agronomy

Article

Hierarchical Detection of *Gastrodia elata* Based on Improved YOLOX

Xingwei Duan [1], Yuhao Lin [1], Lixia Li [1], Fujie Zhang [1,*], Shanshan Li [1] and Yuxin Liao [2]

[1] Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650500, China; 20212214059@stu.kust.edu.cn (X.D.); 20202114016@stu.kust.edu.cn (Y.L.); lilixia2012@kust.edu.cn (L.L.); 20212214014@stu.kust.edu.cn (S.L.)
[2] Faculty of Electric Power Engineering, Kunming University of Science and Technology, Kunming 650500, China; 20202205021@stu.kust.edu.cn
* Correspondence: 20030031@kust.edu.cn

Abstract: Identifying the grade of *Gastrodia elata* in the market has low efficiency and accuracy. To address this issue, an I-YOLOX object detection algorithm based on deep learning and computer vision is proposed in this paper. First, six types of *Gastrodia elata* images of different grades in the *Gastrodia elata* planting cooperative were collected for image enhancement and labeling as the model training dataset. Second, to improve feature information extraction, an ECA attention mechanism module was inserted between the backbone network CSPDarknet and the neck enhancement feature extraction network FPN in the YOLOX model. Then, the impact of the attention mechanism and application position on model improvement was investigated. Third, the 3 × 3 convolution in the neck enhancement feature extraction network FPN and the head network was replaced by depthwise separable convolution (DS Conv) to reduce the model size and computation amount. Finally, the EIoU loss function was used to predict boundary frame regression at the output prediction end to improve the convergence speed of the model. The experimental results indicated that compared with the original YOLOX model, the mean average precision of the improved I-YOLOX network model was increased by 4.86% (97.83%), the model computation was reduced by 5.422 M (reaching 3.518 M), the model size was reduced by 20.6 MB (reaching 13.7 MB), and the image frames detected per second increased by 3 (reaching 69). Compared with other target detection algorithms, the improved model outperformed Faster R-CNN, SSD-VGG, YOLOv3s, YOLOv4s, YOLOv5s, and YOLOv7 algorithms in terms of mean average precision, model size, computation amount, and frames per second. The lightweight model improved the detection accuracy and speed of different grades of *Gastrodia elata* and provided a theoretical basis for the development of online identification systems of different grades of *Gastrodia elata* in practical production.

Keywords: *Gastrodia elata*; YOLOX; target detection; ECA; DS Conv; EIoU

Citation: Duan, X.; Lin, Y.; Li, L.; Zhang, F.; Li, S.; Liao, Y. Hierarchical Detection of *Gastrodia elata* Based on Improved YOLOX. *Agronomy* **2023**, *13*, 1477. https://doi.org/10.3390/agronomy13061477

Academic Editor: Silvia Arazuri

Received: 12 April 2023
Revised: 13 May 2023
Accepted: 24 May 2023
Published: 26 May 2023

1. Introduction

As a geographical indication protection product of Yunnan, *Gastrodia elata* is a type of rare Chinese medicinal material, and it is widely used in the treatment of spasms, vertigo, paralysis, epilepsy, tetanus, asthma, and immune dysfunction [1]. At present, the market classification of *Gastrodia elata* is only performed according to its size, which is a sign of the quality of traditional Chinese medicine and the basis for commodity pricing. In the market, different grades of *Gastrodia elata* have large price differentiation and component content differences [2,3]. The weight and appearance of *Gastrodia elata* are the main factors in market classification, and the classification methods mainly include manual sorting and mechanical sorting. The former relies on experience and manual operation, with strong subjective factors, low sorting efficiency, and high labor intensity, and it often causes misclassification; the latter mainly adopts the way of weighing with a single classification

standard. Therefore, it is of great significance to classify different grades of *Gastrodia elata* quickly and accurately.

Currently, with the development of computer vision, deep learning has been widely applied in agriculture [4–6], medicine [7–9], and other fields. Though the classification of *Gastrodia elata* takes both weights and shape into consideration, it is only sorted by manual experience or only considering weight, leading to low sorting accuracy and a heavy workload. The application of computer vision technology based on deep learning to the detection and classification of agricultural products, such as fruits [10,11], vegetables [12,13], and Chinese medicinal materials [14,15], provides a reference for the visual classification of *Gastrodia elata*.

For instance, Wang et al. [16] proposed an improved object detection network I-YOLOv4-Tiny based on YOLOv4-Tiny to realize precise and rapid identification of blueberry fruit maturity in the complex natural environment. The convolutional attention module CBAM was added to the neck network FPN. Experimental results indicated that the trained I-YOLOv4-Tiny target detection network achieved an average accuracy of 97.30% on a blueberry data set. By comparing YOLOv4-Tiny, YOLOv4, SSD-MobileNet, and Faster R-CNN object detection networks, the average accuracy can reach 96.24% in complex scenes with unequal occlusion and illumination. The average detection time was 5.723 ms, and the memory occupation was only 24.20 M, which can meet the requirements of blueberry fruit recognition accuracy and velocity. Deng et al. [17] constructed a lightweight object detection model (CDDNet) to identify carrots for classification, and a carrot classification approach was proposed based on the smallest enclosing rectangle (MBR) fitting and convex polygon approximation. The experimental results showed that the precision of the proposed CDDNet was 99.82% for the two-category classification (normal, flawed) and 93.01% for the four-category classification (normal, bad, abnormal, and fibrous root). The classification precision of MBR fitting and convex polygon approximation were 92.8% and 95.1%, respectively, indicating that the method can detect carrot defects quickly and precisely. Xu et al. [18] designed an improved YOLOv5 detection network to identify jujube maturity by integrating Stem, RCC, Maxpool, CBS, SPPF, C3, PANet, and CIoU loss networks, which improved the detection accuracy of jujube in complex environments to 88.8% and frames per second (FPS) to 245. The improved network model YOLO-Jujube was proven to be suitable for the identification of jujube maturity. Wang et al. [19] proposed a maturity detection approach for millet spicy green pepper in a complex orchard environment. Based on the improved YOLOv5s model, the convolutional layer of the cross-phase part (CSP) in the backbone network was replaced by GhostConv, the attention mechanism (CA) module was added, and the path aggregate network (PANet) in the neck network was replaced by the bidirectional feature pyramid network (BiFPN) to improve detection accuracy. The experimental results indicated that the maximum mAP of the modified model was 85.1%, and the minimum model size was 13.8 MB. Li et al. [20] proposed a modified YOLOX object detection model called YOLOX-EIoU-CBAM, which was applied to identify the maturity category of sweet cherries quickly and accurately in natural environments. The convolutional attention module (CBAM) was added to the model to consider the different maturity characteristics of sweet cherries. Meanwhile, the replacement of the loss function with an Efficient IoU loss makes the regression of the prediction box more precise. The experimental results indicated that compared with the YOLOX model, the mAP, recall rate, and F-score of this method were increased by 4.12%, 4.6%, and 2.34%, respectively, and the model size and single picture extrapolation time were basically the same.

Studies have shown that deep learning can effectively solve difficult object recognition problems in complex environments, attributed to its high robustness and generalization ability [21,22]. Particularly, the YOLO model can be improved according to the target characteristics and application scenarios to improve the model performance. Therefore, it is necessary to realize efficient and accurate sorting of different grades of *Gastrodia elata*

by the deep learning model, which is of great practical significance for promoting the development of the whole *Gastrodia elata* industry.

