agriculture

MDPI

Article

Improved Cotton Seed Breakage Detection Based on YOLOv5s

Yuanjie Liu [1,2,*,†], Zunchao Lv [1,2,†], Yingyue Hu [1,2], Fei Dai [2,3] and Hongzhou Zhang [1,2]

[1] College of Mechanicaland Electrical Engineering, Tarim University, Alar 843300, China
[2] Modern Agricultural Engineering Key Laboratory at Universities of Education Department of Xinjiang Uygur Autonomous Region, Tarim University, Alar 843300, China
[3] College of Mechanical and Electrical Engineering, Gansu Agricultural University, Lanzhou 730070, China
* Correspondence: 120050023@taru.edu.cn
† These authors contributed equally to this work.

Citation: Liu, Y.; Lv, Z.; Hu, Y.; Dai, F.; Zhang, H. Improved Cotton Seed Breakage Detection Based on YOLOv5s. *Agriculture* **2022**, *12*, 1630. https://doi.org/10.3390/agriculture12101630

Academic Editor: Raul Morais

Received: 31 August 2022
Accepted: 26 September 2022
Published: 7 October 2022

Abstract: Convolutional neural networks have been widely used in nondestructive testing of agricultural products. Aiming at the problems of missing detection, false detection, and slow detection, a lightweight improved cottonseed damage detection method based on YOLOv5s is proposed. Firstly, the focus element of the YOLOv5s backbone network is replaced by Denseblock, simplifying the number of modules in the backbone network layer, reducing redundant information, and improving the feature extraction ability of the network. Secondly, the collaborative attention (CA) mechanism module is added after the SPP pooling layer, and a large target detection layer is reduced to guide the network to pay more attention to the location, channel, and dimension information of small targets. Thirdly, Ghostconv is used instead of the conventional convolution layer in the neck feature fusion layer to reduce the amount of floating-point calculation and speed up the reasoning speed of the model. The CIOU loss function is selected as the border regression loss function to improve the recall rate of the model. Lastly, the model was verified using an ablation experiment and compared with the YOLOv4, Yolov5s, and SSD-VGG16 network models. The accuracy, recall rate, and map value of the improved network model were 92.4%, 91.7%, and 98.1%, respectively, and the average recognition time of each image was 97 fps. The results show that the improved network can effectively solve the problem of missing detection, reduce false detection, and have better recognition performance. This method can provide technical support for real-time and accurate detection of damaged cottonseed in a cottonseed screening device.

Keywords: Denseblock; collaborative attention; Ghostconv; CIOU loss function; YOLOv5s

1. Introduction

The quality of cottonseed is related to the germination rate. Fine screening of cottonseed is of great significance to improve cotton yield. As an important industrial pillar of agriculture in China, Xinjiang cotton accounted for 89.5% of the total cotton yield in 2021, and the total sowing area was 3759.26 acres. Delinted cottonseeds produced by ginning machines, delinting machines, centrifugal rollers, hoists, polishing machines, and other processes of cottonseed processing can easily lead to a large number of damaged seeds. Only by selecting and grading cottonseeds can we get higher-quality cotton and its byproducts. The quality of cottonseed directly affects the yield of cotton. The traditional manual and machine vision screening methods have the problems of low efficiency, poor accuracy, unrealizable real-time monitoring, and inability to identify damage to population cotton species. Therefore, nondestructive, rapid, accurate, and efficient detection of cottonseed is of great significance for improving cotton quality and promoting agricultural development in Xinjiang. In recent years, convolutional neural networks in deep learning have been widely used in the field of agricultural intelligent detection. Compared with traditional machine vision, the convolutional neural network has faster detection speed and higher accuracy. The target detection algorithm in deep learning is divided into two categories.

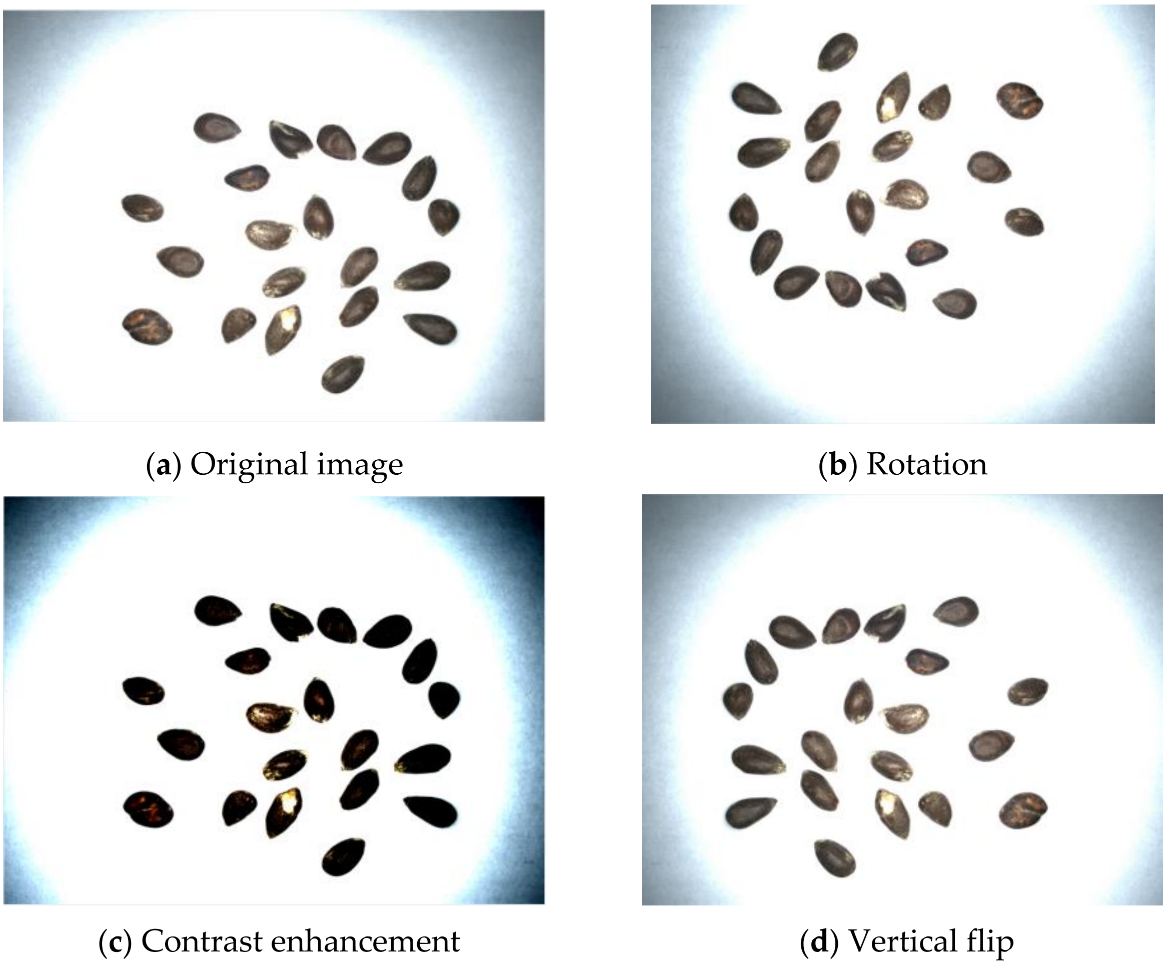
One is a two-stage detection algorithm, which detects first and then classifies, such as faster R-CNN and R-CNN [1].In the first stage, an a priori frame is generated by the region to be detected. In the second stage, the convolution network is used to classify and identify. However, the network depth is large, and the detection speed is slow. The other is a stage detection algorithm, which converts the border problem into a regression problem without generating the a priori frame, such as SSD and YOLO. The one-stage algorithm has the advantages of high precision, fast efficiency, short training speed, and low requirements for the hardware system of the equipment, which can better adapt to different environments. Therefore, it is more applicable in the field of agricultural intelligent detection. Deep learning has been widely studied in nondestructive testing of crop seeds. However, it is difficult to extract effective features due to different types of shadows, light, and targets in complex environments; hence, the robustness and adaptability of the detection algorithm need to be improved. Wentong Wu et al. [2] combined the YOLO algorithm with the multiscale anchor mechanism of faster R-CNN to improve the detection ability of the YOLOv5 algorithm for small and medium targets and achieve high adaptability to different size images. Jin Zhaozhao et al. [3] proposed an improved YOLOv3 network to solve the problems of high missed detection rate and low average accuracy in real road scenes. The network upgraded the existing output and added a 104 × 104 feature detection layer, which improved the accuracy of the network model and reduced the missed detection rate. However, the improved model increased the depth of the network model and reduced the detection speed. Tanvir Ahmad et al. [4] proposed a convolutional neural network based on an improved YOLOv1 network model in the field of target detection. The network modified the loss function to make it more flexible and reasonable in optimizing the network error. In addition, a network pyramid pooling layer was added, which improved the average accuracy of target detection on the Pascal VOC data. In the study by Zhang Chengliang et al. [5] the image segmented by the multichannel fusion algorithm was input into the improved YOLOv4 network model for classification and recognition. In conclusion, YOLO can achieve fast and accurate target location and recognition. Nondestructive detection methods based on deep convolutional neural networks are mostly used for crop morphology prediction [6–10]. Considering the huge differences in shape, size, and damage location of cottonseed and the expectation of obtaining higher detection results, although the method based on deep learning has high accuracy, it also has disadvantages such as high complexity, many parameters, large network scale, high computational cost, and insufficient real-time performance.

