applied sciences

MDPI

Article

Camellia oleifera Fruit Detection Algorithm in Natural Environment Based on Lightweight Convolutional Neural Network

Zefeng Li [1], Lichun Kang [1,*], Honghui Rao [1,*], Ganggang Nie [1], Yuhan Tan [2] and Muhua Liu [1]

[1] College of Engineering, Jiangxi Agricultural University, Nanchang 330045, China; 18761033538@163.com (Z.L.); nieganggang1997@163.com (G.N.); suikelmh@sohu.com (M.L.)
[2] College of Land Resources and Environment, Jiangxi Agricultural University, Nanchang 330045, China; tyh15380915601@163.com
* Correspondence: jxbblscgg@163.com (L.K.); rhh58@jxau.edu.cn (H.R.)

Abstract: At present, *Camellia oleifera* fruit harvesting relies on manual labor with low efficiency, while mechanized harvesting could result in bud damage because flowering and fruiting are synchronized. As a prerequisite, rapid detection and identification are urgently needed for high accuracy and efficiency with simple models to realize selective and intelligent harvesting. In this paper, a lightweight detection algorithm YOLOv5s-Camellia based on YOLOv5s is proposed. First, the network unit of the lightweight network ShuffleNetv2 was used to reconstruct the backbone network, and thereby the number of computations and parameters of the model was reduced to increase the running speed for saving computational costs. Second, to mitigate the impact of the lightweight improvement on model detection accuracy, three efficient channel attention (ECA) modules were introduced into the backbone network to enhance the network's attention to fruit features, and the Concat operation in the neck network was replaced by the Add operation with fewer parameters, which could increase the amount of information under features while maintaining the same number of channels. Third, the Gaussian Error Linear Units (GELU) activation function was introduced to improve the nonlinear characterization ability of the network. In addition, to improve the ability of the network to locate objects in the natural environment, the penalty index was redefined to optimize the bounding box loss function, which can improve the convergence speed and regression accuracy. Furthermore, the final experimental results showed that this model possesses 98.8% accuracy, 5.5 G FLOPs computation, and 6.3 MB size, and the detection speed reached 60.98 frame/s. Compared with the original algorithm, the calculation amount, size, and parameters were reduced by 65.18%, 56.55%, and 57.59%, respectively. The results can provide a technical reference for the development of a *Camellia oleifera* fruit-harvesting robot.

Keywords: *Camellia oleifera* fruit; YOLOv5s; detection; lightweight

Citation: Li, Z.; Kang, L.; Rao, H.; Nie, G.; Tan, Y.; Liu, M. *Camellia oleifera* Fruit Detection Algorithm in Natural Environment Based on Lightweight Convolutional Neural Network. *Appl. Sci.* **2023**, *13*, 10394. https://doi.org/10.3390/app131810394

Academic Editor: DaeEun Kim

Received: 28 July 2023
Revised: 12 September 2023
Accepted: 15 September 2023
Published: 17 September 2023

1. Introduction

Camellia oleifera fruit is one of the four woody oil-bearing plants in the world and a unique oil-bearing crop in China. Due to the large plantation area of *Camellia oleifera* fruit forests and the complex terrain, the current harvesting relies heavily on manual labor, with high labor intensity and certain risks [1]. Recently the aging population has led to a shortage of rural labor in China; consequently, the cost of *Camellia oleifera* fruit harvesting is gradually increasing. The machine replaces manual harvesting, which has important practical significance and application value in realizing cost reduction and efficiency increase in the Camellia industry. The object detection algorithm is one of the core technologies of harvesting robots; hence, the research of object detection algorithm with lightweight, high efficiency, and strong robustness in the natural environment is the premise of realizing the information and automation of *Camellia oleifera* fruit harvesting.

Deep learning techniques are widely used in intrusion detection, disease diagnosis, object detection, etc. [2–4]. For a long time, we have needed simple and efficient detection techniques to assist intelligent machines in making the right decisions. However, quick and accurate detection of the object in real time to help people in a short time to complete the task is an important issue. Most of the existing research focuses on improving the accuracy of algorithms while some important aspects are ignored—for example, the literature [4] in the object detection field demonstrates the potential of neural networks and symmetry in improving the efficiency of waste management; however, it has not focused on inference speed, nonlinear expressivity, and the ability to localize objects in complex environments, while guaranteeing the accuracy, which is a key factor to focus on when applying object detection algorithms.

With the continuous development of deep learning techniques, the application of deep learning to image recognition scenes has become an important research direction in the field of agriculture and forestry. At present, the research on *Camellia oleifera* fruit detection tends to be more and more mature in China. Tang et al. designed an agricultural robot harvesting system and applied the YOLOv4-tiny algorithm to detect *Camellia oleifera* fruit images collected by a binocular stereo-vision camera. The model achieved an average precision of 92.07% with a weight of 29 MB and an average of 31 ms to detect each fruit image, but the addition of 1×1 convolution as well as 3×3 convolution increased the load while ignoring the computational power of the GPU [5]. Lv et al. proposed an improved YOLON object detection algorithm based on YOLOv3 to detect *Camellia oleifera* fruit at nighttime and improved the detection accuracy by adding a light adjuster (LA) module at the input side and a night prior knowledge (NPK) module at the feature fusion layer. Experiments showed that the precision of the improved model was 94.00%, and the mean average precision at 0.5 (mAP@0.5) was 94.37%, which could meet the requirements of the harvesting robot for detecting *Camellia oleifera* fruit at night [6]. Wang et al. used Mask R CNN algorithm to detect *Camellia oleifera* fruit in natural scenes, and the results showed that the segmentation accuracy of the model was 89.85%, and the mAP@0.5 was 89.42% [7]. Song et al. conducted a study on the detection method of *Camellia oleifera* fruit in the natural environment using YOLOv5s, and the experiments showed that the precision was 90.73%, the F_1 score was 94.4%, the mAP@0.5 was 98.71%, the single image detection time was 12.7 ms, and the model weight was 14.08 MB [8]. Chen et al. used the Faster R CNN algorithm to detect *Camellia oleifera* fruit in the natural environment, and this method achieved a precision of 98.92%, a F_1 score of 96.04%, a mAP@0.5 of 92.39%, and the average detection time per image of 0.21 s [9].

In summary, convolutional neural networks have been widely used in agroforestry, and there has been a certain research foundation for *Camellia oleifera* fruit detected in either daytime or nighttime. Although the above research has made great progress, there is less research on lightweight detection algorithms for *Camellia oleifera* fruit-harvesting robots. Due to the great advantages of lightweight detection algorithms in inference speed, number of parameters, model deployment, and other aspects, more and more researchers prefer to use lightweight algorithm models for fruit and vegetable crop detection tasks [10–13]. In order to solve the existing *Camellia oleifera* fruit detection algorithms with complex structures, large memory requirements, and inference delays, some researchers deal with the above problem by drastically simplifying the network structure at the expense of accuracy and reducing the robustness of the algorithm. This paper conducts further research on lightweight *Camellia oleifera* fruit detection algorithms that have not been studied in depth, and it can provide agricultural robotics research organizations and management with an accurate, fast, and easy-to-use object detection solution with low device memory requirements to better serve agriculture. In this paper, the major work can be briefly summarized as follows.

1. A dataset containing 4750 images of *Camellia oleifera* fruit was created and expanded to 19,000 images by data enhancement means.

2. The original YOLOV5s was improved, including the backbone and the neck network lightweight improvements, activation function optimization to improve the nonlinear expressivity, and loss function optimization to improve the ability to localize objects.
3. The effectiveness of the improved method is verified by ablation experiments, and the overall performance of the improved model is evaluated and compared with mainstream algorithms.

2. Materials and Methods

2.1. Datasets Acquisition

The dataset was collected from the Jiangxi Academy of Forestry on 27 October 2021 and 1 October 2022, respectively. Sunlight in the natural environment has a great influence on the detection. To better adapt the algorithm model to the natural environment, the image acquisition was carried out on a sunny day with a clear atmosphere, sufficient light, and high variability of light conditions. The image capture equipment was an IPHONE13 smartphone with a 36 mm wide-angle lens, a sensitivity of ISO 50, image resolution of 3024 pixels × 4038 pixels, and the distance between the lens and the *Camellia oleifera* fruit during the capture process was approximately 300 mm to 1500 mm. A total of 5750 images of eight species of *Camellia oleifera* fruit were collected, including Changlin 3, Changlin 4, Changlin 18, Ganyong 5, Ganyong190, Gan 195, Gan 447, and Gan 83-4.

