agriculture

MDPI

Article

Detection of Famous Tea Buds Based on Improved YOLOv7 Network

Yongwei Wang, Maohua Xiao, Shu Wang, Qing Jiang, Xiaochan Wang and Yongnian Zhang *

Engineering College, Nanjing Agricultural University, Nanjing 210031, China; 9203010615@stu.njau.edu.cn (Y.W.)
* Correspondence: hczyn@njau.edu.cn

Abstract: Aiming at the problems of dense distribution, similar color and easy occlusion of famous and excellent tea tender leaves, an improved YOLOv7 (you only look once v7) model based on attention mechanism was proposed in this paper. The attention mechanism modules were added to the front and back positions of the enhanced feature extraction network (FPN), and the detection effects of YOLOv7+SE network, YOLOv7+ECA network, YOLOv7+CBAM network and YOLOv7+CA network were compared. It was found that the YOLOv7+CBAM Block model had the highest recognition accuracy with an accuracy of 93.71% and a recall rate of 89.23%. It was found that the model had the advantages of high accuracy and missing rate in small target detection, multi-target detection, occluded target detection and densely distributed target detection. Moreover, the model had good real-time performance and had a good application prospect in intelligent management and automatic harvesting of famous and excellent tea.

Keywords: famous and excellent green tea; bud detection; improved YOLOv7 algorithm; attention mechanics

check for updates

Citation: Wang, Y.; Xiao, M.; Wang, S.; Jiang, Q.; Wang, X.; Zhang, Y. Detection of Famous Tea Buds Based on Improved YOLOv7 Network. *Agriculture* **2023**, *13*, 1190. https://doi.org/10.3390/agriculture13061190

Academic Editors: Hongbin Pu and Filipe Neves Dos Santos

Received: 21 April 2023
Revised: 19 May 2023
Accepted: 2 June 2023
Published: 3 June 2023

1. Introduction

China is the most important tea producer in the world, with a tea plantation area of 305.9 million hectares and annual output of 260.9 million tons, accounting for 62.6% and 50.1% of the global tea plantation area and output, respectively [1]. Famous and excellent tea is favored by the people because of its high drinking value and economic value [2]. High-quality tea usually has strict requirements on tenderness and number of leaves, and different grades of high-quality tea often have different requirements on the number of leaves and buds, usually famous and excellent tea picking only pick a bud and a leaf [3]. Famous and excellent tea picking has strong seasonality, short picking cycle and high labor intensity, which is a labor-intensive operation. With the rapid development of the tea industry, the contradiction between the timeliness of famous and excellent tea picking and the shortage of labor force of manual picking is increasingly prominent [4]. In recent years, some famous and excellent tea picking equipment was used for picking tea gardens, but it has some shortcomings, such as imprecise mechanical picking technology and poor quality mechanical tea [5]. In the complex environment of tea gardens, the rapid and accurate detection of young leaves of famous and excellent tea based on vision is the key task to realize automatic picking of famous and excellent tea.

Research on the bud detection of famous and excellent tea is mainly divided into two methods. The first method is the segmentation method based on the physical characteristics of famous and excellent tea [6–10], which mainly takes the shape, color, texture and other physical characteristics of famous and excellent tea as the basis for identifying and segmenting young leaves. Then, traditional methods such as threshold segmentation and watershed segmentation are used to separate and extract the tender leaves from the complex environment. This method is greatly affected by the environment and has a small scope of application. The other is the detection method based on neural network [11–14]. By training the marked famous and excellent tea dataset, the weight model is obtained and then used

to detect the tender buds. At present, it is widely used in related agriculture and agronomy fields. YOLO (you only look once) is a target detection algorithm, which has high precision and high efficiency, and it can directly predict the location and attribute of the target in the whole image. Marco Sozzi et al. [15] applied target detection to yield prediction of white grape, compared detection effects of YOLOv3, YOLOv3-Tiny, YOLOv4, YOLOv4-Tiny, YOLOv5x and YOLOv5s, and finally found that the YOLOv5x model, considering bunch occlusion, was able to estimate the number of bunches per plant with an average error of 13.3% per vine. The YOLOv4-tiny model has a better combination of accuracy and speed, which should be considered for real-time grape yield estimation. YOLOv3 model is affected by a false positive–false negative compensation, which decreases the RMSE. Angelo Cardellicchio et al. [16] used the YOLOv5 model to test the phenotypic traits of tomato plants. The train used a challenging dataset acquired during a stress experiment conducted on multiple tomato genotypes, considering the particular challenges of the input images in terms of object size, similarity between objects and their color. The results demonstrated that the models achieve relatively high scores in identifying nodes, fruit and flowers. Dandan Wang et al. developed an accurate apple fruitlet detection method with small model size based on a channel pruned YOLOv5s deep learning algorithm [17]. The experimental results showed that the channel pruned YOLOv5s model provided an effective method to detect apple fruitlets under different conditions. The recall, precision, F1 score and false detection rate were 87.6%, 95.8%, 91.5% and 4.2%, respectively; the average detection time was 8 ms per image; and the model size was only 1.4 MB. It can be used to help growers optimize their orchard management. Compared with the traditional physical method, the deep learning algorithms have the advantages of high identification accuracy, strong robustness and less influence by environmental factors, so it is appropriate for the detection task of famous and excellent tea.