Therefore, a method to improve the YOLOv5s lightweight network model is proposed [11–14]. At present, no public cottonseed dataset can be found in the network. Therefore, this study used a cmos industrial digital camera with the MV-CE120-10UC series USB3.0 interface of Hikvision, and selected Xinluzao 78# as the research object. Aiming at the problems of a large number of parameters, missing detection, and false detection, the network model is improved and optimized. Firstly, the Denseblock module is used to replace the focus module in the backbone network layer, so that it can better retain the feature information of small targets and improve the accuracy of the network model. Secondly, the collaborative attention mechanism is introduced after the SPP pooling layer, which can not only avoid dimension reduction but also have cross-channel interaction [15–18]. Meanwhile, the complexity of the model is reduced, the feature expression ability is enhanced, and the model recall rate is improved. Lastly, a large target detection layer is reduced in the head structure, and the number of model parameters is reduced. The CIOU loss function is used as the border regression loss function to enhance the learning ability of the network structure, the optimization ability of the YOLOv5s model, and the detection speed of the model, enabling the detection model to better adapt to small target detection [19–22].

2. Materials and Methods

2.1. Image Data Acquisition and Preprocessing

A total of 10,000 Xinjiang Xinluzao 78# cotton seeds were selected for research. The image acquisition equipment was a Hikvision mv-ce120-10uc USB 3.0 cmos industrial digital camera, with a frame rate of 31.9 fps and resolution of 2000 × 1518. The camera lens was a Hikvision mvl-mf0828m-8mp 8 mm 800 W pixel, and the light source was an annular adjustable light source LED with a maximum power of 9W. The ratio of intact cottonseed to damaged cottonseed was 5:1. A total of 500 sets of cottonseed images were collected. In order to improve the generalization ability of the model and prevent the overfitting caused by the lack of image data during model training, 500 sets of cottonseed images were enhanced to 1500 sets by rotation, flipping, contrast enhancement, and other methods. The distribution of datasets is shown in Table 1. Before the model training, the position and category of cottonseeds were marked using a wizard assistant, and the marked XML file was converted into a TXT file with position and category information. The number of cottonseed categories in the dataset was two. As shown in Figure 1, the surface of the intact cottonseed was smooth, the color was brown, and the labeling information was good. Damaged cottonseeds had surface potholes and white markings. Through data enhancement, the cottonseed images in the dataset had a different shooting angle, shooting distance, resolution, and shooting light. All the differences mentioned above reflect the complexity and diversity of the dataset, greatly increasing the difficulty of detection. Thus, the robustness and generalization of the algorithm model could be verified.



(**a**) Original image

(**b**) Rotation

(**c**) Contrast enhancement

(**d**) Vertical flip

Figure 1. Cottonseed data enhancement.

Table 1. Data sets.

Projects	Annotation Information	Training Sets		Testing Set	
	Intact Cotton Seeds (Good) **Broken Cotton Seeds (Bad)**	**Number of Images**	**Number of Cotton Seeds**	**Number of Images**	**Number of Cotton Seeds**
Before data expansion		450	9000	50	1000
After data expansion		1350	24,000	150	3000

2.2. Damaged Cottonseed Detection Network

2.2.1. YOLOv5s Network Structure

Firstly, the input image was divided into S × S grids, and the targets in the grids were detected. Then, the boundary box, position confidence, and probability vector of the targets in the grids were predicted in one stage.

Backbone: The image was enhanced using Mosaic at the input. The image size was automatically filled, and the anchor frame adaptive image size was automatically calculated. The processed image entered the backbone network from the focus structure. The focus module could slice the image, as shown in Figure 2. The input channel of the image was expanded fourfold, and then the scale slice was spliced using Conv to obtain the feature fusion image. After the downsampling operation, the image size was reduced to 320 × 320 × 64. The focus module retained the original image information as much as possible, reduced the floating-point calculation, and improved the training speed of the model.

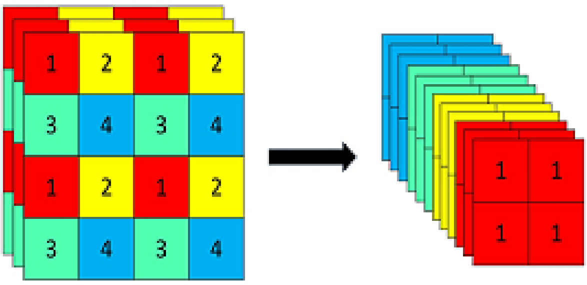


Figure 2. Focus.

Feature fusion layer (neck): Multiscale feature map fusion was completed using the FPN and PAN structure. The FPN structure uses upsampling to transfer and fuse the feature information from top to bottom. The PAN structure uses downsampling to transfer and integrate feature information from bottom to top, so that the YOLOv5 network model can obtain more abundant feature information.

Output-side detection layer (head): There were three detection layers on the output side, with sizes of 20 × 20, 40 × 40, and 80 × 80, corresponding to large, medium, and small targets, which contained the confidence of the detection target and the location information of the prior box. The Diou_nms non-maximum suppression (nms) was used to eliminate the prior box of occlusion overlap, such that the prior box of the maximum confidence was retained and the detection target information was output. The YOLOv5s network model is shown in Figure 3.

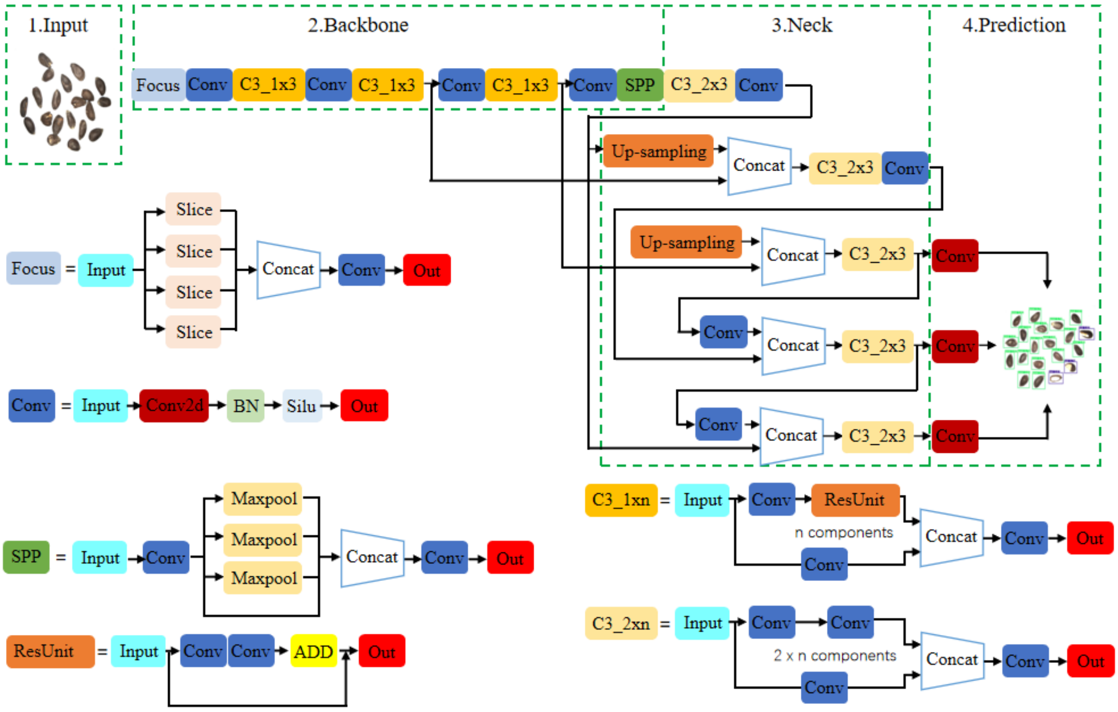


Figure 3. YOLOv5s algorithm network architecture flowchart.