2.2. Images Filtrating and Preprocessing

In order to improve the robustness and generalization of the algorithm model during the detection task in the natural environment, the diversity and richness of the dataset were important means to ensure those properties. From the collected 5750 images, 600 images were selected under different environments such as down-light, back-light, single fruit, multiple fruit, dense, branch and leaf shading, etc. In addition, images with poor quality and high repetition were excluded, resulting in a total of 4750 valid images. Considering the small number of images in the dataset, the use of data enhancement methods can increase the variability of the training data, prevent overfitting, reduce the sensitivity of the algorithm model to images, and improve the generalization ability [14,15]. The common geometric transformation methods for dataset enhancement are flipping, cropping, scaling, noise, Gaussian blur, HSV contrast adjustment, brightness adjustment, saturation adjustment, histogram equalization, etc. [16–18]. The samples in the dataset were processed by contrast enhancement, color enhancement, brightness enhancement, and the addition of salt and pepper noise (as shown in Figure 1), while the dataset was enriched by using the Mosaic data enhancement strategy in the training phase. A total of 19,000 extended images were obtained, and the training (13,300 pieces), validation (3800 pieces), and test sets (1900 pieces) were divided 7:2:1.

2.3. Camellia Oleifera Fruit Object Detection Algorithm

YOLO (You Only Look Once) Convolutional Neural Network (CNN) is an end-to-end one-stage object detection network model. Compared with two-stage algorithms, such as Faster R CNN and Mask R CNN [19–21], YOLO has faster inference speed, lighter weight, and higher detection accuracy, making it easier to deploy on mobile devices for detection tasks. Currently, the YOLO has been updated to YOLOv8; the accuracy of YOLOv8 compared with YOLOv5 is much higher, but the model weight and the amount of computation are also improved, and variations in the number of parameters and computation volumes, especially on edge devices, cannot be ignored. Thus, YOLOv5 is more suitable for the object of this paper.

YOLOv5 integrates the advantages of YOLOv1-YOLOv4 algorithms and develops them to meet the practical applications in engineering [22–25]. Moreover, its size is about one-tenth of YOLOv4, while the difference is that the network size is determined by the depth factor and width factor. Taking the computational cost and detection accuracy of mobile devices into account, YOLOv5s, with its fast inference speed, high detection

accuracy, and only 13.7 M model weight, is suitable for this study and is optimized to improve its weight and model structure, so that it can be better applied to the task of detecting *Camellia oleifera* fruit in the natural environment.

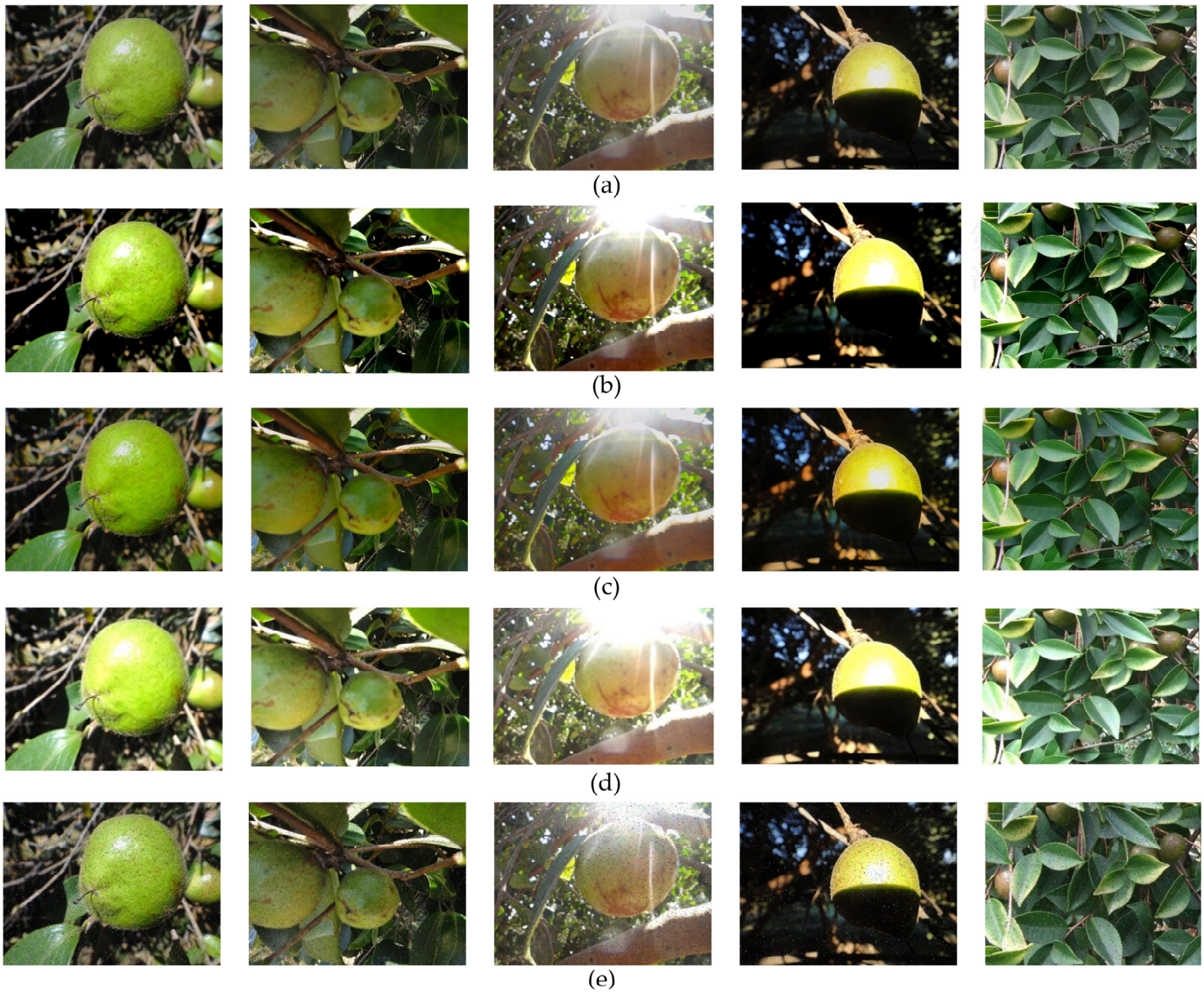


Figure 1. Data augmentation: (**a**) original images; (**b**) contrast enhancement; (**c**) color enhancement; (**d**) brightness enhancement; (**e**) salt and pepper noise.

2.3.1. YOLOv5s-Camellia Detection Model

The improved lightweight YOLOv5s-Camellia algorithm model consists of three main parts: the backbone network (feature extraction network), the neck network (feature fusion network), and the detection head. The flowchart of the model is shown in Figure 2.

In this paper, the Mosaic data enhancement image processing method and the adaptive anchor box generation strategy of the original YOLOv5s were retained, both in the network training stage, for random scaling, random cropping, alignment of different images generated by the stitching to achieve rich object background. During training, the network model output predicted boxes based on the nine initial anchors set for the feature map, calculated the difference between them and the ground truth boxes of the object, and then updated the network parameters by backpropagation to adaptively calculate the best anchor values in different samples. In order to reduce the model parameters and improve the efficiency of the harvesting robot, this paper tries to rebuild the backbone network using Shuffle

Block, the network block of ShuffleNetV2 [26,27], a lightweight classification network. The improved network structure is shown in Figure 3.

