Article

Based on the Improved Yolov5 Cotton Top Bud Recognition Algorithm

Guangze Xi , Jianping Zhou * , Yan Xu , Xiaorong Wang

Posted Date: 27 January 2023

doi: 10.20944/preprints202301.0486.v1

Keywords: YOLOv5s; Gz-ShuffleNetv2;XIoU; Attention Mechanis; Interface system

Article

Based on the Improved YOLOv5 Cotton Top Bud Recognition Algorithm

Guangze Xi [1], Jianping Zhou [1,*],Yan Xu [1] and Xiaorong Wang [2]

[1] Department of Mechanical Engineering, Xinjiang University, Wulumuqi 830000, China
[2] Engineering Application Center, Xinjiang University, Wulumuqi 830000, China
* Correspondence: 1945097519@qq.com

Abstract: Aiming at the problems that the parameters of YOLOv5s model are too large and the computing resources of development board memory are limited, a new target detection method based on deep learning YOLOv5 algorithm model is proposed. First, the lightweight module Gz-ShffleNetv2 is used to construct cotton top bud feature extraction unit, which reduces the number of parameters and improves the running speed. It can be better applied in image classification, speed up detection and meet the requirements of mobile end of development board. Secondly, in order to solve the problem of decreasing detection accuracy caused by lightweight, BotNet and C3SE attention mechanism are added to focus on specific areas of cotton terminal bud. Combined with YYG loss function XIOU boundary frame regression loss, more feature information and rich feature MAP were obtained to further improve the accuracy of target detection. Through analysis of research and experimental results, the average accuracy map reached 91.3% under the Windos system NVIDIA Geforce RTX 2060 SUPER detection. While maintaining high precision identification, the number of parameters of YOLOv6 and YOLOv7-tiny network model is reduced by 83% and 53%, respectively, and the detection accuracy is increased by 1.2% and 7.7%, respectively. Compared with YOLOv5 reasoning image, the speed of image is increased by 0.035s, and the detection accuracy of MAP_0.5:0.95 is increased by 1%. At the same time, PyQt5 and YOLOv5 target detection algorithms are used to design a cotton top bud identification system, which makes cotton top bud detection more intuitive and convenient for subsequent hardware development and use.

Keywords: YOLOv5s; Gz-ShuffleNetv2; XIoU; Attention Mechanis; Interface system

1.Introduction

Cotton is an important raw material of Chinese textile industry, mechanical cotton toppling machine in the process of operation due to the different level of cotton plants, resulting in incorrect toppling, cotton toppling machine effect is not ideal, in order to solve the large area of damage cotton plants caused by mechanical cotton toppling machine, resulting in cotton yield reduction, to ensure the cotton yield, it is necessary to intelligent toppling of cotton top. With the development of intelligent agriculture, intelligent toppling machine is imperative. Intelligent toppling machine can first accurately identify the cotton top bud, and then the executing agency can effectively remove the cotton plant top bud. How to accurately identify the cotton top bud is the main research problem of this paper, which is of great significance to the field of agricultural cotton planting. The deep learning recognition algorithm is divided into One-stage and Two-stage. One-stage extracts features from input images through convolutional neural network, and then predicts the classification and positioning of targets. The detection speed is fast and can avoid false positives, but the detection effect on small objects is not very good. Its main representative network algorithms include YOLO[1,2] series, SSD[3] series, etc. In the Two-stage, the main candidate areas are developed, and then the convolutional neural network is used to predict the target and location. The main advantages are high accuracy, slow speed, long training time, and relatively high false positive rate. Its main representative model algorithms are RCNN[4–6] series. With the continuous development of target detection algorithms, relevant agricultural recognition algorithms have also been proposed. Liu Junqi et al in literature [7] used gray processing, image denoising, edge detection and feature extraction to preprocess the image of cotton plants, and then carried out image segmentation, and finally used BP neural network for recognition. However, the accuracy of this method is not high enough and the recognition accuracy is low. In literature [8], Liu Haitao proposed a method for accurate identification of cotton top bud in complex environment based on YOLOv4[9] network. The number of parameters of YOLOv4 model is too large, the training time is long and the requirements for hardware deployment are high. Literature [10–12] uses the network model MobileNetV2[13] lightweight backbone network replacement and attention mechanism to realize the lightweight research on the network algorithm YOLO series, which has improved the complexity of the network model, but has not given a good consideration to the accuracy. Based on the lightweight of ShuffleNetv2[14], this paper improved the lightweight model of Gz-ShuffleNetv2, and replaced the Relu activation function by adding the Hard-swish activation function. The Gz-ShuffleNetv2 lightweight model was added to the YOLOv5s trunk, and C3SE, BotNet attention mechanism network and XIOU loss function were added in order to avoid the decrease of accuracy while the number of parameters was reduced to improve the accuracy of cotton top bud recognition, while maintaining a faster recognition speed。