2.2.2. Optimization of Backbone Network Layer

In YOLOv5s, the backbone mainly includes the focus, C3, Conv, and SPP. In the configuration file, depth_multiple represents the number of scaling factor control modules in the C3 module. Due to a large number of convolution operations, the feature information of small targets will gradually decrease or even disappear, making the identification of small targets more difficult. In order to solve the above problems and prevent gradient explosion, the feature extraction level in the backbone was simplified, and the number of C3 modules in the backbone was changed from (x3, x9, x9, x3) to (x2, x6, x6, x2); the optimized model could effectively solve the problems of large parameters and loss of feature information caused by the convolution kernel, Therefore, the training and convergence speed of the network were accelerated.

2.2.3. Denseblock Module

The core of Denseblock is the use of a large amount of Densenet in the network, where the expressions for Densenet are shown in Equations (1) and (2).

$$X\varphi = H\varphi[X\varphi_{-1}] + X\varphi_{-1}. \tag{1}$$

$$X\varphi = H\varphi([X_0, X_1, \ldots, X_{\varphi-1}]). \tag{2}$$

In Equation (1), $X\varphi$ represents the output, H_φ represents the nonlinear transformation of the network, and the residual network of the φ_{-1} layer is represented by $X_{\varphi_{-1}}$. The network output is composed of $X\varphi_{-1}$ layers of nonlinear transformation output layer and $X\varphi_{-1}$ layers. The output adopts the absolute value addition method without changing the number of channels. In DenseNet, it can be seen from Equation (2) that the output feature maps of X_0 to $X\varphi_{-1}$ layers are spliced by network, and then the nonlinear transformation is carried out. Each layer of input comes from the output of all previous layers, i.e., channel accumulation. In this way, the scale between feature maps is the same, which can extract feature map information more effectively.

The Denseblock structure is shown in Figure 4. There are H_1,H_2, H_3, and H_4 (BN layer, Relu activation function, and convolution layers Conv1 × 1 and Conv3 × 3) nonlinear transformation output layers. Densenet is a convolutional neural network with high connection characteristics, where there is a direct connection between any two levels. The input of each level in the network is the sum of all previous outputs, and the learning characteristics of this level are directly transmitted to the subsequent levels. This can not only reduce the computation via dimension reduction, but also integrate the characteristics of each channel. At the same time, the feature extraction ability ratio is enhanced, and the splicing fusion efficiency of feature extraction is accelerated. More importantly, due to the regularization of the network, it can also effectively reduce overfitting in smaller training sets.

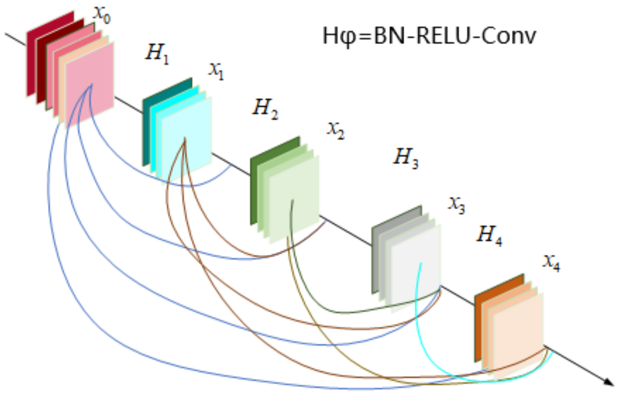


Figure 4. Denseblock module structure diagram.

Replacing the focus structure with Denseblock can double downsampling and extract more scale feature information in the image. The DenseNet connection method can avoid gradient explosion, and it does not need to extract a large amount of irrelevant feature map information. It can not only reduce valley sink information in the feature extraction layer, but also reduce the occurrence of overfitting and underfitting. Adding a small number of parameters and floating-point calculation in the network can improve the ability of network feature extraction and strengthen the training of small target detection. In addition, since the final output combines all the feature outputs of the previous layer, the feature information of small targets can be better maintained during the downsampling period. The subsequent convolution operation strengthens the extraction of small target information again. The improved network model has better detection performance for the training of small targets such as cottonseed.

2.2.4. Embedded Collaborative Attention Mechanism

The general attention mechanism processes the feature information in maximum pooling or average pooling, which will lose the spatial information in the network. Adding the collaborative attention mechanism after the network pooling layer (SPP) can integrate the position information and channel information in the feature map, reduce the calculation of the network, and convert the tensor feature information of the data into a separate feature vector channel. The collaborative attention mechanism can obtain the cross-channel direction and position information, so that the network model can more accurately locate and identify the important feature information in the process of feature extraction, and then improve the sensitivity of the network to the feature map information. The collaborative attention mechanism is shown in Figure 5.

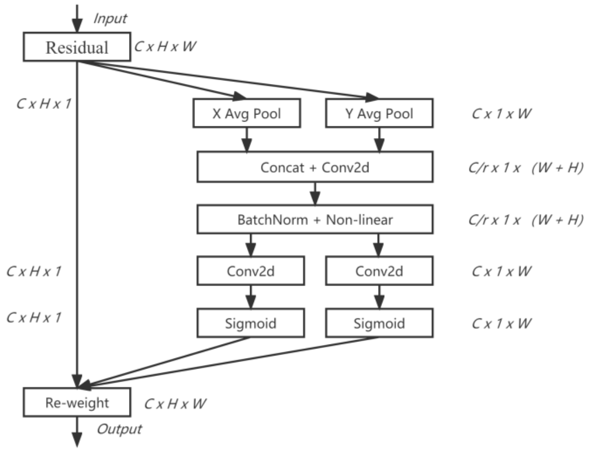


Figure 5. Coordinate attention.

In order to get the attention of the width and height of the image and encode the accurate position information, the global pooling of the width and height direction of the input feature map is carried out to obtain the feature map in these two directions. Then, the feature maps in these two directions are spliced together and sent to the convolution module with a 1 × 1 convolution kernel. Then, the feature map after batch normalization is transformed by nonlinear activation function to obtain the feature map f with channel dimension reduction.

$$\mathrm{f} = \delta\left(\mathrm{F}_1\left(\left[\mathrm{Z}^{\mathrm{h}}, \mathrm{Z}^{\mathrm{w}}\right]\right)\right), \tag{3}$$

where δ represents the nonlinear activation function, and f represents the intermediate feature map encoding spatial information in the width and height directions, $\left[\mathrm{Z}^{\mathrm{h}}, \mathrm{Z}^{\mathrm{w}}\right]$ represents stitching along the spatial dimension, and F_1 represents the stitching feature map after pooling.

Equations (4) and (5) express the feature graph f with a 1 × 1 convolution kernel according to the original w and h, which obtains the same number of channels as the original feature graphs F_h and F_w. Here, σ represents the sigmoid activation function, after which F_h and F_w obtain the attention weight g^h in the height and g^w in the width direction, respectively.

$$\mathrm{g}^{\mathrm{h}} = \sigma\left(\mathrm{F}_{\mathrm{h}}\left(\mathrm{f}^{\mathrm{h}}\right)\right). \tag{4}$$

$$\mathrm{g}^{\mathrm{w}} = \sigma(\mathrm{F}_{\mathrm{w}}(\mathrm{f}^{\mathrm{w}})). \tag{5}$$

The attention weight in the height and width direction of the feature map obtained in Equations (4) and (5) is calculated by multiplying and weighting the original feature map, such that the attention weight in the width and height direction of the original feature map is embedded. Its output expression is as follows:

$$\mathrm{y}_\mathrm{c}(\mathrm{i},\mathrm{j}) = \mathrm{x}_\mathrm{c}(\mathrm{i},\mathrm{j}) \times \mathrm{g}_\mathrm{c}^\mathrm{h}(\mathrm{i}) \times \mathrm{g}_\mathrm{c}^\mathrm{w}(\mathrm{i}). \tag{6}$$

2.2.5. Ghostconv Convolution

After the introduction of the collaborative attention mechanism in the YOLOv5s network, the detection accuracy of damaged cottonseed can be significantly improved, but the real-time and lightweight requirements are difficult to meet. In order to reduce the amount of floating-point calculation, the Ghostconv convolution is used to replace the ordinary convolution in the feature fusion layer. The Ghostconv convolution is divided into two parts. The principle is as follows: first, partial one-dimensional convolution is generated; then the part of the information in the ordinary convolution layer is used to generate the feature map with fewer channels while using less calculation. A simple linear operation is used to convolution the channel feature map so as to obtain more feature map information of the spectrum. Lastly, the feature map generated by the one-dimensional convolution and linear operation is spliced into a new feature map. Using Ghostconv can eliminate a large amount of redundant information in the fusion part of the feature map and accelerate the reasoning speed of the network model.

As shown in Figure 6b, the input is the characteristic feature map $X\epsilon R^{c\times h\times w}$, where c is the channel, h is the height, and w is the width. The feature map information Y' is obtained via conventional convolution of X.

