

## Article

# Detection of Gannan Navel Orange Ripeness in Natural Environment Based on YOLOv5-NMM

Binbin Zhou, Kaijun Wu and Ming Chen \* 

College of Information Technology, Shanghai Ocean University, Shanghai 201306, China; m220951671@st.shou.edu.cn (B.Z.); walt@shou.edu.cn (K.W.)

\* Correspondence: mchen@shou.edu.cn

**Abstract:** In order to achieve fast and accurate detection of Gannan navel orange fruits with different ripeness levels in a natural environment under all-weather scenarios and then to realise automated harvesting of Gannan navel oranges, this paper proposes a YOLOv5-NMM (YOLOv5 with Navel orange Measure Model) object detection model based on the improvement in the original YOLOv5 model. Based on the changes in the phenotypic characteristics of navel oranges and the Chinese national standard GB/T 21488-2008, the maturity of Gannan navel oranges is tested. And it addresses and improves the problems of occlusion, dense distribution, small target size, rainy days, and light changes in the detection of navel orange fruits. Firstly, a new detection head of  $160 \times 160$  feature maps is constructed in the detection layer to improve the multi-scale target detection layer of YOLOv5 and to increase the detection accuracy of the different maturity levels of Gannan navel oranges of small sizes. Secondly, a convolutional block attention module is incorporated in its backbone layer to capture the correlations between features in different dimensions to improve the perceptual ability of the model. Then, the weighted bidirectional feature pyramid network structure is integrated into the Neck layer to improve the fusion efficiency of the network on the feature maps and reduce the amount of computation. Lastly, in order to reduce the loss of the target of the Gannan Navel Orange due to occlusion and overlapping, the detection frame is used to remove redundancy using the Soft-NMS algorithm to remove redundant candidate frames. The results show that the accuracy rate, recall rate, and average accuracy of the improved YOLOv5-NMM model are 93.2%, 89.6%, and 94.2%, respectively, and the number of parameters is only 7.2 M. Compared with the mainstream network models, such as Faster R-CNN, YOLOv3, the original model of YOLOv5, and YOLOv7-tiny, it is superior in terms of the accuracy rate, recall rate, and average accuracy mean, and also performs well in terms of the detection rate and memory occupation. This study shows that the YOLOv5-NMM model can effectively identify and detect the ripeness of Gannan navel oranges in natural environments, which provides an effective exploration of the automated harvesting of Gannan navel orange fruits.



**Citation:** Zhou, B.; Wu, K.; Chen, M. Detection of Gannan Navel Orange Ripeness in Natural Environment Based on YOLOv5-NMM. *Agronomy* **2024**, *14*, 910. <https://doi.org/10.3390/agronomy14050910>

Academic Editors: Valerio Cristofori and Yanbo Huang

Received: 19 March 2024

Revised: 9 April 2024

Accepted: 24 April 2024

Published: 26 April 2024

**Keywords:** Gannan navel orange; ripeness; YOLOv5; weighted bidirectional feature pyramid network; attention mechanism; Soft-NMS



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The Gannan navel orange is a world-famous high-quality fruit with thin and smooth skin, juicy flesh, sweet taste, and rich nutrition, which is known as the “Renowned fruit of China”. Ganzhou City, Jiangxi Province, where it originates, has become the world’s first navel orange planting area, the world’s third-largest annual output, and the largest navel orange producing area in China. At present, Gannan navel oranges are usually picked manually, which greatly reduces the picking efficiency and increases the labour cost. With the development of orchard industrialisation and scale, the continuous expansion of navel orange planting scale and production and the realisation of large-scale mechanised picking is an urgent problem to be solved. Therefore, there is an urgent need to construct and

improve a target detection system with high accuracy to identify whether Gannan navel oranges are ripe or not and combine it with agricultural robots to achieve mechanised harvesting of Gannan navel oranges.

With the continuous development and optimisation of computer vision and target detection-related algorithms, the accuracy of fruit detection has been greatly improved [1]. Since the 1970s, some scholars have applied traditional computer vision techniques to the non-destructive detection of fruits and have achieved certain results [2]. Liu et al. [3] used a histogram of oriented gradients (HOGs) descriptors to train support vector machine (SVM) classifiers to reduce the effect of different light levels on tomato recognition. Li et al. [4] accomplished fruit recognition in green tomatoes when the background colour of the leaf stem is similar, using fusion Fast Normalized Cross Correlation (FNCC) and Hough Transform detection. Kurtulus et al. [5] detected immature citrus by using colour features with the Gabor filtering process. However, most of these methods are based on the combination of threshold segmentation [6], colour space transformation [7], and chemical detection [8] for fruit detection, which results in a complex decision-making process and low recognition efficiency, leading to poor robustness and generalisation in natural environments.

In recent years, deep learning has been a research hotspot in the field of agricultural information technology, and its advantages, such as high recognition accuracy and strong generalisation ability, have been widely reflected in fruit ripeness recognition research. Ashtiani et al. [9] proposed a method of fine-tuning the CNN model by using migration learning to detect the ripeness of mulberries, which achieved good results, but the model consumed too much time and was inefficient in detection. Appe et al. [10] used a DCNN model based on VGG16 to detect the ripeness of tomatoes. However, this experiment lacked image detection in complex backgrounds, and the model's generalisation effect on datasets with different backgrounds and complexities needs to be improved. The You Only Look Once (YOLO) algorithm [11], which is a single-stage target detection model, has gained prominence in the field of target detection since its introduction and has the characteristics of high accuracy and speed.

Several researchers and scholars have utilized the YOLO-based model for fruit detection. Fu et al. [12] developed a fast and accurate kiwifruit detection method based on YOLOv3-tiny [13], with an average precision (AP) of 90.05% and an inference time of 29.4 fps, but the model's weights are large and the detection accuracy needs to be further improved to achieve the desired results. Parico et al. [14] used YOLOv4-tiny to generate a robust real-time pear fruit counter for a mobile application, which recorded more than 50 fps and an AP value of 94.19%. However, it had an associated weight size of 22.97 MB, which means that a high computational cost is still required. Yu et al. [15] used an improved YOLOv7-based ripeness detection for pineapples with an *mAP* value of 95.82%, but the improved model leaves much to be desired in terms of detection performance in dense and highly occluded scenarios and is unable to achieve faster detection speed and deployment on low-power computing devices. It can be seen that deep learning convolutional neural network has a big advantage in target detection. It can quickly and accurately achieve the detection task of recognizing the ripeness of fruits in complex environments [16].