2. Improved YOLOv5S network

YOLOv5s netork model continues to follow the network structure of YOLO series, which is mainly composed of Input, Backbone, Neck and Head. Figure 1 shows the network structure of YOLOv5s. New improved methods are added on the basis of YOLOv4 in the Input input part. These include Mosaic data enhancements, which combine four images to create a new image and increase generalization. Adaptive picture scaling, according to different input sizes of the picture can be adjusted to a fixed size of 640×640 pictures. Focus module is used by Backbone network for slicing operation, and the number of parameters in the first three convolutional layers of network model YOLOv3 [17] is reduced compared with that in network model Yolov3 [17], which reduces the computation and increases the sensitivity field. The new YOLOv5 version starts replacing the C3 neckCsp module with the neckCsp module. The Inception network, in order to reduce the number of parameters, replaces a large-size convolution with multiple small-size convolution, such as 3x3 = 3x1+1x3, which works better than structured convolutional cores in deep network situations. Similarly, the Bottleneck network structure also reduces the number of parameters. Compared with the network module, the C3 network module shorties a 1×1 Conv, and removes a BN layer and activation function. The C3 module improves reasoning speed and reduces the number of parameters. In the new improved model, CBRM module mainly replaces the original Focus module, and Gz-Shuffle_block module is used to replace all modules of Backbone department for lightweight. Meanwhile, C3SE module and BotNet module are used to replace C3 module in the neck. The improved YOLOv5s model is shown in Figure 1 below.

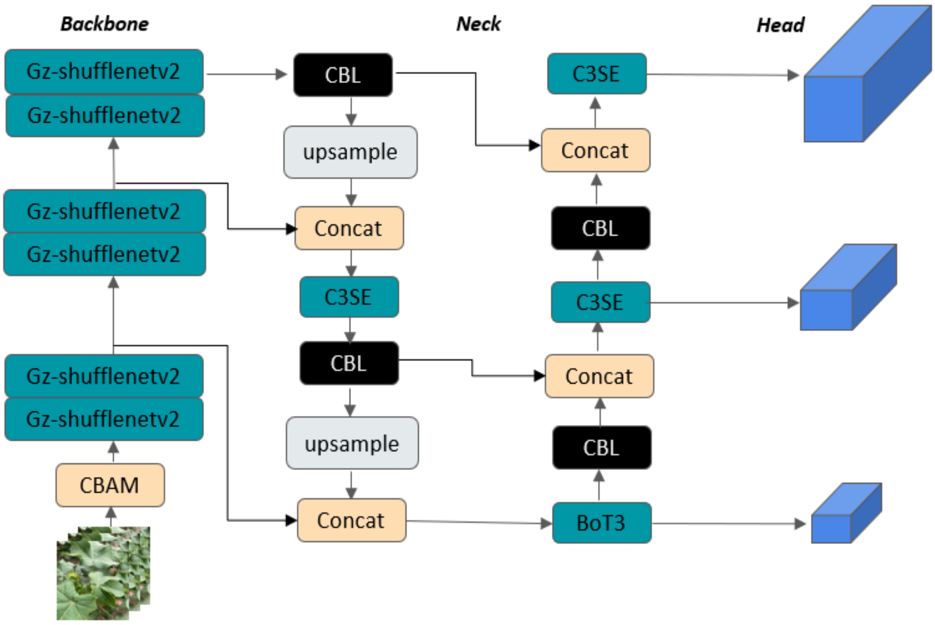


Figure 1. Improved YOLOv5s network structure.

SPPF module can integrate feature maps of different receptive fields. Secondly, the network structure of PANet[15] and FPN[16] is used in the Neck part to better combine features extracted from shallow layer and deep layer, which can avoid the problem of insufficient detailed features taken from deep layer in the process of down-sampling and further strengthen feature extraction. The Head part corresponds to the small, medium and large target feature maps output by the Neck part respectively

2.1. Gz-ShuffleNetv2 Network module

The original model of YOLOv5 is easy to miss detection of small targets in cotton top bud identification, and the detection effect of small targets is not very ideal. The original model of YOLOv5 has a large number of parameters and high requirements on hardware performance, causing certain difficulties in deployment to the development board. In order to solve the problem that there is no information exchange between different groups caused by GConv and only cotton top bud target feature extraction is carried out in the same group, Shufflenetv2 designed Channel-Shuffle operation to shuffle the information of different channels and realize the information exchange of different groups, as shown in Figure 2. The Gz-ShuffleNetv2 network module uses the Hard-swish activation function to replace the original ReLu activation function. The output values of ReLu activation function for the cotton top bud data set are all greater than 0. When the value is less than 0, it will inevitably cause the death of neurons, namely the disappearance of gradient, resulting in the failure to update the weight. When the activation function is activated by Hard-swish, it has the characteristics of no upper bound, lower bound and non-monotonicity. Only when the input value of cotton data set is less than -3, can the gradient disappear be avoided, which greatly reduces the shortcomings caused by ReLu activation function. Meanwhile, the calculation speed is faster, the calculation cost is low and the cost performance is higher.

