detailed information, which can make up for the deficiency of deep features to some extent, if only shallow features are used to detect targets, it will cause serious false detection and missed detection due to the lack of guidance of high-level semantic information. Thus, shallow features and deep features are fused, and then the object position is predicted on the fused feature maps of different scales.

In YOLOv7, if the size of the input image is $640 \times 640 \times 3$, the feature maps of $80 \times 80 \times 255$, $40 \times 40 \times 255$ and $20 \times 20 \times 255$ are obtained after $8\times$, $16\times$ and $32\times$ down-sampling, respectively, and the network finally detects the target on these three feature maps with different scales. Among the feature maps of these three scales, the smallest local receptive field is the eight times downsampled feature map; if the feature map is mapped to the original input image, each pixel corresponds to the 8×8 area of the original image. For the target with low resolution, the receptive field of the feature map obtained by eight times downsampling is still too large, and it is easy to lose the position and detail information of some small targets, which leads to its poor performance on the small target dataset. In order to improve the current situation of missed detection, we optimized the structure of YOLOv7 and added a new prediction branch. After three times of up-sampling, the size of the detection layer corresponding to the new feature layer is 160×160 , which can be used to detect defect targets with a size of more than 4×4 and can effectively improve the detection effect of small lesions.

The feature fusion structure used in the YOLOv7 algorithm is PANet, which aggregates different levels of features through a top-down and bottom-up bidirectional fusion network. In order to integrate more shallow information into the feature enhancement part, some information extracted from the Backbone is also integrated into the FPN process to improve the recognition accuracy of the model. In the process of feature enhancement, in order to minimize the information loss caused by the deepening of the network and the increase in convolution modules, the original model continuously upsamples and downsamples the loaded information, which undoubtedly loses much detailed information and reduces the accuracy of the model in detecting small targets. On the basis of the previous improvement, the modules on the Backbone are connected by leaps to retain more detailed information and reduce the loss of information. The improved feature fusion results are shown in Figure 3. The red arrow represents upsampling, the blue arrow represents downsampling, the middle two boxes are YOLOv7's FPN module and PAN module, the last box is the detection layer part, and the red small box is the newly added P2 detection layer on the original basis.

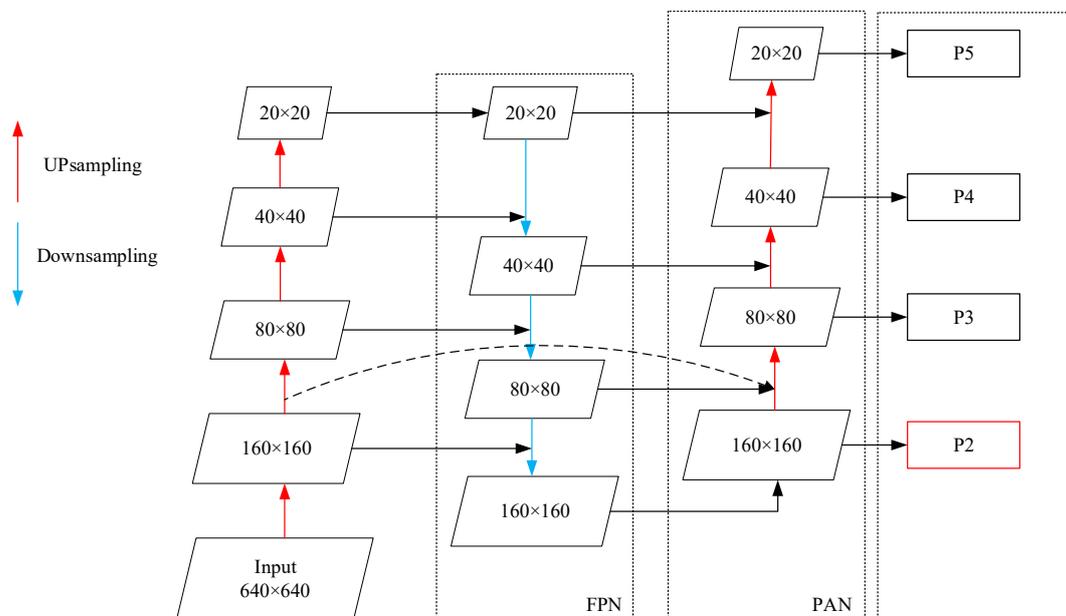


Figure 3. FPN + PAN structure after adding detection layer P2.

In summary, on the one hand, this structure increases the depth of the network so that the network can better learn the multi-level feature information of the target and enhance the network’s ability to detect multi-scale targets in complex environments. On the other hand, it can transfer more shallow features to deep features so that the network can obtain more small target information and improve the detection ability of the model for small targets.

3.2. Improvement Based on E-ELAN

E-ELAN in YOLOv7 is improved based on ELAN. Compared with the ELAN network, E-ELAN adds the use of group convolution to increase the cardinality of features. This structure can enhance the feature extraction ability, improve the use of parameters and calculation efficiency and improve the learning ability of the network without destroying the original gradient path. E-ELAN has a good feature extraction ability when extracting the information of major lesions but does not carry out much calculation when extracting the information of small targets. As a result, the detection cannot guarantee the complete accuracy of the lesion category.

The attention mechanism inside the E-ELAN module can help to understand the decision-making process of the model, but it may not provide a completely accurate explanation or deal with complex scenarios. Thus, this paper introduces asymmetric convolution ACNet [34]. The core idea of ACNet is to decompose the original convolution, which sums up the outputs of two convolutional branches with horizontal and vertical kernels, respectively, thereby reducing the number of parameters and computations while ensuring comparable accuracy. During the training phase, the feature extraction layer adopts a parallel structure of 3×3 conventional convolution kernels and asymmetric convolution kernels for feature extraction and then combines the extracted feature maps with channels to improve the richness of semantic information within the feature maps. In the detection stage, in order to operate the model using convolutional kernels with consistent parameters while minimizing computational complexity, the parallel convolutional kernels in the AC module are fused to obtain their equivalent convolutional kernels, and the convolutional kernels are used for feature extraction. Moreover, when the distribution parameters of the box convolution are uneven, the weight of the center cross position is greater, further enhancing the expression ability.

ACNet uses three parallel $d \times d$, $1 \times d$ and $d \times 1$ cores to replace the original $d \times d$ core, as shown in Figure 4 (let $d = 3$):

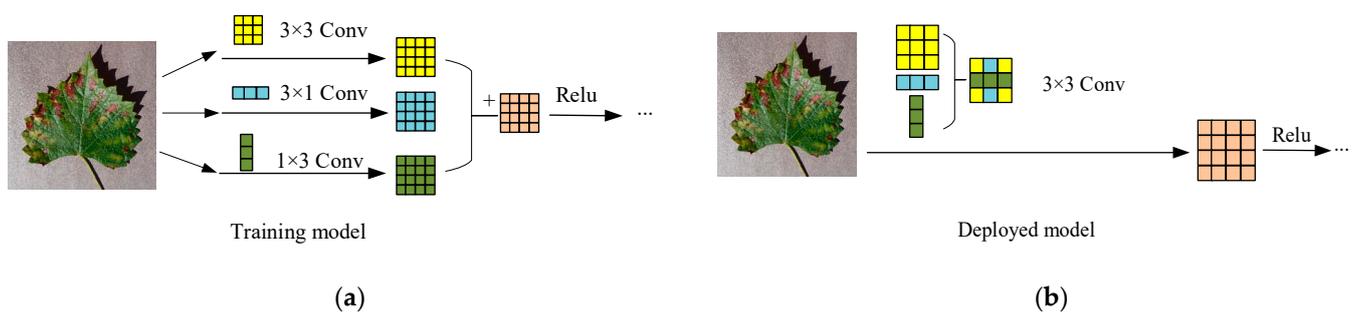


Figure 4. Schematic diagram of AC structure training and deployed process: (a) ACNet training model; (b) ACNet deployed model.