However, with the increase in researchers' attention, Wu et al. [18] found that YOLO has the disadvantage of insufficient frame positioning and difficult to distinguish overlapping detection objects. Famous and excellent tea has a small bud shape and high density, which also have the same problems in detection, and the emergence of attention mechanism can effectively settle the above problems. The attention mechanism can obtain a weight through module calculation and multiply it with input information to achieve the purpose of focusing on important information with high weight and ignoring irrelevant information with low weight. It directly establishes the dependency relationship between input and output without cycling, making the parallelization degree enhanced, the running speed greatly improved and the weight automatically adjusted. So that important information can be selected in different situations, it has higher scalability and robustness. It achieved good results in the detection of famous and excellent tea and other agricultural fields, and it was widely used in the optimization of the model. Liu Tianzhen et al. [19] added SE Block to the YOLOv3 network, and compared with the YOLOv3 model, the F1 score increased by 2.38 percentage points and mAP increased by 4.78 percentage points. Yang et al. applied CBAM Block to wheat detection, and the results showed that the model could effectively overcome the field environmental noise and achieve the accurate detection and counting of wheat ears with different density distributions [20]. The average accuracy of wheat ears detection increased to 94%, 96.04% and 93.11%, respectively. To compare the effect of SE, CBAM and ECA attention modules on the network in the YOLO v5 network model for the posture detection of meat geese, Liu Yingying et al. [21] proved that YOLOv5+ECA had better stability and was more suitable for the posture detection of meat geese in complex scenarios in farms. Fang Mengrui et al. added the CBAM module to YOLOv4-tiny adopted bidirectional feature pyramid network (BiFPN) to integrate feature information of different scales. It was found that the F1 score of the improved Yolov4-Tiny-tea model was 12.11, 11.66 and 6.76 percentage points higher than that of the YOLOv3, YOLOv4 and YOLOv5l network models, respectively [22]. Fu et al. introduced the channel attention-asymmetric spatial pyramid pool (CA-ASPP) module to improve the detection of weak and weak pod targets [23]. The precision of the improved YOLOv5 model increased by about 6%, and the

precision of POD number in the 200 soybeans population reached 88.14%. Bao et al. [11] proposed an improved AX-RetinaNet target detection and recognition network for automatic detection and recognition of tea diseases in natural scene images. AX-RetinaNet took the improved X-module multi-scale feature fusion module and added SE Block in the network. Compared with the original network, the mAP, recall rate and recognition accuracy increased by nearly 4%, 4% and nearly 1.5%, respectively. However, it was also found that adding the attention mechanism had the opposite effect for some networks, such as SSD and EfficientNet.

Through research and experiments, it can be found that SE Block [24], CBAM Block [25], ECA Block [26] and CA Block [27] have different degrees of improvement in the detection of different crops, and the improvement effect is related to the position in the model. However, there was no research to compare the effects of four kinds of attention mechanism modules in different positions in the YOLOv7 network on parameters such as the recognition accuracy rate and recall rate of famous and excellent tea.

Therefore, this study focused on the influence of SE Block, ECA Block, CBAM Block and CA Block on the recognition accuracy, recall rate and F1 score in different positions of YOLOv7 network for famous and excellent tea detection. The purpose of this study was to select the most suited network for the detection of famous and excellent tea by comparison.

2. Materials and Methods

2.1. Data Acquisition

Longjing 43 Tea Garden in Juyuanchun, Yangzhou, has a large planting area, standard tea garden and relatively flat shed surface, which meet the experimental requirements. Therefore, the object of our study was Longjing 43 in famous and excellent green tea. In this research, we collected images of famous tea buds in Juyuanchun tea garden in Yangzhou. In order to improve the fitness of the model and the environment, we chose to face the tender buds at a total of 5 shooting angles of $\pm 30^{\circ}$ and $\pm 60^{\circ}$ with the tender buds, respectively, during the shooting, as shown in Figure 1.

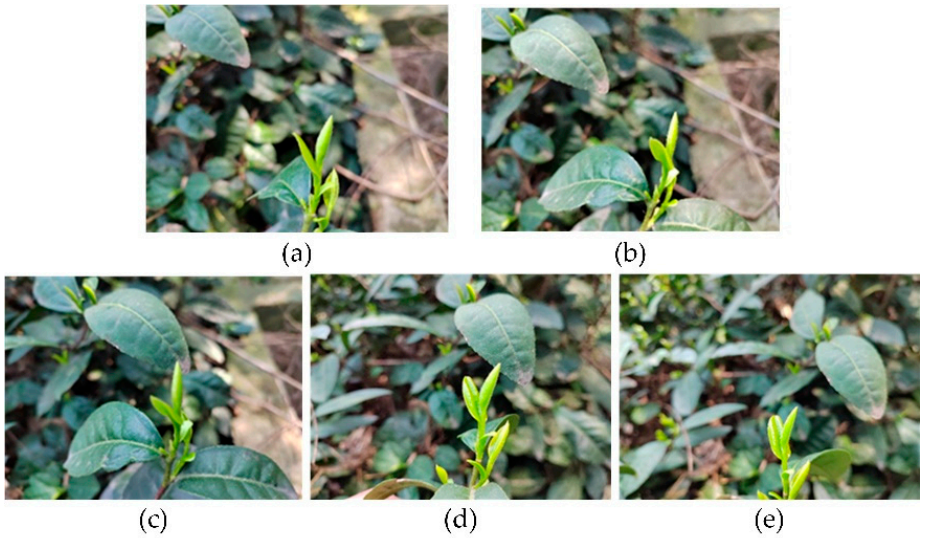

Figure 1. Multi-view bud image. (**a**) front view, (**b**) rotate counterclockwise by 30 degrees, (**c**) rotate counterclockwise by 60 degrees, (**d**) turn it clockwise by 30 degrees, (**e**) turn it clockwise by 60 degrees.