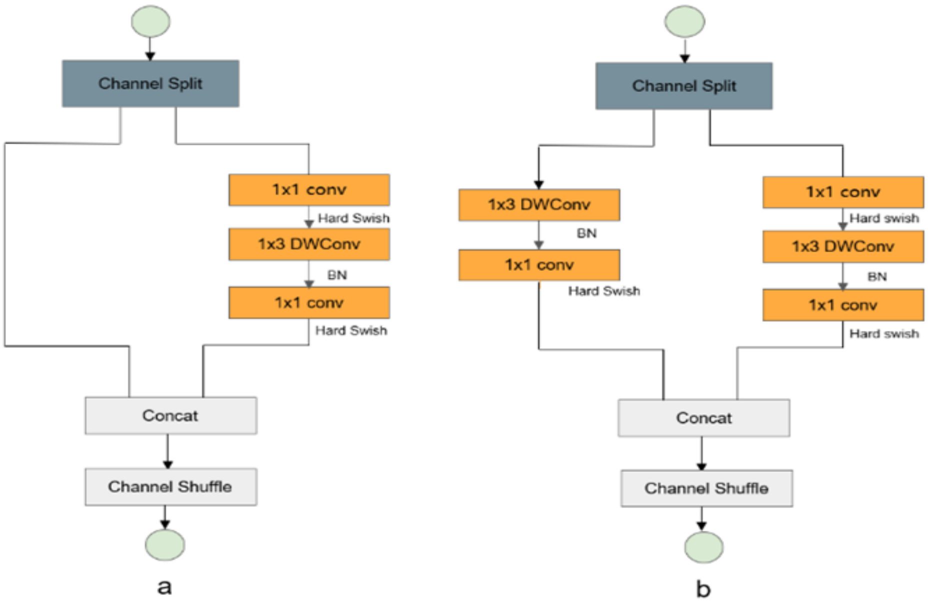


Figure 2. Gz-ShuffleNetv2 network module.

2.2. C3SE attention mechanism module

C3SE attention mechanism network carried out global average pooling on the input cotton top bud data feature map, transformed the cotton top bud data feature map into 1×1× channel number, and then adjusted the pooled cotton top bud feature map again by using the full connection layer and activation function to change the weight of each cotton top bud feature map, and then multiplied with the input cotton top bud feature. The overall task is to give weight to each cotton top bud feature map, focusing on the most important top bud feature.

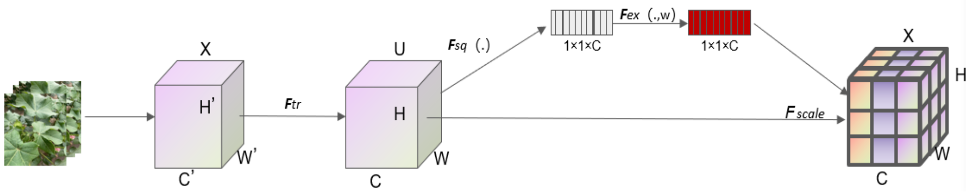


Figure 3. C3SE attention mechanism network module.

2.3. BotNet Attention mechanism module

BoTNet is a Bottleneck structure that is simple and powerful compared to ResNet [26] convolutional neural networks, which tend to incorporate multiple self-attention mechanisms into computer vision tasks of target detection. Bottleneck 3*3 convolution in the ResNet50 network is replaced with a Multi-Head Self-Attention (MHSA) structure, which is Bottleneck 3*3, as shown in Table 1. A ResNet library, which introduces MHSA structures, is a Bottleneck Transf ormer (BoT). The multi-head self-attention layer includes relative position coding and multi-head self-attention model. The multi-head self-attention model is not only used for semantic recognition, but also can be combined with other capabilities to be applied to the YOLOv5 network model for cotton top bud classification. During training, relative position coding and multi-head self-attention can be used to enable the network model to learn more features and details of pictures [27], thus improving the network performance.

2.4. Loss function improvement

The loss function of YOLOV5 is composed of position, confidence and class loss function, as shown in Equation (1); CIOU[17] loss function, as shown in Equation (2); v, as shown in Equation (3), measures the similarity between the length and width ratio of the predicted frame and the real frame, where Equation (4) a is the weight function. w^{gt} is the true frame width， h^{gt} Is the true height of the box， w is the width of the forecast frame, h is the height of the forecast frame, c is the minimum external frame containing the forecast frame and the real frame, ϱ is the center point of the forecast frame b and the center point of the real frame b^{gt} Distance between。When the width of the real box is infinitely large and the height infinitely small, and the prediction box is infinitely close to 0, v is 1 at this time, the gradient can be updated continuously. The closer the real box is to the prediction box, the smaller v will be. If the true aspect ratio is the same as the true aspect ratio, and the real frame and the prediction frame are concentric, the CIOU will degenerate into IOU loss function. When the prediction frame and the real frame do not intersect, the loss will be 0 and the gradient cannot be updated iterately, resulting in the reduction of the target detection rate. In Formula (5) EIOU[18] loss, the influence factors of aspect ratio of CIOU are disassembled to calculate the length h and width w of real frame and forecast frame respectively, as shown in Formula (5), which consists of three parts, including the loss of overlapping, center distance and width. The overlap and middle distance parts were not changed, and the CIOU method was continued to be used. The width and height loss was added to solve the problem that the width and height of the CIOU could not increase or decrease at the same time, so that the width and height difference between the target box and the forecast box was minimized. The introduced EIOU loss function converges faster, ϱ is the center point w of the predicted frame width and the center point of the true frame width w^{gt} Distance between， c_w， c_h Respectively are the width and height of the minimum external frame covering the predicted frame and the real frame. EIOU loss function cannot take into account the advantages of CIOU loss function. In the training, the effect of Eiou loss function is not as good as that of CIOU. XIOU loss function can complement CIOU loss function and EIOU loss function. Meanwhile, both sides degenerate into IOU loss function, which makes it impossible to regression and more difficult to identify the blocked cotton top bud.

$$Loss = L_{CIOU} + L_{obj} + L_{cls} \tag{1}$$

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

$$V = \frac{4}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{\omega}{h}\right)^2 \tag{3}$$

$$a = \frac{U}{(1 - IOU) + v} \tag{4}$$

$$L_{EIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(\omega, \omega^{gt})}{c_\omega^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \tag{5}$$