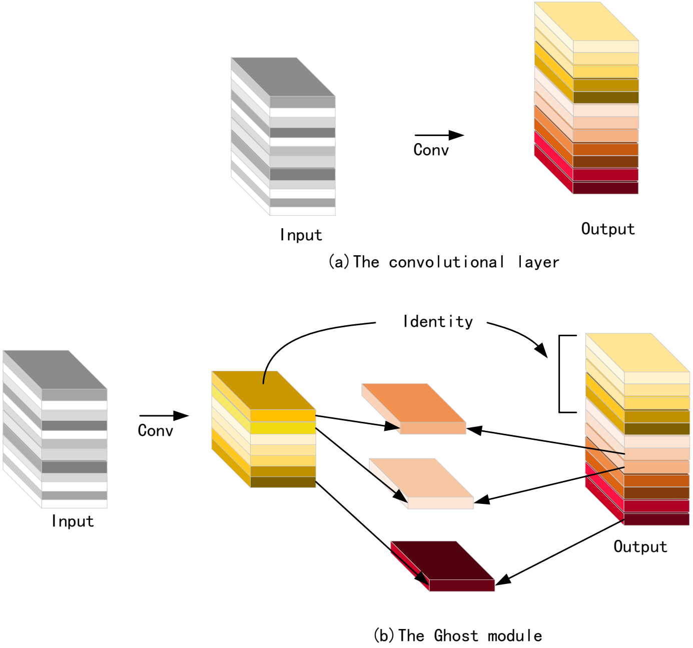


Figure 6. Ghostconv convolution.

$$Y' = X * f', \tag{7}$$

where * is the convolution operation, and f' is the convolution filter of this layer. $Y'\epsilon R^{h'\times w'\times n}$ is the output feature map of the n-channel. The feature map of each channel in Y' is represented by y'_i, and $\Phi_{i,j}$ represents a simple linear transformation. Finally, the feature map y_{ij} is obtained as a function of $\Phi_{i,j}$. The expression is as follows:

$$y_{ij} = \Phi_{i,j}\left(y'_i\right). \tag{8}$$

Next, the Y' and y_{ij} output features are spliced. The original intrinsic feature map information can be preserved by identity, and the intrinsic feature map information to the final output feature information can be obtained by linear operation on each channel, which consumes less computation than ordinary convolution.

2.2.6. Reduce the Detection Layer

The input size of YOLOv5s is set to 640 × 640. YOLOv5s has three recognition and detection layers, with corresponding detection sizes of 20 × 20,40 × 40 and 80 × 80, which are used to identify large, medium and small detection targets respectively. The cotton seed in this study belongs to the small target detection, because its small size will not change greatly in the picture. Based on this, the recognition and detection layer of YOLOv5s is simplified from large, medium and small to medium and small, and the Conv convolution layer of the 21st layer of feature fusion layer, the Concat splicing of the 22nd layer, and the C3 module of the 23rd layer are removed to reduce the model parameters and calculation, and enhance the generalization of model training and detection speed.

2.2.7. Optimization of Loss Function

The loss function in YOLOv5 consists of three parts: confidence loss l_{obj}, category loss l_{cls}, and position loss of prior and real frames l_{box}. The overall loss formula is as follows:

$$\text{loss} = l_{obj} + l_{cls} + l_{box}. \tag{9}$$

YOLOv5 uses the binary cross-entropy loss function to calculate the category probability and confidence loss, and then uses the measurement loss calculated by IOU to replace the regression loss. The IOU calculation formula is as follows:

$$\text{IOU} = \frac{|A \cap B|}{|A \cup B|}, \tag{10}$$

where A and B represent the prior prediction and real prediction frames, respectively. The intersection of AB is the module, and the union of AB is denominator. When there is no intersection between A and B, the value of IOU is 0, which cannot map the real distance between the prior prediction box and the real prediction box; the gradient is 0 and cannot be optimized. Therefore, the CIOU loss function is introduced into YOLOv5 as a border predictor regression loss function, as shown in Equation (11).

$$\text{GIoU} = \text{IOU} - \frac{|C - (A \cup B)|}{|C|}. \tag{11}$$

However, the GIOU loss function is unable to judge when the a priori box coincides with the true box completely. As shown in Figure 7, the loss value cannot be judged when the IOU value between the a priori box and the prediction box is 1.

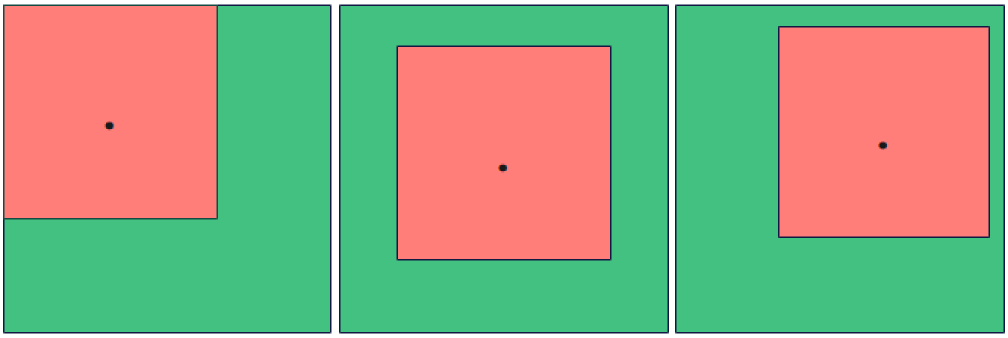


Figure 7. Same GIOU value.

Therefore, the CIOU loss function is used as the border regression loss function. As shown in Equation (12), using CIOU as the loss function can better judge the center distance, overlapping area, center point, and aspect ratio between the real frame and the prior frame. When the real frame overlaps with the prior frame, considering the center distance information of the prior frame and the width–height ratio information of the real frame, the distance between the two centers can be directly measured to optimize the boundary regression results.

$$L_{CIOU} = 1 - \text{IOU} + \frac{\rho^2(b, b^{gt})}{c^2} + av, \tag{12}$$

where the center point of the prior frame is represented by b, b^{gt} represents the center point of the real box, the Euclidean distance is represented by ρ, c represents the distance of the diagonal line in the minimum enclosing rectangle between the prior box and the real box, a represents the weight information, and v represents the aspect ratio coefficient. The calculation formulas are shown in Equations (13) and (14).

$$\mathrm{v} = \frac{4}{\pi^2}\left(\left(\arctan\frac{\omega^{\mathrm{gt}}}{\mathrm{h}^{\mathrm{gt}}}\right) - \arctan\frac{\omega}{\mathrm{h}}\right)^2. \quad (13)$$

$$\mathrm{a} = \frac{\mathrm{v}}{(1 = \mathrm{IOU}) + \mathrm{v}}. \quad (14)$$

The CIOU loss function is used to calculate the loss function of the model, which solves the problem of the same loss value of the overlapping area between the prior frame and the real frame in the target detection. It makes the calculation of the loss value in the network model more accurate, enhances the authenticity of the loss function predicted by the border regression, and improves the detection performance of the model.

2.2.8. Improved YOLOv5s Network Structure

An excessive number of parameters in the backbone network will slow down the network detection speed. In order to control the network depth, the number of original C3 modules is reduced, and the number of model parameters is reduced. Denseblock is used to replace the focus module. Although a small number of parameters and calculations are added, the extraction ability of small target feature information is improved, the gradient descent in the network is optimized, and the gradient explosion is prevented. Secondly, the collaborative attention mechanism is added after the SPP pooling layer, which can make full use of the channel position information to obtain the feature map with attention weight in the width and height directions, so as to improve the weight ratio of channel position in the splicing map of important feature information, and enhance the generalization of the model and the ability of feature extraction. Then, Ghostconv is used in the neck to replace the common convolution in the feature fusion structure, which can learn more small-scale feature information and reduce the false and missed detection of the model under dense targets. The CIOU loss function as the loss function of the border regression can better reduce the error between the prior box and the prediction box, as well as improve the model recall rate. Lastly, cottonseeds represent small targets for detection, thus reducing the large target feature-scale detection layer in Head, which is conducive to improving the accuracy of the model. The improved network structure is shown in Figure 8.

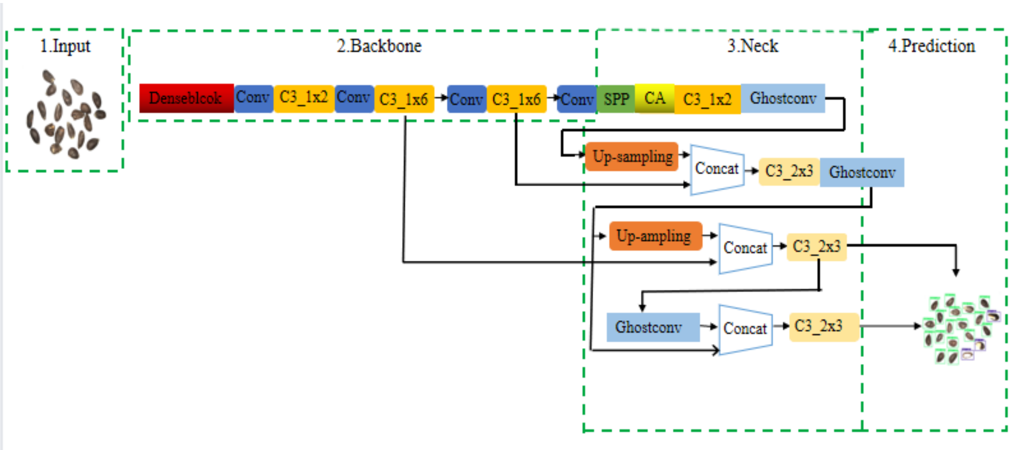


Figure 8. Improved YOLOv5s model diagram.