ACNet is divided into the training stage and the reasoning stage. In the training phase, ACNet replaces every 3×3 convolution in the existing network with 1×3 , 3×1 and 3×3 convolution and finally fuses the calculation results of these three convolution layers to obtain the output of the convolution layer. In the reasoning stage, ACNet mainly fuses three convolution kernels. In this part, the fused convolution kernel parameters are used to initialize the existing network; in the reasoning stage, the network structure is exactly the

same as the original network, except that the network parameters adopt the parameters with stronger feature extraction ability, that is, the fused convolution kernel parameters.

The E-ELAN network in YOLOv7 is improved, and shortcut connection, 1×3 and 3×1 convolution structures are added between ELAB modules to enrich the features extracted from the network, as shown in Figure 5. The module with a plus sign in the circle is used to add up the weights of the branches

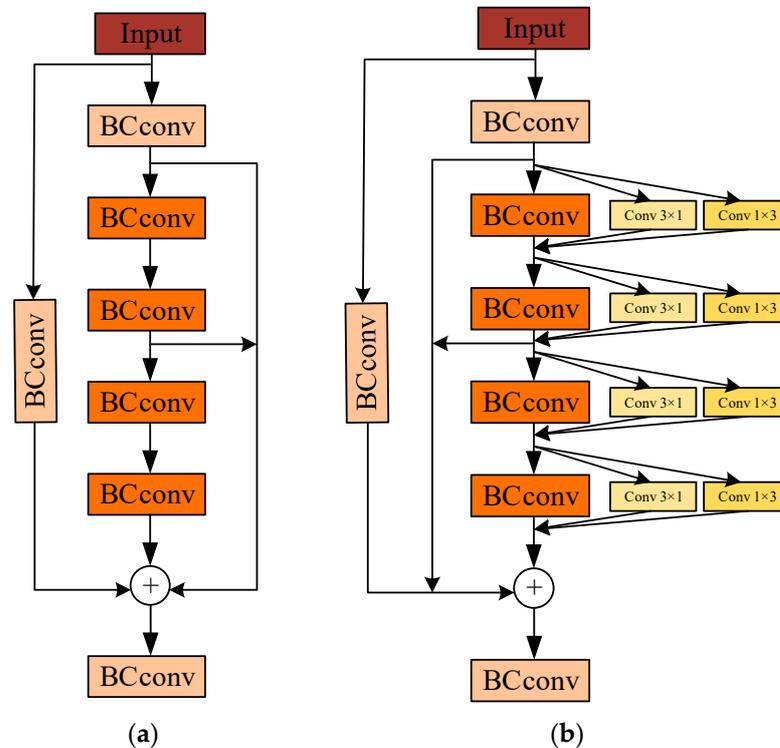


Figure 5. E-ELAN model and improved E-ELAN model: (a) E-ELAN model; (b) improved E-ELAN model.

E-ELAN in YOLOv7 is improved, and 3×1 and 1×3 convolution structures are added between modules to make the network extract more abundant features to improve the detection accuracy of the model.

3.3. Improvement Combined with Channel Attention

The attention mechanism establishes dynamic weight parameters by making relevant and irrelevant choices on information features to strengthen important information and weaken useless information. Among them, the Squeeze and Exclusion (SE) [35] channel attention mechanism is widely used for its simplicity and high efficiency. However, the attention mechanism of the SE channel ignores the importance of location information. In the Convective Block Attention Module (CBAM) [36], GAP (global average pooling) and GMP (global maximum pooling) are combined to improve performance, thereby increasing the efficiency of deep learning algorithms and improving some defects of traditional deep learning, but it needs more computing resources and has higher computational complexity. Coordinate attention (CA) [37] is a flexible and efficient attention mechanism that can not only obtain cross-channel information but also obtain location information and help the model to locate and identify the target of interest more accurately without introducing much computational overhead. The CA module completes the weight distribution of feature information through two stages: coordinate information embedding and coordinate attention generation. CA attention mechanism performs GAP operation on the input feature map in the height and width directions to obtain the feature map in two directions. However, the simplicity of GAP makes it difficult to capture the complex information of

various inputs. The existing CA channel attention mechanism divides the input feature map into horizontal and vertical parts and calculates the corresponding weights, respectively. However, this method cannot balance the global information and leads to the neglect of important spatial information.

In order to overcome the limitations of the existing CA attention module, this paper makes the following improvements:

- (1) Add the GMP pathway to the GAP pathway of the original CA attention mechanism. The input feature map contains many targets with different sizes, so it is a better choice to average the pool first and then maximize the pool. This is because average pooling can capture all the information in the feature map, smooth the image and reduce noise better. Then, the maximum pooling is carried out, which can emphasize the most prominent features of the target while retaining the size transformation and further extracting effective information about the target.
- (2) Calculate the weight by processing the input feature map. Adaptive average pooling and full connection layer operation are performed on the input feature map to generate a two-dimensional Softmax vector, which is used to represent the mixed proportion of horizontal and vertical attention weights. This method can effectively balance the global and local information, as shown in Figure 6.