XIOU loss function improves the precision of cotton top bud. XIOU loss is combined with regression of prediction frame and real frame to avoid the problem that the center point overlapped and the aspect ratio was the same, which degenerates into IOU loss function and causes the boundary frame regression to be impossible, so the boundary frame regression can be better completed as follows: (6)

$$L_{XIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(\omega, \omega^{gt})}{c_\omega^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} + av \tag{6}$$

3. Analysis of experimental results

3.1. Experimental environment and its configuration

Cott on top bud data sets were obtained through on-site shooting in the cotton planting experimental site in Shihezi City, Xinjiang Uygur Autonomous Region of the People's Republic of

China, and 3103 data sets were obtained through data enhancement, which were divided into single target top bud, multi-target top bud, and blocked top bud, which were respectively shot under sunny, cloudy and other lighting conditions. Training set verification Set 8: Data enhancement is used to improve generalization in classification. The distribution of data sets is shown in Figure 4 below. pytorch, a deep learning framework, is used to build an experimental operating environment for the study of cotton top bud pictures. The CPU is Xeon Bronze 3106, the GPU is Geforce RTX 2060 SUPER, Win-dows is selected as the main operating system, python-3.7 programming language is used to call the libraries needed by CUDA, etc., for the classification training and testing of cotton top bud image data set.

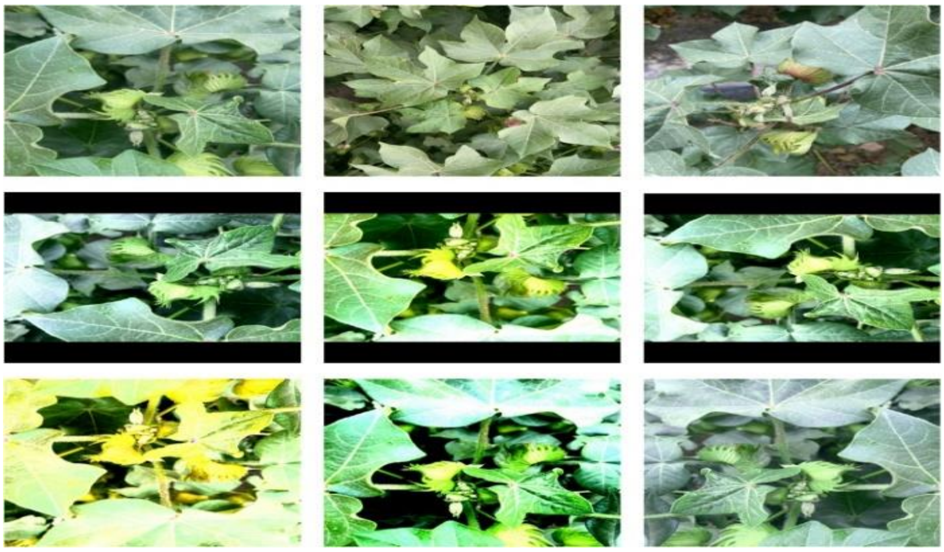


Figure 4. C3SE attention mechanism network modul.

3.2. Evaluation index

The performance of the deep learning object detection algorithm is mainly reflected by Precision (Equation (7), Recall (Equation (8), and mean Average Precision (MAP).

$$Precision \ = \frac{TP}{TP+FP} \tag{7}$$

$$Recall \ = \frac{TP}{TP+FN} \tag{8}$$

TP represents the correct boxes predicted from positive samples; FP represents the boxes predicted from positive samples by negative samples; FN represents the number of negative samples predicted from positive samples; TP+FP represents the correct boxes predicted from positive samples plus the boxes predicted from negative samples, namely, the number of prediction boxes. TP+FN means the number of correct boxes predicted by positive samples plus the number of negative samples predicted by positive samples, that is, the number of marked boxes. The higher the MAP, the better the prediction performanceFurther notes on the usage of the dataset that will help other researchers to quickly get their hands on the dataset and work with it.

3.3. Object detection model detection compariso

Through the experimental analysis of the three models using the same data set, it is concluded that compared with YOLOv6[19] and YOLOv7[20], the improved YOLOv5s significantly reduces the number of parameters and model size, and the MAP accuracy is 7.7% higher than that of Yolov7-Tiny. The number of parameters of YOLOv7-tiny is too large and the training time is too long. Therefore, it is ideal to use the method in this paper to improve YOLOv5s.

Table 1. Network model comparison diagram.

Model	MAP（%）	Params/M	Model size /MB
Textual method	91.3	2.90	6.70
YOLOv6s	90.4	37.8	38.02
YOLOv7-tiny	83.5	6.0	12.2

3.4. Effect comparison of different loss function models

By comparing the improved YOLOv5 model with the traditional loss functions DIOU[21], CIOU and EIOU, the test data are obtained in Tables 3 and 4. Through the test verification and comparative analysis, it can be concluded that the research method in this paper reduces the number of parameters while adding Gz-ShuffleNetv2. By joining C3SE and BotNet attention mechanism network, we can maintain high accuracy and improve the detection speed. Achieve the effect of embedding embedded devices。

Table 2. Network model comparison diagram.