2.3. Model Training Metrics

The commonly used evaluation indices in target detection models are precision P (precision), recall R (recall), single-category precision (AP), average precision (map@0.5), F1, and model size (MB). AP refers to the area under the PR curve, while the precision and recall rates and the area surrounded by the coordinate axis are calculated using the integral. Map values can be obtained by summing AP values of a single category and dividing them into the number of total categories. Map@0.5 indicates the average value of each AP category when the confidence is 0.5. The specific calculation formulas are shown in Equations (15)–(19).

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{15}$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{16}$$

$$\mathrm{AP} = \frac{\sum \mathrm{P}}{\mathrm{Num}(\mathrm{TotalObjects})}. \tag{17}$$

$$\mathrm{mAP} = \frac{\sum \mathrm{AP}}{\mathrm{N}(\mathrm{class})}. \tag{18}$$

$$\mathrm{F1} = 2 \times \frac{\mathrm{PR}}{\mathrm{P} + \mathrm{R}}. \tag{19}$$

The true positive index in Equations (15)–(19) is TP, and the false positive index is FP. The number of predicted categories of cottonseed in model training is two.

2.4. Experimental Equipment and Parameter Settings

The device was a Windows 10 with x64 bit operating system, 16 GB of memory, NVIDIA GeForce RTX 3060 graphics card, Intel (R) Core (TM) i5-10400 F processor, and 2.90 GHz CPU processing speed. The PyTorch deep learning framework(Pytorch 1.9.0, Python 3.7, CUDA 11.3). IDE for PyCharm Community Edition 2021.9. In order to obtain a better damage detection model based on YOLOv5s, the model parameters were fine-tuned. Table 2 shows the improved YOLOv5s tuning parameters.

Table 2. YOLOv5s training parameters.

Parameters	Values
Input size	640 × 640
Batch size	16
Classes	2
Epoch	300
Learning rate	0.01
Termination learning rate	0.2
IOU threshold	0.5
Optimizer	SGD
Prediction box size	[10, 13, 16, 30, 33, 23] [30, 61, 62, 45, 59, 119]
Scale layer	[128, 256]

2.5. Training Results of Population Cottonseed Damage Model Based on YOLOv5s

In training and testing, the loss curve includes position loss, confidence loss, and classification loss, as shown in Figure 5. A lower value of the loss function indicates a smaller error and a more accurate forecast [23–26]. Trust loss is a measure of the learning degree of cottonseed damage characteristics. A smaller loss function indicates a higher accuracy. Classification loss represents an algorithm that can accurately predict whether cottonseed is damaged. A lower loss value indicates that it is more accurately classified. In this article, because the cottonseeds to be tested were either intact or damaged, the value was two. In the 100 epochs of damaged cottonseed detection, the loss value decreased rapidly and the convergence speed of the training curve increased. The experimental results show that the accuracy rate, recall rate, and mAP value of the model are greatly improved during this period. As the training progressed, the slope of the training curve gradually decreased and reached a stable state after 250 epochs. Therefore, this paper took the training output of 300 iterations as the target identification mode. Figure 9a,b shows the relevant training index and PR curve of the model without data Enhancement, while Figure 9c,d shows the relevant training index and PR curve of the model after data enhancement.

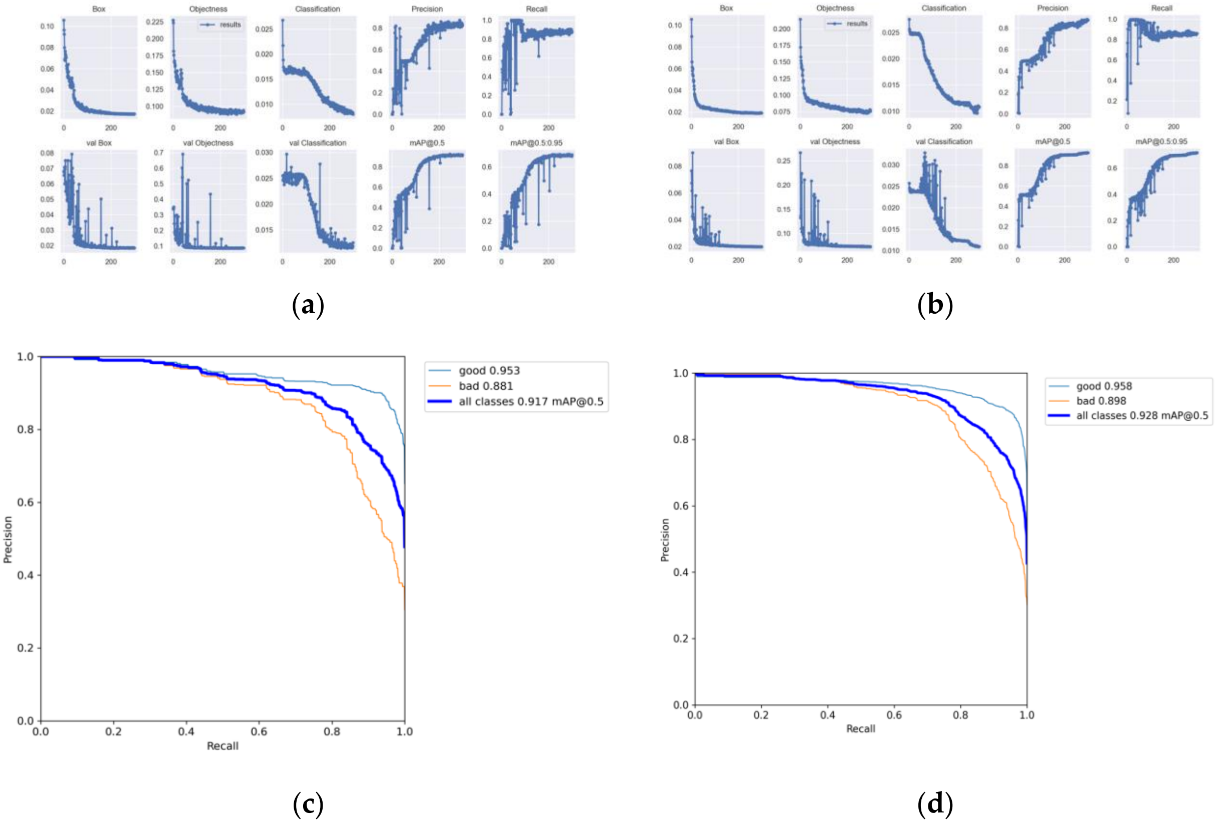
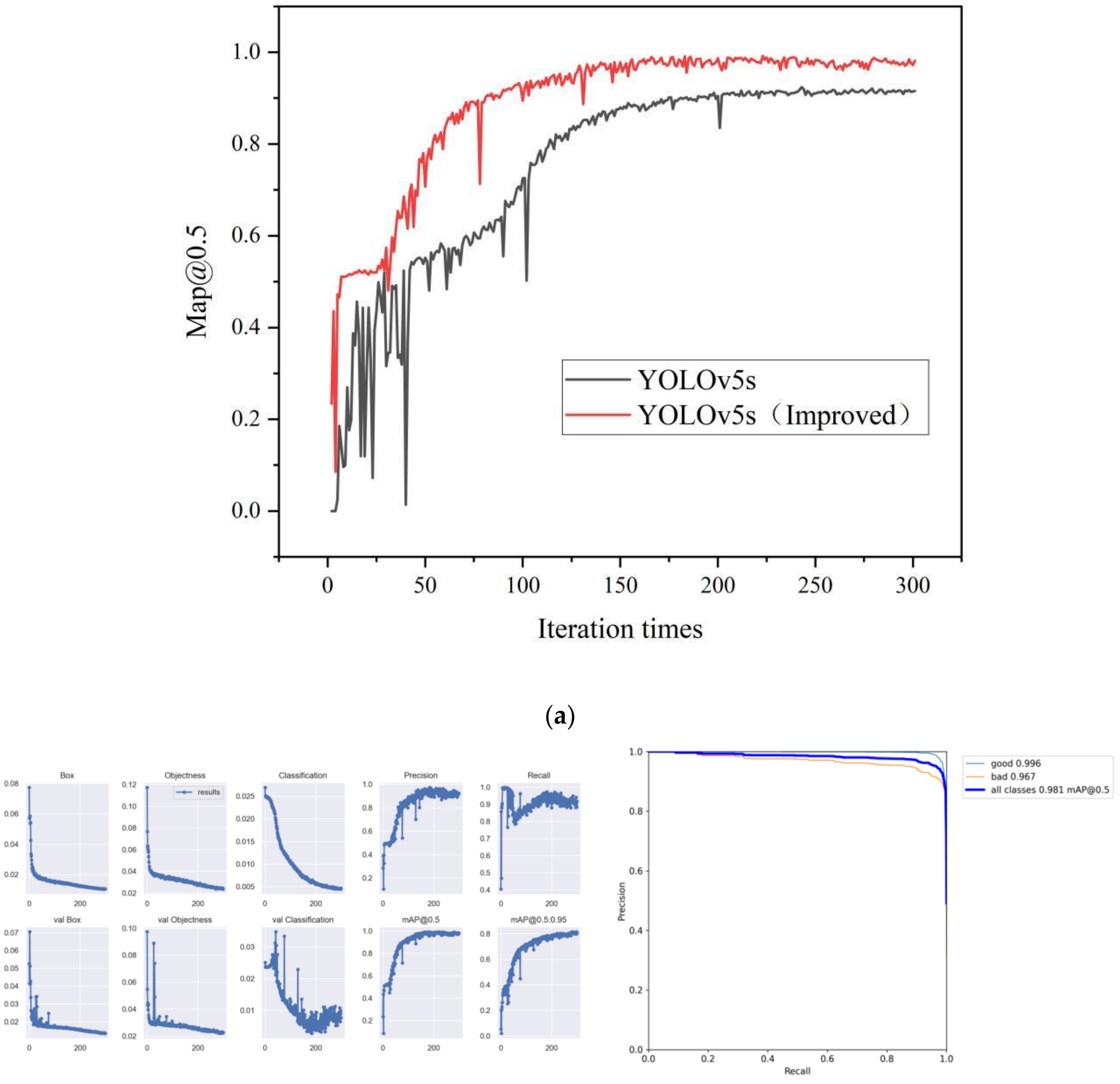