At present, major target detection algorithms conduct target detection of large sample sizes and multiple categories, and their network structure may not be suitable for all projects. Two-stage detection algorithms [23,24] require a large amount of computation, which is difficult to meet the requirement of real-time detection in an ordinary hardware environment. The one-stage detection algorithm [25] has a fast detection speed and can realize real-time detection, but its detection accuracy is slightly lower than that of two-stage detection algorithms. The most classic algorithm in this category is YOLO (You only look once) series [26], among which the YOLOX target detection algorithm is a relatively new and widely used target detection algorithm at present [27]. However, due to its complex network structure, it requires high computing power when applied to devices. To achieve real-time detection and classification of *Gastrodia elata* in complex environments, it is crucial to improve the model to reduce the number of parameters, improve the model accuracy, and keep the model size small.

To date, no research has conducted research on *Gastrodia elata* grade detection model based on deep learning and computer vision. To meet the requirements of rapid and accurate detection of different grades of *Gastrodia elata*, a YOLOX-based target detection algorithm I-YOLOX for different grades of *Gastrodia elata* was proposed in this study. To enhance the feature expression and improve the detection performance of *Gastrodia elata* grade, the ECA attention mechanism module [28] was introduced into the transmission process of the feature layer between the backbone network and the neck enhancement feature extraction network in the YOLOX model, and the improvement effect of different attention mechanisms on the network was compared. In the neck enhancement feature extraction network FPN [29] and the head network, depthwise separable convolution was replaced to reduce the calculation of model parameters, and the improvement effects of different replacement positions were analyzed. The EIoU loss function [30] was used to improve the model convergence effect and make the prediction box regression more precise. Finally, the feasibility and reliability of the proposed method were validated on the *Gastrodia elata* data set.

By establishing the improved YOLOX target detection model of different grades of *Gastrodia elata*, the problem of grade discrimination relying on manual experience and low recognition efficiency was solved, which provided the foundation for the construction of the sorting system of *Gastrodia elata* later.

The subsequent sections are structured as follows: Section 2 introduces the establishment of the *Gastrodia elata* image data set and the detailed content of the I-YOLOX classification detection algorithm proposed in this study. Section 3 evaluates the performance of the I-YOLOX network through experiments. Section 4 summarizes the work of this study and points out the shortcomings and prospects of this study.

2. Materials and Methods

In order to identify different grades of *Gastrodia elata* quickly and effectively, the image data set of *Gastrodia elata* was established, and the network model was improved by different methods. An improved YOLOX target detection algorithm for different grades of *Gastrodia elata* was proposed, which improved the detection effect of *Gastrodia elata*.

2.1. Image Data Acquisition of Gastrodia elata

The images of different grades of *Gastrodia elata* were taken and collected, and the image data set of *Gastrodia elata* was established by image screening and enhancement.

2.1.1. Test Materials

Yunnan is one of the origins of *Gastrodia elata*, and Yiliang County, Zhaotong City, is the representative and main producing area of Yunnan Xiaocaoba *Gastrodia elata*. Different grades of Gastrodia have a large price difference. In this study, Xiaocaoba *Gastrodia elata* was

taken as the research object, and data collection was conducted in Xiaocaoba Changtong *Gastrodia elata* Cooperative. After measuring and weighing each Gastrodia, more than 200 *Gastrodia elata* of each grade were selected to collect image information, including Grade S, Grade 1, Grade 2, Grade 3, Grade 4, and Bad Grade. The six grades of *Gastrodia elata* are shown in Figure 1.

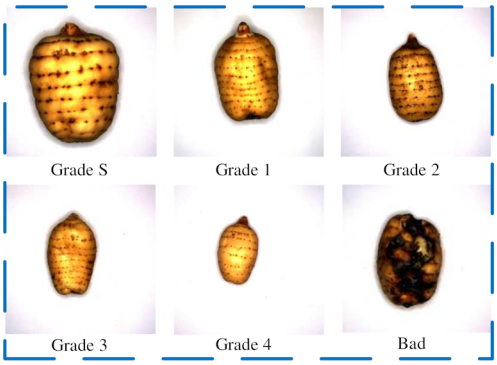


Figure 1. Various grades of *Gastrodia elata*.

2.1.2. Grading Standards of *Gastrodia elata*

According to the local standard of *Gastrodia elata* of Yunnan Province (DB53T1077-2021), fresh *Gastrodia elata* was divided into five grades: Grade S, Grade 1, Grade 2, Grade 3, and Grade 4. In the process of growth, digging, and storage of fresh *Gastrodia elata*, there are Bad *Gastrodia elata*, such as mildew, moths, skin damage, etc. In this paper, to better study the classification of fresh *Gastrodia elata*, Bad *Gastrodia elata* is added to the above five grades as one grade, and there are six grades in total. The grading specifications of fresh *Gastrodia elata* are listed in Table 1.

Table 1. The grading specifications of fresh *Gastrodia elata*.

Grade	Number (pcs/kg)	Weight (g/pcs)	Shape
S	≤4	≥250	Length: 10–13.5 cm, width: 5.5–8 cm, thickness: 4.5–6.5 cm, length-width ratio: 1.47–2.48
1	≤5	≥200	Length: 8.5–13 cm, width: 5–7 cm, thickness: 4–6 cm, length-width ratio: 1.23–2.59
2	≤7	≥150	Length: 8–13 cm, width: 4.5–6.5 cm, thickness: 3.5–6 cm, length-width ratio: 1.29–2.62
3	≤10	≥100	Length: 7.5–12.5 cm, width: 3.5–6 cm, thickness: 3–5.5 cm, length-width ratio: 1.31–2.90
4	>10	<100	*Gastrodia elata* not belonging to Grade S, Grade 1, Grade 2, or Grade 3 are of this grade.
Bad	\	\	mildew, moth-eaten, breakage

2.1.3. Data Collection

In this study, the Hikvision industrial camera (MV-CA050-20GC, 5 million pixels, CMOS Gigabit Ethernet industrial array camera, resolution of 2592 × 2048, Hikvision, Hikvision) was used for image acquisition. The camera was set up with an adjustable bracket, and the camera was fixed 25 cm away from the horizontal plane to take pictures. The model of the light source is JL-HAR-110W, the power is 5.9 W, and the installation height is 25 cm away from the horizontal plane. Material surface brightness is about 2.3×10^4 Lux. All images were captured under the same camera height, the same white background plate, and the same light source brightness.

As shown in Figure 2, industrial cameras were used in this study to take images of *Gastrodia elata* of six grades, namely, Grade S, Grade 1, Grade 2, Grade 3, Grade 4, and Bad *Gastrodia elata*, with more than 200 *Gastrodia elata* in each grade. A total of 800 images of each grade were collected, and a total of 4800 images were saved in the .jpg format.

2.2. Image Data Screening, Enhancement, and Dataset Establishment

The main purpose of this study is to improve the classification recognition accuracy of *Gastrodia elata*. First, the collected image data were screened, and the images that were not suitable for the detection algorithm caused by damaged image data files and image blurring caused by external factors in the collection process were screened. As shown in Table 2, more than 800 original images of each grade of *Gastrodia elata* were collected. After screening, 800 images of each grade of *Gastrodia elata* were retained, and a total of 4800 images of six grades were collected. Then, the images of each grade in the data set saved after screening were enhanced.