However, although algorithms such as deep learning convolutional neural networks and YOLO models can detect different targets quickly and accurately, complex and changing natural environments still pose a challenge for fruit detection, such as leaf occlusion, fruit overlap, light changes, brightness changes, target size, and shooting distant views, all of which affect fruit ripeness detection precision and accuracy [17]. In addition, the existing fruit ripeness detection studies mainly focus on crops such as apple [18], tomato [19], jujube [20], mango [21], and oil palm [22]; as an economic fruit crop of China's National Geographic indication products, there are few related studies on Gannan navel orange.

In summary, in view of these limitations, the main objective of this study is to achieve fast, accurate, and non-destructive detection of Gannan navel orange and its ripeness to improve the accuracy of ripeness detection of Gannan navel orange under the environments of shading [23], fruit overlapping [24], light variation [25], and target densification [26]

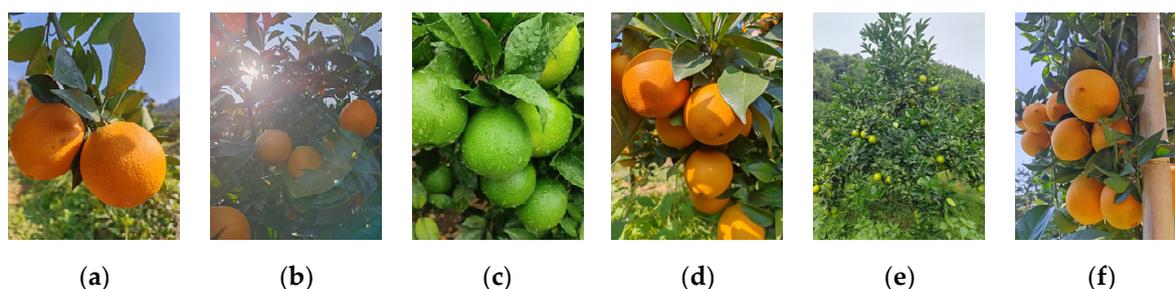
and to support the visual detection technology for selective harvesting of Gannan navel orange fruits in agricultural production. To this end, this paper proposes a Gannan navel orange ripeness target detection model based on YOLOv5-NMM. In the backbone layer of the model, we incorporate the CBAM attention mechanism, which improves the feature extraction and perception ability of the model. In the neck layer of the model, we incorporate the weighted bidirectional feature pyramid network structure in order to reduce the computational volume and improve the fusion efficiency of the images. In the prediction layer, we add a detection header to improve the detection accuracy of dense small targets. The Soft-NMS algorithm is used in the prediction stage at the end to reduce the missed detection of overlapping prediction frames. Finally, the feasibility and reliability of the method in this paper are verified on a home-made dataset.

The subsequent sections are structured as follows: Section 2 describes the preparation of the Gannan navel orange dataset and the improved algorithms to be used in the YOLOv5-NMM model proposed in this study. Section 3 evaluates the performance of the YOLOv5-NMM model through experiments. Section 4 summarises and discusses the work of the present study and points out the shortcomings of the present study, as well as the outlook for the future.

## 2. Materials and Methods

### 2.1. Image Acquisition

The experimental image acquisition site was located in the navel orange orchard, Anxi Town (115°41' E, 25°12' N), Ganzhou City, Jiangxi Province, with the Newhall variety of Gannan navel oranges as the research object, which was collected at the end of September and the beginning of December 2023, and the image acquisition equipment was a HONOR 70 smartphone, which captured 3000 images of Gannan navel oranges in different maturation periods, including 1500 mature Gannan navel oranges and 1500 immature Gannan navel oranges. The resolution was  $3072 \times 4096$  pixels and the image format was JPG. The types of collected images include single-target images, multi-target images, smooth light images, backlight images, branches, leaf shade images, images on a rainy day, no branch and leaf shade images, heavy fruit images, etc. Figure 1 shows some of the collected images under different light, environment, angle, and other factors.



**Figure 1.** Some Gannan navel orange samples in different environments. (a) forward light; (b) backward light; (c) rainy; (d) overlap; (e) long-distance vision; (f) dense.

### 2.2. Dataset Production

In order to identify Gannan navel orange fruits at different mature stages, the fruits were divided into two categories according to different growth stages, one of which was the mature navel orange. The other is immature navel-oranges (including the young fruit stage and expansion stage). The format of the experimental dataset is the YOLO dataset annotation format, and LabelImg 1.8.6 software is used to label Gannan navel orange with different maturity levels in the images (a total of 12,967 Gannan navel orange fruit labels are annotated, among which 7280 labels are mature Gannan navel orange labels and 5687 labels are immature Gannan navel orange labels). After all the labelling was completed, the 3000 Gannan navel orange photos were randomly divided into 2400 as the

training set, 300 as the validation set, and 300 as the test set in the ratio of 8:1:1. The basic information of the dataset is shown in Table 1.

**Table 1.** The number of pictures and labels of Gannan navel orange with different maturity.

Dataset	Number of Pictures	Labels of Mature	Labels of Immature
training set	2400	6060	4624
validation set	300	628	463
test set	300	592	600
Total	3000	7280	5687