Model	MAP@0.5(%)	MAP_0.5:0.95(%)	Params/M	Model size /MB
Shufflenetv2+XIOU	91.2	46.2	3.7	7.97
Shufflenetv2+DIOU	90.9	47.6	3.7	7.97
Shufflenetv2+EIOU	89.1	49.1	3.7	7.97
Shufflenetv2+CIOU	90.4	44.3	3.7	7.97
Textual method	91.3	50.2	2.9	6.70

Table 3. Loss function model effect comparison.

Backbone+Neck	MAP@0.5(%)	MAP_0.5:0.95(%)	Params/M	FPS
Shufflenetv2+4C3SE	90.6	46.2	3.29	45
Shufflenetv2+BoT+3C3SE	90.3	49.2	3.21	45
Shufflenetv2+BoT+3C3SE	90.3	49.2	3.21	45
Shufflenetv2+C3SE+BoT+2C	90.8	45.3	3.27	43
Shufflenetv2+2C+BoT+C3SE	91.1	43.5	3.21	47
Textual method	91.3	50.2	2.95	47

3.5. Evaluation index

YOLOv5s target detection network structure has four network models, which are YOLOv5s, YOLOv5n, YOLOv5m and YOLOv5l, respectively. Through experimental comparison on coco128 data set, the author concludes that YOLOv5s has the best detection accuracy and speed. At the same time, his network depth and width is also the best. Gz-ShuffleNetv2, MobileNetv3[22], GhostNet [23], EIOU and YOLOv5s were selected as independent variable modules, and the control variable method was adopted to study and compare the improvement effects of each method on the YOLOv5s network model. Table 4 shows the ablation results.

Table 4. Ablation experiment results.

YOLOv5s	T-method	EIOU	Shuf-v2	Ghostnet	Mo-V3	XIOU	MAP@0.5(%)	MAP_0.5:0.95(%)	Params/M
√							90.8	45.9	7.01
		√	√				89.1	49.1	3.79
				√			91.0	47.3	3.68
		√		√			90.3	47.4	3.68
		√			√		89.1	49.2	3.54
			√				90.4	47.1	3.79
			√			√	91.2	49.0	3.79
	√						91.3	50.2	2.91

As can be seen from Table 5, ShuffleNetv2 network and EIOU network module have the same effect on average precision MAP as network model YOLOv5s network module. Although ShuffleNetv2 has fewer parameters and model parameters than GhostNet and MobileNetv3, Compared with the original YOLOv5, the average accuracy MAP is basically the same and the detection speed is better. Therefore, ShuffleNetv2 and XIOU are selected as the combination for target detection. Table 6 shows the fusion experiment with the original YOLOv5s model based on different experimental modules. Ghostnet and MobileNetv3 methods were used respectively to take YOLOv5s as independent variable modules, and the fusion method of different modules was adopted to study the improvement effect of each combination module on the YOLOv5s network, and to detect the average accuracy AP of each cotton plant terminal bud. The AP calculation method (average index accuracy) is shown in Formula 8. It can be intuitively seen from the table that AP value of Multiple plants is higher, while that of Shelter plants is lowest

$$AP = \int_0^1 P（R）\mathrm{dR} \tag{8}$$

Table 5. MAP test results of different terminal bud targets.

Model	Single Plant	Multiple plants	Shelter plants
YOLOv5s	0.958	0.978	0.811
Shufflenetv2+XIOU	0.937	0.978	0.821
MobileNetV3+EIOU	0.948	0.984	0.764
Ghostnet	0.947	0.954	0.806
Ghostnet＋EIOU	0.937	0.969	0.804
Shufflenetv2	0.937	0.979	0.802
Textual method	0.954	0.980	0.805

3.6. Experimental analysis of target detection

When the superparameters were the same, batch-size was set to 20, the original YOLOv5s network model was compared with the improved YOLOv5s model for training and testing, and the PR curve as shown in the following figure was obtained when the threshold of IOU was 0.5. The area enclosed by the corresponding curve and the horizontal and vertical coordinates is AP of this type. The closer the curve is to the upper right corner, the larger the AP will be and the better the effect will be. It can be seen from Figure 4 that the four thin curves from top to bottom respectively represent Single Plant, Multiple plants and Shelter plants of cotton top bud. It can be seen from Figure 4 that the identification accuracy of Multiple plants is the highest, basically covering the whole area. Single Plant increased by 1% compared with the original network, and Shel-ter plants increased by 1.0% compared with the original network. Figures 3 and 4 more intuitively show the effect before and after the improvement of the network model.

Figure 5 Compares the MAP of YOLOv5s model before and after improvement by using the obtained data, and takes the average accuracy of all categories with IOU threshold of 0.5 to 0.95. The blue and black lines are the YOLOv5s mAP_0.5 curves before and after the improvement, and the red and green curves are the YOLOv5s MAP_0.5 curves before and after the improvement. According to the experimental results of curve comparative analysis, the MAP threshold is within the range of 0.5. The lightweight YOLOv5s curve is roughly the same as the original network MAP, with a 1% increase in MAP threshold between 0.5 and 0.95.