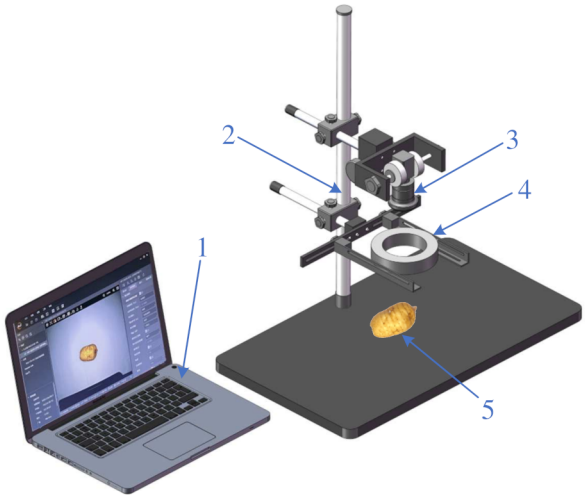


Figure 2. Image acquisition test bench. (**1**) Computers; (**2**) Image acquisition bracket; (**3**) Camera; (**4**) Light source; (**5**) Fresh *Gastrodia elata* samples.

Table 2. Sample number of *Gastrodia elata* data set.

Grade	Number of Original Pictures	Number of Pictures after Screening	Number of Expand Pictures	Data Set			Total
				Training Set	Verification Set	Test Set	
S	828	800	800	1280	160	160	1600
1	867	800	800	1280	160	160	1600
2	857	800	800	1280	160	160	1600
3	816	800	800	1280	160	160	1600
4	823	800	800	1280	160	160	1600
Bad	821	800	800	1280	160	160	1600
Total	5012	4800	4800	7680	960	960	9600

There was only a single *Gastrodia elata* in each image collected, and there were no other types of targets. Therefore, Mosaic and MixUp data enhancement strategies for YOLOX were not adopted in this study because they have good effects for detecting multi-target and multi-type images. For the *Gastrodia elata* dataset in this study, it is more suitable for random processing of the original image data through flipping, translation, rotation, changing brightness, and adding noise, thus expanding the number of images, enhancing the diversity of image information and samples, and avoiding overfitting in the training process. The image after the sample image enhancement is shown in Figure 3.

In this paper, by image enhancement, the images of each grade of *Gastrodia elata* were expanded at an equal ratio, the dataset was expanded to 9600, and the training set, verification set, and test set were divided at a ratio of 8:1:1. The sample quantity of the *Gastrodia elata* dataset is presented in Table 2. Additionally, LableImg software version 1.8.1 was used to annotate the image dataset, and XML files containing the dataset name, center point (x_c, y_c) coordinate information, label name, and other information were obtained. Then, the XML files were converted into annotation files in the txt format required by the YOLO training model through Python programming. In this way, the image dataset of *Gastrodia elata* was established.

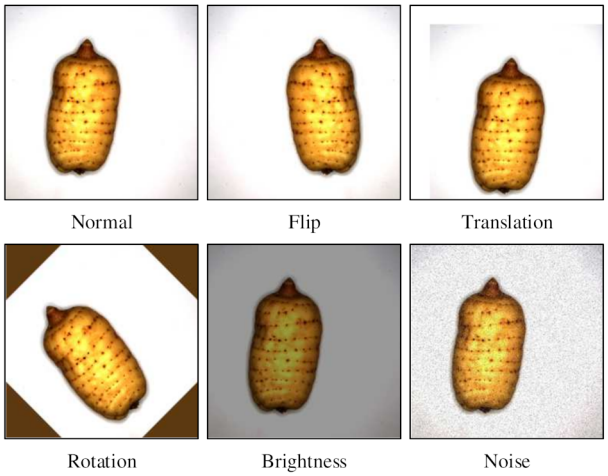


Figure 3. The image enhancement of the *Gastrodia elata* dataset.

2.3. Improved YOLOX Gastrodia Classification Recognition Model

By introducing the ECA attention mechanism, replacing the depthwise separable convolution, and using the EIoU position regression function calculation to improve the original YOLOX network structure, an improved YOLOX *Gastrodia elata* classification recognition model was established.

2.3.1. I-YOLOX Network

At present, the main target detection algorithm network structure is diverse and complex, with different applicability. YOLOX, as one of the most advanced real-time object detection algorithms, has the advantages of high detection precision and flexible deployment. Considering the problems of model deployment and cost control, the YOLOXs target detection network is selected as the basic network model in this paper to reduce model storage occupation and improve the identification speed.

The network structure of YOLOX consists of four parts: the input terminal, the backbone feature extraction network, the enhanced feature extraction network, and the head prediction network. Specifically, the input terminal performs data enhancement and adaptive image scaling for the input image. The backbone feature extraction network adopts residual convolution operation to extract feature maps at different levels [31] and adopts the CSPDarknet53 structure for the main part. Then, the extracted features are fused by jumping connections, which alleviates the problem of gradient disappearance caused by adding depth in the deep learning network, reduces parameter redundancy, and improves model accuracy. In the network structure, the feature pyramid network structure is used to enhance the semantic features from top to bottom, and the path aggregation network structure is used to enhance the positioning features from bottom to top. Then, feature fusion is performed by combining the information of different scales to achieve a better feature extraction effect. The head prediction network forgoes the coupled head method of the YOLO series and uses the decoupled head as the detection head to support two branches, classification and regression, where the former obtains category information, and the latter obtains detection frame information and confidence information. Finally, the information is integrated into the prediction stage. After decoupling, different branches of the detection head have independent parameters, so directional reverse optimization can be performed according to the loss function to accelerate the convergence speed and improve the precision of the model.

YOLOX re-adopted the idea of Anchor free. Instead of using clustering algorithm to obtain prior boxes, it used simOTA to flexibly match positive samples for objects of different sizes. This method solves the problems that anchor-based detection method requires artificial design of Anchor frames and a large number of anchor frames in the training process cause huge computation. Since the shapes and sizes of different grades

of *Gastrodia elata* vary, the model detection box needs to adapt to the target detection of various scales of *Gastrodia elata*. Therefore, simOTA positive and negative samples in YOLOX target detection algorithm are more suitable for hierarchical detection of different grades of *Gastrodia elata*. The data enhancement strategy enhanced the image information and sample diversity and improved the generalization ability of target detection model of *Gastrodia elata*.

In this study, only the classification of *Gastrodia elata* was involved. To enhance the feature expression and improve the detection performance of *Gastrodia elata* grade, as shown in Figure 4, the following improvements were made to YOLOX: the ECA attention mechanism module is added to the basic YOLOX network to enhance image feature extraction; in the YOLOX FPN network and head network, depthwise separable convolution is used instead of normal convolution to further reduce the number model parameters. EIoU loss is used to replace the boundary box loss function in the original YOLOX model to make the regression of the prediction frame more precise.

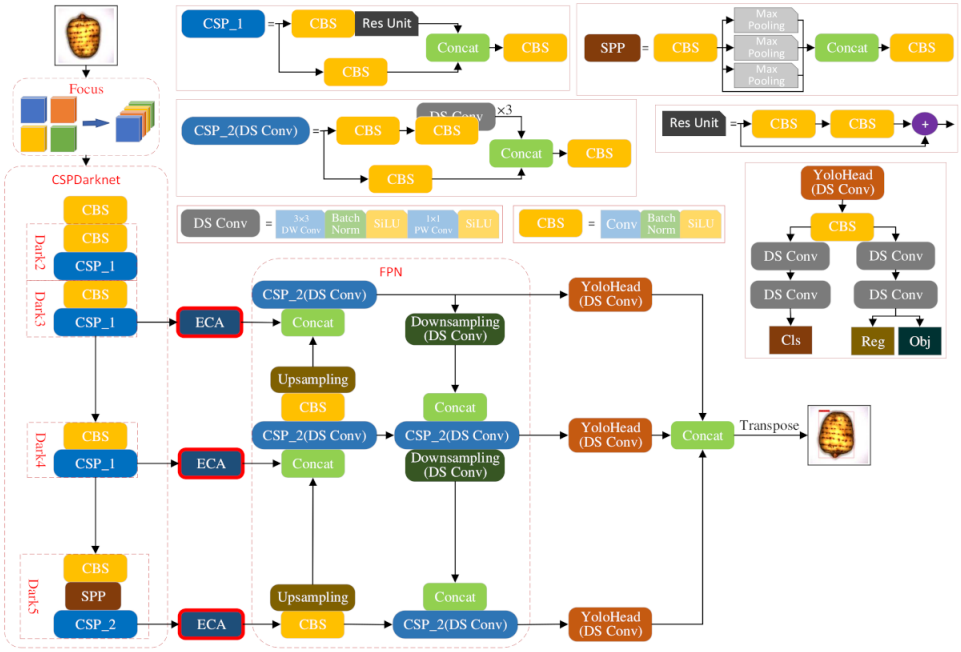


Figure 4. I-YOLOX network structure.