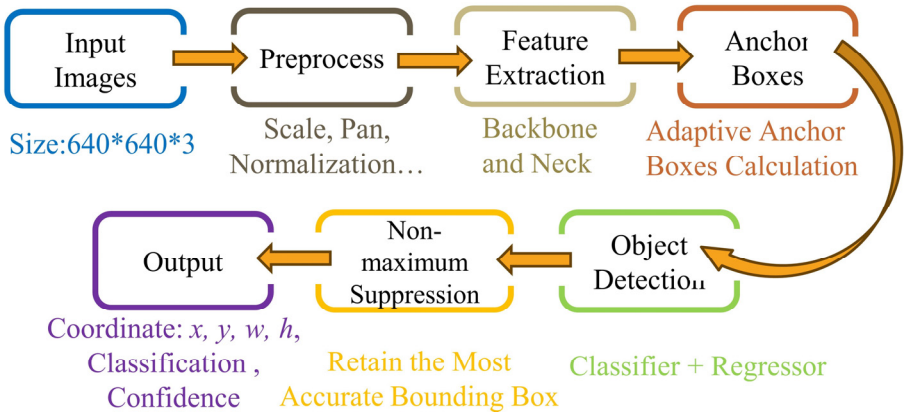
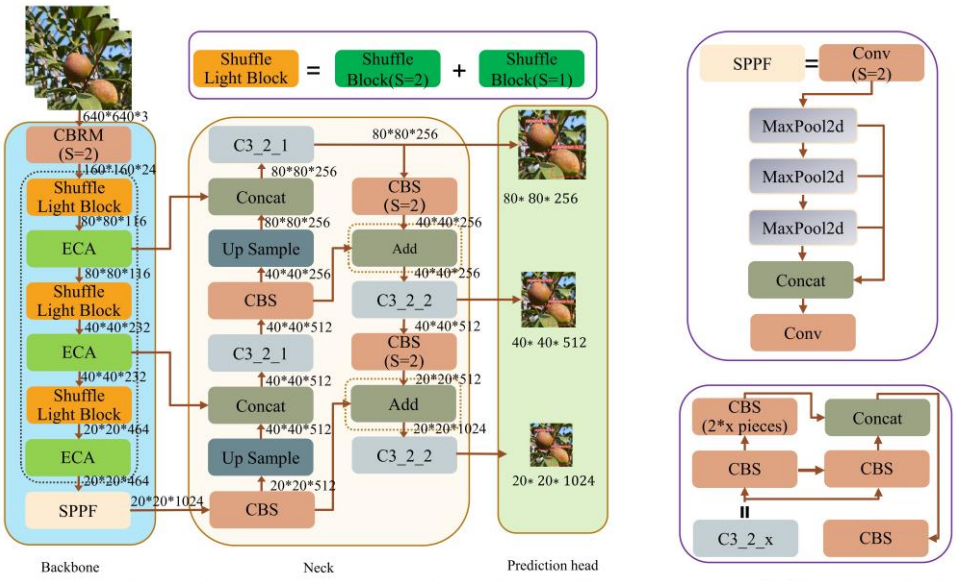


Figure 2. Flowchart of YOLOv5s-Camellia methodology.



Note: The dotted rectangular boxes are improved parts. Conv stands for two-dimensional convolution. CBRM is a convolutional module composed of Conv, BN (batch normalization), ReLU activation function, and pooling layer (Maxpool). Shuffle Light Block represents a lightweight network unit module. ECA is an efficient channel attention module. SPPF is a pooling operation. Concat indicates feature concatenation. CBS is a convolutional module composed of Conv, BN (batch normalization), and SiLU activation functions. Add represents the addition of feature maps, with the number of channels unchanged. Upsample indicates feature upsampling. MaxPool2d represents a maximum pooling layer.

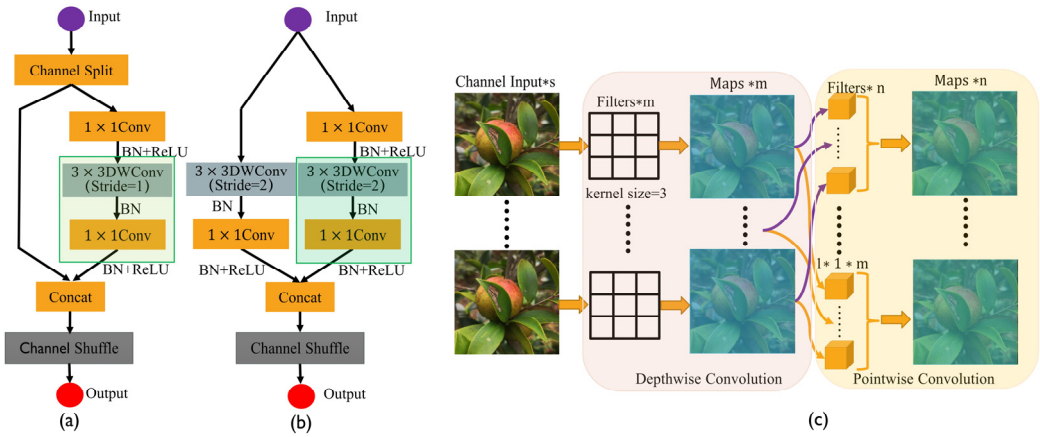
Figure 3. Improved lightweight YOLOv5s-Camellia network structure.

The spatial pyramid pooling fast (SPPF) module is retained at the end of the backbone network to fuse the local and global detail features of the *Camellia oleifera* fruit to enrich the feature map information [28]. The neck network continues the combined structure of the Feature Pyramid Network (FPN) and the Path Aggregation Network (PAN) of the original YOLOv5s [29,30], aggregates and refines the granularity of feature map images in different feature layers, while the introduction of the Add operation in the PAN instead of the Concat operation increases the amount of information under the features, reducing the parameters and keeping the number of feature map channels unchanged. This paper promotes further fitting improvements based on YOLOv5s to optimize model performance.

2.3.2. Lightweight Improvements to the Backbone Network for YOLOv5

On the input side, the image is scaled to 640 pixels × 640 pixels and then fed into the backbone network for feature extraction operation. The original YOLOv5s network

model uses CSP-Darknet53 (Cross Stage Partial Darknet) as the backbone network to extract feature maps and obtains three different 80 × 80 × 256, 40 × 40 × 512, and 20 × 20 × 1024 scales of feature layers. Although the CSP-Darknet53 network is very deep to extract more detailed features, such as the shape, color, and texture of the *Camellia oleifera* fruit, while avoiding gradient disappearance and network degradation, each C3 (concentrated-comprehensive convolution module, one of the components of the backbone network) contains extensive network parameters, which limits the computing power and storage capacity of mobile devices, which cannot meet the real-time detection requirements. Therefore, this paper intends to achieve the lightweight YOLOv5s-Camellia network model by improving the backbone network of the original YOLOv5s, which consisted of CBRM, Shuffle Light Block, ECA, and SPPF. As shown in Figure 4, where the Shuffle Block (the network unit of ShuffleNetV2) with stride 1 is the basic unit and stride 2 is the downsampling unit, both blocks follow the channel mixing operation and the depth separable convolution of the ShuffleNetV1 [31–33].



Note: The green rectangular box is the depthwise separable convolution network, BN means the batch normalization operation, ReLU is the activation function, 1 × 1Conv stands for the two-dimensional convolution, Concat means the dimension addition, 3 × 3DWConv means the depthwise convolution with a convolution kernel of 3.

Figure 4. The structure of Shuffle Block: (**a**) base unit; (**b**) downsampling unit; (**c**) the structure of depthwise separable convolution.

In the basic unit shown in Figure 4a, when the feature maps are input, the channel slicing operation is first performed to divide the network channels into two branches, and the number of channels in each branch is equal to c/2. Branch 1 is passed down directly without any processing, and branch 2 is passed through 1 × 1 Conv, 3 × 3 DWConv (Stride = 1, DWConv is the depthwise convolutional part of the depthwise separable convolution), and 1 × 1 Conv (the 1 × 1 convolution at this point is the pointwise convolution part of the depthwise separable convolution) in order, and then the splicing operation is performed with branch 1. The size of the feature maps remains the same, the number of channels is added to achieve feature fusion, and finally, a channel blending operation is performed to fuse the information between the channels. Since the number of input feature channels in the convolution layer is the same as the number of output feature channels, i.e., there is no difference in the number of convolution kernels, the memory access cost during the convolution operation is reduced. In the downsampling unit of Figure 4b, the channel cutting operation is canceled, so the feature maps are directly input into two branches with stride 2 for height and width dimensionality reduction. Then, the splicing operation is performed after the output, which halves the height and width of the feature maps and doubles the number of channels, increasing the network width and improving the feature extraction capacity of the network without significantly increasing the computation. Finally, channel randomization is performed to strengthen the information fusion between channels.

Depthwise separable convolution differs from an ordinary convolution, as shown in Figure 4c, and consists of two parts: depthwise convolution and pointwise convolution. When the number of input image channels is "s", depthwise convolution first processes the spatial information in the aspect direction where each 3 × 3 two-dimensional convolution kernel is responsible for processing only one channel individually, and the number of channels in the generated feature maps is exactly the same as the number of input channels, with the parameter only 1/s of the ordinary convolution. However, it does not efficiently use the effective information at the same location in different feature layers. When the number of channels of the feature maps generated by depthwise convolution is "m", the number of channels of the final output feature maps is "n". To compensate for the feature loss, "n" sets of 1 × 1 × m convolution kernels are introduced on the feature maps generated after depthwise convolution to perform the pointwise convolution operation. New feature maps are generated by weighted combination, and the essence of depthwise separable convolution is to reduce the parameters and memory requirement of the convolution operation.