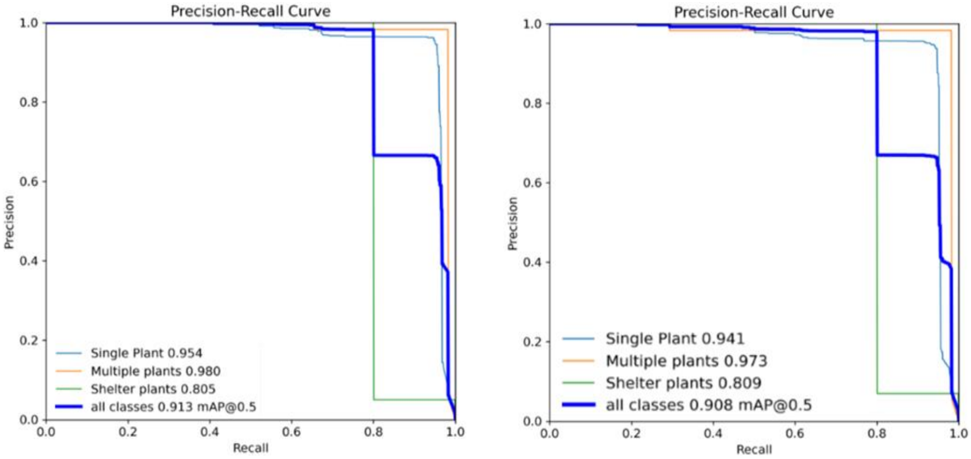


Figure 5. Improved YOLOv5s PR map.

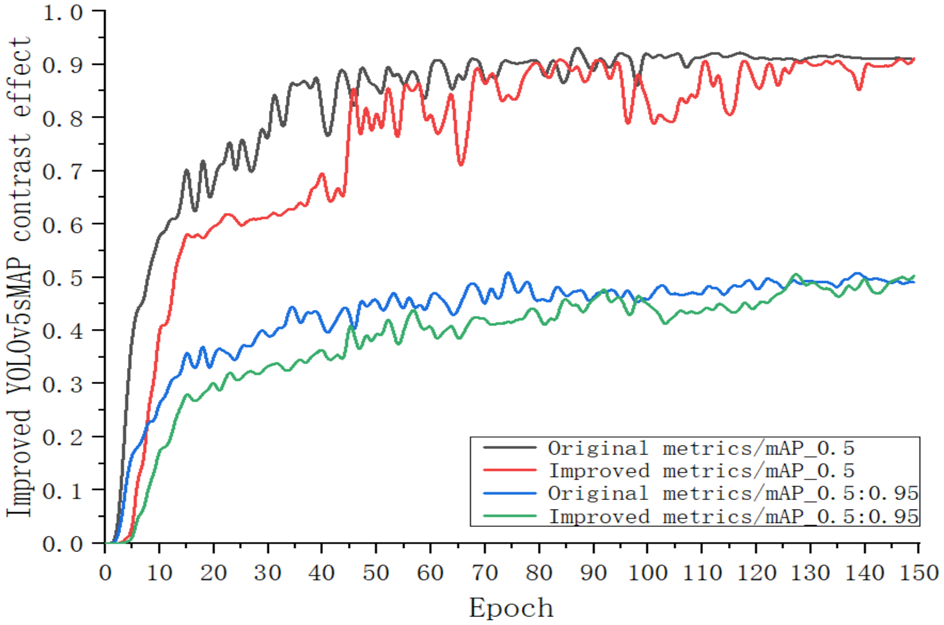


Figure 6. MAP Average precision comparison map.

Figure 7 shows the convergence of the loss function of the improved YOLOv5s network model. It can be seen from the figure that the convergence of three loss verification functions is achieved without over-fitting, and the target detection effect is effective.

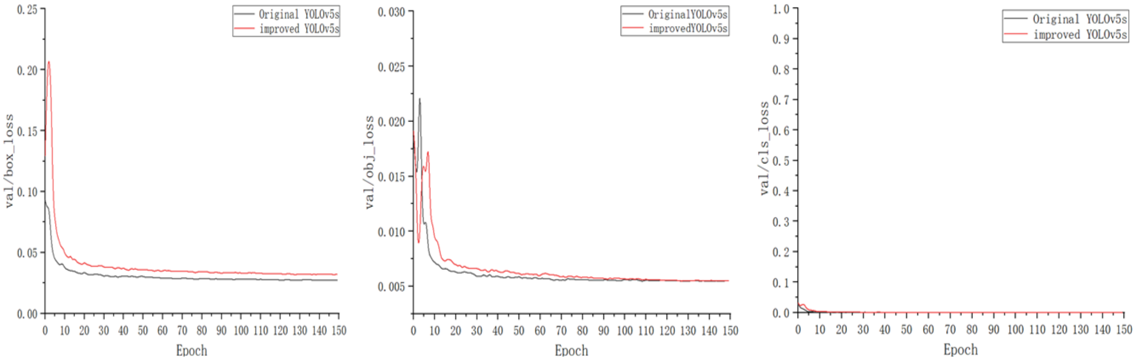


Figure 7. Loss function diagram.

3.7. Different models test the effect of the image

By comparing the YOLOv5 network model before and after the improvement, it can be seen that the cotton top bud identified by the original YOLOv5 model did not fully identify the top bud, and the confidence level was low. The improved YOLOv5 detection model could not only identify the top bud, but also improve the confidence level. The target detection speed was maintained while the number of parameters was greatly reduced. When it is implanted into mobile embedded devices, it can meet the requirements of dynamic tracking, high-precision detection and improve detection speed, so as to achieve good operation results.