Due to the strong seasonality of the picking of famous and excellent tea, the quality of the dataset was the best in a few days before Tomb-Sweeping Day. Therefore, the shooting time of this time was 2 April 2023, and considering the influence of light on the dataset, a total of 1049 images were taken from 9 to 11 a.m. and 3 to 5 p.m., respectively. The resolution of the photos was 3904 × 2928. Figure 2 shows the overall layout of the tea garden, and Figure 3 shows the images of the dataset taken.

Figure 2. Overall layout of tea garden.

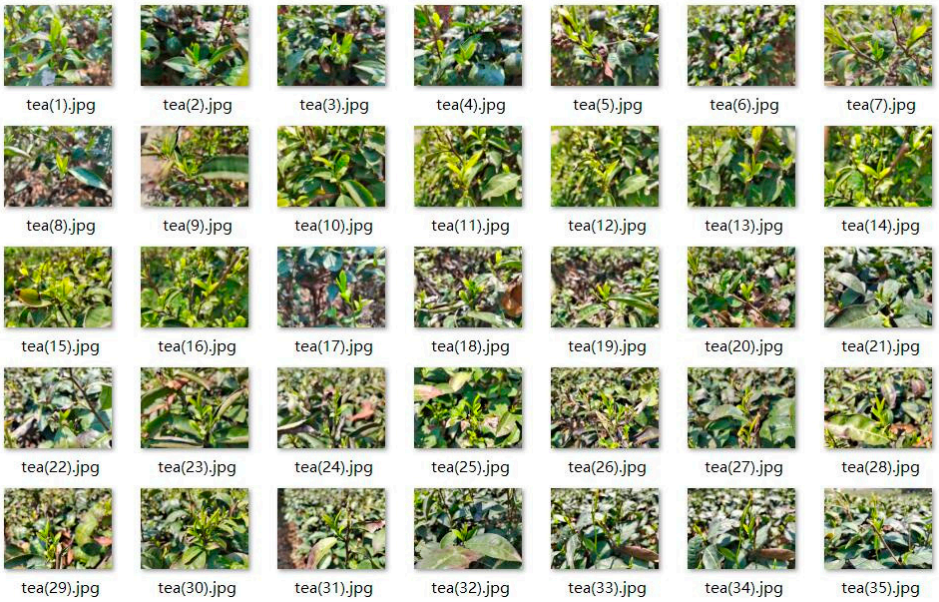

Figure 3. Dataset image.

2.2. Data Enhancement

In order to improve the robustness of the trained model, we used different processing methods to make data enhance for the famous and excellent tea image. In the research, we adopted the following methods:

(1) First, we adjusted the brightness of the image. To be specific, we raised the brightness of the image to 1.3 times and decreased it to 0.7 times compared with the original image, respectively. During our shooting time, the way of dealing with brightness could reflect the change from brightest to darkest during mechanical picking. Through this operation method, our model will be more suitable for the complex tea garden environment with changeable light;

(2) Then, we adjusted the contrast of the image. To be specific, the contrast of the images was increased by 1.2 times and weakened by 0.8 times, so that the sharpness, gray level and texture details of the famous and excellent tea images could be better expressed [20];

(3) Finally, we rotated the image taken. We thought rotation of 30 degrees can better reflect the detection of famous and excellent tea when the machine picks. This operation can enhance the adaptability of the detection model to shoot from different angles.

In the dataset, 200 images were randomly selected each time for brightness enhancement, brightness reduction, contrast enhancement, contrast reduction, horizontal flip and random angle flip. After processing, the number of datasets reached 2049. Figure 4 reflects the results of data enhancement.

Figure 4. Data enhancement. (**a**) Original image, (**b**) brightness enhanced by 30%, (**c**) reduce brightness by 30%, (**d**) contrast enhancement 20%, (**e**) contrast reduction by 20%, (**f**) rotate counterclockwise by 30 degrees.

2.3. Data Annotation

There are mainly two labeling methods for famous and excellent tea. One is to label the front view of famous and excellent tea squarely, as shown in Figure 5, and the other is to label famous and excellent tea images with the front view, side view and top view, respectively, as shown in Figure 6. Both methods can only be picked when the front view images are detected in the picking process. In addition, multi-angle labeling will increase the difficulty of control in mechanical picking and reduce the detection rate, which is unnecessary in the picking process of famous and excellent tea. Therefore, we chose the first labeling method, which only labeled the front view images. We annotated the dataset using the software of labeling. Specifically, the tender bud in the image was selected with a square frame and the label of tea was added. The format of the annotation file was selected in PascalVOC format, and the corresponding xml file was generated after saving. The software will save the image size, label name, target location and other information, as shown in Figure 5.

Figure 5. Front view of labeling.

Figure 6. Multi-view of labeling. (**a**) Red reflects the annotation of side view, cyan reflects the annotation of front view. (**b**) Yellow reflects the annotation of top view.