Figure 9. YOLOv5s training results. (**a**)YOLOv5s network training results; (**b**)YOLOv5s Network Training Results After Data Enhancement; (**c**)YOLOv5s PR curve; (**d**)YOLOv5s PR curve after data enhancement.

3. Analysis of Results

3.1. Model Training Results

The average precision (IOU = 0.5) curve of the improved model after 300 rounds of training is shown in Figure 10. The ordinate is the map@0.5 value, while the abscissa is the number of iterations. Figure 10a represents the indicator for each part of the improved YOLOv5s, while Figure 10b is the PR curve. It can be seen that, in the first 100 rounds, the improved YOLOV5s model converged faster than the original YOLOV5s model, and the training curve of the model tended to be stable after 250 rounds. After 300 rounds, the training effect of the original model and the improved model was significantly different. There was no overfitting or underfitting in either model. However, compared with the original model, the improved YOLOV5s model had a significant improvement in the average accuracy of map@0.5.



(**a**)

(**b**) (**c**)

Figure 10. Model map comparison diagram. (**a**) Model comparison before and after YOLOv5s improvement; (**b**) YOLOv5s improved training results; (**c**) YOLOv5 s improved PR curve diagram.

3.2. Analysis of Test Results under Different Models

In order to verify the performance of the improved YOLOv5 algorithm, the improved YOLOv5s network was compared with the original YOLOv5s, YOLOv4, and SSD-VGG16 networks on a test set of 150 cottonseed images. The accuracy, recall rate, map value, F1, detection speed, size, and parameters were used as evaluation indicators. The performance comparison of the three detection networks for group damaged cottonseed is shown in Figure 11.

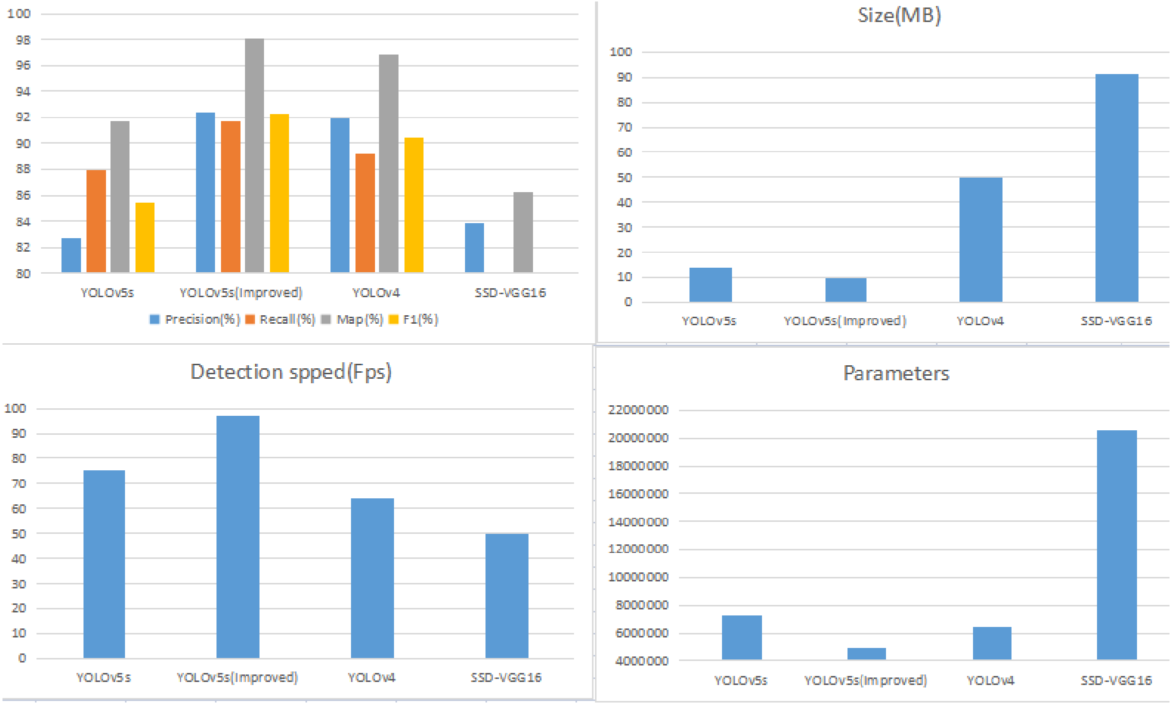


Figure 11. Performance comparison of four target detection networks for group cottonseed damage detection.

Figure 11 shows that the accuracy, recall, map, and F1 values of the improved YOLOv5 s were 92.4%, 91.7%, 98.1%, and 92.1%, respectively, exceeding the three compared networks in performance. Compared with the original Yolov5, the precision was improved by 9.7%. The map value was improved by 6.4%, which indicates that the improved Yolov5s network was better than other detection models in identifying damaged cottonseeds. In terms the recognition speed of the model, the inference time per image of the improved YOLOv5s was 0.067 s (97 fps), which was the fastest among the four networks (1.5 times, 1.94 times, and 1.29 times that of SSD-VGG16, YOLOv4, and YOLOv5s, respectively). Moreover, in terms of the scale of the network model, the size of the improved YOLOv5s network model was only 9.88 MB, and the number of parameters was 4.98 × 10^6, which was much smaller than the other three network models. The size of YOLOv4 network model was 50.2MB, and the number of parameters was 6.4 × 10^6. The size of the SSD-VGG16 network model was 91.1MB, and the number of parameters was 2.05 × 10^7. The size of the YOLOv5s network model was 14.2MB, and the number of parameters was 7.23 × 10^6.

Obviously, the improved YOLOv5s network had the smallest size. On the premise of improving accuracy and detection speed, the damage detection effect of cottonseeds was better. The improved YOLOv5s network had the highest precision, recall, map, and F1 values compared to the other three network algorithms. The reduction in network depth and width improved the accuracy. The reduction in a large amount of irrelevant semantic information in the model improved the network speed and reduced the network size. In general, as the depth and width of the network model increased, the expression and learning capability of the network did not necessarily increase, and the performance did not necessarily improve. When the performance reached a certain level, the performance improvement was not significant with a further increase in the depth and width of the network, and the number of computations and parameters increased significantly. Therefore, greatly increasing the depth and width of the network would not make the network more practical.

The test results of 150 cottonseed images were statistically analyzed. There were 3000 cottonseeds, including 2500 lossless cottonseeds and 500 damaged cottonseeds. The improved YOLOv5s model misjudged 31 cottonseeds without missing detection. The original YOLOv5s model misjudged 84 cottonseeds and missed 40 cottonseeds. The YOLOv4 model misjudged 79 cottonseeds and missed 30 cottonseeds. The SSD-VGG model detected 30 misjudged cottonseeds and 20 missed cottonseeds. Figure 11 shows the misjudgments and omissions under different models.