2.3.2. The Network Structure Improvement of the ECA Attention Mechanism

The attention mechanism imitates the biological vision mechanism. By rapidly scanning the global image, the regions of interest can be selected, more attention resources can be invested, and other useless information can be suppressed, thus improving the efficiency and accuracy of visual information processing.

The ECA attention mechanism module is a channel attention module, which is commonly used in visual detection models. It can enhance the channel feature of the input feature graph, eliminate the full connection layer, avoid dimension reduction, and capture cross-channel interactions effectively. The final output of the ECA module does not change the size of the input characteristic pattern. It can be regarded as an improved version of the SE attention module: it solves the problem that dimension reduction in the SE attention module brings side effects to the channel attention mechanism, and captures the dependency between all channels has low efficiency.

As shown in Figure 5, the ECA module performs global averaging pooling on input feature graphs, making the feature graphs change from a matrix of size [H,W,C] to a vector of size [1,1,C]. After the global averaging pooling layer, one-dimensional 1×1 convolution is used to obtain a cross-channel mutual information. The size of the convolution kernel is

adjusted by an adaptive function, which allows layers with a larger number of channels to interact more across channels. The adaptive function is represented below:

$$k = \left| \frac{\log_2(c)}{\gamma} + \frac{b}{\gamma} \right| \quad (1)$$

where, $\gamma = 2$ and $b = 1$.

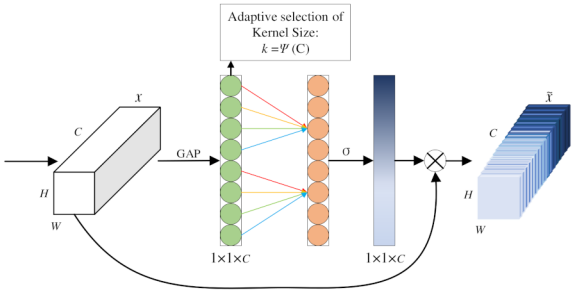


Figure 5. The schematic diagram of the ECA attention mechanism module.

The adaptive function is applied to one-dimensional 1×1 convolution to obtain the weight of each channel of the characteristic pattern. Finally, the normalized weight is multiplied by the initial input characteristic pattern channel by channel to generate the weighted characteristic pattern with channel attention.

2.3.3. Structure Improvement of the Feature Fusion Network Based on Depthwise Separable Convolution

The depthwise separable convolution is composed of depthwise (DW) convolution and pointwise (PW) convolution. Similar to normal convolution, this configuration can be used to extract characteristics, but it has a smaller parameter number and lower work cost than normal convolution. The common convolution in YOLOX's enhanced feature extraction network FPN and the head network was replaced by depthwise separable convolution to further compress the model and improve its computational efficiency.

As shown in Figure 6, normal convolution is to perform convolution operation on the input characteristic pattern and the corresponding convolution kernel of each channel, then add them to output features. The calculation amount P_1 is:

$$P_1 = D_k D_k M N D_w D_h \quad (2)$$

where D_k represents the size of the convolution kernel; M and N represent the number of channels of input and output data, respectively; and D_w and D_h represent the width and length of output data, respectively.

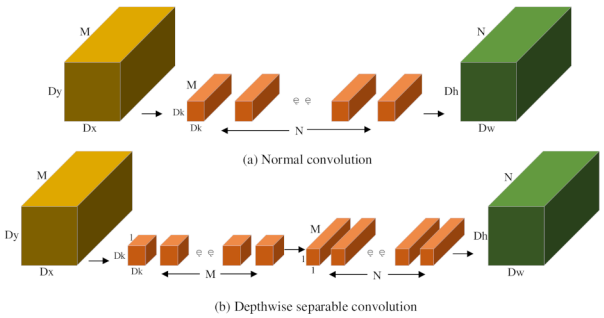


Figure 6. Normal convolution and Depthwise separable convolution.

Depthwise separable convolution changes the one-step operation of ordinary convolution into 3 × 3 depthwise convolution and 1 × 1 pointwise convolution, and its calculation amount P_2 is:

$$P_2 = (D_k D_k M + MN) D_w D_h \tag{3}$$

Therefore, the arithmetical ratio of depth-wise separable convolution to normal convolution is:

$$\frac{P_2}{P_1} = \frac{(MD_w D_h)(D_k^2 + N)}{D_k^2 MND_w D_h} = \frac{1}{N} + \frac{1}{D_k^2} \tag{4}$$

Generally, D_k is set to 3, and the network computation amount and parameter number are reduced by about 1/3 after replacing normal convolution with depthwise separable convolution.

The lightweight improvement of the network structure can greatly reduce the parameter number and calculation amount of the model; however, it will cause a loss of detection accuracy. Therefore, it is necessary to further optimize the model to improve the detection precision of the model.

2.3.4. Adopting the EIoU Position Regression Loss Function

The EIoU loss consists of three parts: IoU loss, distance loss, and height-width loss (overlapping area, center distance, and height-width ratio). The height-width loss directly minimizes the difference between the height and width of the predicted object boundary box and the real boundary box to achieve a faster convergence rate and better positioning results.

As shown in Figure 7, the calculation formula for the EIoU loss is as follows:

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} = 1 - IoU(A,B) + \frac{\rho^2(b, b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \tag{5}$$

where w^c and h^c are the width and height of the minimum enclosing rectangle of the predicted boundary box and the real boundary box, respectively. ρ is the Euclidean distance between two points.

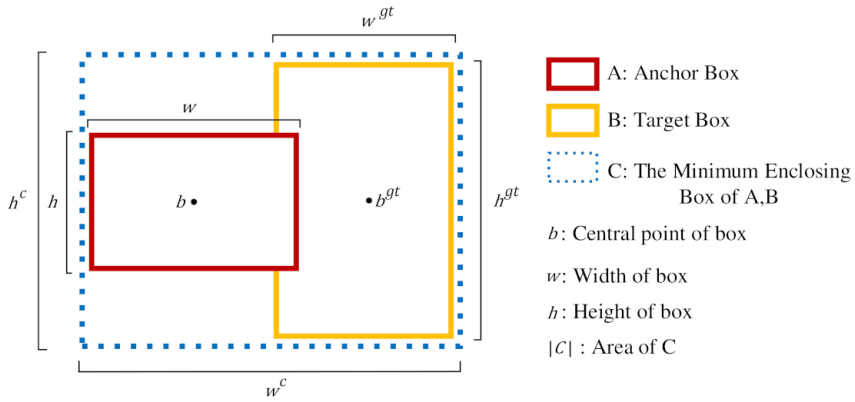


Figure 7. The schematic diagram of the EIoU loss calculation.