2.4. Excellent Tea Detection Algorithm

2.4.1. Yolov7 Algorithm

Figure 7 shows the structure diagram of YOLOv7 network. It was mainly composed of input, backbone feature extraction network, enhanced feature extraction network and head. The input module scales the input image to a unified pixel size to reduce the amount of computation [28]. The backbone module is composed of CBS, ELAN and MP modules. The CBS module is composed of batch normalization layer (BN), conv and silu activation function, and it is used to extract multi-scale information of images. ELAN module is composed of multi-branch convolution. It improves the learning ability of the network without destroying the original gradient path. MP module integrates MaxPool and convolution two dimensions of information, improving the feature extraction ability of the network.

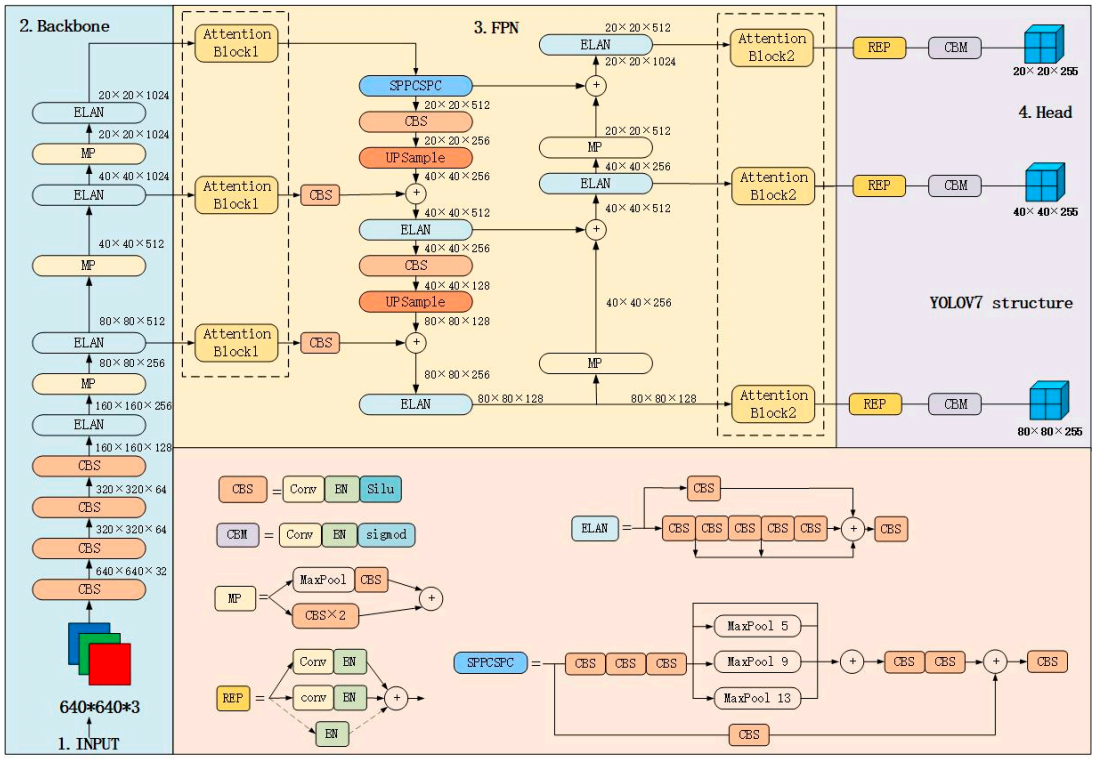

Figure 7. Yolov7 network structure diagram.

The enhanced feature extraction module is composed of path aggregation feature pyramid network (PAFPN) structure. By introducing a bottom-up path, it was easier for

the information from the bottom layer to be transferred to the top layer, thus realizing the efficient fusion of features at different levels, improving the accuracy of positioning information. In the SPPCSPC structure, SPP module, composed of CBS and 4 different sizes of maximum pooling, can better distinguish different sizes of the target through different sizes of maximum pooling; the CSP module, composed of two parts, was used for conventional processing and the above part was used for SPP module processing. Finally, the two modules merged together, which can reduce half of the amount of computation, making the processing speed faster, obtaining higher precision.

The head module was composed of REP structure and CBM structure. The REP block (RepVGG Block) was composed of two parallel convolutional layers (Conv), batch normalization layers (BN) and one batch normalization layer (BN) path. The CBM structure was composed of the convolutional layer (Conv), batch normalization layer (BN) and sigmoid activation function. The REP (RepVGG Block) structure adjusted the number of image channels for 3 features of different scales output by PAFPN, including P3, P4 and P5, and it was then used to predict the confidence, category and anchor frame through CBM structure.

2.4.2. Introduction of Attention Mechanism

The attention mechanism can make the network pay more attention to the bud target. It was found that the attention mechanism modules, such as SE Block, ECA Block, CBAM Block and CA Block, play a significant role in improving the model recognition effect. For the input features, CBAM module first learns the weight information of each channel through a shared multilayer perceptron (MLP) and sigmoid function, and then, through a hollow convolution [20] with convolution kernel of 3×3 and expansion coefficient of 2 and sigmoid function, learns the weight information of each point in the space. SE Block obtains the channel weight information after passing through the full connection layer twice and sigmoid function. The ECA block changes the two fully connected layers into one-dimensional convolution, and obtains the channel weight information after passing the sigmoid function, which has a good ability to obtain cross-channel information. The CA block divides channel attention into two 1-dimensional feature coding processes, which aggregates features along two spatial directions, respectively. In this way, remote dependencies can be captured along one spatial direction. At the same time, accurate location information can be retained along the other spatial direction, and then, the generated feature map was encoded into a pair of direction-aware and position-sensitive attention maps, which can be applied complementary to the input feature map to enhance the representation of the object of attention. The schematic diagram of the network structure of the four modules is shown in Figure 8.