2.3.3. Efficient Channel Attention Mechanism

After the backbone network was lightened and improved, the network depth became shallower and the number of convolutional kernels was reduced, which can make the improved network weaker in extracting the features of *Camellia oleifera* fruit, thus affecting the final detection accuracy. In view of this, this paper introduced the efficient channel attention (ECA) mechanism in the backbone network to enhance the channel features and achieved a significant improvement in the model performance with a smaller number of operations and parameters [34]. The ECA module uses a nondegenerate local cross-channel information interaction strategy to improve the detection performance of the model, achieving significant performance gains with a small increase in parameters.

The structure of the ECA block is shown in Figure 5. First, the input feature maps are compressed by global average pooling (GAP) to obtain the aggregation feature of 1 × 1 × C. Then, the 1-dimensional convolutional kernel of size 3 is used to learn the channel features from the aggregated features to achieve local cross-channel interaction and extract important feature information of the *Camellia oleifera* fruit. Finally, the channel-normalized weights obtained with the sigmoid function are multiplied by the original, element by element, to output the feature maps with channel attention.

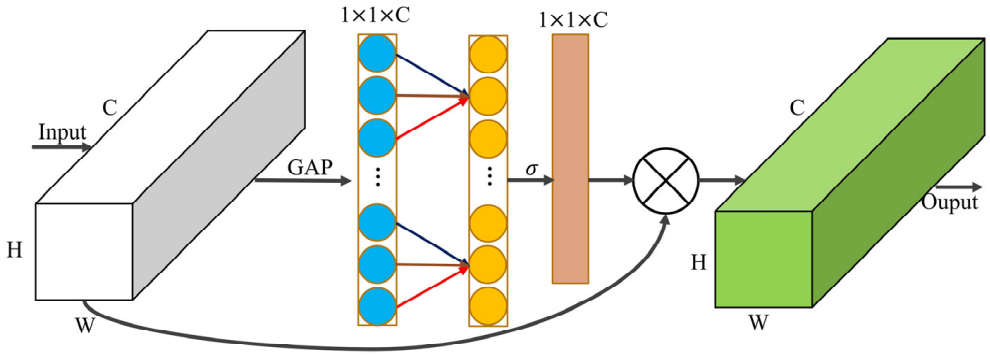


Figure 5. The Structure of efficient channel attention block.

2.3.4. Improved PAN of the Neck Network

YOLOv5s network extracts a total of three feature layers for object detection when the input image size is (640, 640, 3), feature 1: (80, 80, 116), feature 2: (40, 40, 232), and feature 3: (20, 20, 464) are generated, which are located in the top layer, middle layer, and bottom layer of the backbone network, respectively. The neck network uses the FPN and the PAN to build the feature fusion module, which fuses the information from the shallow, middle, and deep layers to facilitate the extraction of better features. The PAN uses a bottom-up path

enhancement strategy, which facilitates the transfer of feature information from the lower layer to the upper layer. The Concat operation in PAN is replaced by the Add operation with a smaller number of parameters, which improves the network's ability to refine the image granularity of feature maps from different layers. Both the Concat and the Add are methods of fusing channel information under the integration of different feature maps by the network model: the former stitching of feature tensor of equal size will expand the two tensor dimensions, both by adding two feature maps of the same size and merging the number of channels; the latter adds the feature tensor with the same dimension, and the dimension remains unchanged, which increases the amount of feature information under the channel while reducing the number of parameters, and the increase of information under each dimension is beneficial to the detection task.

2.3.5. Activation Function Optimization

The activation function is a crucial part of a neural network, responsible for adding nonlinear factors to the output of neurons in the previous layer so that the network model can fit nonlinear functions, thus improving the network model characterization ability. In order to alleviate the accuracy degradation of the backbone network after lightweight improvement and to improve the model generalization ability and detection accuracy, the Gaussian Error Linear Units (GELU) activation function is introduced in this paper [35,36]. This function introduces the idea of stochastic regularity, a stochastic regular transformation method, which is essentially a fusion of Dropout, Zoneout, and ReLU, as shown in Equation (1).

$$GELU(x) = xP(X \leq x) = x\Phi(x) \tag{1}$$

$\Phi(x)$ is the cumulative probability function of the normal distribution, as shown in Equation (2).

$$\Phi(x) = \int_{-\infty}^{x} \frac{e^{-(X-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma} dX \ \ (X \leq x) \tag{2}$$

where X is a random variable and x is a real number, μ is 0 and σ is 1 represent the mean and variance of the normal distribution, respectively; $\Phi(x)$ changes with the current neuron input x, the larger the x value, the more likely the activation output will be maintained, while the smaller the x, the more likely the activation output will be set to 0. The neuron is regularized with the dropout method so that the weight of each node is not too large, mitigating network overfitting. Since the value of Equation (2) cannot be derived directly by calculation, it can be calculated by the approximation formula of the GELU activation function of the standard normal distribution, as shown in Equation (3).

$$GELU(x) = 0.5x\left[1 + \tanh\left(\sqrt{2/\pi}\left(x + 0.044715x^3\right)\right)\right] \tag{3}$$

The smooth activation function has better generalization performance and network optimization ability, thus improving the characterization ability of the network model, as can be seen in Figure 6. The ReLU activation function used in the original YOLOv5s network can only achieve nonlinearity when the input is less than 0, and the gradient is 0, which is prone to neuron necrosis. The GELU activation function has basically linear output when the input $x > 0$ is large; when the input $x < 0$ is small, the output is 0; when the input x tends to 0, the output is nonlinear. From the curve of the derivative function, it can be seen that the GELU function is fully derivable and it is not easy to cause gradient explosion and gradient disappearance, which is good for keeping small negative values and keeping the network gradient flow in a stable state. Based on this, the GELU activation function was introduced into the Shuffle Block in Figure 4a to replace the ReLU in the original module and improve the detection accuracy of the model.

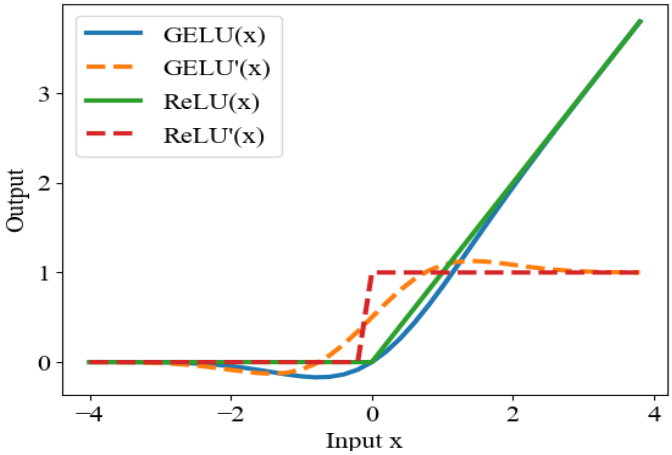


Figure 6. Image of activation function and derivative function.

2.3.6. Loss Function Improvement

Camellia oleifera fruit's precise position in the detection task is a key technology for harvest informatization, which usually uses bounding box regression for localization and a rectangular box to predict the location of the object in the image, and continuously adjusts the position and size of the rectangular predicted box to make it match the ground truth box more accurately. The whole process neural network needs to correct and refine the predicted box position by the loss function. The original YOLOv5s model uses the CIoU (complete intersection over union) loss function, which considers the distance between the center point of the ground truth box and the predicted box, and the diagonal distance of the minimum wrapped box of the two boxes, but it does not consider the angle between the ground truth box and the predicted box, which affects the regression accuracy. It is more complicated to measure the aspect ratio, which reduces the convergence speed, so the SIoU (Scylla intersection over union loss) loss function is introduced to the improved model [37,38], adding the vector angle loss penalty term between the ground truth box and the predicted box. Finally, the loss function specifically includes angle loss, distance loss, shape loss, and intersection over union ratio (IoU) loss. Figure 7 shows a schematic diagram of the SIoU bounding box loss function.

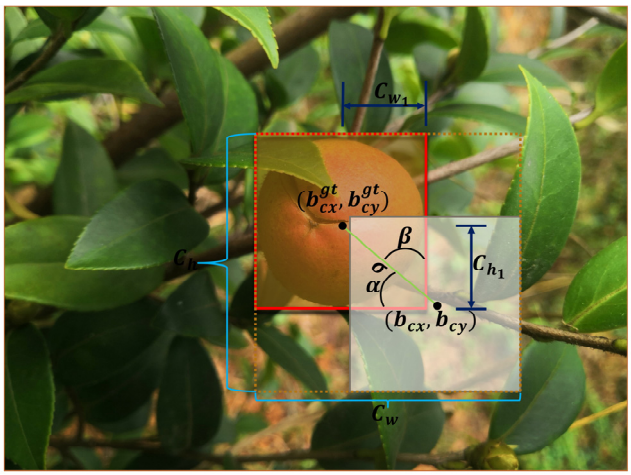


Figure 7. Schematic diagram of SIoU bounding box loss function.