EIoU loss is an improvement of the CIoU loss. Based on CIoU, the aspect ratio is disassembled, and the loss item of the aspect ratio is divided into the difference between the width and height of the predicted frame and the width and height of the real minimum external frame, which improves the convergence effect and the regression precision. Meanwhile, by adding focal focusing high-quality anchor frames, the sample imbalance problem in the boundary box regression task is resolved, i.e., the contribution of many anchor frames with little registration with the object frame to the optimization of B Box regression is reduced, making the course of regression centered on superior anchor frames.

2.4. Transfer Learning

Since there is no public *Gastrodia elata* data set, there are few research on visual classification of *Gastrodia elata* at home and abroad. In order to expedite model training and improve model generalization, this study loaded the pre-training parameters of the model in VOC2012 data sets based on the think of transfer learning. During the training, the front-end pre-training weight network layer should be frozen first, and only the back-end network should be retrained, and the parameters updated. After thawing this part of the training layer, the weight can be effectively retained.

2.5. Test Environment and Parameter Setting

All tests in this paper were completed in the laboratory workstation; the workstation model is DELL-P2419H. Hardware configuration: The CPU processor is Inter Core i7-9700F CPU @ 3.70 GHz, the CPU processor core is 16, the multi-threading is 32, the running memory is 64 GB, the GPU processor is Quadro P5000, 16 G video memory, 2560 CUDA cores, and the operating system is Windows 10. Pytorch1.3.2 deep learning environment of GPU version, compiled by Python3.8 version and CUDA11.0 version. All model training and testing are worked in the same hardware environment.

The image input size is 640 × 640. The model optimizer selected SGD, and the learning rate was set to 0.01. The total number of iterations is set to 300. In the first 50 rounds of training, the network trunk was frozen, the batch size was set to 32 times, and only the later network layers were trained. In the last 250 rounds, the batch size was set to 16 times, and the thread was set to 8.

2.6. Model Evaluation Index

In this paper, precision rate (P), recall rate (R), harmonic mean ($F1$), Average Precision (AP), and Mean Average Precision (mAP) were used to evaluate the detection precision of the proposed *Gastrodia elata* grade detection model and the lightweight degree of the model and the number of frames per second (FPS) were used to judge the real-time capability of the model. The calculation formulas of P, R, $F1$, AP, and mAP are shown in Equations (6)–(10).

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \times 100\% \tag{6}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \times 100\% \tag{7}$$

$$F1 = \frac{2\mathrm{P}\cdot\mathrm{R}}{\mathrm{P} + \mathrm{R}} \times 100\% \tag{8}$$

$$AP = \int_0^1 \mathrm{P}\cdot\mathrm{Rd}(R) \times 100\% \tag{9}$$

$$mAP = \frac{\sum_{C=1}^{C} \mathrm{AP(C)}}{C} \times 100\% \tag{10}$$

where TP, FP, and FN represent the number of true cases, false positive cases, and false negative cases, respectively; true cases represent the actual positive cases that are divided into positive cases by model classification; false positive cases represent the actual negative cases that are divided into positive cases by model classification; false negative cases represent the actual positive cases that are divided into negative cases by model classification; and C is the number of detection categories. This study needs to identify the Grade S, Grade 1, Grade 2, Grade 3, Grade 4, and Bad *Gastrodia elata*, i.e., C = 6.

3. Results and Analysis

The effect of different improvement methods on model detection was verified by experiments, and the optimal classification recognition model of *Gastrodia elata* was obtained by analyzing the test results.

3.1. The Influence of Attention Mechanism Applied to Different Positions on Model Detection Effect

Based on the original YOLOX network model, the ECA attention mechanism is applied to different positions in the feature fusion network, as shown in Figure 8, and its influence on the model detection ability was analyzed. Position 1 is in the middle of the process in which the three effective feature layers $80 \times 80 \times 256$ (feature layer 1), $40 \times 40 \times 512$ (feature layer 2), and $20 \times 20 \times 1024$ (feature layer 3) obtained from the backbone feature extraction network CSPDarknet are separately input into the enhanced feature extraction network FPN for feature fusion. In the enhanced feature extraction network FPN, the obtained valid feature layer is used for further characteristic extraction. Then, feature fusion is performed on the extracted characteristics through the adaptive bidirectional fusion channel with up-sampling and down-sampling methods. Position 2 is the connection position after each up-sampling and down-sampling.

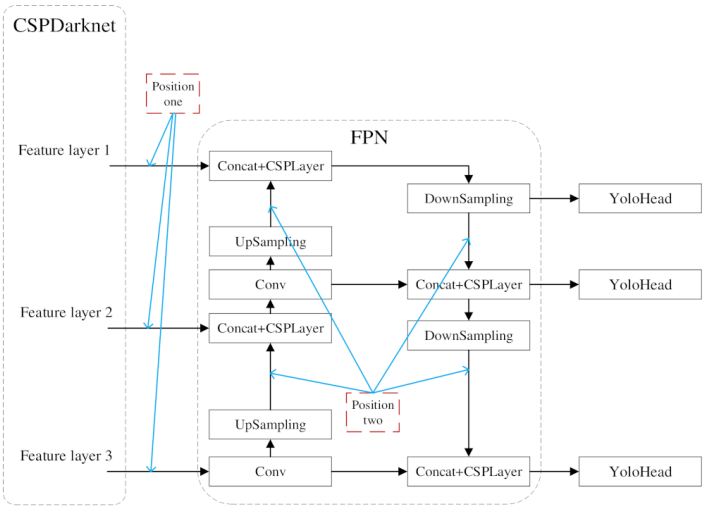


Figure 8. The feature fusion network applies coordinate attention mechanism.

The results are listed in Table 3. It can be seen that the attention mechanism ECA improves the mAP of the model by 2.84% at position 1 and 2.23% at position 2, indicating that ECA can improve the model detection performance to varying degrees at different positions in the feature fusion. Since position 1 is at the intersection of different scale information of the backbone network and the feature extraction network, compared with position 2, adding ECA to the feature extraction network after each up-sampling and down-sampling operation can enable the attention mechanism to obtain richer feature information in the information embedding stage. Meanwhile, ECA is added to both position 1 and position 2. In this case, the amount of model computation is slightly increased, and the mAP of the model is increased by 1.1%. The improvement of model detection performance is not as good as that of adding an attention mechanism in a single location. Additionally, as shown in Table 2, ECA applied to different positions occupied a smaller size of the model, and the model had basically the same size as the original YOLOX model.

Table 3. Comparison of detection ability to apply attention mechanism to different positions.

Applied Position	mAP (%)	Parameters (M)	Model Size (MB)
None	92.97%	8.94	34.3
Position 1	95.81%	8.94	34.3
Position 2	95.20%	8.94	34.3
All	94.07%	8.95	34.3

3.2. Influences of Different Attention Mechanisms on Model Detection Effects

Based on the original YOLOX network model, different attention mechanisms were applied to position 1, as shown in Figure 2, to compare the influence of different attention mechanisms on the model detection ability. It can be seen from Table 4 that, when ECA was applied to the model, compared with SE, CBAM, and no attention mechanism, the mAP obtained was the highest, which was 95.81%. After applying the SE, CBAM, and ECA attention mechanisms, the mAP increased by 0.63%, 0.48%, and 2.8%, respectively, indicating that applying different attention mechanisms in the model feature fusion network improved detection accuracy to varying degrees, so the attention mechanism should be selected according to different objects and tasks. The ECA attention mechanism module introduced in this paper can enhance channel features in the input feature graph, and the use of 1×1 convolution learning channel attention information helped to avoid the dimension reduction problem and effectively capture cross-channel interactions. Meanwhile, only a few parameters can be involved to effectively improve the model for image feature extraction of *Gastrodia elata*, obtain better detection results, and achieve a good effect even when there are water stains in the *Gastrodia elata* image. Additionally, as shown in Table 4, the increase in the model size caused by applying different attention mechanisms in the feature network was relatively small. SE and CBAM increased the model by 0.2 MB compared with the original model size without applying attention mechanisms, while the ECA maintained the same model size as the original one. After applying different attention mechanisms, SE and ECA increased the model's image detection speed, i.e., the FPS, by 1, while CBAM reduced the FPS by 5. Combined with Table 4, the following observations can be obtained: when the model size and computing power are constrained, the detection performance of the model can be improved by inserting appropriate attention mechanism into the model. The ECA attention mechanism module contributes to the optimal detection ability of the model.