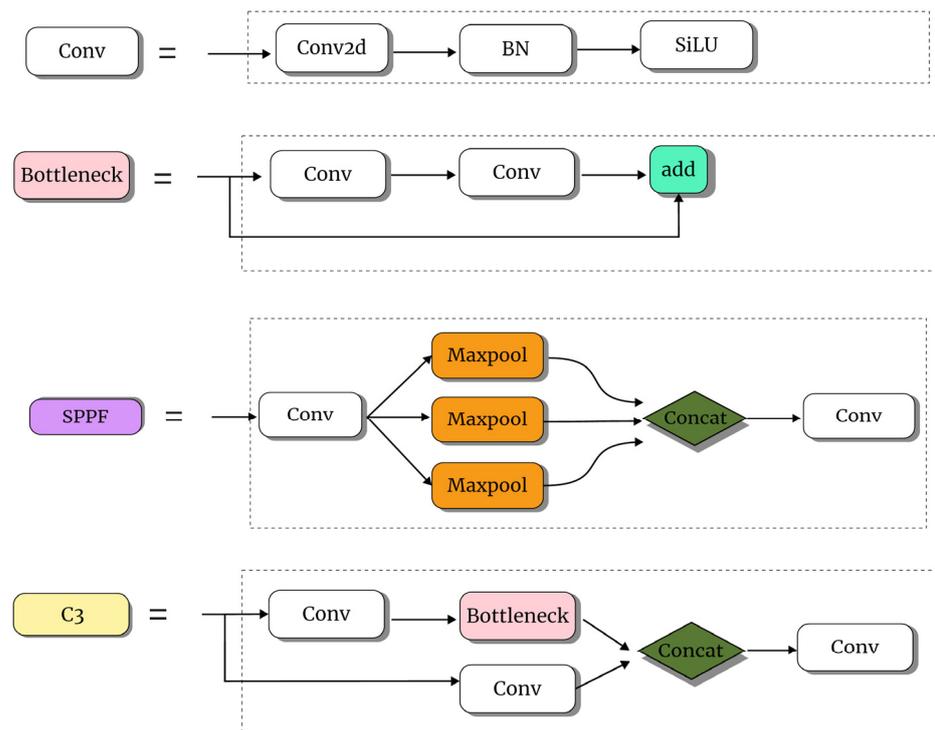
### 2.3. YOLOv5 Object Detection Algorithm

The YOLO target detection algorithm is used in this experiment; after its development in recent years, the YOLO version has been updated to YOLOv9, and YOLOv5, among many versions, has a relatively lightweight network structure, which is more suitable for deployment on mobile devices. It uses CSPDarknet53 as the backbone network and also optimises the header network to reduce the number of parameters and calculations in the model and improve the efficiency of the model. The YOLOv5 model is mainly composed of four parts: input, backbone, neck and prediction layer. The input layer establishes the image processing strategy and anchor frame generation mechanism, and the image processing uses Mosaic data enhancement and adaptive calculation of optimal anchor frame values for different training sets. The backbone layer mainly uses the Conv, C3, and SPPF basic structure of the input image for feature extraction. Different levels of feature maps are extracted by applying convolutional operations, and features are fused by cross-stage connection to reduce parameter redundancy and improve model accuracy. The neck layer uses a feature pyramid network structure to enhance semantic features from top to bottom, fully integrating the semantic information of deep and shallow features in feature maps at different scales and constructing a Path Aggregation Network structure to enhance feature information from the bottom up. The prediction layer generates the category probability and location information of the predicted targets and applies three detection heads to predict the large, medium, and small targets of the image in three different scales of feature maps, and the structure of the network is shown in Figures 2 and 3. YOLOv5 has achieved excellent results in many target detection competitions after its release and has been used and recommended by many researchers and developers [27–29]. As one of the popular algorithms in the field of target detection, YOLOv5 has a total of five base network models, which are YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, and the five models have the same network architecture, which only differs in depth, as well as width. Among the single-stage target algorithms, YOLOv5s has good detection accuracy and fast detection speed, so YOLOv5s is used in this paper for Gannan navel orange detection.

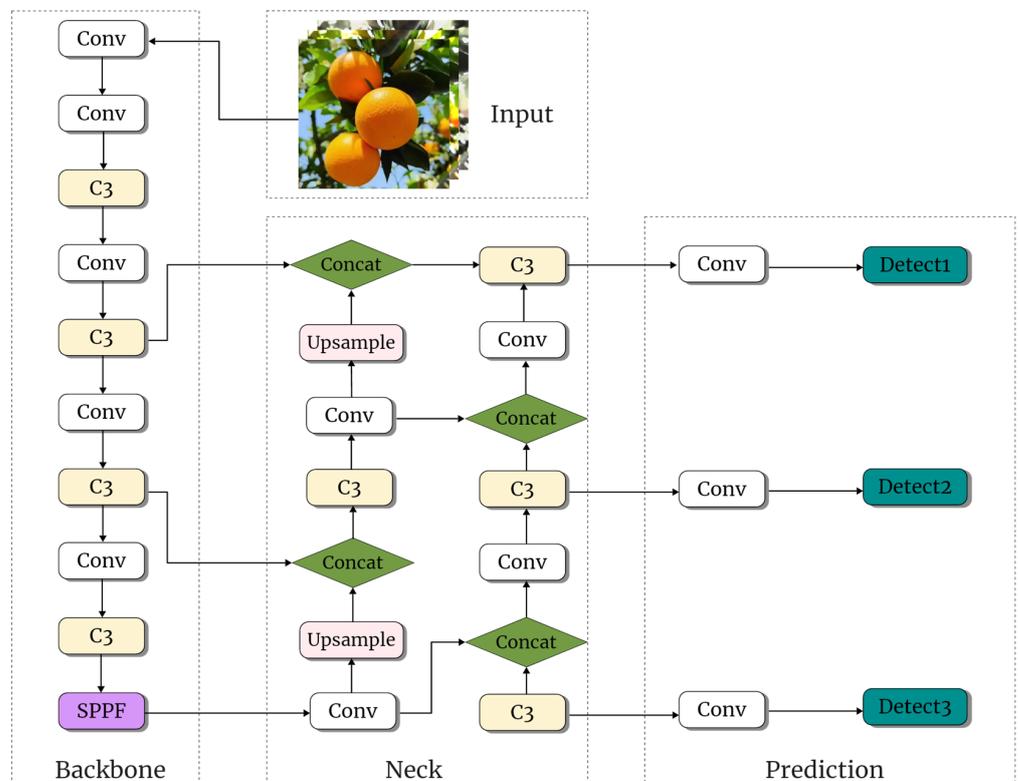
### 2.4. Improved YOLOv5-NMM Ripeness Detection Model for Gannan Navel Orange

#### 2.4.1. Improve the Multi-Scale Detection Layer

Due to the large number of small target Gannan navel oranges in the dataset, the original YOLOv5s uses an  $80 \times 80$  size detection head, which is not able to accurately detect small-sized Gannan navel oranges with different maturity levels. The feature map generated by YOLOv5s in the P2 layer is  $160 \times 160$ , which contains more shallow semantic features of the Gannan navel orange in small size. Therefore, in this paper, Upsample and Concat algorithms are applied to fuse the P2 layer of the backbone structure so as to generate shallow semantic features of Gannan navel orange. Thus, the generated deep semantic features of Gannan navel orange contain richer small-size, contour texture information. And we construct the detection head of the  $160 \times 160$  feature maps in the prediction so as to improve the detection accuracy of small target size Gannan navel orange with different maturity.