The angular loss ξ is defined in Equation (4).

$$\xi = 1 - 2\sin^2\left[\sin^{-1}\left(\frac{c_{h_1}}{\sigma}\right) - \frac{\pi}{4}\right] \quad (4)$$

$$c_{h_1} = \max\left(b_{c_y}^{gt}, b_{c_y}\right) - \min\left(b_{c_y}^{gt}, b_{c_y}\right) \tag{5}$$

$$\sigma = \sqrt{\left(b_{c_x}^{gt} - b_{c_x}\right)^2 + \left(b_{c_y}^{gt} - b_{c_y}\right)^2} \tag{6}$$

where c_{h_1}, σ are the height difference and distance between the center point of the ground truth box and the predicted box, respectively; b_{c_x},b_{cy} represent coordinates of the center point of the bounding box; and $b_{c_x}^{gt}$,$b_{c_y}^{gt}$ are coordinates of the center point of the ground truth box [39].

The angle loss calculation is applied to the distance loss calculation by redefining the distance loss Δ, as shown in Equation (7).

$$\begin{aligned} \Delta &= \sum_{t=x,y} \left(1 - e^{-\gamma\rho_t}\right) \\ &= 2 - e^{-(2-\xi)\rho_x} - e^{-(2-\xi)\rho_y} \\ &= 2 - e^{-(2-\xi)\left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w}\right)^2} - e^{-(2-\xi)\left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h}\right)^2} \end{aligned} \tag{7}$$

where c_w, c_h are the width and height of the minimum outer rectangle of the ground truth box and the predicted box, and ρ_x, ρ_y are the distance loss.

The shape loss is defined as ϵ in Equation (8).

$$\begin{aligned} \epsilon &= \sum_{t=w,h} \left(1 - e^{-\omega_t}\right)^\theta \\ &= \left(1 - e^{-\omega_w}\right)^\theta + \left(1 - e^{-\omega_h}\right)^\theta \\ &= \left(1 - e^{-\frac{|w - w^{gt}|}{\max(w, w^{gt})}}\right)^\theta + \left(1 - e^{-\frac{|h - h^{gt}|}{\max(h, h^{gt})}}\right)^\theta \end{aligned} \tag{8}$$

where (w, h) and (w^{gt}, h^{gt}) are the width and height of the ground truth box and the predicted box. The weight factor θ controls the degree of attention to shape loss in the network model, to avoid too much attention to shape loss, and to reduce the movement of the predicted box, generally taking the best 4.

The existing loss functions are based on additive implementation, and the principle of $L_i = L_{IOU} + R_i$ is followed in constructing the penalty term, the final penalty term R_i contains the influence factors angle loss ξ, distance loss Δ and shape loss ϵ, and its bounding box regression loss function is calculated by Equation (9).

$$\begin{aligned} L_{box} &= L_{IOU} + R_{box} \\ &= (1 - IoU) + \frac{\Delta + \epsilon}{2} \end{aligned} \tag{9}$$

where L_{IOU} is the loss of intersection over union(IoU), R_{box} is the penalty term, which refers to some restrictions on some parameters in the loss function to reduce the complexity of the network.

The SIoU loss function takes into account not only the overlapping area of the predicted box and the ground truth box, the distance between the two boxes, and the aspect ratio but also the angle between the two boxes, which limits the degree of freedom of the predicted box during the training, improves the regression accuracy of the network as well as the convergence speed, and makes the network better adapted to the natural environment assessment.

3. Experiments and Analyses

3.1. Experimental Platform Construction

The algorithm models involved in this paper were all run on the same platform. The experimental platform is configured as shown in Table 1. Stochastic gradient descent (SGD) was used as the optimizer to speed up the training process, the initial learning rate of the

experiment was set to 0.001, the batch size was 32, the weight recession coefficient was set to 0.001, and the momentum factor was set to 0.92 with 100 iterations.

Table 1. Experimental platform configuration.

Designation	Environment Configuration
Operating System	Windows11
CPU	Intel(R) Core(TM) i5-9400F
GPU	GeForceRTX3070Ti(8GB)
Development Framework	Pytorch1.7.1
Development Environment	Anaconda, Python3.9, CUDA11.3, OpenCV4.5.2

3.2. Evaluation Indicators

In this paper, the following are used in calculations: precision (P), recall (R), mean average precision (mAP), floating point of operations (FLOPs), number of parameters (Params), model size (MB), comprehensive evaluation index (F_1 score), the detection time of a single image as the improved network evaluation index. P, R, F_1, and mAP are calculated as follows:

$$P = \frac{T_P}{T_P + F_P} \times 100\% \tag{10}$$

$$R = \frac{T_P}{T_P + F_N} \times 100\% \tag{11}$$

$$F_1 = \frac{2PR}{P + R} \times 100\% \tag{12}$$

$$mAP = \frac{\sum_{C=1}^{C} AP(C)}{C} \tag{13}$$

where T_P is the positive sample predicted as positive class; F_P is the negative sample predicted as positive class; F_N is the positive sample predicted as negative class.

3.3. Ablation Experiments

In response to the existing problems of the complex network structure of the *Camellia oleifera* fruit object detection algorithm, slow inference speed of mobile terminal, and high cost of memory usage, we conducted experiments and analysis on the improved method based on YOLOv5s. The experimental results are shown in Table 2.

Table 2. Lightweight ablation comparison experiment.

Model	FLOPs/(G)	Parameters	Size/MB	Layers	mAP@0.5/%
YOLOv5s	15.8	7,012,822	14.5	213	98.4
YOLOv5s + ShuffleNet	5.3	2,898,610	6.2	184	96.9
YOLOv5s + ShuffleNet + ECA	5.7	3,137,679	6.7	187	98.1
YOLOv5s + ShuffleNet + ECA + Add	5.5	2,973,839	6.3	187	98.2
YOLOv5s + ShuffleNet+ECA + Add + GELU	5.5	2,973,839	6.3	187	98.5

Yolov5s + ShuffleNet indicates that the backbone network of the original YOLOv5s.

The YOLOv5s network model has been lightened and improved, and the FLOPs, number of parameters, size, and number of network layers have been significantly reduced, according to the ablation comparison test in Table 2. It can be seen that after the backbone network was reconstructed by the unit of ShuffleNetv2, the number of convolutional kernels was reduced due to the lightweight improvement of the network, while the network's ability to extract the detailed features of *Camellia oleifera* fruit was weakened, and the mAP@0.5 was reduced by 1.5%. In order to improve the model accuracy, three ECA channel attention modules were introduced in the backbone network to enhance the channel features of

the input feature map, which increased the mAP@0.5 by 1.2% with an 8.25% increase in computation. The Concat splicing operation in the PAN was changed to the Add operation. Add dimensional fusion increased the amount of information in each dimension of a feature map without expanding the tensor dimension and achieved a 0.1% improvement in mAP@0.5 while reducing the number of model parameters and computational effort. In addition, the GELU activation function was introduced into the Shuffle Block in Figure 3 to optimize the network, which enhanced the nonlinear expressivity of the network. The final detection mAP@0.5 reaches 98.5%, compared with the YOLOv5s model before the improvement. In terms of FLOPs, the number of parameters and model size were reduced by 65.19%, 57.59%, and 56.55%, respectively, and the number of network layers was reduced by 26 layers, and the mAP@0.5 was increased by 0.1 percentage point, which improved the performance of *Camellia oleifera* fruit real-time detection.

In order to verify that the SIoU loss function outperforms the CIoU in the *Camellia oleifera* fruit dataset and can perform the task of real-time detection in the complex natural environment after the improvement of the backbone network, the PAN and the optimization of the activation function, the loss values between the ground truth box and the predicted box of the model training phase and the validation phase were selected as the evaluation index of the convergence performance of the improved model, respectively, and the loss curves of the final version of the improved model after 100 iterations are shown in Figure 8.

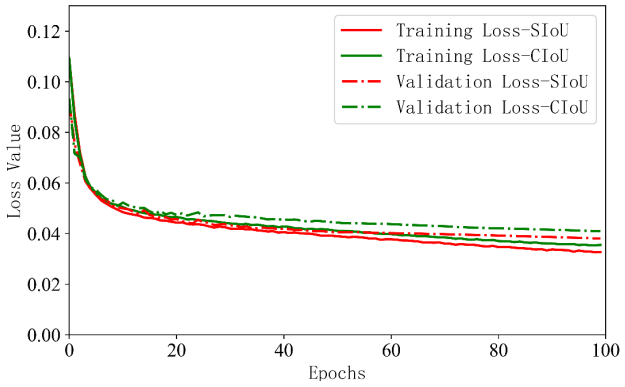


Figure 8. Loss curve of the bounding box.