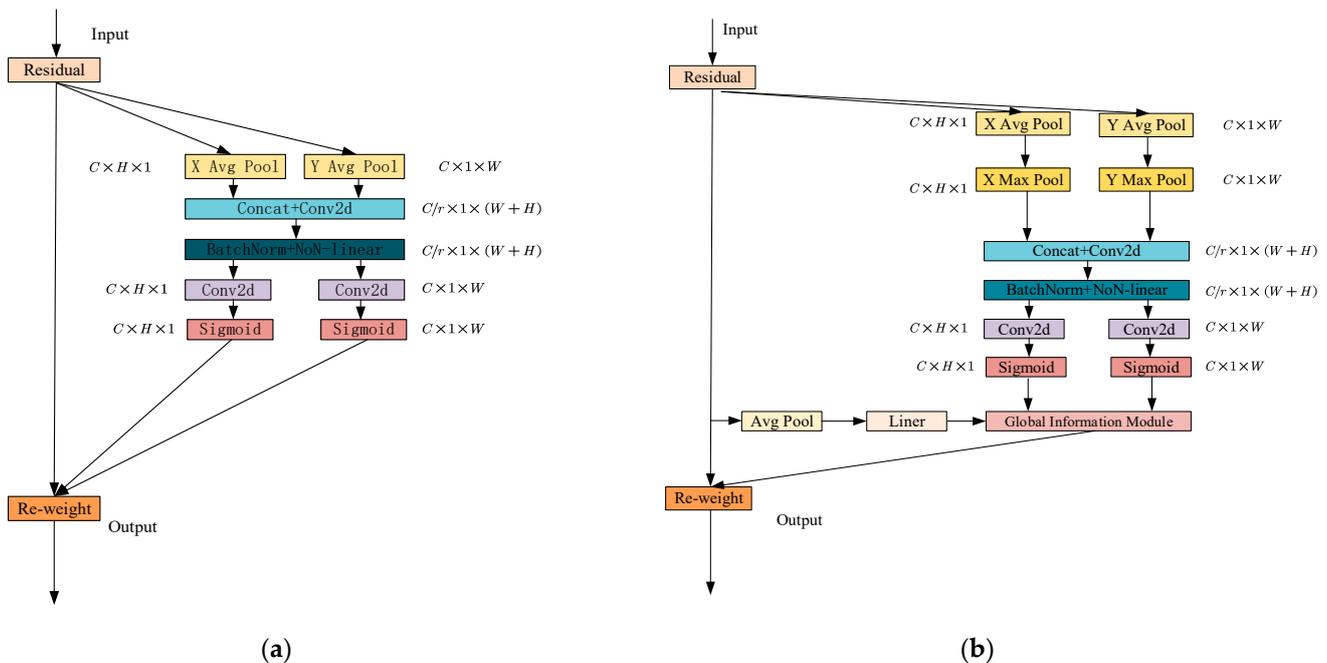


Figure 6. CA attention and improved CA attention: (a) CA attention; (b) improved CA attention.

3.4. Loss Function

The loss function in the YOLOv7 network model is shown in Formula (1):

$$L_{object, Loss} = L_{loc, Loss} + L_{conf, Loss} + L_{class, Loss} \tag{1}$$

where $L_{loc, Loss}$ represents positioning loss, $L_{conf, Loss}$ represents confidence loss and $L_{class, Loss}$ represents classification loss. The coordinate loss is calculated using CIoU [38]. In order to describe the aspect ratio of the real frame and the predicted frame, when the aspect ratio is the same, the penalty item value is 0, which means that the difference in the aspect ratio is calculated as the penalty item, but it cannot represent the real difference between the length and width. Different from the CIoU loss function, the SIoU [39] loss function calculates the relationship of length and width of the real box and the prediction box separately, rather than calculating the difference of the aspect ratio, avoiding the loss of penalty terms when

the aspect ratio of the prediction box and the real box is equal, and ensuring the stability of convergence, as shown in Figure 7.

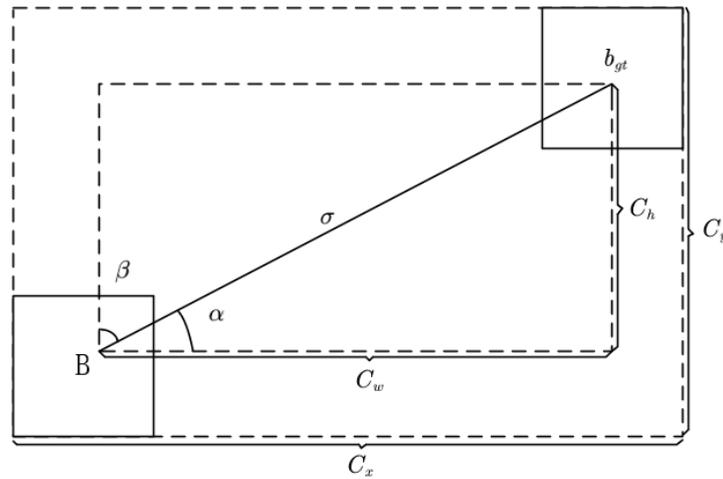


Figure 7. SIOU loss function model.

c_x and c_y is the length and width of the real frame, c_w and c_h represent the difference in width and height between the center point of the target box and the prediction box, B is the prediction frame, b_{gt} is the real frame, and σ represents the distance between the center point of the target box and the prediction box.

In order to solve the direction angle regression problem of the target box, the SIOU loss function introduces angle perception, distance loss and shape loss on the basis of the original IoU to improve the speed and accuracy of reasoning during the training of the detection model.

The angle perception added by SIOU first predicts whether the angle is greater than 45° and determines the use of minimization β or α . The calculation of angle cost is shown in Equation (2):

$$\Lambda = 1 - 2 \times \sin^2\left(\arcsin(x) - \frac{\pi}{4}\right) \tag{2}$$

where

$$x = \frac{c_h}{\sigma} = \sin(\alpha) \tag{3}$$

$$\sigma = \sqrt{\left(b_{c_x}^{gt} - b_{c_x}\right)^2 + \left(b_{c_y}^{gt} - b_{c_y}\right)^2} \tag{4}$$

$$c_h = \max\left(b_{c_y}^{gt} - b_{c_y}\right) - \min\left(b_{c_y}^{gt} - b_{c_y}\right) \tag{5}$$

$$c_w = \max\left(b_{c_x}^{gt} - b_{c_x}\right) - \min\left(b_{c_x}^{gt} - b_{c_x}\right) \tag{6}$$

In the formula, $b_{c_x}^{gt}$ and $b_{c_y}^{gt}$ represent the coordinates of the center point of the target box; b_{c_x} and b_{c_y} represent the coordinates of the center point of the prediction box. The distance cost represents the distance between the center points of the predicted box and the actual box. Based on the angle cost defined above, SIOU redefines distance cost as follows.

$$\Delta = \sum_{t=x,y} \left(1 - e^{-\gamma \rho_t}\right) \tag{7}$$

$$\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w}\right)^2 \tag{8}$$

$$\rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h} \right)^2 \quad (9)$$

$$\gamma = 2 - \Lambda \quad (10)$$

The shape cost Ω is defined as follows:

$$\Omega = \sum_{t=w,h} (1 - e^{-w_t})^\theta \quad (11)$$

$$w_w = \frac{|w - w_{gt}|}{\max(w - w_{gt})} \quad (12)$$

$$w_h = \frac{|h - h_{gt}|}{\max(h - h_{gt})} \quad (13)$$

θ controls the degree of attention to shape error, and its selection is related to the specific dataset. If $\theta = 1$, shape error will be optimized, thereby reducing the free movement of the predicted box and the actual box shape. The regression loss of the target box is written as follows:

$$L_{SloU} = 1 - IoU + (\Delta + \Omega)/2 \quad (14)$$

4. Results

4.1. Experimental Environment

The experimental environment for the CPU is Intel(R) Core(TM) i7-9700k, the GPU is NVIDIA GeForce RTX 2080 Ti with 24 GB memory, RAM is DDR4 with 64 GB, the CUDA version is 1.12.1, and the deep learning framework is Pytorch 1.7.