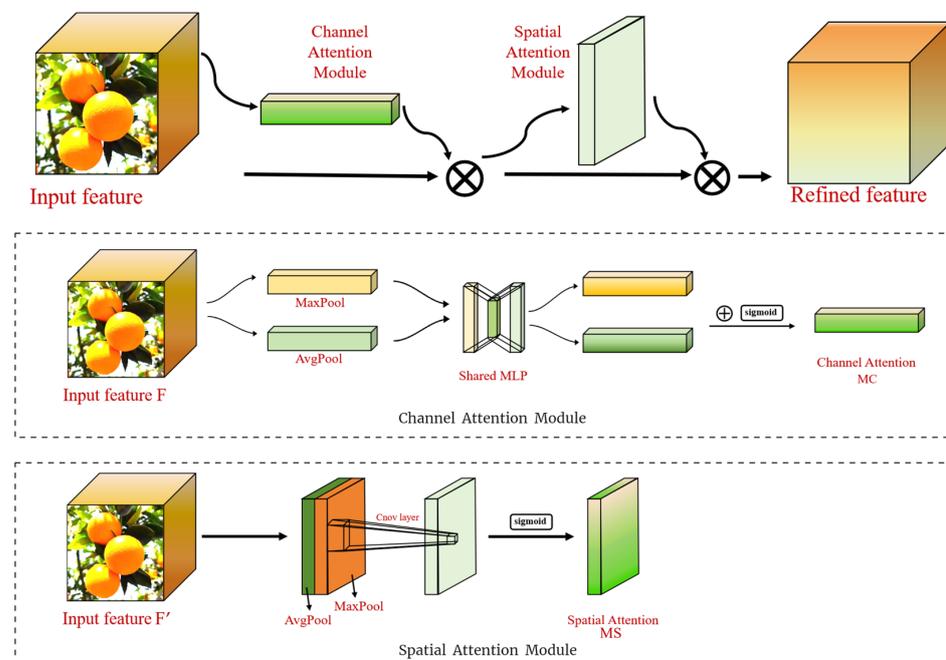
**Figure 2.** The component of the YOLOv5 model. Conv denotes the convolution operation, C3 denotes the feature extraction module, SPPF denotes the spatial pyramid pooling structure, upSample is the upsampling operation, and Concat is the feature fusion function.



**Figure 3.** Schematic of the overall structure of YOLOv5 model.

#### 2.4.2. Add the CBAM Attention Module

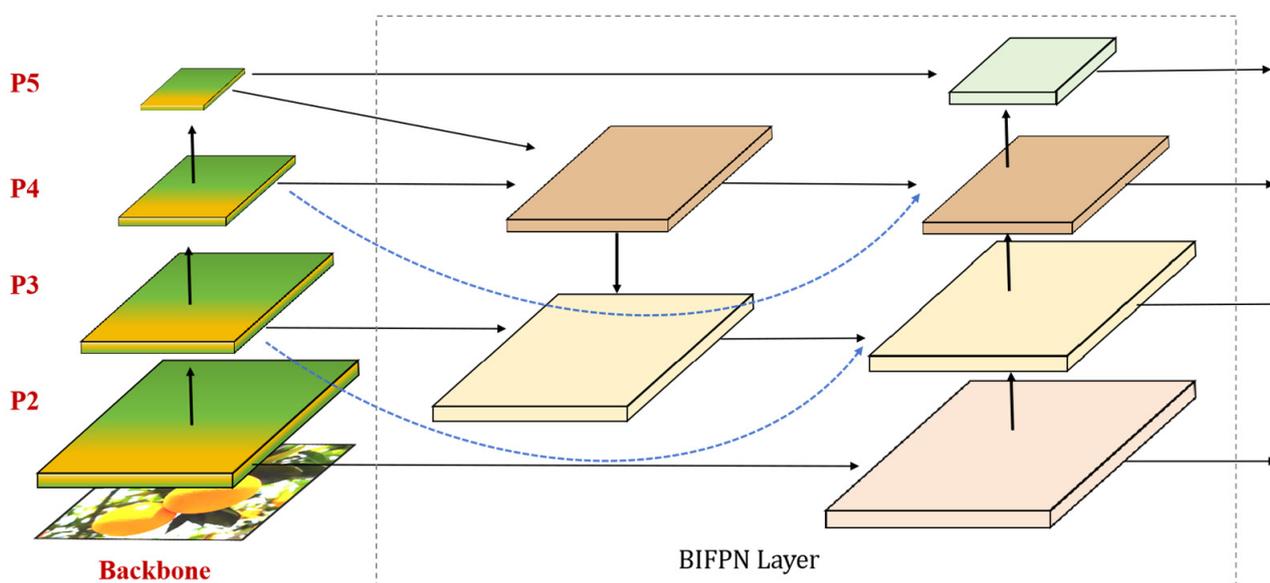
In order to better extract the ripeness characteristics of Gannan navel oranges under complex conditions and improve the model detection ability, the CBAM [30] attention mechanism is introduced, which consists of a channel attention module (CAM) and spatial attention module (SAM), and its structure is shown in Figure 4. In CAM, the input feature map  $F$  is subjected to the pooling operation, and then the number of channels is compressed through the fully connected layer. It is activated by the ReLU activation function and then expanded to the original number of channels through a fully connected layer to obtain two activated feature vectors for feature merging. Finally, the attention weight  $MC$  of the included channel is obtained using the softmax function. Finally, the feature vector is multiplied with the original feature map  $F$  to obtain the feature map  $F'$ , which is passed to SAM. In SAM,  $F'$  undergoes a pooling operation to obtain 2 2D vectors, which are then subjected to a splicing and convolution operation and undergo a sigmoid function to obtain the spatial attention weights  $MS$  of the input features, and finally, the weights  $MS$  are multiplied with the corresponding elements of the input features to obtain the final feature map. The CBAM module can automatically obtain the importance of each feature space and feature. The CBAM module can automatically obtain the importance degree of each feature space and channel by learning and then assign different weights according to the importance degree to multiply the attention map with the input feature map for adaptive feature optimisation, which overcomes the limitations of traditional convolutional neural networks in dealing with information of different scales, shapes, and orientations and at the same time, suppresses noise and other irrelevant information. Since CBAM is a lightweight and general-purpose module, the overhead of the module can be ignored and seamlessly integrated into all C3 architectures in the backbone layer for end-to-end training together with the underlying convolutional module, which can effectively avoid the influence of light, environment, brightness, background, and other factors in the recognition process of Gannan navel oranges and improve the ability to perceive the features of Gannan navel orange fruits to achieve all-weather picking.