In order to verify the detection performance of the improved YOLOv5s, it was compared with the SSD-VGG16, YOLOv4, and YOLOv5s network models. The performance of the four network models on the same population of cottonseed images is shown in Figures 12 and 13. SSD-VGG16 suffered from missed detection, while the original YOLOv5s and YOLOv4 suffered from false detection. The improved YOLOv5s outperformed the other three network models in terms of image detection. Under the guarantee of improved accuracy, the model was lightweight. The results show that the improved YOLOv5s network is more suitable for ensuring the accuracy of detection in complex environments, and it has strong robustness.

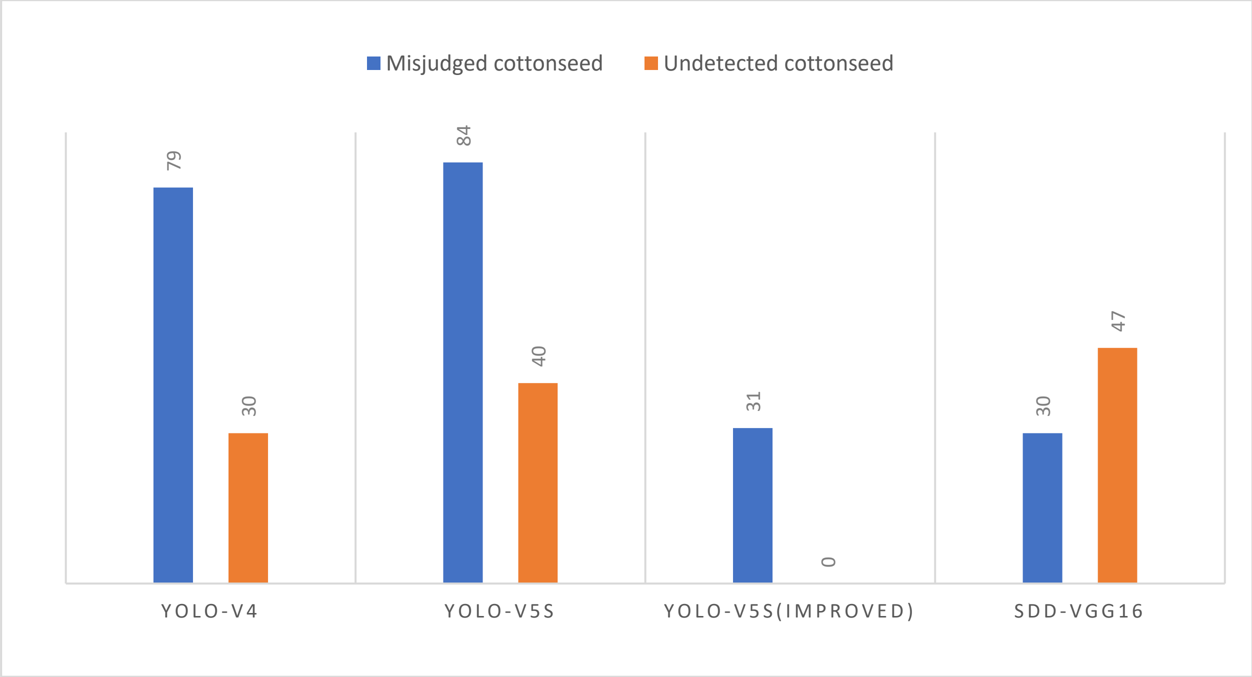
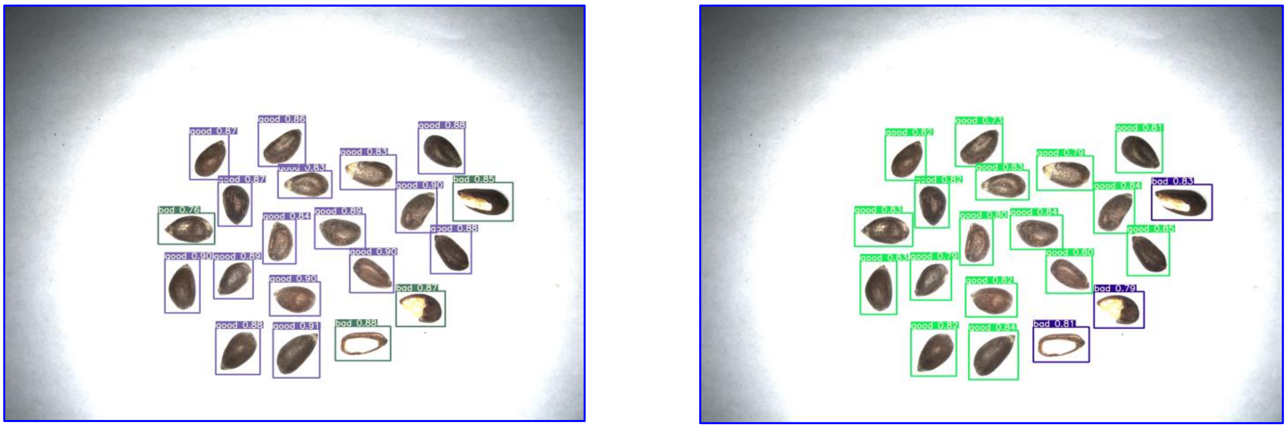


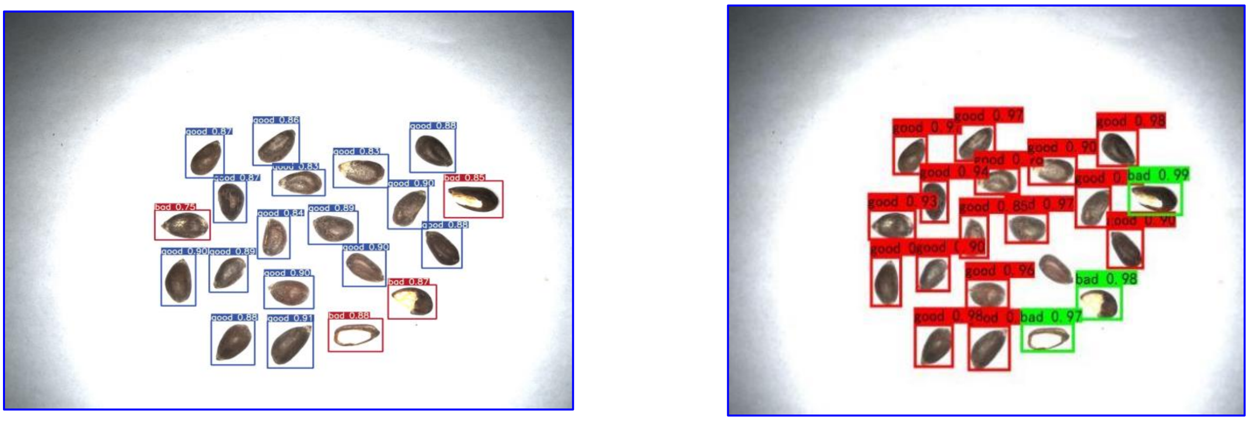
Figure 12. Different model misclassification cases.



(**a**) Test results of improved YOLOV5s model

(**b**) Test results of YOLOV5s model

Figure 13. *Cont.*



(**c**) Test results of YOLOV4model (**d**) Test results of SSD-VGG16model

Figure 13. Detection of different models.

3.3. Ablation Experiments

The ablation experiment results are shown in Table 3. The image samples were expanded, and the focus module was replaced by Denseblock. The collaborative attention mechanism was added, and the large target detection layer was reduced in the neck feature fusion layer. Ghostconv was used to replace the ordinary convolution to reduce the redundant information of the fusion feature layer. The CIOU loss function was used as a border regression loss function to improve the extraction ability of network features, strengthen the training of small target feature information, and improve the accuracy, recall rate, and map value of the model. Compared with YOLOv5s, the average accuracy of the above improved method increased by 6.4%.

Table 3. Results of ablation experiments.

Models	Expanded Data	Denseblock	CoordAtt	Neck	CIOU Loss Function	Ghostconv	P	R	Map (%)
YOLOv5s							82.7	87.9	91.7
YOLOv5s	√						87.6	86.0	92.8
YOLOv5s	√	√					86.1	86.5	93.4
YOLOv5s	√	√	√				87.8	90.1	94.5
YOLOv5s	√	√	√	√			87.4	90.3	95.0
YOLOv5s	√	√	√	√	√		86.1	90.7	95.3
YOLOv5s	√	√	√	√	√	√	92.4	91.7	98.1

4. Conclusions

Aiming at the problems of low detection accuracy, high missed detection rate and false detection rate, and slow detection speed of YOLOv5s for cottonseed, an improved lightweight damaged cottonseed detection method based on YOLOv5s was proposed.