As shown in Figure 8, the SIoU loss function has better performance in terms of convergence speed and optimized loss value, both in the training and validation phases. It converges faster and the loss value between the ground truth box and the predicted box is smaller because it takes into account the vector angle between the ground truth box and the predicted box. Consequently, it solves the directional matching problem of direction matching between the ground truth box and the predicted box. Moreover, the SIoU loss function optimizes the convergence performance of the network model and improves the detection and regression accuracy of the improved model.

In order to verify the effect of introducing the SIoU loss function on the detection accuracy of the network model, the mean average precision curves of the network model with the SIoU and CIoU loss functions are compared and analyzed after 100 iterations. As shown in Figure 9, the mAP@0.5 values of the network model with the introduction of the SIoU loss function represented by the gray curve are higher than those in the network model with the introduction of the CIoU loss function represented by the brown color from the beginning to the end of training. The comparison shows that the introduction of the SIoU loss function has improved the regression accuracy of the model in the prediction process.

3.4. Analysis of Improved Model Results

The core of lightweight network design is to reduce the computational and spatial complexity of the model while ensuring accuracy as much as possible and to improve the

detection speed of the model. In order to verify the effectiveness of the model improvement, this paper further analyzed the performance of precision, recall, mAP, and F_1 score under three different influencing factors, such as the backbone improvement, loss function improvement, and PAN improvement, and the results are shown in Table 3.

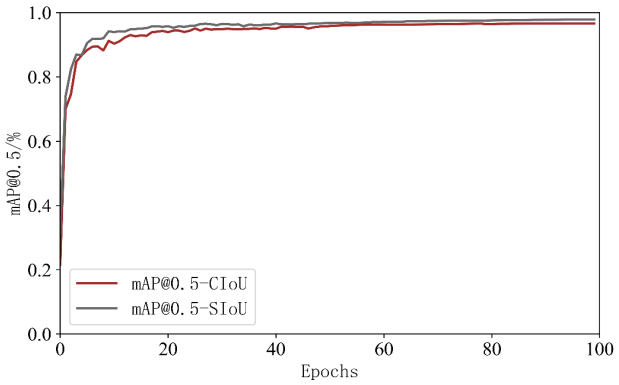


Figure 9. mAP@0.5 curve.

Table 3. Analysis of model effectiveness.

Model	Improvement Factors			Evaluation Indicators			
	Backbone Improvement	SIoU	PAN Improvement	P/%	R/%	mAP@0.5/%	F_1 Score
YOLOv5s	×	×	×	98.2	97.7	98.4	97.94
YOLOv5s-B	√	×	×	96.3	95.4	98.1	95.84
YOLOv5s-B-P	√	×	√	97.2	95.1	98.5	96.14
YOLOv5s-Camellia	√	√	√	98.6	97.4	98.8	97.99

Backbone improvement includes the introduction of the unit of ShuffleNetV2, ECA and GELU. "×" means not to use the improvement factor, "√" means to use the improvement factor.

From Table 3, we can see that the final improved model YOLOv5s-Camellia in this paper has the highest precision, mAP@0.5, and F_1 scores of 98.6%, 98.8%, and 97.99%, respectively, which are 0.4%, 0.3% and 0.05% higher than the original YOLOv5s, verifying that the improvement measures in this paper have positive impact. However, the recall of the model has decreased by 0.3%, which is not much different from the original YOLOv5s network. The reason is that after the improvement of the lightweight of the backbone network, the depth of the network becomes shallow, which leads to the reduction of the model's ability to distinguish between positive and negative samples. In conclusion, the final improved model of this paper has better overall performance.

3.5. Performance Comparison of Different Object Detection Algorithms

In order to verify the superiority of the improved network relative to other object detection algorithms, this paper analyzed several current mainstream algorithms, including the two-stage object detection algorithm Faster RCNN, YOLOv5s, YOLOv4-tiny, and the YOLOv5s-EfficientNet network, run under the experimental environment configuration in Table 1, and the results are shown in Table 4.

As can be seen from Table 4, the two-stage object detection algorithm Faster RCNN is slow in inference because it has two stages of candidate region generation and feature extraction. In addition, the fully connected network used by this model occupies a large number of parameters and does not meet the real-time detection requirements. Compared with the other YOLO series algorithms in Table 4, YOLO v4-tiny has the largest model size, which seriously affects the computation, the YOLOv5s-EfficientNet algorithm, which is based on the lightweight improvement of YOLOv5s, does not differ much in the model

volume and detection accuracy performance indexes, but its overall performance is still inferior to the final improved algorithm YOLOv5s-Camellia in this paper. The final improved model in this paper improves the mAP@0.5 value by 0.4 percentage points compared with the original YOLOv5s network, increases the detection speed by 57.74%, reduces the average single image detection time and model volume by more than 50%, meets the real-time detection requirements, and saves the computational cost.

Figure 10 compares the actual detection performance of different object detection algorithms in Table 4. From the detection effect graph, it can be seen that the YOLOv5s-Camellia model proposed in this paper has high regression accuracy, the best detection effect for fruit overlap and small targets in the distant view, and the edge detection ability for *Camellia oleifera* fruit is stronger than other target detection algorithms. In summary, the detection algorithm proposed in this paper has stronger robustness in the natural environment.

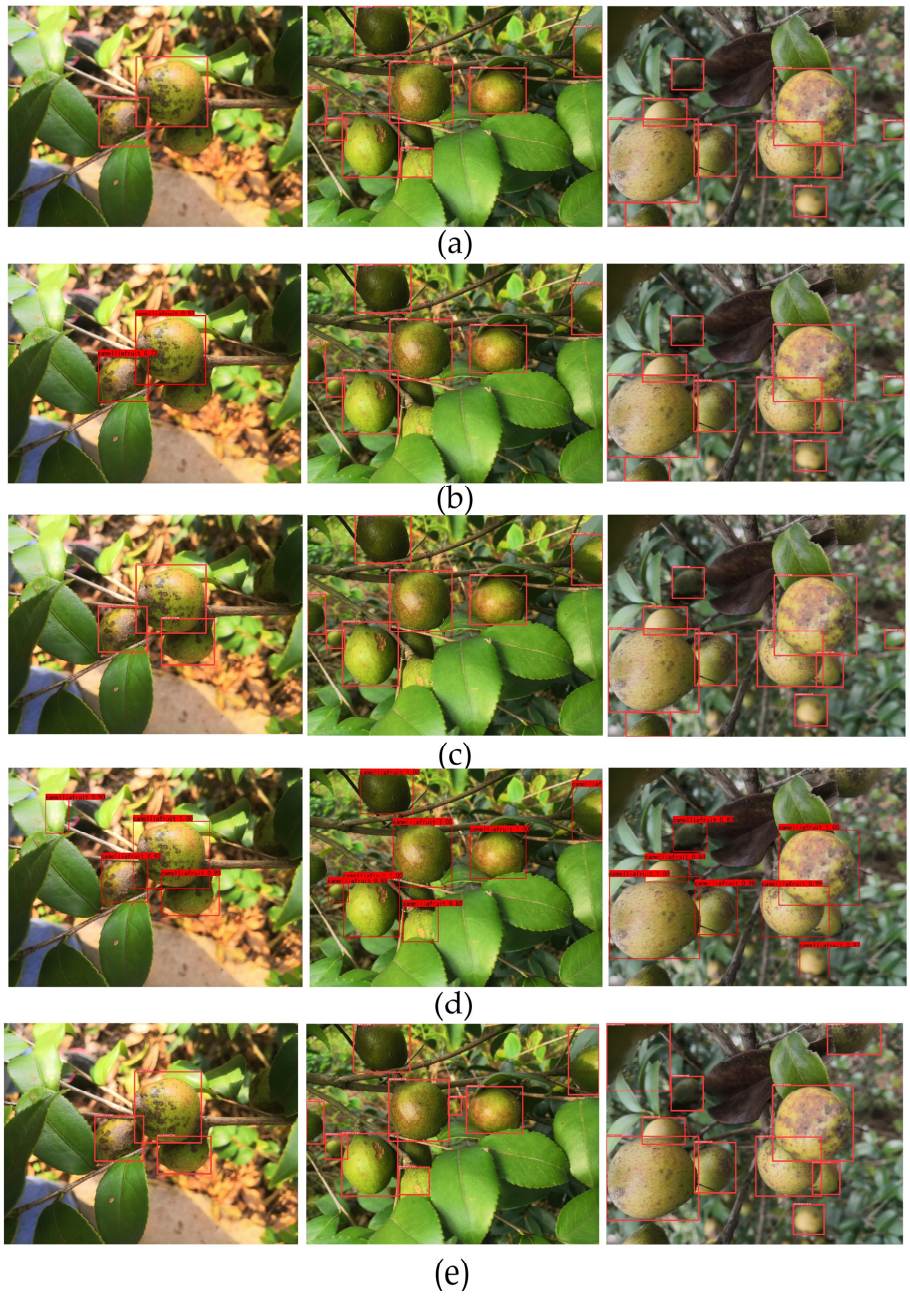


Figure 10. Comparison of recognition results of four target detection algorithms: (**a**) YOLOv5s, (**b**) YOLO v4-tiny, (**c**) YOLOv5s-EfficientNet, (**d**) Faster R CNN, and (**e**) YOLOv5s-Camellia.