**Figure 4.** The CBAM attention module consists of a channel attention module and a spatial attention module: Shared MLP is a shared fully connected layer, Avgpool is a global average pooling operation, and Maxpool is a global maximum pooling operation.

### 2.4.3. Introducing Bidirectional Feature Pyramid Network

A bidirectional feature pyramid network structure (BiFPN) [31] is introduced in the neck layer to replace the path aggregation Network (PANet) [32] structure, which solves the problems of insufficient information transfer and inaccurate feature fusion in the traditional feature pyramid network. The network introduces an adaptive weight adjustment mechanism, which can automatically learn the importance of each feature map to the final detection result, thus fusing the feature maps more accurately, and its structure is shown in Figure 5. By introducing bidirectional connections between the pyramid levels, the information can flow bottom-up and top-down in the network at the same time to better allow the features to be fully interacted and fused between the different levels, to solve the problems of poor information flow and feature loss, and to extract richer feature representations without increasing the number of parameters. It can effectively solve the problem of the recognition process of Gannan navel orange over the influence of image resolution, serious occlusion, network width and depth, small targets, and other factors on feature extraction and, at the same time, reduce the amount of computation for recognising the ripeness of Gannan navel orange and improve the accuracy and efficiency of the model.



**Figure 5.** The BiFPN structure in YOLOv5-NMM. A jump connection is added between the original input and output nodes to incorporate more functionality at no additional cost. P2–P5 are feature maps at different scales.

### 2.4.4. Replace the Non-Maximum Suppression Algorithm

In order to ensure recall, current object detection algorithms have multiple prediction frame outputs for the same real object. Since redundant prediction frames affect the detection accuracy, a non-maximal suppression algorithm (NMS) is required. This algorithm filters out the overlapping prediction frames to obtain the best-predicted output. For the traditional NMS method, among all the prediction frames of the same category in an image, if the prediction score of a frame is higher, this frame will be prioritised, and other frames of the same category overlapping with it and exceeding a certain *IOU* (Intersection Over Union) threshold will be discarded (i.e., their confidence scores will be set to zero). Although this method is simple and effective, in the actual picking environment, when the Gannan navel orange fruit is more dense and a considerable part of the fruit has overlapping occlusion, the prediction frames belong to several different fruits themselves; due to their overlapping with each other, the lower-scoring borders among them may be suppressed as the detection frames of the same fruit, which results in missed detection. Therefore, replacing the traditional non-maximum suppression algorithm is improved to

a soft non-maximum suppression (Soft-NMS) algorithm [33]; this algorithm uses a decay function to calculate the confidence level of the current detection frame, the formula is shown in Equation (1).

$$S_i = \begin{cases} s_i & , \quad IOU(M, b_i) < N_t \\ s_i(1 - IOU(M, b_i)) & , \quad IOU(M, b_i) \geq N_t \end{cases} \quad (1)$$

The parameter  $b_i$  is the pending frame,  $S_i$  is the bi-frame update score, and  $N_t$  is a manually set threshold, generally taken as 0.5.  $IOU$  is the ratio of the predicted bounding box to the true bounding box, and  $M$  is the current highest-scoring frame. The Soft-NMS algorithm uses a decay function to calculate the confidence level, instead of directly performing violent zeroing. Firstly, the detection frame with the highest confidence is identified from the detection frames, and the overlap degree  $IOU$  between this detection frame and the current detection frame is calculated. If the  $IOU$  is larger than the set threshold, a decay function is used to calculate the detection frame confidence  $S_i$ , and the larger the  $IOU$  is, the greater the degree of inhibition will be until all the detection frames are processed. The improved algorithm can effectively improve the situation of missed detection, as well as retain more frames to improve the detection capability of dense targets.

### 2.5. Test Environment Configuration and Network Parameter Setting

The experimental hardware and software environments in this paper are as follows: model training and testing with Windows 10 operating system, CPU is Intel Xeon Platinum 8255C (Intel Corporation, Santa Clara, CA, USA), GPU is NVIDIA RTX3080 (NVIDIA Corporation, Santa Clara, CA, USA), with 10 GB of video memory, 40 GB of running memory; in order to improve the speed of the network training, the GPU is used for acceleration, the CUDA version is 11.1, and the software environment is Python3.8. All comparison experiments are run in the same environment.

The model underwent training with a batch size of 16, the test iteration epoch is 200 rounds, and the initial learning rate is 0.001. The training and testing of this experiment are completed on the AutoDL Server (Beijing Seetatech Technology Corporation, Beijing, China), which adopts a high-performance hardware configuration and an optimised deep learning framework and provides perfect security measures so that it can excellently complete the model training and prediction tasks.

### 2.6. Experimental Evaluation Indicators

The process of detecting the ripeness of Gannan navel orange needs to consider the detection precision and speed. In terms of model detection precision, Precision ( $P$ ), Recall ( $R$ ), and  $AP$  are selected as evaluation indexes. In terms of model detection performance, mean average precision ( $mAP$ ), model weight size, and number of parameters were chosen as evaluation indicators to test the model performance. When robotic picking is carried out in the natural environment, in order to prevent the peel damage of fruits in the process of robotic arm picking, the picking action is more soothing. So there is no high requirement for the model detection rate. Therefore, this experiment takes the mean average precision mean value  $mAP$  as the first evaluation index, and the formula of its relevant evaluation index is shown in (2)–(5).

$$P = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$AP = \int_0^1 P(R) dR \times 100\% \quad (4)$$