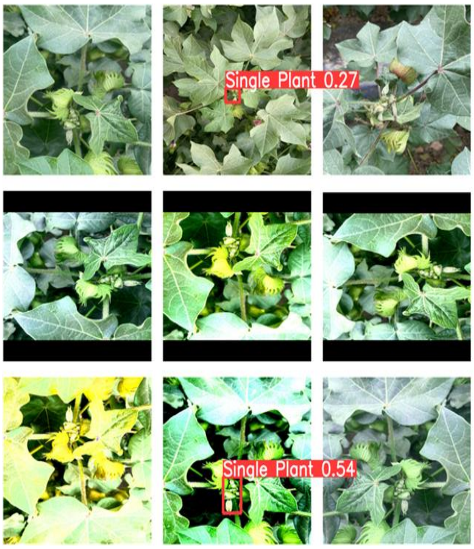


Figure 8. Original yolov5 model.

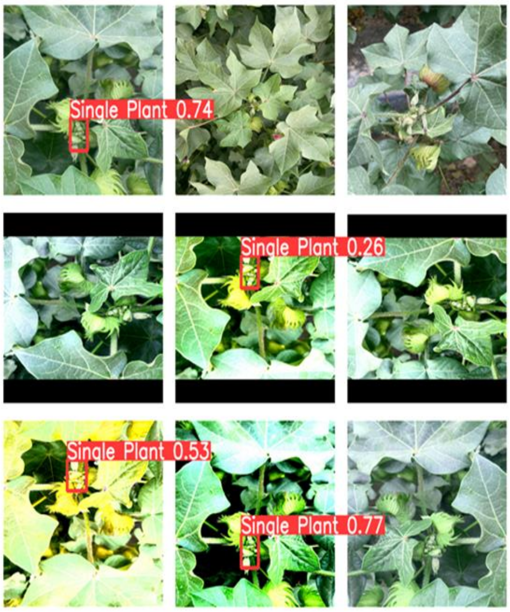


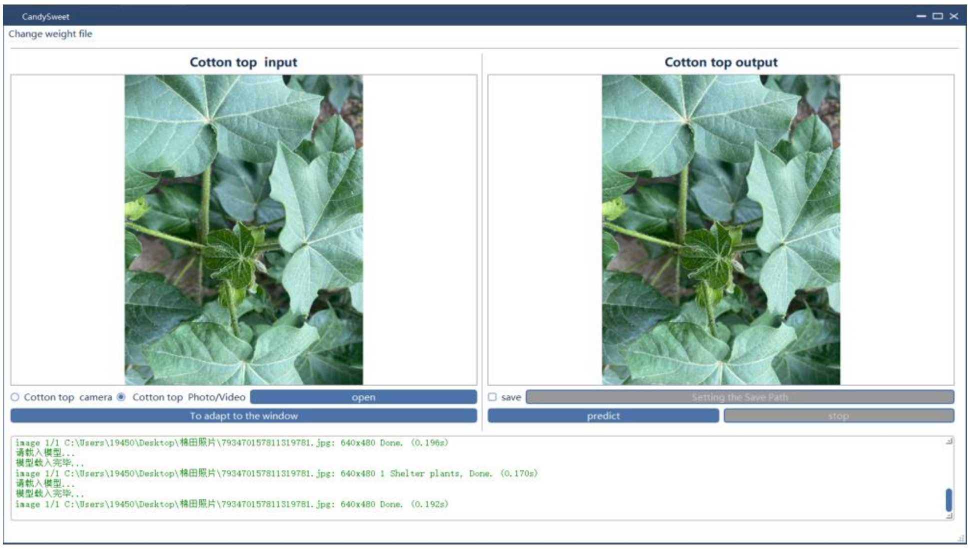
Figure 9. Improves the yolov5 model.

doi:10.20944/preprints202301.0486.v1

4. Experimental analysis of interface system

4.1. Interface system introduction

When the YOLOv5s target detection model makes inference analysis on pictures or videos, it cannot analyze the data intuitively. Based on this problem, an interactive cotton top bud detection interface system is developed by using PyQt5 framework to improve the YOLOv5s target detection model. The interface of the cotton top bud detection system mainly consists of four parts. Weight model replacement interface, screen display interface, function area interface, test result output area interface, as shown in Figure 10。



```
image 1/1 C:\Users\19450\Desktop\棉田照片\793470157811319781.jpg: 640x480 Done. (0.196s)
请载入模型...
模型载入完毕...
image 1/1 C:\Users\19450\Desktop\棉田照片\793470157811319781.jpg: 640x480 1 Shelter plants, Done. (0.170s)
请载入模型...
模型载入完毕...
image 1/1 C:\Users\19450\Desktop\棉田照片\793470157811319781.jpg: 640x480 Done. (0.192s)
```

Figure 10. Cotton terminal bud detection interface system.

Main Window is a basic category in PyQt5 frame, which can set the size of basic interface components and place other controls. In this paper, QLabel is used to design two screen interfaces for cotton top bud detection system. The left side is used to input cotton top bud picture or video interface, and the right side is used to output waste detection interface. When the camera is turned on to process data, the original shot picture is displayed on the left side, while the right side is the real-time cotton top bud detection interface. The function area includes the control of the left input interface, click the "enable" button on the left interface, open means to select the path of the input image or video and open the camera, after selecting the right window interface and click the predict button to predict. The stop button can be controlled in the middle of the process. The save path can be selected to save the predicted video or picture. In the output area of detection results, category information of detection results and time system of image processing can be observed more intuitively, and it is convenient to understand the detection process。

4.2. Interface system test

Through the image processing functional area of the cotton top bud detection system, the improved YOLOv5s network model algorithm was used to detect and verify the cotton top bud images. The improved YOLOv5 model and the original YOLOv5s model were selected to detect the blocked images. After the improvement, the reasoning image speed of the YOLOv5s network model is accelerated by 0.018s and the confidence is also increased by 11% to achieve the detection effect.