4.2. Dataset Description

The dataset is grape leaf disease on the public dataset Plant Village [40]; labeling was used as the tool for labeling datasets. This experiment extracts blight, black rot and black measles disease and uses the labeling toolbox to manually label the grape disease spots. In order to improve data diversity and solve the model overfitting caused by data imbalance, this paper expanded the dataset by mirroring, flipping and other operations, as shown in Table 1. The first column is the serial number of grape leaf diseases, the second column is the name of grape leaf diseases, and the third column is the number of original pictures of each type of grape leaf disease in the Plant Village dataset. The original dataset is prone to overfitting in the training process of the deep learning model, so this paper expands the dataset to increase the generalization ability of the model, thereby alleviating the overfitting problem. In addition, the expanded dataset has better robustness and enhanced the stability of the model, and the fourth column is the number of expanded pictures in the Plant Village dataset; using python scripts, the original image is expanded by translation, flipping, rotation, mirroring, scaling and noise. Column 5, column 6 and column 7 are the numbers of images of the training set, the verification set and the test set, respectively.

Table 1. Dataset information.

Series	Class	Number of Original Images	Number of Expanded Images	Number of Training Set	Number of Validation Set	Number of Test Set
1	blight	1076	2908	2037	290	581
2	black rot	1180	2320	1624	232	464
3	black measles	1383	2528	1770	253	505

By reading the related articles in the field of crop leaf disease detection, it was found that most of the training sets account for 70% of the total dataset, the verification sets account for 10%, and the test sets account for 20%, so this paper divides them according to this ratio. After inputting the dataset into the model, it was found that the indexes such as accuracy, recall and average accuracy tend to be stable within 150 to 180 rounds; therefore, this paper iterates 200 times. When training a deep learning model, the batch size directly affects the required computing resources and memory. A large batch size may overflow the GPU memory, while a small batch size may fail to make full use of the GPU's computing power. In this case, we chose a batch size of four to achieve a balance between compute resources and memory. The learning rate determines the amplitude of weight updating in each iteration of the model. The improved YOLOv7 model is a relatively complex model containing a large number of parameters and layers. For a model of this scale, a higher initial learning rate can accelerate the updating of parameters and the convergence of the model. This can provide better training stability and convergence speed, and the learning rate of 0.01 can help the model find a good solution space faster. Cycle learning rate is a learning rate adjustment strategy that allows the learning rate to change dynamically during the training process. By setting a periodic learning rate, the step size can be adjusted at different stages of training to better optimize the model. In this paper, the cycle learning rate is set at 0.01. This is in order to better adapt to the needs of model training and help the model find the best step size in different training stages to improve the training efficiency and performance. Learning rate momentum is a hyperparameter used to accelerate the gradient descent process, usually between 0 and 1, in which a larger momentum value can accelerate the convergence of the model. Through experiments and tuning, we found that in the model, setting the learning rate momentum to 0.937 can help the model converge faster and improve the stability of training, so we chose to set the learning rate momentum to 0.937. The weight attenuation coefficient is 0.0005. The weight attenuation coefficient is used to regularize the loss function to reduce the possibility of overfitting the model. The weight parameter in the model can be penalized to be closer to zero, and in many deep learning practices, 0.0005 is considered a relatively good experience value for different models and tasks.

4.3. Evaluating Indicator

In this paper, four performance indicators, missed detection rate (M), precision rate (P), recall rate (R) and mean average precision (mAP), are used to verify the effectiveness of the network. mAP refers to the average of average accuracy AP of all categories. Generally, the threshold is set to 0.5; that is, the prediction box with IoU greater than 0.5 is valid, which is represented by mAP@0.5.

4.4. Experimental Results and Analysis

Firstly, the influence of SIoU loss function on the network is verified. YOLOv7 is trained with the original CIoU loss function and SIoU loss function. The visualization result of the loss function is shown in Figure 8. The loss value of using SIoU is smaller than that of CIoU, and the stability is also improved to some extent. Therefore, it is more important to use SIoU as the bounding box loss function of the dataset in this paper to improve the performance of the network model.

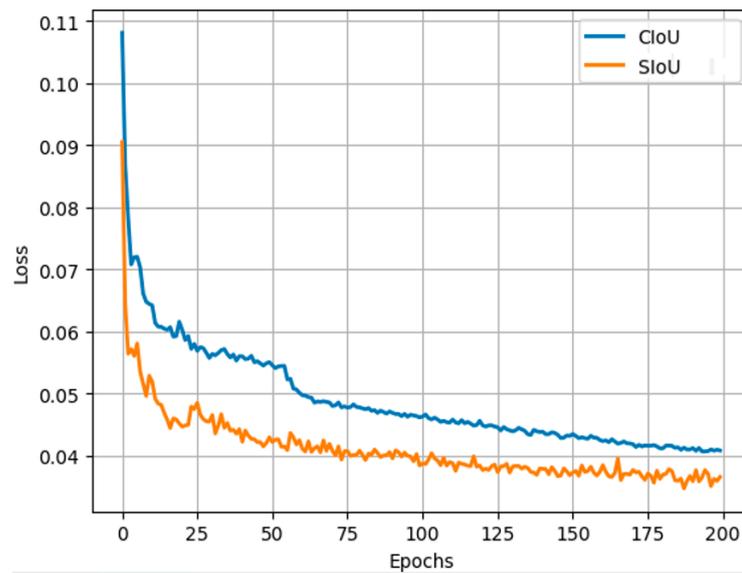


Figure 8. Loss visualization results diagram.

In order to verify the effectiveness of the method proposed in this paper in improving the accuracy of the YOLOv7 network, we compared it with other object detection models, including Faster-RCNN, YOLOv3, SSD and YOLOv7, as shown in Table 2.

Table 2. Detection results of different models.

Method	Blight (%)	Black Rot (%)	Black Measles (%)	mAP (%)
Faster-RCNN	89.6	89.6	89.9	89.7
SSD	89.3	89.8	90.6	89.9
YOLOv5	89.6	89.8	91.5	90.3
YOLOv7	89.9	91.4	91.1	90.8
YOLOv8	89.5	89.7	91.1	90.1
Improved YOLOv7	93.5	93.7	93.3	93.5

The mAP of the improved model is the highest, reaching 93.5%, which is 3.8%, 3.6%, 3.2%, 2.7% and 3.4% higher than that of Faster-RCNN, SSD, YOLOv5, YOLOv7 and YOLOv8, respectively. Compared with YOLOv7, the mAP of blight, black rot and black measles increased by 3.6%, 1.7% and 2.2%, respectively. The mAP of the improved YOLOv7 is the largest, which shows that adding a small target detection layer improve CA attention can enhance the feature extraction ability of YOLOv7 and can solve the problem of low accuracy in identifying small disease spots in the early stage of grape leaf diseases.

In order to further analyze the performance of different classes, the precision, recall, mAP@0.5 and mAP@0.5:0.95 of each class of the improved network are analyzed as shown in Table 3. The detection accuracy of black rot and blight is higher than that of black measles, mainly because the size and shape of black rot and downy mildew spots are regular, which is beneficial to model detection.