Table 4. Comparison of the detection abilities of different attention mechanisms.

Attention Mechanism	mAP (%)	FPS	Model Size (MB)
None	92.97%	66	34.3
SE	93.60%	67	34.5
CBAM	93.45%	61	34.5
ECA	95.81%	67	34.3

3.3. Structure Improvement of Feature Fusion Network Based on Depthwise Separable Convolution

In the structure of the YOLOX object detection network, the effective feature layer obtained by the enhanced feature extraction network FPN is used for further feature extraction, and feature fusion is performed on the extracted features through the adaptive bidirectional fusion channel of up-sampling and down-sampling. The head network is the final prediction structure of the target detection network. Different from other YOLO versions, the head network of YOLOX has decoupled detection heads. Classification and regression are divided into two parts for processing, then the results are integrated into the final prediction stage, which can greatly improve the convergence speed of the network.

To further reduce the number of parameters of the model, as shown in Figure 7, in the FPN and head networks of the YOLOX target detection network structure, the normal convolution with a convolution kernel size of 3×3 in the CSP_2 module, downsampling

module, and YoloHead module is replaced by depthwise convolution with a convolution kernel size of 3 × 3 and point by pointwise convolution with a convolution kernel size of 1 × 1, which can reduce the number of parameters and the operation cost and improve the target detection accuracy. As shown in Table 5, when the normal convolution of the FPN network and head network was replaced by depthwise separable convolution, the mAP of the model decreased by 3.79% and 1.77%, respectively, compared with the original model. The number of parameters was reduced by 1.556 M and 3.866 M, and the model size was reduced by 14.7 MB and 5.9 MB, respectively. When the normal convolution in both the FPN and head network was replaced by the depthwise separable convolution, the mAP of the model decreased by 1.33%, which was the least. Meanwhile, the calculation amount and the model size were reduced the most, which were reduced by 5.422 M and 20.6 MB, respectively. In addition, the size of the network structure occupied by adding the alternative depthwise separable convolution is basically the same as that of the original model. Combined with Table 5, the following observations can be drawn: when the 3 × 3 normal convolution in the FPN and head networks were replaced by the depthwise separable convolution at the same time, the improved model had the least loss of detection accuracy and the best lightweight degree.

Table 5. Comparison of the detection ability of the network structure in different parts by alternative depthwise separable convolution.

Alternate Position	mAP (%)	Parameters (M)	Model Size (MB)
None	92.97%	8.940	34.3
FPN	89.18	7.384	19.6
Head	91.20	5.074	28.4
All	91.31	3.518	13.7

3.4. Ablation Test of the Improved YOLOX Model

Ablation experiments were conducted to verify the effect of combining improved mechanisms or strategies on model performance. By using the improved model to identify the grade of *Gastrodia elata*, the effectiveness of different improved models was verified.

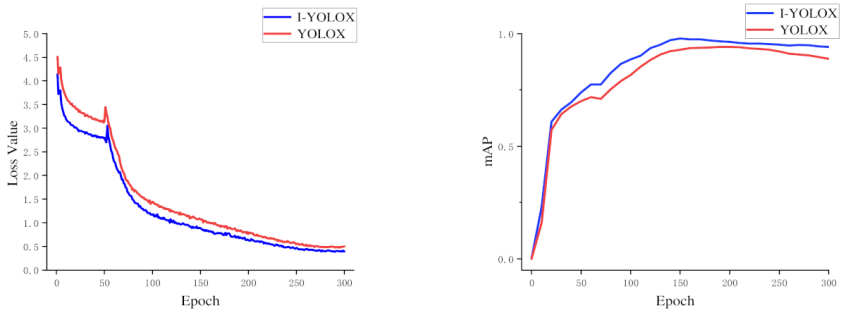
As shown in Table 6, the ECA attention mechanism was added to the YOLOX network, the normal convolution was replaced by depthwise separable convolution, and the EIoU loss function was used to perform an ablation experiment analysis. When only one type of improvement was carried out on the original network, only the replacement of depthwise separable convolution among the three improvement methods reduced the mAP of the model by 1.33%, the number of parameters was greatly reduced by 5.422 M, and the model size was reduced by 20.6 MB. The addition of the ECA attention mechanism and the use of the EIoU loss function improved the mAP of the model by 2.84% and 1.18%, respectively, while the number of parameters and the model size did not increase. When the three improved methods were combined in pairs, the I-YOLOX network model with the three improved methods achieved the highest mAP, which increased by 4.86%. Moreover, the speed of model detection, i.e., the FPS, was the highest, which was 69. Compared with other models, it had the smallest model size, which was 13.7 MB. The ablation test showed that the improvement of the grade detection model of *Gastrodia elata* had positive effects, and the improved I-YOLOX network model achieved the best effect in identifying the grade of *Gastrodia elata*. As shown in Figure 9, the loss curve of the improved I-YOLOX network model converged faster and decreased more gently than that of the original YOLOX model. Meanwhile, the mAP curves rose faster and had higher values. The detection accuracy of each grade of *Gastrodia elata* is shown in Figure 10. It can be seen that the detection effect of the improved network model was improved at all levels of *Gastrodia elata*, among which the detection accuracy of Grade 2 *Gastrodia elata* was significantly improved by 23% compared with the original model. It was verified that the improved model enhanced the extraction of important features of *Gastrodia elata* and efficiently improved the detection accuracy of the model.

Table 6. Ablation test.

Models	ECA	DS Conv	EIoU	mAP (%)	Parameters (M)	FPS	Model Size (MB)
The improved YOLOX				92.97	8.940	66	34.3
	√			95.81	8.940	67	34.3
		√		91.31	3.518	65	13.7
			√	94.15	8.940	66	34.3
	√	√		93.07	3.518	66	13.7
	√		√	95.44	8.940	68	34.3
		√	√	89.70	3.518	65	13.7
	√	√	√	97.83	3.518	69	13.7

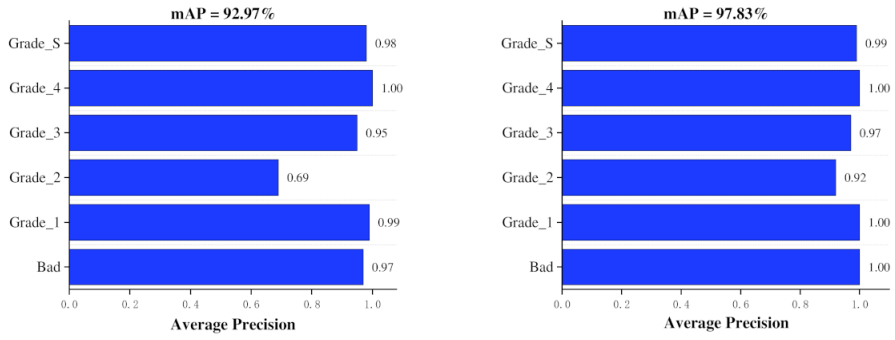
3.5. Comparative Test of Different Target Detection Algorithms

To evaluate the effect of the improved I-YOLOX network model in identifying *Gastrodia elata* grade, the same dataset was used. The target detection algorithms Fast-R-CNN, SSD, YOLOv3s, YOLOv4s, YOLOv5s, YOLOXs, and YOLOv7 were trained, respectively, under the same test conditions. The training rounds were set to 300. After the optimal weight was obtained, the tests were carried out on the same test set. Finally, the detection performance of seven target detection network models was compared.