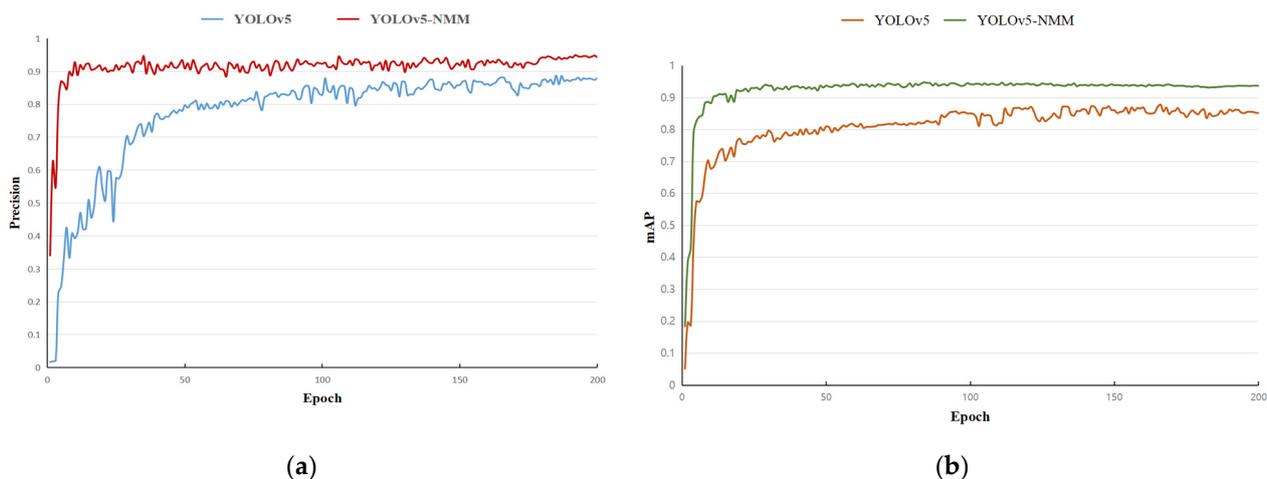
$$mAP = \frac{1}{k} \sum_{i=1}^k AP_i \times 100\% \quad (5)$$

where  $TP$  denotes the number of samples predicted by the model to be positive and are positive samples,  $FP$  denotes the number of samples predicted by the model to be positive but are negative samples,  $FN$  denotes the number of samples predicted by the model to be negative samples but are positive samples,  $k$  denotes the number of categories of the data,  $AP_i$  denotes the average precision of detection objectives for category  $i$ , and  $mAP$  denotes the mean of all its categories of  $AP_i$ .

### 3. Results

#### 3.1. Analysis of Model Training Results

In order to better represent the detection performance of the YOLOv5-NMM model, we compare the original YOLOv5 model with the improved YOLOv5-NMM model. The curves of the changes in the performance indicators of the two models during the training process are shown in Figure 6. As can be seen from the change curve of precision in Figure 6a and the change curve of  $mAP$  in Figure 6b, during the iterative training process, the change in the model training process is constantly and rapidly rising in the first 50 rounds, and the indexes gradually converge and stabilize in the training process after that. After 200 rounds of training, the precision value and  $mAP$  value finish convergence, indicating that the model has reached the fitting state. Since  $mAP$  integrates the accuracy and recall of Gannan navel orange ripeness of two categories, the weight value at the optimal number of rounds of  $mAP$  is selected as the model weights for this model. At this time, the YOLOv5-NMM model P is 93.2% and  $mAP$  is 94.2%. In contrast, the YOLOv5 model has a P of 88.3% and an  $mAP$  of 87.1%, which shows that the improved YOLOv5-NMM model has a greater improvement in terms of precision and  $mAP$  values. In addition, we also compared with the fruit ripening models in the literature mentioned in the introduction and found that the map of most of these fruit detection algorithms is below 90%. The YOLOv5-NMM model is in the leading position in terms of precision and  $mAP$  value. Some examples of Gannan navel orange detection are shown in Figure 7, from which we can see that the algorithm in this paper is able to accurately detect the ripeness of Gannan navel oranges under different environmental conditions. In summary, this experiment has achieved the expected purpose. The YOLOv5-NMM model is able to accurately detect the ripeness of fruits and has a better detection effect on small targets, multiple targets, branch and leaf shading, heavy fruits, and the effect of light.



**Figure 6.** Curve of each performance index of the two models. (a) Precision curve; (b)  $mAP$  curve.



Figure 7. Selected samples were detected based on the YOLOv5-NMM model.

### 3.2. Analysis of Ablation Experiment Results

To further validate the performance of the YOLOv5-NMM model, under the same experimental conditions, ablation experiments were set up for the Gannan navel orange ripeness detection dataset to validate the performance of the five groups of networks, and the results are shown in Table 2. Group 1 is the basic YOLOv5s network model, and Group 2 adds a multi-scale detection layer on top of Group 1, and the *mAP* value of the improved model improves by 2.2%, which indicates that the improved multi-scale detection layer can improve the detection accuracy of small-size and dense Gannan navel oranges; Group 3 introduces the CBAM attention mechanism on the basis of Group 2, and the *mAP* value is improved by 1.4%, and its results prove that the CBAM attention mechanism can better capture the key information of the complex scene in the image of Gannan navel oranges and improve the accuracy and robustness of the task; Group 4 introduces the BiFPN. The *mAP* value of the improved model is improved by 0.6%, and the number and size of model parameters do not increase, which indicates that BiFPN can effectively enhance the transfer of features and information fusion between different network structure layers and can obviously improve the detection accuracy of the YOLOv5 algorithm while having fewer parameters and computational complexity, which makes it suitable for embedded devices and practical deployments. The last group replaces the original non-extremely large value suppression algorithm on the basis of Group 4, and the improved Group 5 constitutes the YOLOv5-NMM model, whose *mAP* value reaches 94.2%, which indicates that the improved Soft-NMS can effectively improve the situation of leakage detection, and increase the detection ability for dense targets. Through the analysis of the above ablation test data, it is found that the improved model can achieve better detection results.

Table 2. Results of ablation experiments.