In order to verify the effectiveness of this algorithm for small targets, 1551 images in the test set are selected to verify this algorithm. Among them, there are 571 images of blight disease with 1730 small targets, 478 images of black rot with 1356 small targets and 502 images of ring spot disease with 2098 small targets. As can be seen from Table 4, the recognition rate of the three diseases was significantly improved by an improved YOLOv7 algorithm. Compared with the original YOLOv7 model, the recognition rate of the three diseases was increased by 6%, 8% and 7%, respectively.

Table 3. Performance analysis of YOLOv7 and improved YOLOv7.

Class	Method	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
blight	YOLOv7	89.9	92.6	89.9	55.3
	Improved YOLOv7	93.5	95.6	93.5	58.5
black rot	YOLOv7	91.4	92.9	91.4	52.5
	Improved YOLOv7	93.6	94.7	93.7	55.4
black measles	YOLOv7	81.6	79.3	91.1	43.4
	Improved YOLOv7	84.3	81.9	93.3	45.4

Table 4. Comparison of small target detection quantity.

	Small Target Quantity	YOLOv7	Improved YOLOv7
blight	1730	1570	1675
black rot	1356	1223	1339
black measles	2098	1803	1954

In order to verify the advantages of the improved model in the identification of grape leaf diseases, the detection results for each of the three diseases are shown in Figure 9 by YOLOv7 and improved YOLOv7, respectively. According to the comparison, it can be found that the original YOLOv7 model in the figure has many missed detections in the detection of small disease spots, and the discrimination of a single disease spot is low. The missed detection rate of the improved YOLOv7 algorithm for grape leaf lesion detection is obviously reduced, and the regional location of the lesion is more accurate. At the same time, the disease spots of different leaves can be detected simultaneously in the natural environment, and the disease spots in the edge area of the image also have a good detection effect, as shown in Figure 9.

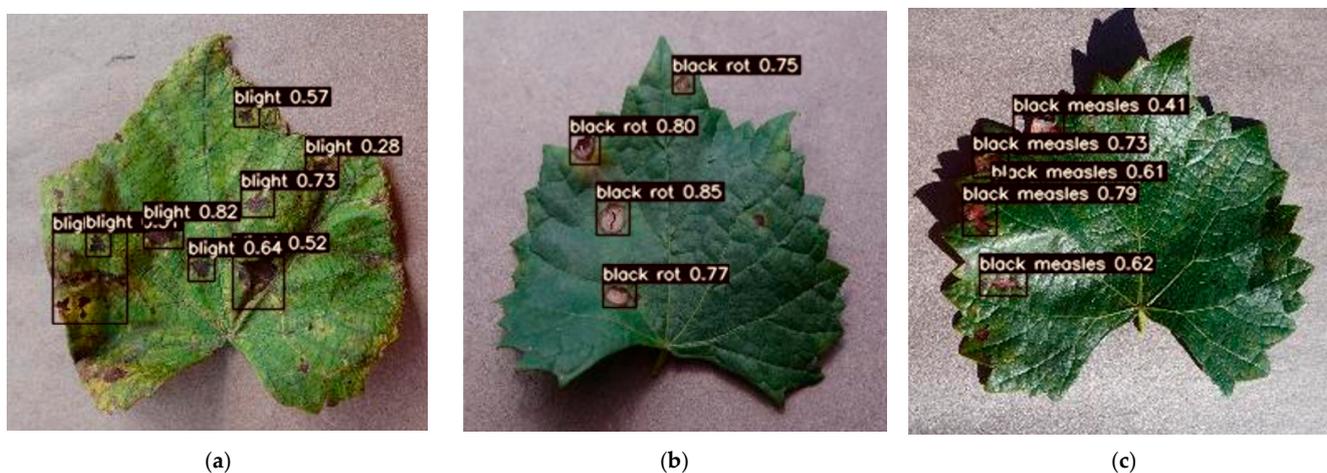


Figure 9. Cont.

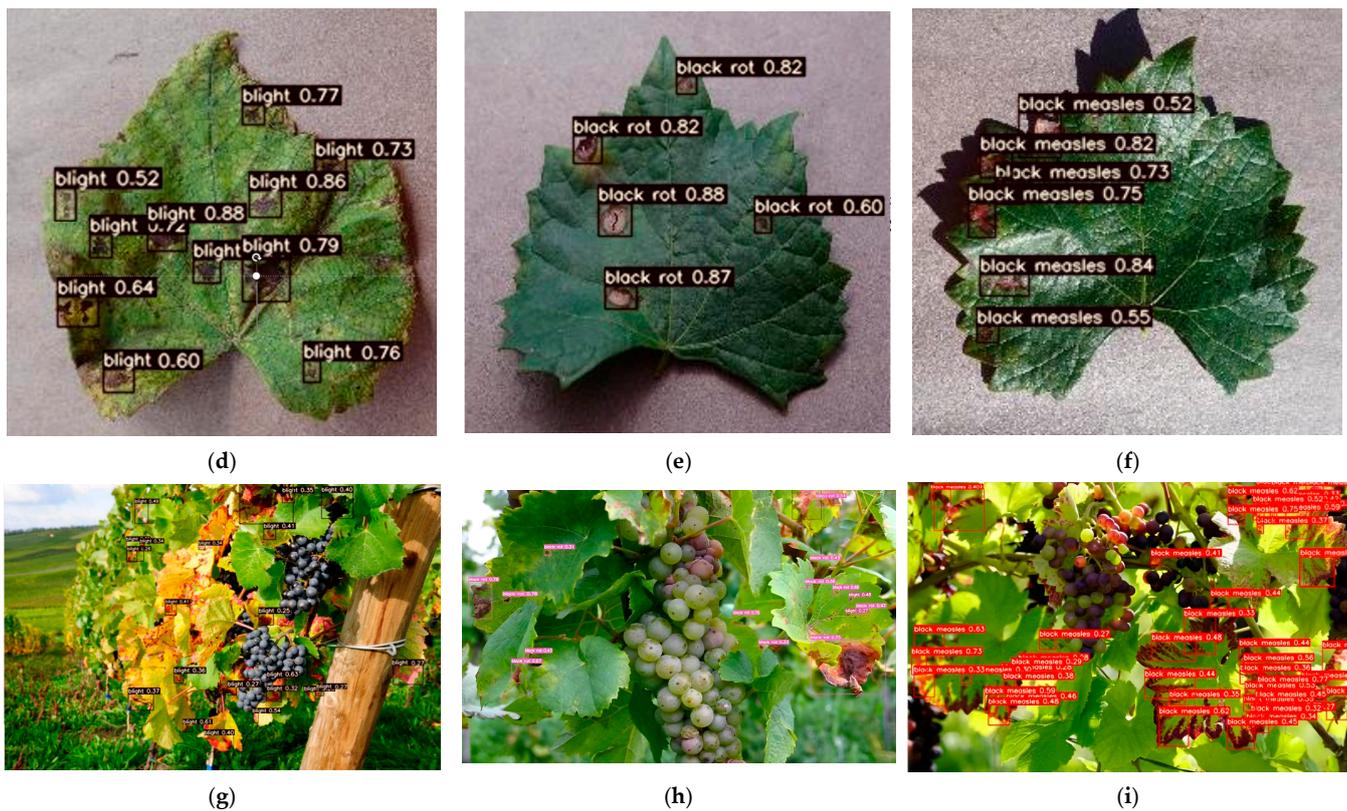


Figure 9. The disease detection results of YOLOv7 and improved YOLOv7: (a) Blight detected by YOLOv7; (b) Black rot detected by YOLOv7; (c) Black measles detected by YOLOv7; (d) Blight detected by improved YOLOv7; (e) Black rot detected by improved YOLOv7; (f) Black measles detected by improved YOLOv7; (g) The improved YOLOv7 detects blight under natural conditions; (h) The improved YOLOv7 detects black rot under natural conditions; (i) The improved YOLOv7 detects black measles under natural conditions.