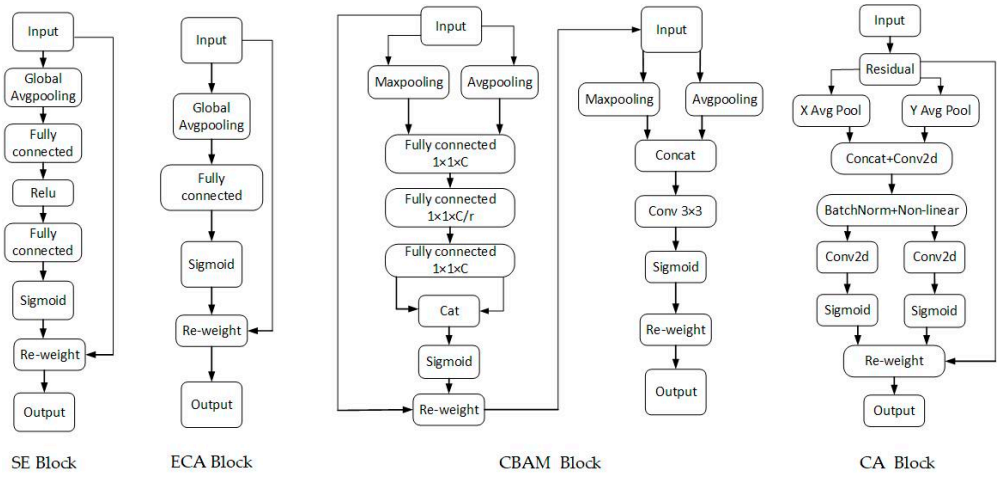

Figure 8. Attention mechanism module.

In order to better adapt to the complex scene of the tea garden, four attention mechanism modules were added to the front and back positions of the enhanced feature extraction network, named Attention Block1 (AB1) and Attention Block2 (AB2), as shown in Figure 7. In the experiment, we compared the recognition effects of adding four attention modules in the above position.

2.5. Training Environment and Parameter Configuration

2.5.1. Experimental Platform and Environment Configuration

This research was based on Pytorch deep learning framework and used cloud server AI-GCI-02PM3 for training. The operating system was Linux, Ubuntu18.04, CPU was Intel 8358P@2.6/10 core, and graphics card (GPU) was NVIDIA GeForce RTX 3090. The video memory was 24 GB, and the deep learning Conda environment was configured as pytorch-1.8.0+python3.8+cuda11.1.

2.5.2. Training Parameter Settings

The number of images in the pre-training stage was 631, and the labels number of buds was 1723 during the annotation. The recognition effect of YOLOv7 model was compared with 4 different attention mechanisms, which were added at different positions. According to the ratio of training set: verification set = 9:1 and training verification set: test set = 9:1, the dataset was randomly divided into training set, verification set and test set. In the improved YOLOv7 network training process, we proceeded with 50 generation freeze training, 300 generation thaw training in turn. In parameter, batch (batch size) was set to 8, the initial learning rate set to 0.001, the minimum learning rate set to 0.00001, using the sgd optimizer, and momentum parameter was set to 0.937. We used the cosine annealing function to dynamically reduce the learning rate and turned off the Mosaic data enhancement method. We set the confidence level to 0.5 and the intersection ratio size used for non-maximum suppression to 0.3. The loss function consisted of three parts: Reg (rectangular box regression prediction) part, 0bj (confidence prediction) part and Cls (classification prediction) part. The Reg part uses CIOU Loss, and the 0bj part and Cls part uses BCE Loss (cross entropy loss).

The number of images in the formal training stage was 2049 and the labels number of buds was 6055 during the annotation. The parameter settings were the same as those in the pre-training stage. We selected the YOLOv7+CBAM model and the YOLOv7+ECA model, which had the preferable recognition effect in the pre-training process, and compared them with the recognition effect of YOLOv7 model in the formal training stage.

2.6. Evaluation Index

In this study, precision (P) was used to represent the percentage of buds correctly identified by the model; recall (R) was used to represent the coverage of bud targets identified in the images; mean average precision (mAP) represents the sum of all classes divided by all classes; F1 score was used to evaluate the performance of the method by balancing the weights of precision and recall; the frames per second (FPS) about the detection time of a single image was used to evaluate the actual bud recognition speed of the model. These parameters were used as evaluation indicators to evaluate the trained model. The relevant calculation formulas are as follows:

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

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

$$F1 = \frac{2P \times R}{P + R} \times 100\% \quad (3)$$

$$\mathrm{mAP} = \frac{1}{C}\sum_{k=i}^{N} P(k)\Delta R(k) \quad (4)$$

In the formula: true positives (TP) means that both the detection result and the true value are a famous tea bud; in other words, the number of famous tea buds is detected correctly. False positives (FP) indicates the detection result is famous tea buds, and the true value is the background; in other words, the number of famous tea buds is counted incorrectly. False negatives (FN) means that the detection result is the background, and the true value is the famous tea buds; in other words, the number of famous tea buds are not counted.

"TP + FP" refers to the total number of famous tea buds detected, and "TP + FN" refers to the total number of famous tea buds in an image. C is the number of categories, N represents the number of all pictures in the test set, $P(k)$ represents the precision when k pictures can be recognized, and $\Delta R(k)$ represents the change of the recall value when the number of recognized pictures changes from $k-1$ to k [20].