ID	Add One Layer	CBAM	BiFPN	Soft-NMS	Precision (%)	Recall (%)	<i>mAP</i> (%)	Param (M)	Size of Model (MB)
1	×	×	×	×	88.3	84.7	87.1	7.03	14.6
2	✓	×	×	×	91.2	86.8	89.3	7.12	14.8
3	✓	✓	×	×	91.9	87.4	90.7	7.18	15.1
4	✓	✓	✓	×	92.4	87.9	91.3	7.13	15.0
5	✓	✓	✓	✓	93.2	89.6	94.2	7.19	15.3

### 3.3. Comparative Analysis of Different Detection Models

In order to obtain a high-precision, lightweight, reliable and easy-to-deploy ripeness detection model for Gannan navel orange and to qualitatively evaluate the detection capability of the improved YOLOv5-NMM model, under the same experimental conditions, algorithmic comparisons were made between the improved model and mainstream Faster R-CNN, YOLO v3, YOLOv7-tiny, and the original YOLO v5 models for the Gannan Navel orange image in the test set. The statistical results of the performance index of each model are shown in Table 3, and the detection results are mainly from the comparison of the two aspects of detection accuracy and algorithm size, in which the highest detection accuracy is the *mAP* of 94.2% for YOLOv5-NMM, which is higher than that of the other algorithms Faster R-CNN, YOLOv3, YOLOv7-tiny, and the original YOLO v5 models by 6.1, 8.0, 7.1, and 3.9 percentage points. Meanwhile, in terms of model size, the improved YOLOv5-NMM model weight size is only 15.3 M, and its performance is obviously due to other models.

**Table 3.** Performance comparison of different target detection models.

Model	Precision (%)	Recall (%)	<i>mAP</i> (%)	Size of Model (MB)
Faster R-CNN	86.7	82.6	88.1	216.4
YOLO v3	87.9	83.8	86.2	121.7
YOLOv5s	88.3	84.7	87.1	14.6
YOLOv7-tiny	85.8	81.9	90.3	17.6
YOLOv5-NMM	93.2	89.6	94.2	15.3

## 4. Discussion

Aiming at the problem of the maturity detection of Gannan navel oranges in the natural environment, this paper is based on the YOLOv5 target detection network and improves the detection accuracy of small-sized navel oranges with different maturity in Gannan by improving the multi-scale detection layer. In order to improve the perception ability of the target and realize the detection in complex scenes, a lightweight channel attention module CBAM is introduced. The bidirectional feature pyramid network is added to better fuse the hierarchical features of different scales. Finally, the non-maximum suppression algorithm was improved to improve the missed detection and retain more boxes to improve the detection ability of dense targets.

In the evaluation process of the model, this project set up five groups of networks for ablation experiments and carried out a quantitative analysis of the experimental results. Experimental analysis shows that the improved YOLOv5-NMM model has good performance in terms of detection accuracy and calculation amount. The *mAP* value increased from 87.1% to 94.2%, which indicates that the original YOLOv5 model has certain limitations in the recognition of fruit maturity, while the YOLOv5-NMM model can effectively deal with the recognition of fruit maturity in environments such as small targets, overlapping, dense, and illumination changes. At the same time, we also found that the number of parameters and model size did not grow, which provides good conditions for our next deployment on mobile devices. In the process of comparison with the current mainstream algorithms, by comparing with the Faster R-CNN, YOLOv3 model, YOLOv7-tiny model, and the original YOLOv5 model, the YOLOv5-NMM model has achieved good results on the Gannan Navel-orange dataset, while other algorithms have poor performance due to insufficient accuracy or too large model.

At the same time, we compared the YOLOv5-NMM model with the algorithms in the cited references and found that the requirements of detection algorithms for different kinds of fruits are slightly different. For example, the detection of jujube needs to focus on how to detect and recognize the small targets, and the research on grapes focuses on the study of the dense targets. The YOLOv5-NMM model algorithm in this study draws the advantages of these algorithms. Therefore, the recognition of the maturity of Gannan navel orange

under the environment of small targets, overlapping, dense, and light changes has been comprehensively improved.

## 5. Conclusions

In summary, the focus of this paper was to investigate the use of target detection technology to identify the maturity of Gannan navel orange. The mechanized harvesting of Gannan navel oranges under an all-weather environment is realized and the development and growth of the navel orange industry are promoted through the method of precision agriculture. This includes improved detection capabilities under conditions such as small targets, shading, heavy fruit, dense distribution, brightness, and light environments. It can also be lightweight and deployed in embedded devices and has achieved good results in comparison with existing mainstream detection algorithms. This is of great practical significance and value to the development of the Gannan navel orange industry.

However, due to the high similarity between the blurred immature Gannan navel oranges at the edge in the image background and the green leaf background contour, there are still few samples that recognize the background as immature Gannan navel oranges in the recognition process, which affects the detection accuracy. In addition, the recognition rate of the improved Gannan navel orange ripeness detection model needs to be improved under dark and low light conditions, and the sensitivity of the model to low light needs to be further strengthened. To address the above problems, subsequent research is needed to gradually improve the performance and reliability in practical applications by increasing the training sample data, improving the algorithm structure, and adjusting the model parameters. Finally, migrating this YOLOv5-NMM model to other fruit ripeness recognition will be another focus of future research, and we have already conducted small-scale tests on other fruits with good results. In the next step, we will also conduct in-depth research on embedding the Gannan navel orange ripeness target detection model into an agricultural picking robot with a high-definition camera combined with a picking robotic arm, which can really realize efficient mechanized picking operations.

**Author Contributions:** Conceptualization, B.Z.; methodology, B.Z.; software, B.Z.; validation, M.C. and K.W.; formal analysis, B.Z.; investigation, B.Z. and M.C.; resources, K.W.; data curation, B.Z.; writing—original draft preparation, B.Z.; writing—review and editing, B.Z.; visualization, B.Z.; supervision, K.W.; project administration, M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data are contained within the article.

**Acknowledgments:** This research was supported by Shanghai Ocean University.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Barth, R. Vision Principles for Harvest Robotics: Sowing Artificial Intelligence in Agriculture. Doctoral Dissertation, Wageningen University and Research, Wageningen, The Netherlands, 2018; Volume 83-02.
2. He, W.B.; Wei, A.Y.; Ming, W.Y.; Jia, H.J. Survey of Fruit Quality Detection Based on Machine Vision. *Comput. Eng. Appl.* **2020**, *56*, 10–16.
3. Liu, G.X.; Mao, S.Y.; Kim, J.H. A Mature-Tomato Detection Algorithm Using Machine Learning and Color Analysis. *Sensors* **2019**, *19*, 2023. [[CrossRef](#)] [[PubMed](#)]
4. Li, H.; Zhang, M.; Gao, Y.; Li, M.; Ji, Y. Green ripe tomato detection method based on machine vision in greenhouse. *Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 328–334.
5. Kurtulmus, F.; Lee, W.S.; Vardar, A. Green citrus detection using ‘eigenfruit’, color and circular Gabor texture features under natural outdoor conditions. *Comput. Electron. Agric.* **2011**, *78*, 140–149. [[CrossRef](#)]
6. Zhang, W.; Huang, S.; Wang, J.J.; Liu, L.Z. A segmentation method for wheat leaf images with disease in complex background. *Comput. Eng. Sci.* **2015**, *37*, 1349–1354.
7. Ji, W.; Zhao, D.; Cheng, F.Y.; Xu, B.; Zhang, Y.; Wang, J.J. Automatic recognition vision system guided for apple harvesting robot. *Comput. Electr. Eng.* **2012**, *38*, 1186–1195. [[CrossRef](#)]