Through field experiments, we can better understand the performance of the algorithm in actual use and provide us with more datasets in real scenarios, which is also very beneficial to the training and verification of the algorithm. Therefore, we took three images using a Canon EOS 90D mid-range DSLR camera with a pixel size of 1280×720 , it is produced in Shenzhen, Guangdong, China. Grape disease scenes were detected under natural conditions, enhancing the credibility of grape disease image identification under natural imaging conditions. As can be seen from Figure 9, this algorithm can not only detect on the test set but also has a good detection effect in complex environments such as background occlusion under natural conditions and similar colors between disease spots and background.

5. Discussion

In order to verify the effectiveness of the improved CA attention in this paper, we compared it with CA attention, as shown in Figure 10 and Table 5.

Table 5. Performance analysis of CA and improved CA.

Method	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
CA	87.8	88.4	91.5	51.3
Improved CA	89.3	88.8	92.4	51.9

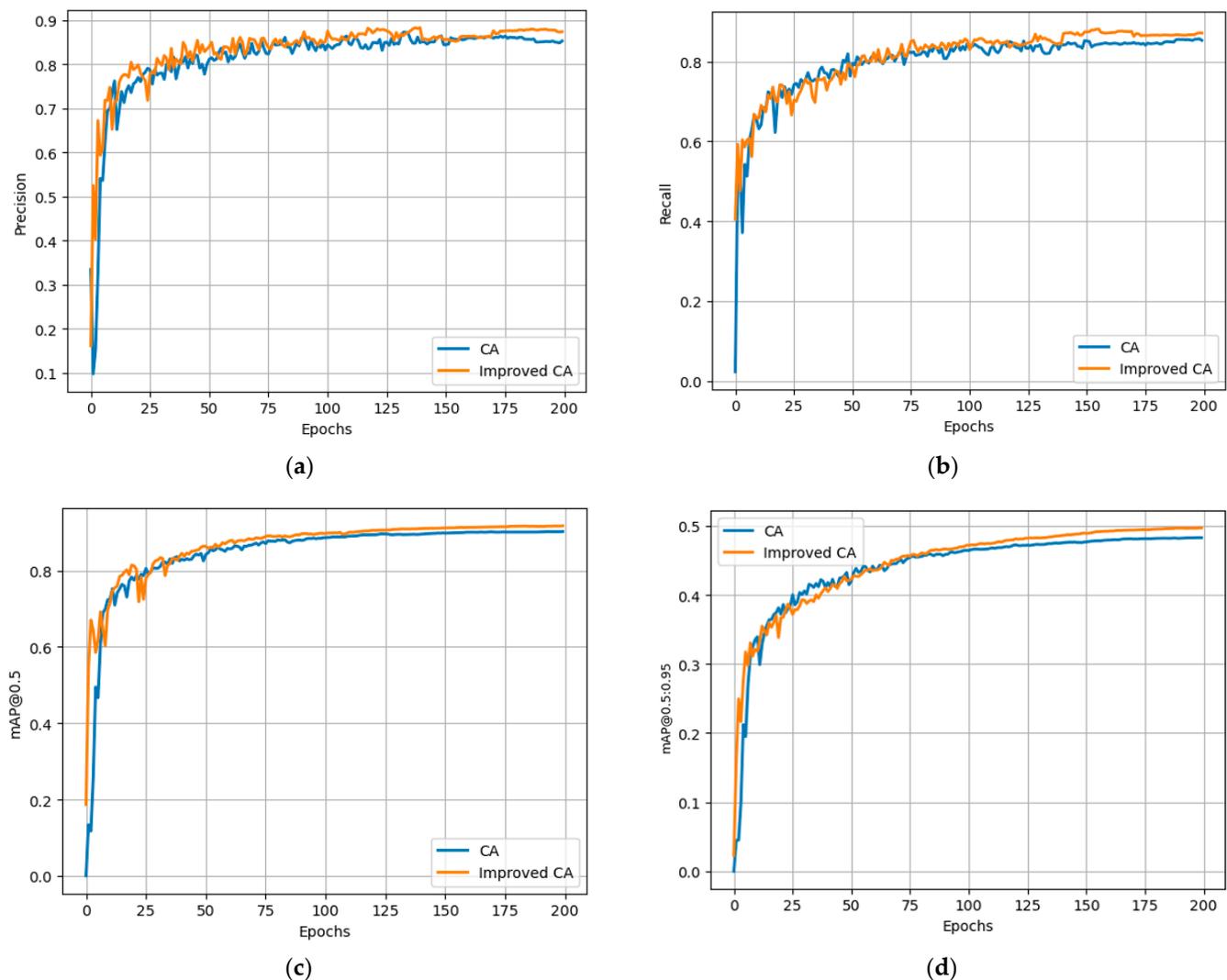


Figure 10. Performance comparison chart of CA and improved CA algorithms: (a) precision; (b) recall; (c) mAP@0.5; (d) mAP@0.5:0.95.

As can be seen from the table, compared with the original CA model, the precision, recall, mAP@0.5 and mAP@0.5:0.95 of the improved algorithm are increased by 1.5%, 0.4%, 0.9 and 0.6%, respectively.

As can be seen from Figure 10, precision, recall, mAP and mAP@0.5:0.95 all increase to varying degrees, indicating that CA has a stronger feature extraction ability after improved attention. Moreover, in rounds 10 to 25, the curves of the model with improved CA were smoother than those with CA, indicating that the stability of the model was further enhanced.

The mean of various objects was taken as the overall performance evaluation index and compared with the original YOLOv7, as shown in Figure 11. It shows that the P2 detection layer constructed by this algorithm improves the accuracy of small targets, the improved E-ELAN enhances the ability to extract small target features, the improved CA attention enhances the ability to extract key information of disease spots, and the SiOU enhances the positioning ability of the network, which means that the improved model can improve the image detection accuracy of grape leaf diseases and greatly improve the detection effect of small disease spots. The improved YOLOv7 was compared with the original YOLOv7, as shown in Figure 11. As can be seen from Figure 11, precision, recall, mAP and mAP:0.95 all increase to varying degrees, and most of the four indexes of the improved YOLOv7 are higher than YOLOv7 in 200 rounds, indicating that the improved

YOLOv7 has stronger feature extraction capability. It can solve the problem of low accuracy in the early stage of grape leaf disease.