(**a**) (**b**)

Figure 9. Comparison of the training results of the model before and after improvement. (**a**) Loss curve, (**b**) mAP curve.



(**a**) (**b**)

Figure 10. The mAP of the model for identifying each grade of *Gastrodia elata*. (**a**) YOLOX, (**b**) I-YOLOX.

As shown in Table 7, compared with other detection networks, the improved I-YOLOX network model achieved better performance in terms of P, R, and F1. It can be seen from Figure 11a that the improved network model improved the mAP compared with other detection networks. Compared with YOLOv5s, the improved network model improved the mAP by 9.95%, reaching 97.83%. It can be seen from Figure 11b that the improved

network model obtained greater FPS than other detection networks. The FPS of the single-object detection algorithm was significantly improved than that of the dual-object detection algorithm. Compared with Fast-R-CNN, the maximum improvement was 60. The improved network model achieved the highest FPS of 69. It can be seen from Figure 11c that the improved network model reduced the number of parameters compared with other detection networks. Compared with Fast-R-CNN, it was reduced by 133.273 M, and the number of parameters of the improved network model was the least, which was 3.518 M. Moreover, Figure 11d indicates that compared with other detection networks, the model size of the improved network model was reduced by the most (230.3 MB) compared to YOLOv4s. The number of parameters of the improved network model was the least, which was 13.7 M.

A comprehensive analysis of the above results shows that the improved I-YOLOX network model can not only improve the detection precision but also reduce the number of parameters and model size, thus obtaining a lightweight model. As shown in Figure 12, the improved I-YOLOX network model can be used to identify and detect the six grades of *Gastrodia elata*.

Table 7. The evaluation indexes of different target detection models.

Models	Dataset of Fresh *Gastrodia elata*						
	P (%)	R (%)	F1 (%)	mAP (%)	Parameters (M)	FPS	Model Size (MB)
Fast R-CNN	84.69	98.33	90.70	97.58%	136.791	9	108
SSD-VGG	92.89	96.05	94.20	97.80%	24.280	60	93.1
YOLOv3s	95.12	96.36	95.30	95.91%	61.535	41	235
YOLOv4s	91.91	92.71	91.80	91.74%	63.949	32	244
YOLOv5s	91.66	88.22	88.30	87.88%	7.077	67	27.2
YOLOXs	91.56	91.87	90.20	92.97%	8.940	66	34.3
YOLOv7	90.62	91.77	90.30	93.38%	37.222	28	142
I-YOLOX	93.98	94.69	93.80	97.83%	3.518	69	13.7

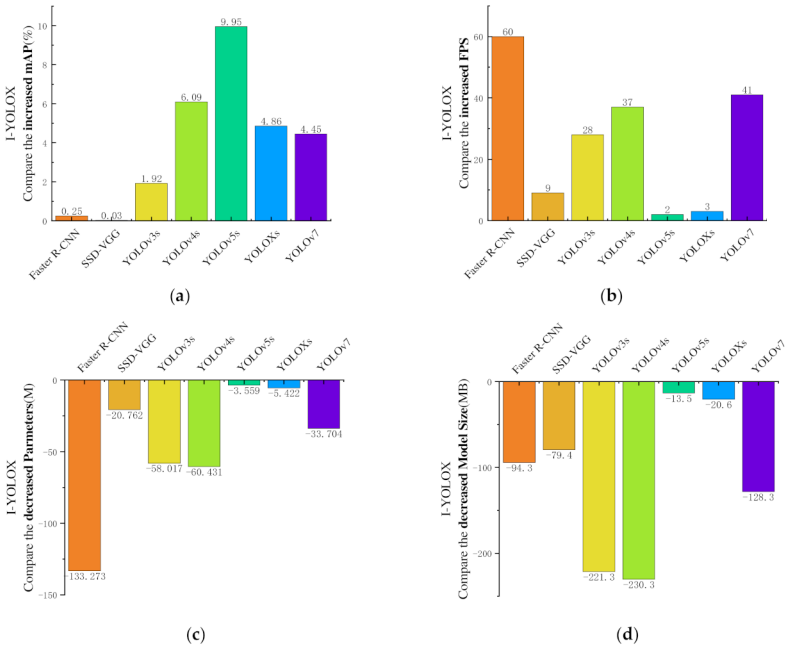


Figure 11. The training effect of the improved target detection algorithm compared with other target detection algorithms. (**a**) mAP, (**b**) FPS, (**c**) Parameters, (**d**) Model size.

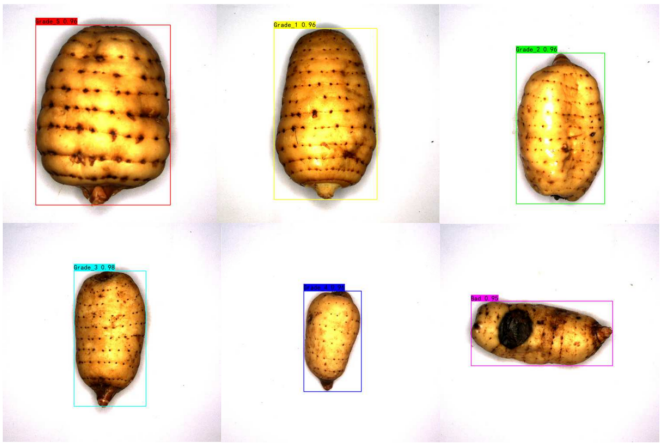


Figure 12. Target detection results of six grades of *Gastrodia elata*.

4. Conclusions

At present, there are few studies on the visual classification of *Gastrodia elata*. To fill this research gap and provide technical support for the classification of *Gastrodia elata*, this paper collected and established the *Gastrodia elata* image dataset and improved the detection accuracy and efficiency of different grades of *Gastrodia elata* by improving the deep learning network model.

In this study, an improved YOLOX network target detection model called I-YOLOX was proposed for the recognition and detection of different grades of *Gastrodia elata*. By analyzing the influence of model improvement and transfer learning on the model performance, the conclusions are as follows.

By adding the ECA attention mechanism module between the backbone network and the neck network to enhance feature extraction of image information and improve model characterization ability, depthwise separable convolution was used to replace normal convolution in the FPN network and the head network to reduce the number of parameters and model size, and a better EIoU loss function was adopted. The model can obtain faster convergence speed and better prediction frame regression accuracy. Meanwhile, the modeling effects of different attention mechanisms, improved positions, improved modules, and different typical target detection networks were explored on the *Gastrodia elata* dataset. The results indicate that compared with the original YOLOX model, the mAP of the improved I-YOLOX model increased by 4.86%, the number of parameters was reduced by 5.422 M, the FPS per increased by 3, and the model size was reduced by 20.6 MB. It shows that the improvement enhanced the detection precision and speed while reducing the calculation amount.

In this study, an improved *Gastrodia elata* grade detection model called I-YOLOX was proposed to identify and detect the six grades of *Gastrodia elata*, which was improved from the aspects of detection accuracy, model complexity, detection speed, etc. This study provides a reference for the deployment and application of the model in the complex environment of *Gastrodia elata* sorting devices in the later stage and extends the use of deep learning models in other fields of Chinese medicinal materials. It promotes the development of the *Gastrodia elata* sorting standards and processing industry and lays a theoretical foundation for the subsequent establishment of automatic *Gastrodia elata* sorting systems.