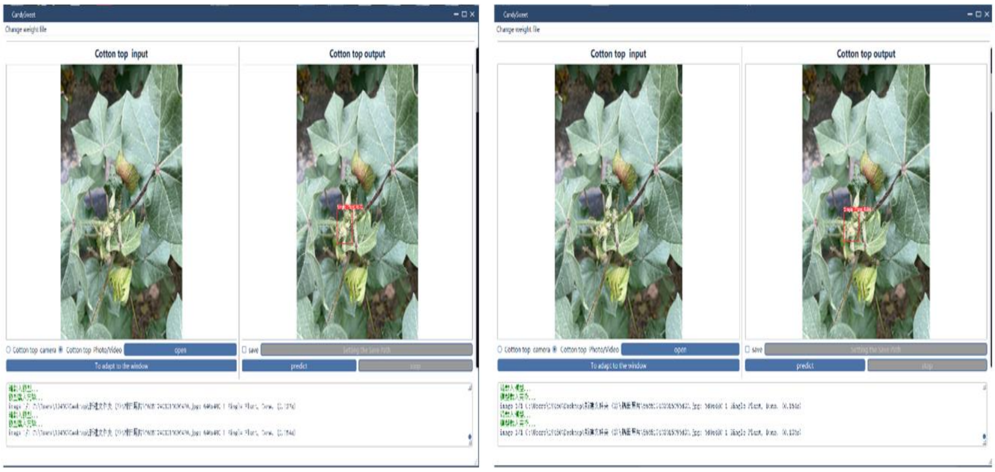


Figure 11. YOLOv5S model improved front and rear interface system.

5. Conclusions

By improving the lightweight model of YOLOv5 and replacing the loss function with XIOU loss function, the prediction box can be returned to the real box more quickly, and BotNet and C3SE attention mechanisms can be added to focus more on specific areas of cotton top bud. It can be seen that the improved YOLOv5s network model detection system has a significant improvement effect. The detection accuracy can be improved, and the image reasoning detection time is accelerated by 0.018s. Compared with the original YOLOv5s model, more parts of the cotton top bud that are blocked are also identified, and the confidence is greatly improved. The lightweight model achieves the effect.

Author Contributions: Conceptualization, X.Z. and X.R.; methodology, X.Z.; software, X.Y.; validation, X.Z., Z.J. and X.Y.; formal analysis, X.Z.; investigation, W.X. and Z.J.; resources, K.L., X.Z. and Z.J.; data curation, X.Z., X.Y. and Z.J.; writing—original draft preparation.

Funding: National Natural Science Foundation of China (51765063) Cotton Laser Topping Machine Research, Development and Application Project (YTHSD2022-05) Innovation and Entrepreneurship Education Reform Project (202101133033).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References

1. De Carolis B, Ladogana F, Macchiarulo N. Yolo trashnet: Garbage detection in video streams[C]//2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS). IEEE, 2020: 1-7.
2. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:.
3. Anguelov D Liu, Erhan D, et al. SSD: Single shot multibox detector[J]. Springer, Cham, 2016:
4. Girshick R, Jeff D, Trevor D. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on computer vision and Pattern Recognition, 2014: 2-6.
5. Li J, Liang X, Shen S M, et al. Scale-aware fast r-cnn for pedestrian detection[J]. IEEE Transactions on Multimedia, 2015, PP(99): 1-1
6. Girshick R, Scale-aware fast r-cnn for pedestrian detection[J]. Computer Science, 2015
7. LIU Junqi Research on Cotton plant top identification system [D]. Shihezi University,2009
8. Liu Haitao, Han Xin, LAN Yubin, Yi Lili, Wang Baoju, Cui Lihua. Precision identification of cotton terminal bud based on YOLOv4 network [J]. Journal of Agricultural Science and Technology of China,202,24(08):99-108. (in Chinese
9. Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020

10. Li Renying, Qian Huifang, Guo Jiahao. Lightweight target detection algorithm based on M-yolov4 model [J]. Foreign Electronic Measurement Technology, 222,41(4):15-21
11. Wang Chen, Yuan Qingni, Bai Huan. Light Weight Target detection algorithm for warehouse goods [J]. Laser & Optoelectronics Progress :1-10[2022-06-07].
12. Qin Weiwei, Song Tainian, Liu Jieyu. Remote Sensing military target detection algorithm based on lightweight Yolov3 [J]. Computer Engineering and Applications,2021,57(21): 263-269
13. Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.
14. Ma N, Zhang X, Zheng H T, et al. ShuffleNet v2: Practical guidelines for efficient cnn architecture design[J]. European Conference on Computer Vision, 2018
15. Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 390-391
16. He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916
17. Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 12993-13000..
18. Zhang Y F, Ren W, Zhang Z, et al. Focal and efficient iouloss for accurate bounding box regression[J]. Computer Vision & Pattern Recognition. 2021
19. Li C, Li L, Jiang H, et al. YOLOv6: a single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv:2209.02976, 2022..
20. 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[J]. arXiv preprint arXiv:2207.02696, 2022..
21. Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 12993-13000.
22. Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1314-132
23. Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1580-1589.

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.