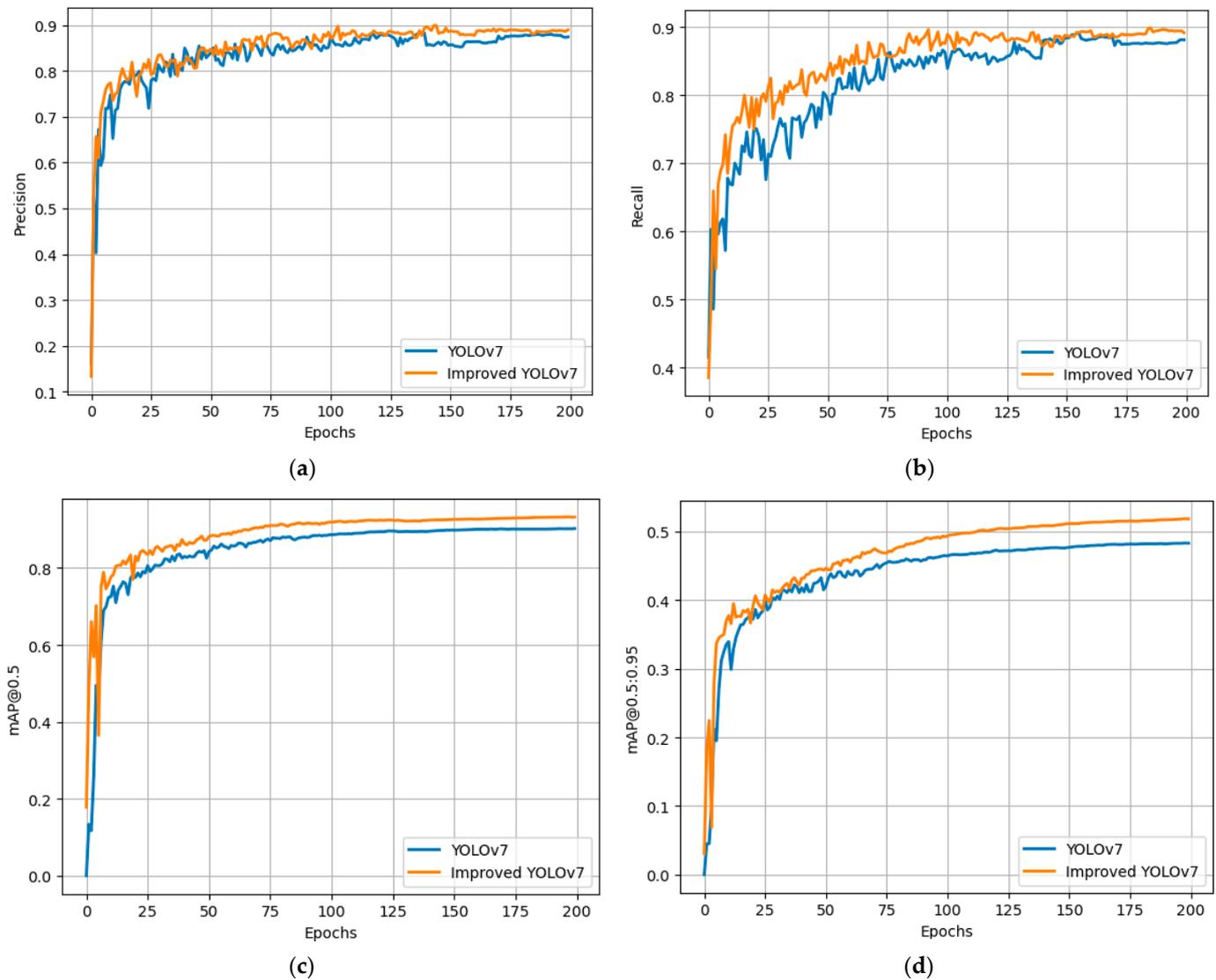


Figure 11. Performance comparison between YOLOv7 and the improved YOLOv7: (a) precision; (b) recall; (c) mAP@0.5; (d) mAP@0.5:0.95.

In order to verify the model detection effect after adding different improvements, the improved feature fusion network, improved E-ELAN module, improved CA attention module and Siou loss function are ablated and analyzed. The comparison results of ablation experiments are shown in Table 6. As can be seen from the table, the mAP was significantly improved after the improvement of the feature fusion network, which shows that the low-level and high-level feature fusion and the addition of detection heads have played a very good role and significantly improved the detection performance. Adding improved E-ELAN and CA can obviously improve mAP, which shows that adding these two modules is effective. After using the Siou loss function, the model diagram is slightly improved. Finally, it can be seen that the Siou loss function not only accelerates the convergence of the model but also improves the detection performance of the model. In the table, “×” indicates that it was not added in YOLOv7, and “√” indicates that it was added in YOLOv7.

Table 6. Ablation experimental results.

Experiment	SIoU	Improved Feature Fusion Network	Improved E-ELAN	Improved CA	mAP@0.5
1	×	×	×	×	90.8
2	✓	×	×	×	90.9
3	✓	✓	×	×	91.7
4	✓	×	✓	×	91.3
5	✓	×	×	✓	92.5
6	✓	✓	✓	×	92.3
7	✓	×	✓	✓	92.9
8	✓	✓	×	✓	92.6
9	✓	✓	✓	✓	93.5

6. Conclusions

In this paper, three common diseases of grape leaves, namely black rot, black rot and black measles, were taken as the research objects. An improved YOLOv7-based grape leaf disease detection algorithm was proposed to counteract the problem of poor detection effect of small spots. The algorithm first constructs a new feature layer and introduces PANet to improve the E-ELAN structure. Secondly, Maxpool is introduced into the original CA attention to obtain more abundant feature information, and the global information module is introduced into the CA attention to assign different weights to deep image features so as to obtain channel information and position information and improve the recognition effect of the model. Finally, the loss function CIoU is replaced by SIoU to enhance the positioning ability of the network. Compared with the original YOLOv7, the algorithm proposed in this paper increases the mAP by 2.7% and has a good detection effect in both natural and simple environments.

This paper aimed to detect three kinds of diseases in grape leaves. In order to verify the robustness of the model proposed in this paper, it is necessary to increase the categories in the grape disease dataset to obtain a more comprehensive effect. In addition, the algorithm can be applied to the detection of other agricultural diseases. Building an agricultural disease detection platform through mobile phones can also provide farmers with convenient and efficient disease detection services.

Author Contributions: Conceptualization, M.Y., X.T. and H.C.; methodology, M.Y. and X.T.; software, X.T.; validation X.T.; writing—review and editing M.Y., X.T. and H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the high-end foreign experts' introduction program (G2022012010L).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: <https://github.com/spMohanty/PlantVillage-Dataset> (1 February 2018).