2.7. Results and Analysis

(1) We randomly selected 631 images from the captured images to form the pre-training dataset. We used the pre-training dataset to conduct training under the same environment, and compared the recognition effect parameters of the networks with different attention mechanisms added at different positions of the YOLOv7 network. The recognition effect parameters of the model were shown in Table 1. The changes of loss value and mAP of YOLOv7+CBAM network in the training process are shown in Figure 9.

Table 1. Recognition effect parameters of different networks.

Model	P/%	Recall/%	F1 Score/%	Detection Speed/FPS
AB1=SE	85.75	77.32	0.82	54.92
AB2=SE	86.97	81.56	0.83	49.74
AB1=ECA	89.04	84.43	0.84	58.37
AB2=ECA	86.48	81.97	0.84	55.09
AB1=CBAM	89.06	84.70	0.87	60.03
AB2=CBAM	87.11	81.97	0.84	58.79
AB1=CA	87.16	82.38	0.85	56.26
AB2=CA	85.36	81.23	0.83	54.23
No additions	87.33	83.12	0.84	57.37

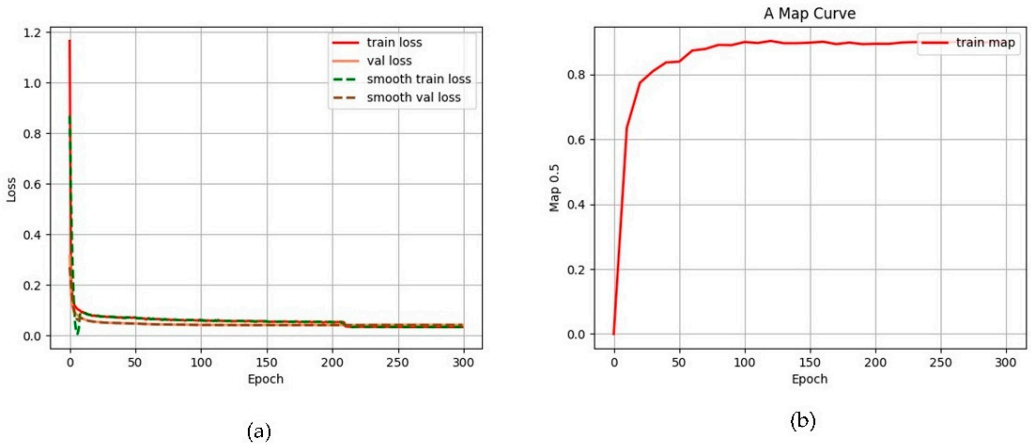

Figure 9. YOLOv7+CBAM network diagram. (**a**) Loss value change curve during training. (**b**) mAP curve change during training.

By comparing the recognition effects of different attention modules placed at different positions in the YOLOv7 network, it can be seen that after adding CBAM Block and ECA Block, YOLOv7 can achieve certain improvement in the parameters of P, Recall, F1 score and detection speed. Overall, it had a better recognition effect.

(2) In order to compare the influence of the number of images used for training on the network recognition effect, we used 1049 images taken and 1000 images after data enhancement during the formal training phase. The total images number of the dataset was 2049, and the training parameters remained unchanged compared with the pre-training phase. We selected the network of AB1=CBAM and AB1=ECA. The recognition effect of YOLOv7+CBAM network, YOLOv7+ECA network and YOLOv7 network were compared. The recognition effect parameters of the three networks are shown in the following Table 2.

Table 2. Identification effect parameters of different models.

Model	P/%	Recall/%	F1 Score/%	Detection Speed/FPS
YOLOv7	89.23	85.34	0.87	58.21
YOLOv7+ECA	91.83	87.16	0.89	59.64
YOLOv7+CBAM	93.71	89.23	0.91	61.23

(3) The bud images with occlusion, small targets and dense distribution were selected, respectively. The detection effects of the improved attention mechanism network based on the CBAM network, the ECA network and the YOLOv7 network without added attention mechanism were compared, as shown in Figure 10. It was found that the improved YOLOv7 network based on CBAM Block had higher recognition accuracy. Additionally, it had a lower rate of missed detection.

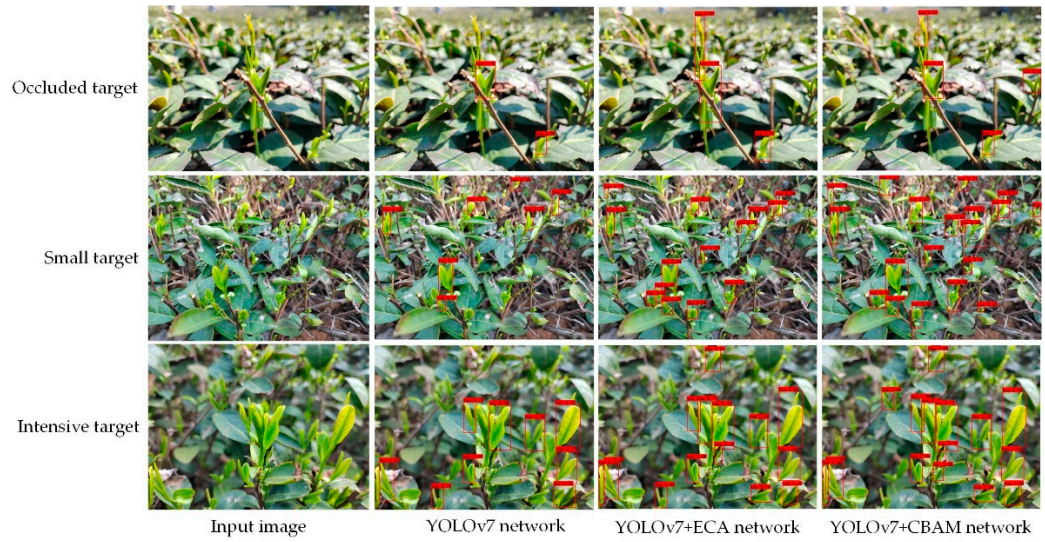

Figure 10. Comparison of recognition effects of different networks.