8. Mallick, M.; Basu, D.; Hossain, S.M.; Das, J. Ethylene sensor based on graphene oxide for fruit ripeness sensing application. *Appl. Phys. A—Mater. Sci. Process.* **2023**, *129*, 140. [[CrossRef](#)]
9. Ashtiani, S.H.M.; Javanmardi, S.; Jahanbanifard, M.; Martynenko, A.; Verbeek, F.J. Detection of Mulberry Ripeness Stages Using Deep Learning Models. *IEEE Access* **2021**, *9*, 100380–100394. [[CrossRef](#)]
10. Appe, S.R.N.; Arulsevi, G.; Balaji, G. Tomato ripeness detection and classification using VGG based CNN models. *Int. J. Intell. Syst. Appl. Eng.* **2023**, *11*, 296–302.
11. 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 (CVPR), Seattle, WA, USA, 27–30 June 2016; pp. 779–788.
12. Fu, L.; Feng, Y.; Wu, J.; Liu, Z.; Gao, F.; Majeed, Y.; Al-Mallahi, A.; Zhang, Q.; Li, R.; Cui, Y. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model. *Precis. Agric.* **2021**, *22*, 754–776. [[CrossRef](#)]
13. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
14. Parico, A.I.B.; Ahamed, T. Real Time Pear Fruit Detection and Counting Using YOLOv4 Models and Deep SORT. *Sensors* **2021**, *21*, 4803. [[CrossRef](#)]
15. Lai, Y.; Ma, R.; Chen, Y.; Wan, T.; Jiao, R.; He, H. A Pineapple Target Detection Method in a Field Environment Based on Improved YOLOv7. *Appl. Sci.* **2023**, *13*, 2691. [[CrossRef](#)]
16. Zhu, Y.-N.; Zhou, W.; Yang, Y.; Li, J.-P. Automatic Identification Technology of Lycium barbarum Flowering Period and Fruit Ripening Period Based on Faster R-CNN. *Chin. J. Agrometeorol.* **2020**, *41*, 668.
17. An, Q.L.; Wang, K.; Li, Z.Y.; Song, C.Y.; Tang, X.Y.; Song, J. Real-Time Monitoring Method of Strawberry Fruit Growth State Based on YOLO Improved Model. *IEEE Access* **2022**, *10*, 124363–124372. [[CrossRef](#)]
18. Yan, B.; Fan, P.; Lei, X.; Liu, Z.; Yang, F. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. *Remote Sens.* **2021**, *13*, 1619. [[CrossRef](#)]
19. Huang, Y.-P.; Wang, T.-H.; Basanta, H. Using Fuzzy Mask R-CNN Model to Automatically Identify Tomato Ripeness. *IEEE Access* **2020**, *8*, 207672–207682. [[CrossRef](#)]
20. Defang, X.; Huamin, Z.; Lawal, O.M.; Xinyuan, L.; Rui, R.; Shujuan, Z. An automatic jujube fruit detection and ripeness inspection method in the natural environment. *Agronomy* **2023**, *13*, 451.
21. Ignacio, J.S.; Eisma, K.N.A.; Caya, M.V.C. A YOLOv5-based deep learning model for in-situ detection and maturity grading of mango. In Proceedings of the 2022 6th International Conference on Communication and Information Systems (ICCIS), Chongqing, China, 14–16 October 2022; pp. 141–147.
22. Suharjito, Junior, F.A.; Koeswandy, Y.P.; Debi, Nurhayati, P.W.; Asrol, M.; Marimin. Annotated datasets of oil palm fruit bunch piles for ripeness grading using deep learning. *Sci. Data* **2023**, *10*, 72. [[CrossRef](#)] [[PubMed](#)]
23. Gai, R.L.; 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](#)]
24. Xu, Z.; Liu, J.; Wang, J.; Cai, L.; Jin, Y.; Zhao, S.; Xie, B. Realtime Picking Point Decision Algorithm of Trellis Grape for High-Speed Robotic Cut-and-Catch Harvesting. *Agronomy* **2023**, *13*, 1618. [[CrossRef](#)]
25. Wang, C.; Han, Q.; Li, J.; Li, C.; Zou, X. YOLO-BLBE: A Novel Model for Identifying Blueberry Fruits with Different Maturities Using the I-MSRCR Method. *Agronomy* **2024**, *14*, 658. [[CrossRef](#)]
26. Fu, X.; Zhao, S.; Wang, C.; Tang, X.; Tao, D.; Li, G.; Jiao, L.; Dong, D. Green Fruit Detection with a Small Dataset under a Similar Color Background Based on the Improved YOLOv5-AT. *Foods* **2024**, *13*, 1060. [[CrossRef](#)] [[PubMed](#)]
27. Zhu, X.K.; Lyu, S.C.; Wang, X.; Zhao, Q.; Soc, I.C. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios. In Proceedings of the 18th IEEE/CVF International Conference on Computer Vision (ICCV), Electr Network, Virtual, 11–17 October 2021; pp. 2778–2788.
28. Jia, W.; Xu, S.Q.; Liang, Z.; Zhao, Y.; Min, H.; Li, S.J.; Yu, Y. Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector. *IET Image Process.* **2021**, *15*, 3623–3637. [[CrossRef](#)]
29. Yao, J.; Qi, J.M.; Zhang, J.; Shao, H.M.; Yang, J.; Li, X. A Real-Time Detection Algorithm for Kiwifruit Defects Based on YOLOv5. *Electronics* **2021**, *10*, 1711. [[CrossRef](#)]
30. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the 15th European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
31. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In *Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020*; pp. 10781–10790.
32. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
33. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS—Improving Object Detection with One Line of Code. In Proceedings of the 16th IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5562–5570.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.