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

References

- Global Grape Production Will Reach 27.3 Million Tons, with China's Production Ranking First. Available online: <http://news.foodmate.net/2023/07/665306.html> (accessed on 8 January 2024).
- Grape Harvest in the Southern Hemisphere, a Few Happy and a Few Sad. Available online: https://m.163.com/dy/article_v2/I610I88M0553OUSU.html (accessed on 8 January 2024).
- Chen, J.; Zhang, D.; Nanekaran, Y.A. Identifying plant diseases using deep transfer learning and enhanced lightweight network. *Multimedia Tools Appl.* **2020**, *79*, 31497–31515. [CrossRef]
- Wang, D.; Chai, X. Application of machine learning in plant diseases recognition. *J. Chin. Agric. Mech.* **2019**, *40*, 171–180. [CrossRef]
- Adeel, A.; Khan, M.A.; Sharif, M.; Azam, F.; Shah, J.H.; Umer, T.; Wan, S. Diagnosis and recognition of grape leaf diseases: An automated system based on a novel saliency approach and canonical correlation analysis based multiple features fusion. *Sustain. Comput. Inform. Syst.* **2019**, *24*, 100349. [CrossRef]

6. Agrawal, N.; Singhai, J.; Agarwal, D.K. Grape Leaf Disease Detection and Classification Using Multi-Class Support Vector Machine. In Proceedings of the 2017 International Conference on Recent Innovations in Signal Processing and Embedded Systems(RISE), Bhopal, India, 27–29 October 2017; pp. 238–244.
7. Tian, Y. Grape disease recognition based on texture feature and support vector machine. *Chin. J. Sci. Instrum.* **2005**, *26*, 606–608.
8. Tian, Y.; Li, T.; Li, C.; Piao, Z.; Sun, G.; Wang, B. Grape disease image recognition method based on support vector machine. *Trans. Chin. Soc. Agric. Eng.* **2007**, *23*, 175–180.
9. Mohan, K.J.; Balasubramanian, M.; Palanivel, S. Detection and recognition of diseases from paddy plant leaf images. *Int. J. Comput. Appl.* **2016**, *144*, 143–156.
10. Padol, P.B.; Sawant, S.D. Fusion Classification Technique Used to Detect Downy and Powdery Mildew Grape Leaf Diseases. In Proceedings of the International Conference on Global Trends in Signal Processing, Jalgaon, India, 22–24 December 2016.
11. Waghmare, H.; Kokare, R.; Dandawate, Y. Detection and Classification of Diseases of Grape Plant using Opposite Colour Local Binary Pattern Feature and Machine Learning for Automated Decision Support System. In Proceedings of the 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 11–12 February 2016; pp. 513–518.
12. Jaisakthi, S.M.; Mirunalini, P.; Thenmozhi, D. Grape Leaf Disease Identification Using Machine Learning Techniques. In Proceedings of the 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 21–23 February 2019; pp. 1–6.
13. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* **2016**, *7*, 1419. [[CrossRef](#)] [[PubMed](#)]
14. Ramcharan, A.; Baranowski, K.; McCloskey, P. Deep learning for image-based cassava disease detection. *Front. Plant Sci.* **2017**, *8*, 1852. [[CrossRef](#)]
15. Liu, B.; Zhang, Y.; He, D.J. Identification of apple leaf diseases based on deep convolutional neural networks. *Symmetry* **2017**, *10*, 11. [[CrossRef](#)]
16. Liu, Z.; Yuan, X.; Weng, J.; Liao, Y.; Xie, L. Recognition of strawberry leaf powdery mildew disease based on convolutional neural networks. *Jiangsu J. Agric. Sci.* **2018**, *34*, 527–532.
17. Too, E.C.; Yujian, L.; Njuki, S. A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* **2019**, *161*, 272–279. [[CrossRef](#)]
18. Gong, A.; Jing, X. Crop disease image recognition based on multi-convolutional neural network model fusion. *J. Comput. Technol. Dev.* **2020**, *30*, 134–139.
19. Long, Y.; Liu, C.; Sun, K. Crop disease recognition based on deep convolutional neural networks. *J. Wuhan Univ. Light Ind.* **2020**, *39*, 17–22.
20. Fuentes, A.; Yoon, S.; Kim, S.C. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors* **2017**, *17*, 2022. [[CrossRef](#)] [[PubMed](#)]
21. Adhikari, S.; Saban Kumar, K.C.; Balkumari, L. Tomato Plant Diseases Detection System Using Image Processing. In Proceedings of the 1st KEC Conference on Engineering and Technology, Lalitpur, Nepal, 27–29 September 2018; Volume 1, pp. 81–86.
22. Liu, T.; Feng, Q.; Yang, S. Grape leaf disease detection method based on convolutional neural network. *Northeast. Agric. Univ.* **2018**, *49*, 73–83.
23. Qi, X. Construction and application research of grape pest and disease recognition model based on deep learning. *Xian Build. Univ. Sci. Technol.* **2022**. [[CrossRef](#)]
24. Jiang, P.; Chen, Y.; Liu, B. Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *IEEE Access* **2019**, *7*, 59069–59080. [[CrossRef](#)]
25. Zhao, J.; Qu, J. A Detection Method for Tomato Fruit Common Physiological Diseases Based on YOLOv2. In Proceedings of the 2019 10th International Conference on Information Technology in Medicine and Education (ITME), Qingdao, China, 23–25 August 2019; pp. 559–563.
26. He, X. *Research on Grape Leaf Disease Recognition Method Based on Deep Learning*; Northwest Agriculture and Forestry University: Xianyang, China, 2020.
27. Zhu, X.; Ma, H.; Ji, J. Detection and recognition analysis of blueberry canopy fruits based on Faster R-CNN. *J. South. Agric. Sci.* **2019**, *51*, 1493–1501.
28. He, Z.; Huang, J.; Liu, Q.; Zhang, Y. Segmentation of apple leaf disease based on asymmetric mixing convolutional neural network. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 221–230.
29. Zhuo, W.; Jian, W.; Wang, X.; Jia, S.; Bai, X.; Zhao, Y. Lightweight detection method of apple in natural environment based on improved YOLO v4. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 294–302.
30. Huang, T.; Huang, H.; Li, Z. Citrus fruit recognition method based on improved YOLOv5 model. *Huazhong Agric. Univ.* **2022**, *41*, 170–177.
31. Zhao, Z.; Guo, G.; Zhang, L.; Li, Y. A new anti-vibration hammer rust detection algorithm based on improved YOLOv7. *Energy Rep.* **2023**, *9*, 345–351. [[CrossRef](#)]
32. Zhuang, J.; Zhang, Y.; Wang, J. A Dragon Fruit Picking Detection Method Based on YOLOv7 and PSP-Ellipse. *Sensors* **2023**, *23*, 3803.
33. Zhang, L.; Zhang, L. MS-YOLOv7: YOLOv7 Based on Multi-Scale for Object Detection on UAV Aerial Photography. *Drones* **2023**, *7*, 188.

34. Ding, X.; Guo, Y.; Ding, G. ACNet: Strengthening the Kernel Skeletons for Powerful CNN via Asymmetric Convolution Blocks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 1911–1920.
35. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
36. Woo, S.; Park, J.; Lee, J.Y. Cbam: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
37. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 13713–13722.
38. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475.
39. Gevorgyan, Z. SIoU loss: More powerful learning for bounding box regression. *arXiv* **2022**, arXiv:2205.12740.
40. Hughes, D.; Salathé, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv* **2015**, arXiv:1511.08060.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.