Table 4. Performance comparison of different target detection algorithms.

Model	Evaluation Indicators			
	mAP@0.5/%	Average Single Image Detection Time/s	Speed/ (Fame/s)	Size/MB
YOLO v5s	98.4	0.035	25.77	14.5
YOLO v4-tiny	89.9	0.025	40.45	23.1
YOLO v5s-EfficientNet	98.3	0.027	33.56	6.3
Faster R CNN	94.3	3.6	0.03	108
YOLOv5s-Camellia	98.8	0.014	60.98	6.3

Yolov5s-EfficientNet indicates that the backbone network of the original YOLOv5s model is reconstructed of EfficientNet V2.

4. Conclusions

In this paper, a lightweight detection algorithm YOLOv5s-Camellia was proposed to meet the requirements of real-time detection, in view of the problems of low accuracy, low detection efficiency, and complex model of *Camellia oleifera* fruit detection algorithm in the natural environment. The main conclusions are as follows:

1. The unit of the ShuffleNetV2 was introduced as the basic unit of the backbone network, which significantly reduced the number of parameters, computation, and size of the model while saving computational resources and cache space.
2. After the model was lightened, the feature extraction ability of *Camellia oleifera* fruit details was weakened, and the detection performance was improved by embedding three efficient channel attention modules in the backbone network while increasing the number of partial parameters.
3. To enhance the neck network's ability and refine the granularity of feature maps, the Concat dimensional stitching in the PAN was replaced with Add dimensional fusion, which increased the amount of information under each dimension while reducing the number of parameters and maintaining the dimension of the feature map tensor.
4. The better nonlinearity of the GELU activation function was used to optimize the model, which improved the characterization ability of the deep neural network. Compared with the ReLU activation function, the nonzero gradient is better able to maintain a smaller negative value, avoiding the problems of gradient disappearance and gradient explosion.
5. By introducing the SIoU loss function, the vector angle loss between the ground truth box and the predicted box was added to the bounding box regression loss, which reduced the model error and improved the convergence speed and bounding box regression accuracy. The final average detection accuracy of the model reached 98.8% and the detection speed was 60.98 frame/s. Compared with other object detection algorithms, the comprehensive performance of the YOLOv5s-Camellia was better and can meet the real-time detection requirements.

5. Discussion

In this paper, we intend to make improvements in three aspects as part of future research: At first, the dataset based on the complex background of *Camellia oleifera* fruit plantations needs to be continuously expanded to include more images taken under different conditions, especially images of small-object *Camellia oleifera* fruit, and the complexity of the context needs to be increased to improve the model detection accuracy. Second, to study the expanded dataset, in future research, our study will incorporate multiple image enhancement methods, thus increasing the diversity of image types within the dataset. Then, the network model needs to be further optimized, so advanced optimization algorithms are introduced to improve the detection accuracy and recall to make the improved model more suitable for the detection requirements of specific scenarios. Finally, the improved detection model should be deployed in the *Camellia oleifera* fruit harvesting robot to

achieve autonomous detection and harvesting, thus demonstrating the practical value of this research.

Author Contributions: Conceptualization, Z.L. and H.R.; methodology, Z.L., H.R. and L.K.; software, G.N.; validation, Z.L., H.R. and L.K.; investigation, Z.L. and H.R.; data curation, Z.L.; writing—original draft preparation, Z.L.; writing—review and editing, H.R., L.K. and Y.T.; visualization, Z.L.; supervision, H.R. and M.L.; project administration, H.R.; funding acquisition, H.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number 52065207); the National Key Research and Development Program of China (grant number 2022YFD2202104); and Jiangxi Provincial Forestry Bureau *Camellia oleifera* Fruit Research Special Project (YCYJZX2023221).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

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

Acknowledgments: Thanks to Jiangxi Academy of Forestry for supporting our work!

Conflicts of Interest: The authors declare no conflict of interest. Accurate detection of key body parts of cattle is of great significance to Precision Livestock.

References

1. Rao, H.H.; Wang, Y.L.; Li, Q.S.; Wang, B.Y.; Yang, J.L.; Liu, M.H. Design and Experiment of *Camellia oleifera* Fruit Layered Harvesting Device. *Trans. Chin. Soc. Agric Mac.* **2021**, *52*, 203–212. Available online: http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20211021&journal_id=jcsam (accessed on 5 October 2022.).
2. Sapra, L.; Sandhu, J.K.; Goyal, N. Intelligent Method for Detection of Coronary Artery Disease with Ensemble Approach. *J. Adv. Commun. Comput. Technol.* **2021**, *668*, 11–18. Available online: https://link.springer.com/article/10.1007/s40747-017-0048-6 (accessed on 7 September 2023).
3. Rana, P.; Batra, I.; Malik, A.; Imoize, A.L.; Kim, Y.; Pani, S.K.; Goyal, N.; Kumar, A.; Rho, S. Intrusion Detection Systems in Cloud Computing Paradigm: Analysis and Overview. *Complexity* **2022**, *2022*, 3999039. [CrossRef]
4. Verma, V.; Gupta, D.; Gupta, S.; Uppal, M.; Anand, D.; Ortega-Mansilla, A.; Alharithi, F.S.; Almotiri, J.; Goyal, N. A Deep Learning-Based Intelligent Garbage Detection System Using an Unmanned Aerial Vehicle. *Symmetry.* **2022**, *14*, 960. [CrossRef]
5. Tang, Y.C.; Zhou, H.; Wang, H.J.; Zhang, Y.Q. Fruit detection and positioning technology for a *Camellia oleifera* C. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision. *J. Expert Syst. Appl.* **2023**, *211*, 118573. [CrossRef]
6. Lv, S.C.; Ma, B.L.; Song, L.; Wang, Y.N.; Duan, Y.C.; Song, H.B. Nighttime detection method of polymorphic *Camellia oleifera* fruits based on YOLON network. *J. N. W. A&F Univ. (Nat. Sci. Ed.)* **2023**, *51*, 1–14. Available online: https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C45S0n9fL2suRadTyEVl2pW9UrhTDCdPD66SiUpAuf8Bk08bs0aFBNWmLyxqJGMvQZI-DYFmge5uZAFqqnoOxDZN&uniplatform=NZKPT (accessed on 22 February 2023).
7. WANG, L.; Hou, Y.F.; He, J. Target Recognition and Detection of *Camellia oleifera* Fruit in Natural Scene Based on Mask-RCNN. *J. Chin. Agric. Mech.* **2022**, *43*, 148–154. Available online: https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C44YLTlOAiTRKibYlV5Vjs7iJTKGjg9uTdeTsOI_ra5_XWoeTkBDDiZyEr12KcpOhrgb9KsyIs7KlfowkNwkNRx5&uniplatform=NZKPT (accessed on 24 October 2022).
8. Song, H.B.; Wang, Y.N.; Wang, Y.F.; Lv, S.C.; Jiang, M. *Camellia oleifera* Fruit Detection in Natural Scene Based on YOLO v5s. *Trans. Chin Soc Agric Mach.* **2022**, *53*, 234–242. Available online: http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20220724&journal_id=jcsam (accessed on 24 October 2022).
9. Chen, B.; Wang, Y.L.; Li, Q.S.; Wang, B.Y.; Yang, J.L.; Liu, M.H. Study on Detection of *Camellia oleifera* Fruit in Natural Environment Based on Faster R CNN. *Acta. Agric. JX* **2021**, *33*, 67–70. Available online: https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C44YLTlOAiTRKibYlV5Vjs7iy_Rpms2pqwbFRRUtoUImHa_XfV1B1hbSSmFX7alT4YCqO1eVRcLO9JnimIKxHru-&uniplatform=NZKPT (accessed on 25 October 2022).
10. Hu, G.R.; Zhou, J.G.; Chen, C.; Li, C.; Sun, L.; Chen, Y.; Zhang, S.; Chen, J. Fusion of The Lightweight Network and Visual Attention Mechanism to Detect Apples in Orchard Environment. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 131–142. Available online: http://www.tcsae.org/nygcxb/article/abstract/20221915?st=search (accessed on 20 December 2022).
11. Wang, Z.; Wang, J.; Wang, X.X.; Shi, J.; Bai, X.P.; Zhou, Y.J. Lightweight Real-time Apple Detection Method Based on Improved YOLOv4. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 294–302. Available online: http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20220831&journal_id=jcsam (accessed on 27 December 2022).