2.8. Visual Recognition of Heatmap

In order to visually explain the detection process of YOLOv7+CBAM network model, the visualization method of gradient-weighted class activation mapping (Grad-CAM) [29] was adopted in this paper. The recognition effect of YOLOv7 network model, YOLOv7+ECA network model and YOLOv7+CBAM network model was compared, respectively. In the Grad-CAM visualization method, the fusion weights of target feature maps are expressed as gradients, and the global average of gradients is used to calculate the weights. After the weights of all feature maps of each category are obtained, the weights are weighted and heatmap is obtained.

The heatmap can intuitively show the focus of attention of the model when extracting features. The warmer the color, the more attention of the model, and the red part (the warmest part) represents the focus of the model. Leafy, occluded and densely distributed buds were plotted using Grad-CAM, respectively, as shown in Figure 11. As can be seen in Figure 11, the YOLOv7+CBAM network model can accurately focus on different types of images and was little affected by background factors, which further proves that the network proposed in this study had a better effect on improving the detection effect of famous and excellent tea.

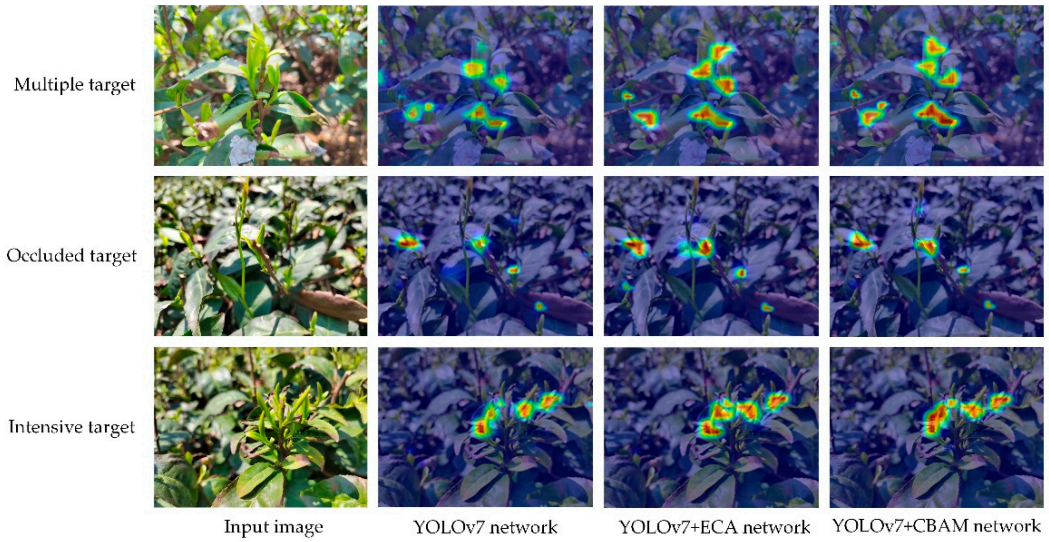

Figure 11. Activation graphs of famous and excellent tea image classes of different models.

3. Conclusions

(1) This study compared the attention mechanisms' effects of the SE, ECA, CBAM and CA blocks on the bud detection of famous and excellent green tea in different positions of YOLOv7 network. It was found that the YOLOv7+CBAM network model had the best recognition effect with the recognition accuracy of 93.71%, recall rate of 89.23% and F1 score of 0.91.

(2) The improved YOLOv7 networks basing CBAM, ECA attention mechanism were compared with the YOLOv7 network for the bud images with occluding, small targets and dense distribution, and it was found that the YOLOv7+CBAM network had better recognition effect on various tender leaves.

(3) Multi-leaf, occluding and densely distributed images containing young tea leaves were drawn by Grad-CAM, respectively. It could be seen that the YOLOv7+CBAM network model could accurately focus on different types of images and was little affected by background factors, which further proved that the network proposed in this study had a better effect on improving the recognition effect of famous and excellent green tea.

Author Contributions: Conceptualization, M.X. and X.W.; data curation, S.W. and Q.J.; writing—original draft, Y.W. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Jiangsu Agricultural Science and Technology Innovation Fund Project (CX (21) 3148), the Fundamental Research Funds for the Central Universities (YDZX2023007) and Key R&D Program of Jiangsu Province (BE2021016).

Data Availability Statement: The data that were used are confidential.

Acknowledgments: The authors would like to thank the help of Juyuanchun Tea Garden for providing us with the experimental site and the funding support above all. Yongwei Wang wants to thank the standing support from the teacher Yongnian Zhang and the standing patience, company and encouragement from Shu Wang, particularly.