Author Contributions: Collected data on *Gastrodia elata*, X.D., L.L. and Y.L. (Yuhao Lin); analyzed the data, X.D., F.Z. and S.L.; wrote the paper, X.D. and F.Z.; drew pictures for this paper, X.D., L.L., Y.L. (Yuhao Lin), S.L. and Y.L. (Yuxin Liao); reviewed and edited the paper. X.D., F.Z., L.L., Y.L. (Yuhao Lin), S.L. and Y.L. (Yuxin Liao) All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Project of Cloud Medicine Hometown (Grant No. 202102AA310045).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

Acknowledgments: The authors thank all the reviewers who participated in the review.

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

References

1. Tang, C.; Wang, L.; Liu, X.; Cheng, M.; Qu, Y.; Xiao, H. Comparative pharmacokinetics of gastrodin in rats after intragastric administration of free gastrodin, parishin and *Gastrodia elata* extract. *J. Ethnopharmacol.* **2015**, *176*, 49–54. [CrossRef]
2. Zhong, R.X.; Chen, Y.B.; Duan, X.Y.; Ji, O.; Wu, C.H.; Wen, H.M.; Wei, T.S.; Feng, G.F. Studies on the Correlation of Commodity Grade with Gastrodin in *Gastrodia elata*. *Asia-Pac. Tradit. Med.* **2016**, *12*, 27–30.
3. Zhang, G.; Wang, J.; Yu, X.; Tian, M.; Liu, X.; Tian, Z.; Guo, Y.; Liu, D. Characteristic Analysis on Mineral Elements Contents of Different Grades and Different Regions of Zhaotong *Gastrodia elata*. *Southwest China J. Agric. Sci.* **2016**, *29*, 1392–1397.
4. Saleem, M.H.; Potgieter, J.; Arif, K.M. Automation in Agriculture by Machine and Deep Learning Techniques: A Review of Recent Developments. *Precis. Agric.* **2021**, *22*, 2053–2091. [CrossRef]
5. Dhiman, P.; Kaur, A.; Hamid, Y.; Alabdulkreem, E.; Elmannai, H.; Ababneh, N. Smart Disease Detection System for Citrus Fruits Using Deep Learning with Edge Computing. *Sustainability* **2023**, *15*, 4576. [CrossRef]
6. Qin, J.; Sun, R.; Zhou, K.; Xu, Y.; Lin, B.; Yang, L.; Chen, Z.; Wen, L.; Wu, C. Lidar-Based 3D Obstacle Detection Using Focal Voxel R-CNN for Farmland Environment. *Agronomy* **2023**, *13*, 650. [CrossRef]
7. Egger, J.; Gsaxner, C.; Pepe, A.; Pomykala, K.L.; Jonske, F.; Kurz, M.; Li, J.; Kleesiek, J. Medical deep learning—A systematic meta-review. *Comput. Methods Programs Biomed.* **2022**, *221*, 106874. [CrossRef]
8. Marullo, G.; Tanzi, L.; Ulrich, L.; Porpiglia, F.; Vezzetti, E. A Multi-Task Convolutional Neural Network for Semantic Segmentation and Event Detection in Laparoscopic Surgery. *J. Pers. Med.* **2023**, *13*, 413. [CrossRef]
9. Song, H.; Yin, C.; Li, Z.; Feng, K.; Cao, Y.; Gu, Y.; Sun, H. Identification of Cancer Driver Genes by Integrating Multiomics Data with Graph Neural Networks. *Metabolites* **2023**, *13*, 339. [CrossRef]
10. Gai, R.; Chen, N.; Yuan, H. A detection algorithm for cherry fruits based on the improved YOLO-v4 model. *Neural Comput. Appl.* **2023**, *35*, 13895–13906. [CrossRef]
11. Fu, L.; Yang, Z.; Wu, F.; Zou, X.; Lin, J.; Cao, Y.; Duan, J. YOLO-Banana: A Lightweight Neural Network for Rapid Detection of Banana Bunches and Stalks in the Natural Environment. *Agronomy* **2022**, *12*, 391. [CrossRef]
12. Flores, P.; Zhang, Z.; Igathinathane, C.; Jithin, M.; Naik, D.; Stenger, J.; Ransom, J.; Kiran, R. Distinguishing seedling volunteer corn from soybean through greenhouse color, color-infrared, and fused images using machine and deep learning. *Ind. Crop. Prod.* **2021**, *161*, 113223. [CrossRef]
13. Anh, P.T.Q.; Thuyet, D.Q.; Kobayashi, Y. Image classification of root-trimmed garlic using multi-label and multi-class classification with deep convolutional neural network. *Postharvest Biol. Technol.* **2022**, *190*, 111956. [CrossRef]
14. Zhu, Y.; Zhang, F.; Li, L.; Lin, Y.; Zhang, Z.; Shi, L.; Tao, H.; Qin, T. Research on Classification Model of *Panax notoginseng* Taproots Based on Machine Vision Feature Fusion. *Sensors* **2021**, *21*, 7945. [CrossRef] [PubMed]
15. Zhang, F.; Lin, Y.; Zhu, Y.; Li, L.; Cui, X.; Gao, Y. A Real-Time Sorting Robot System for *Panax notoginseng* Taproots Equipped with an Improved Deeplabv3+ Model. *Agriculture* **2022**, *12*, 1271. [CrossRef]
16. Wang, L.S.; Qin, M.X.; Lei, J.Y.; Wang, X.F.; Tan, K.Z. Blueberry maturity recognition method based on improved YOLOv4-Tiny. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 170–178. [CrossRef]
17. Deng, L.; Li, J.; Han, Z. Online defect detection and automatic grading of carrots using computer vision combined with deep learning methods. *LWT* **2021**, *149*, 111832. [CrossRef]
18. Xu, D.; Zhao, H.; Lawal, O.M.; Lu, X.; Ren, R.; Zhang, S. An Automatic Jujube Fruit Detection and Ripeness Inspection Method in the Natural Environment. *Agronomy* **2023**, *13*, 451. [CrossRef]
19. Wang, F.; Sun, Z.; Chen, Y.; Zheng, H.; Jiang, J. Xiaomila Green Pepper Target Detection Method under Complex Environment Based on Improved YOLOv5s. *Agronomy* **2022**, *12*, 1477. [CrossRef]
20. Li, Z.; Jiang, X.; Shuai, L.; Zhang, B.; Yang, Y.; Mu, J. A Real-Time Detection Algorithm for Sweet Cherry Fruit Maturity Based on YOLOX in the Natural Environment. *Agronomy* **2022**, *12*, 2482. [CrossRef]
21. Mirhaji, H.; Soleymani, M.; Asakereh, A.; Mehdizadeh, S.A. Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions. *Comput. Electron. Agric.* **2021**, *191*, 106533. [CrossRef]
22. Sun, J.; He, X.; Wu, M.; Wu, X.; Shen, J.; Lu, B. Detection of tomato organs based on convolutional neural network under the overlap and occlusion backgrounds. *Mach. Vis. Appl.* **2020**, *31*, 31. [CrossRef]
23. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]

24. Dai, J.; Li, Y.; He, K.; Sun, J. R-FCN: Object Detection via Region-based Fully Convolutional Networks. *arXiv* **2016**, arXiv:1605.06409.
25. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016, Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; pp. 21–37.
26. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 December 2016; pp. 779–788.
27. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430.
28. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11531–11539.
29. Zhao, Y.; Han, R.; Rao, Y. A New Feature Pyramid Network for Object Detection. In Proceedings of the 2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), Jishou, China, 14–15 September 2019; pp. 428–431.
30. Zhang, Y.-F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146–157. [CrossRef]
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.