The improved YOLOv5s network can be used to detect the damage of cottonseeds. Firstly, the cottonseeds were divided into two categories: intact and damaged. A total of 500 images of cottonseeds were collected and expanded to 1500 to improve the generalization ability of the training model. Secondly, the focus module of the main structure of the YOLOv5s network was replaced by Denseblock, which improved the feature extraction ability of the model and optimized the number of backbone modules. Then, the convolution layer in the neck structure was replaced by Ghostconv to reduce the number of network parameters and accelerate the floating-point calculation. Lastly, the large-scale detection layer was reduced to further reduce the number of network parameters to make the model lightweight. The DIOU loss boundary box loss function of the YOLOv5 network was changed to the CIOU loss function, so as to improve the accurate positioning ability of the prediction box and enhance the convergence effect of the model. The experimental comparison showed that the map value of the improved YOLOv5 network in this paper was 6.4% higher than the original YOLOv5, and the detection speed reached 97 fps per second. The model size occupied 9.88 MB, reflecting good robustness and generalizability. This paper provides an accurate and rapid detection strategy for group cottonseed damage. The problem of missed detection was solved, while meeting the real-time detection requirements in actual production.

This method solves the problem of detection of cottonseed breakage, which is difficult for traditional machine vision. A deep learning convolutional neural network was used to replace the traditional machine vision recognition method. The improved YOLOv5s network model can be used for the detection of cottonseed damage. Compared with the target detection technology of YOLOv4, YOLOv5s, and SSD-VGG16 with respect to cottonseed, the detection accuracy and speed were significantly improved, and the model parameters were effectively reduced. The improved model is more conducive to the deployment of embedded devices. In the future, the damaged cottonseeds will be subdivided into categories for improved detection, which is of great significance for fine screening.

Author Contributions: Writing—original draft preparation, Y.L. and Z.L.; writing—review and editing, Y.H.; resources, F.D.; data curation, Y.L.; visualization, H.Z.; supervision, Y.L.; project administration, H.Z. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China, grant number 61701334, the China Agricultural University-Tarim University Union and Fund, project TDZNLH201703, the Central China Agricultural University -Tarim University Union and Fund, project HNLH202002, the Bingtuan Science and Technology Program, grant number 2018DB001, and the Open project of Key Laboratory of modern agricultural engineering, project TDNG2021101.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Xu, L.; Zhang, K.; Yang, G.; Chu, J. Gesture recognition using dual-stream CNN based on fusion of sEMG energy kernel phase portrait and IMU amplitude image. *Biomed. Signal Process. Control* **2022**, *73*, 103364. [CrossRef]
2. Wu, W.; Liu, H.; Li, L.; Long, Y.; Wang, X.; Wang, Z.; Li, J.; Chang, Y. Application of local fully Convolutional Neural Network combined with YOLO v5 algorithm in small target detection of remote sensing image. *PLoS ONE* **2021**, *16*, e0259283. [CrossRef] [PubMed]
3. Zhao-Zhao, J.; Yu-Fu, Z. Research on Application of Improved YOLO V3 Algorithm in Road Target Detection. *J. Phys. Conf. Ser.* **2020**, *1654*, 012060. [CrossRef]
4. Ahmad, T.; Ma, Y.; Yahya, M.; Ahmad, B.; Nazir, S.; Haq, A.U. Object Detection through Modified YOLO Neural Network. *Sci. Program.* **2020**, *2020*, 8403262. [CrossRef]
5. Jabir, B.; Falih, N. Deep learning-based decision support system for weeds detection in wheat fields. *Int. J. Electr. Comput. Eng.* **2022**, *12*, 816–825. [CrossRef]
6. Tong, Y.; Ma, H.; Zhang, S.; Wu, X.; Chen, W. Research on Object Detection in Campus Scene Based on Faster R-CNN. *J. Phys. Conf. Ser.* **2022**, *2203*, 012050. [CrossRef]
7. Wang, Y.; Cui, G.; Wang, S.; Zhang, J. Preceding Vehicle Detection Based on Optimized Faster R-CNN Algorithm. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2021. [CrossRef]
8. Yi, Z.; Yongliang, S.; Jun, Z. An improved tiny-yolov3 pedestrian detection algorithm. *Optik* **2019**, *183*, 17–23. [CrossRef]
9. Yu, J.; Zhang, W. Face Mask Wearing Detection Algorithm Based on Improved YOLO-v4. *Sensors* **2021**, *21*, 3263. [CrossRef]
10. Hongwen, Y.; Zhenyu, L.; Qingliang, C.; Zhiwei, H.; Wen, L.Y. Detection of facial gestures of group pigs based on improved Tiny-YOLO. *Trans. Chin. Soc. Agric. Eng.* **2019**, *35*, 169–179.

11. Yuanbing, L. Identification and Classification Method of Agricultural Diseases and Insect Pests Based on Yolov3. In Proceedings of the 2021 International Conference on Applied Mathematics, Modeling and Computer Simulation (AMMCS 2021), Wuhan, China, 13–14 November 2021; pp. 538–547.
12. Du, S.; Zhang, P.; Zhang, B.; Xu, H. Weak and Occluded Vehicle Detection in Complex Infrared Environment Based on Improved YOLOv4. *IEEE Access* **2021**, *9*, 25671–25680. [CrossRef]
13. Hou, X.; Ma, J.; Zang, S. Airborne infrared aircraft target detection algorithm based on YOLOv4-tiny. *J. Phys. Conf. Ser.* **2021**, *1865*, 042007. [CrossRef]
14. Wang, G.; Ding, H.; Li, B.; Nie, R.; Zhao, Y. Trident-YOLO: Improving the precision and speed of mobile device object detection. *IET Image Process.* **2021**, *16*, 145–157. [CrossRef]
15. Shi, D.; Tang, H. A New Multiface Target Detection Algorithm for Students in Class Based on Bayesian Optimized YOLOv3 Model. *J. Electr. Comput. Eng.* **2022**, *2022*, 1–12. [CrossRef]
16. Xiao, X.; Huang, J.; Li, M.; Xu, Y.; Zhang, H.; Wen, C.; Dai, S. Fast recognition method for citrus under complex environments based on improved YOLOv3. *J. Eng.* **2022**, *2022*, 148–159. [CrossRef]
17. Wang, H.; Shang, S.; Wang, D.; He, X.; Feng, K.; Zhu, H. Plant Disease Detection and Classification Method Based on the Optimized Lightweight YOLOv5 Model. *Agriculture* **2022**, *12*, 931. [CrossRef]
18. Qi, J.; Liu, X.; Liu, K.; Xu, F.; Guo, H.; Tian, X.; Li, M.; Bao, Z.; Li, Y. An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease. *Comput. Electron. Agric.* **2022**, *194*, 106780. [CrossRef]
19. Fu, D.; Gao, L.; Hu, T.; Wang, S.; Liu, W. Research on Safety Helmet Detection Algorithm of Power Workers Based on Improved YOLOv5. *J. Physics: Conf. Ser.* **2022**, *2171*, 012006. [CrossRef]
20. Zhang, X.; Yan, M.; Zhu, D.; Guan, Y. Marine ship detection and classification based on YOLOv5 model. *J. Phys. Conf. Ser.* **2022**, *2181*, 012025. [CrossRef]
21. Song, Q.; Li, S.; Bai, Q.; Yang, J.; Zhang, X.; Li, Z.; Duan, Z. Object Detection Method for Grasping Robot Based on Improved YOLOv5. *Micromachines* **2021**, *12*, 1273. [CrossRef]
22. Tian, M.; Liao, Z. Research on Flower Image Classification Method Based on YOLOv5. *J. Phys. Conf. Ser.* **2021**, *2024*, 012022. [CrossRef]
23. Zhao, J.; Zhang, X.; Yan, J.; Qiu, X.; Yao, X.; Tian, Y.; Zhu, Y.; Cao, W. A Wheat Spike Detection Method in UAV Images Based on Improved YOLOv5. *Remote Sens.* **2021**, *13*, 3095. [CrossRef]
24. Yao, J.; Qi, J.; Zhang, J.; Shao, H.; Yang, J.; Li, X. A Real-Time Detection Algorithm for Kiwifruit Defects Based on YOLOv5. *Electronics* **2021**, *10*, 1711. [CrossRef]
25. Jabir, B.; Falih, N.; Rahmani, K. Accuracy and Efficiency Comparison of Object Detection Open-Source Models. *Int. J. Online Biomed. Eng.* **2021**, *17*, 165–184. [CrossRef]
26. Zhang, C.; Li, T.; Zhang, W. The Detection of Impurity Content in Machine-Picked Seed Cotton Based on Image Processing and Improved YOLO V4. *Agronomy* **2021**, *12*, 66. [CrossRef]