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

References

1. Liang, Y.; Shin, Y.H.; Zhang, L.-J.; Wang, K.-R. Advances in tea Plant Breeding in China. *Agric. Food* **2019**, *7*, 1–10.
2. Hicks, A. Review of global tea production and the impact on industry of the Asian economic situation. *AU J. Technol.* **2001**, *5*, 227–231.
3. Yang, H.; Chen, L.; Ma, Z.; Chen, M.; Zhong, Y.; Deng, F.; Li, M. Computer vision-based high-quality tea automatic plucking robot using Delta parallel manipulator. *Comput. Electron. Agric.* **2021**, *181*, 105946. [CrossRef]
4. Lu, Y.J. Debiao, Significance and realization of mechanized picking of famous green tea in China. *Chin. Tea* **2018**, *40*, 1–4.
5. Fan, W.Y.H.; Xin, Y.-Y.; Fei, L.; Ting, Z.; Li, C.-H. Chinese tea mechanization picking technology research status and development trend. *Jiangsu Agric. Sci.* **2019**, *47*, 48–51.
6. Fuzeng, Y.L.Y.; Yana, T.; Qing, Y. Tea bud recognition method based on color and shape characteristics. *Trans. Chin. Soc. Agric. Mach.* **2009**, *40*, 119–123.
7. Jian, W. Research on tea image segmentation Algorithm Combining color and region growth Wang Jian. *Tea Sci.* **2011**, *31*, 72–77.
8. Miaoting, C. Recognition and Localization of Famous Tea bud Based on Computer Vision. Master's Thesis, Qingdao University of Science and Technology, Qingdao, China, 2019.
9. Wu, X.; Zhang, F.; Lv, J. Research on tea leaf recognition method based on image color information. *Tea Sci.* **2013**, *33*, 584–589.
10. Tang, Y.; Han, W.; Hu, A.; Wang, W. Design and Experiment of Intelligentized Tea-plucking Machine for Human Riding Based on Machine Vision. *Nongye Jixie Xuebao/Trans. Chin. Soc. Agric. Mach.* **2016**, *47*, 15–20. [CrossRef]
11. Bao, W.; Fan, T.; Hu, G.; Liang, D.; Li, H. Detection and identification of tea leaf diseases based on AX-RetinaNet. *Sci. Rep.* **2022**, *12*, 2183. [CrossRef]
12. Yang, H.; Chen, L.; Chen, M.; Ma, Z.; Deng, F.; Li, M.; Li, X. Tender tea shoots recognition and positioning for picking robot using improved YOLO-V3 model. *IEEE Access* **2019**, *7*, 180998–181011. [CrossRef]
13. Wang, T.; Zhang, K.; Zhang, W.; Wang, R.; Wan, S.; Rao, Y.; Jiang, Z.; Gu, L. Tea picking point detection and location based on Mask-RCNN. *Inf. Process. Agric.* **2023**, *10*, 267–275. [CrossRef]
14. Chen, Y.-T.; Chen, S.-F. Localizing plucking points of tea leaves using deep convolutional neural networks. *Comput. Electron. Agric.* **2020**, *171*, 105298. [CrossRef]
15. Sozzi, M.; Cantalamessa, S.; Cogato, A.; Kayad, A.; Marinello, F. Automatic Bunch Detection in White Grape Varieties Using YOLOv3, YOLOv4, and YOLOv5 Deep Learning Algorithms. *Agronomy* **2022**, *12*, 319. [CrossRef]
16. Cardellicchio, F.S.A.; Dimauro, G.; Petrozza, A.; Summerer, S.; Cellini, F.; Renò, V. Detection of tomato plant phenotyping traits using YOLOv5-based single stage detectors. *Comput. Electron. Agric.* **2023**, *207*, 1077757. [CrossRef]
17. Wang, D.; He, D. Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. *Biosyst. Eng.* **2021**, *210*, 271–281.
18. Wu, D.; Lv, S.; Jiang, M.; Song, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Comput. Electron. Agric.* **2020**, *178*, 105742. [CrossRef]
19. Liu, T.; Teng, G.; Yuan, Y.; Liu, B.; Liu, Z. Winter jujube fruit recognition method in natural scene based on improved YOLO v3. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 17–25.
20. Yang, B.; Gao, Z.; Gao, Y.; Zhu, Y. Rapid Detection and Counting of Wheat Ears in the Field Using YOLOv4 with Attention Module. *Agronomy* **2021**, *11*, 1202. [CrossRef]
21. Liu, Y.; Cao, X.; Guo, B.; Chen, H.; Dai, Z.; Gong, C. Research on Attitude detection Algorithm of meat goose in complex scene based on improved YOLO v5. *J. Nanjing Agric. Univ.* **2022**, 1–12.
22. Fang, M.; Lü, J.; Ruan, J.; Bian, L.; Wu, C.; Qing, Y. Tea bud detection model based on improved YOLOv4-tiny. *Tea Sci.* **2022**, *42*, 549–560.
23. Fu, X.; Li, A.; Meng, Z.; Yin, X.; Zhang, C.; Zhang, W.; Qi, L. A Dynamic Detection Method for Phenotyping Pods in a Soybean Population Based on an Improved YOLO-v5 Network. *Agronomy* **2022**, *12*, 3209. [CrossRef]
24. 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, 18–23 June 2018; pp. 7132–7141.
25. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. Supplementary material for 'ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 13–19.
26. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
27. 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, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.

28. Lang, C.; Chao, J. X-ray image rotating object detection based on improved YOLOv7. *J. Graph.* **2023**, *44*, 324–334.
29. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference On Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.