12. Wang, Y.T.; Xue, J.R. Lightweight Object Detection Method for Lingwu Long Jujube Images Based on Improved SSD. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 173–182. Available online: http://www.tcsae.org/nygcxb/article/abstract/20211920?st=search (accessed on 17 December 2022).
13. Zhang, X.M.; Zhu, D.L.; Yu, M.S. Lightweight Detection Model of Maize Tassel in UAV Remote Sensing Image. *Acta. Agric. Univ. Jiangxiensis* **2022**, *44*, 461–472. Available online: https://www.sciengine.com/JXNYDXXB/doi/10.13836/j.jjau.2022048 (accessed on 16 April 2023).
14. Peng, H.X.; Xu, H.M.; Liu, H.N. Lightweight Agricultural Crops Pest Identification Model Using Improved ShuffleNetV2. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 161–170. Available online: http://www.tcsae.org/nygcxb/article/abstract/20221118?st=search (accessed on 7 September 2023).
15. Li, Z.T.; Sun, J.B.; Yang, K.W.; Xiong, D.H. A Review of Adversarial Robustness Evaluation for Image Classification. *J. Comp. Res. Dev.* **2022**, *59*, 2164–2189. Available online: https://crad.ict.ac.cn/cn/article/doi/10.7544/issn1000-1239.20220507 (accessed on 8 September 2023).
16. Nanni, L.; Paci, M.; Brahnam, S.; Lumini, A. Comparison of Different Image Data Augmentation Approaches. *J. Imaging* **2021**, *7*, 254. [CrossRef]
17. Khalifa, N.E.; Loey, M.; Mirjalili, S. A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artif. Intell. Rev.* **2022**, *55*, 2351–2377. [CrossRef]
18. Guo, Y.K.; Zhu, Y.C.; Liu, L.P.; Huang, Q. Research Review of Space-Frequency Domain Image Enhancement Methods. *Comput. Eng. Appl.* **2022**, *58*, 23–32. Available online: http://cea.ceaj.org/CN/Y2022/V58/I11/23 (accessed on 7 September 2023).
19. Zhou, T.; Jiang, Y.; Wang, X.; Xie, J.; Wang, C.; Shi, Q.; Zhang, Y. Detection of Residual Film on the Field Surface Based on Faster R-CNN Multiscale Feature Fusion. *Agriculture* **2023**, *13*, 1158. [CrossRef]
20. Zhang, X.; Cui, J.; Liu, H.; Han, Y.; Ai, H.; Dong, C.; Zhang, J.; Chu, Y. Weed Identification in Soybean Seedling Stage Based on Optimized Faster R-CNN Algorithm. *Agriculture* **2023**, *13*, 175. [CrossRef]
21. Liu, Y.; Yang, G.; Huang, Y.; Yin, Y. SE-Mask R-CNN: An Improved Mask R-CNN for Apple Detection and Segmentation. *J. Intell. Fuzzy. Syst.* **2021**, *41*, 6715–6725. [CrossRef]
22. Shao, D.; He, Z.; Fan, H.; Sun, K. Detection of Cattle Key Parts Based on the Improved Yolov5 Algorithm. *Agriculture* **2023**, *13*, 1110. [CrossRef]
23. Yao, J.; Qi, J.; Zhang, J.; Shao, H.; Yang, J.; Li, X. A Real-Time Detection Algorithm for Kiwifruit Defects Based on YOLOv5. *Electronics* **2021**, *10*, 1711. [CrossRef]
24. Chen, Z.; Wu, R.; Lin, Y.; Li, C.; Chen, S.; Yuan, Z.; Chen, S.; Zou, X. Plant Disease Recognition Model Based on Improved YOLOv5. *Agronomy* **2022**, *12*, 365. [CrossRef]
25. Hong, W.; Ma, Z.; Ye, B.; Yu, G.; Tang, T.; Zheng, M. Detection of Green Asparagus in Complex Environments Based on the Improved YOLOv5 Algorithm. *Sensors* **2023**, *23*, 1562. [CrossRef]
26. Yang, R.; Lu, X.; Huang, J.; Zhou, J.; Jiao, J.; Liu, Y.; Liu, F.; Su, B.; Gu, P. A Multi-Source Data Fusion Decision-Making Method for Disease and Pest Detection of Grape Foliage Based on ShuffleNet V2. *Remote Sens.* **2021**, *13*, 5102. [CrossRef]
27. Zhou, Y.; Fu, C.; Zhai, Y.; Li, J.; Jin, Z.; Xu, Y. Identification of Rice Leaf Disease Using Improved ShuffleNet V2. *Comput. Mater. Contin.* **2023**, *75*, 4501–4517. [CrossRef]
28. Li, J.; Chen, L.; Shen, J.; Xiao, X.; Liu, X.; Sun, X.; Wang, X.; Li, D. Improved Neural Network with Spatial Pyramid Pooling and Online Datasets Preprocessing for Underwater Target Detection Based on Side Scan Sonar Imagery. *Remote Sens.* **2023**, *15*, 440. [CrossRef]
29. Xie, J.; Pang, Y.; Nie, J.; Cao, J.; Han, J. Latent Feature Pyramid Network for Object Detection. *IEEE Trans. Multimedia* **2022**, *25*, 2153–2163. Available online: https://ieeexplore.ieee.org/document/9684715 (accessed on 2 April 2023). [CrossRef]
30. Yu, H.; Li, X.; Feng, Y.; Han, S. Multiple attentional path aggregation network for marine object detection. *Appl. Intell.* **2023**, *53*, 2434–2451. [CrossRef]
31. Wang, Y.; Liang, Q. Fast 3D-CNN Combined with Depth Separable Convolution for Hyperspectral Image Classification. *J. Front. Comp. Sci. Technol.* **2022**, *16*, 2860–2869. Available online: http://fcst.ceaj.org/CN/10.3778/j.issn.1673-9418.2103051 (accessed on 4 March 2023).
32. Chu, K.; Wang, L.; Ma, D.; Zhang, Z.N. Research on Application of Depthwise Separable Convolution in Android Malware Classification. *Appl. Res. Comp.* **2022**, *39*, 1534–1540. Available online: https://www.arocmag.com/article/01-2022-05-043.html (accessed on 10 September 2023).
33. Jang, J.-G.; Quan, C.; Lee, H.D.; Kang, U. Falcon: Lightweight and accurate convolution based on depthwise separable convolution. *Knowl. Inf. Syst.* **2023**, *65*, 2225–2249. [CrossRef]
34. Song, H.B.; Ma, B.L.; Shang, Y.Y.; Wen, Y.C.; Zhang, S.J. Detection of Young Apple Fruits Based on YOLOv7-ECA Model. *Trans. Chin. Soc. Agric Mac.* **2023**, *54*, 233–242. Available online: http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20230624&journal_id=jcsam (accessed on 11 September 2023).
35. Sarkar, A.K.; Tan, Z.-H. On Training Targets and Activation Functions for Deep Representation Learning in Text-Dependent Speaker Verification. *Acoustics* **2023**, *5*, 693–713. [CrossRef]
36. Guo, W.L.; Liu, F.H.; Wu, W.Y.; Li, C.; Xiao, P.; Liu, C. Wood Surface Defect Recognition Based on ViT Convolutional Neural Network. *Comp. Sci.* **2022**, *49*, 609–614. Available online: https://www.jsjkx.com/CN/10.11896/jsjkx.211100090 (accessed on 15 February 2023).

37. Gevorgyan, Z. SIoU Loss: More Powerful Learning for Bounding Box Regression. *arXiv* **2022**, arXiv:2205.12740.
38. Guo, Y.; Chen, S.; Zhan, R.; Wang, W.; Zhang, J. LMSD-YOLO: A Lightweight YOLO Algorithm for Multi-Scale SAR Ship Detection. *Remote Sens.* **2022**, *14*, 4801. [CrossRef]
39. Long, Y.; Yang, Z.; He, M. Recognizing apple targets before thinning using improved YOLOv7. *Trans. Chin. Soc. Agric. Eng. (Trans. CSAE)* **2023**, *39*, 191–199. [CrossRef]