agriculture

Article

Lightweight Detection Algorithm of Kiwifruit Based on Improved YOLOX-S

Junchi Zhou, Wenwu Hu *, Airu Zou, Shike Zhai, Tianyu Liu, Wenhan Yang and Ping Jiang

College of Mechanical and Electrical Engineering, Hunan Agricultural University, Changsha 410128, China; jcz9858@stu.hunau.edu.cn (J.Z.); zar@stu.hunau.edu.cn (A.Z.); zsk0371@stu.hunau.edu.cn (S.Z.); liutianyu@hunau.edu.cn (T.L.); yangwenhan@stu.hunau.edu.cn (W.Y.); 1233032@hunau.edu.cn (P.J.)
* Correspondence: 1233087@hunau.edu.cn

Abstract: Considering the high requirements of current kiwifruit picking recognition systems for mobile devices, including the small number of available features for image targets and small-scale aggregation, an enhanced YOLOX-S target detection algorithm for kiwifruit picking robots is proposed in this study. This involved designing a new multi-scale feature integration structure in which, with the aim of providing a small and lightweight model, the feature maps used for detecting large targets in the YOLOX model are eliminated, the feature map of small targets is sampled through the nearest neighbor values, the superficial features are spliced with the final features, the gradient of the SiLU activation function is perturbed, and the loss function at the output is optimized. The experimental results show that, compared with the original YOLOX-S, the enhanced model improved the detection average precision (AP) of kiwifruit images by 6.52%, reduced the number of model parameters by 44.8%, and improved the model detection speed by 63.9%. Hence, with its outstanding effectiveness and relatively light weight, the proposed model is capable of effectively providing data support for the 3D positioning and automated picking of kiwifruit. It may also successfully provide solutions in similar fields related to small target detection.

Keywords: YOLOX; small target scale; loss function; feature integration; fruit picking

Citation: Zhou, J.; Hu, W.; Zou, A.; Zhai, S.; Liu, T.; Yang, W.; Jiang, P. Lightweight Detection Algorithm of Kiwifruit Based on Improved YOLOX-S. *Agriculture* **2022**, *12*, 993. https://doi.org/10.3390/agriculture12070993

Academic Editors: Gniewko Niedbała and Sebastian Kujawa

Received: 15 June 2022
Accepted: 7 July 2022
Published: 9 July 2022

1. Introduction

Agriculture is the source of human clothing, food, housing, and transportation; an important foundation for people's lives; the backbone that supports the national economy; and the guarantee for the country's stable development. At present, the application of artificial intelligence in agriculture mainly includes intelligent farm systems with management and decision-making capabilities based on the background of agricultural big data [1], motion obstacle target detection and path recognition [2], crop growth and pest detection [3], weed recognition [4], fruit and vegetable quality detection [5], and automatic picking based on agricultural robots and other related fields.

Fruit-picking robots can automate picking work, effectively resolving issues related to labor shortages, high costs, and low efficiency in the manual picking process [6,7]. Determination of the critical criteria of picking robots involves studying the visual system, while the efficiency and stability of such robots predominantly depend on the speed and accuracy of fruit recognition, along with the accuracy and adaptability in complex environments [8,9]. Therefore, research on visual systems that possess the capability to accurately identify the fruit on trees in complex environments is of substantial value and practical significance for achieving automatic picking and yield estimation.

Numerous scholars across the world have conducted extensive research on target object recognition technology [10–13]. In the field of fruit crop detection in natural environments, feature extraction and recognition have predominately targeted tomato [14,15], apple [16–18], cucumber [19,20], strawberry [21], sugarcane [22], pineapple [23], and various other fruits. Among the various fruits, the planting area and yield of kiwifruit in

particular have continued to increase over time. With its high yield and rich nutritional value, kiwifruit has been widely planted and become popular among consumers. Methods for detection and recognition are predominantly segregated into traditional machine vision methods and deep learning methods. As an example of such machine vision, Cui Yongjie et al. [24] utilized the L*a*b* color space a* channel for kiwifruit image segmentation, and adopted the elliptical Hough transform to fit the contour of a single fruit for segmentation and recognition. In addition, Fu et al. [25] have proposed the use of 1.1R-G color characteristics for nighttime kiwifruit image segmentation, and combined the minimum circumscribed rectangle method and elliptical Hough transform to identify each fruit. However, both methods presented unsatisfactory results for fruit segmentation and unfavorable results for multi-fruit cluster recognition compared to traditional algorithms, such as SIFT [26], HOG [27], and texture extraction algorithms [28–30]. Kiwifruit images in the field environment possess vastly diverse features, complex backgrounds, and substantial differences in morphological features. Traditional machine vision methods are mainly constructed based on experience and are influenced by samples and human subjectivity; hence, they are unable to effectively meet the demands of applications in complex field environments.

Deep learning target detection algorithms have experienced significant leaps in performance and accuracy, and various model networks have substantially enhanced their ability to resist scale changes and translation. Song Zhenzhen et al. [31] have constructed a fast VGG16 model to achieve the detection of kiwifruit in live images by integrating a region proposal network (RPN) and a fast R-CNN network, while Fu Longsheng et al. [32] have proposed a network-based LeNet convolutional neural network deep learning model for multi-cluster kiwifruit images with general applicability to the recognition of multi-cluster kiwifruits. Although, as research on deep learning-based target detection methods has focused more on the construction of a deeper networks for the purpose of enhancing detection accuracy, the associated network models have generally suffered from an overly large number of parameters. This has led to slow detection speeds, meaning that the algorithms can only be run on high-performance graphics processors and generally have high equipment requirements. Concurrently, according to analysis of the growth characteristics of kiwifruit, most targets in kiwifruit detection tasks have predominantly been on small targets (both absolute and relative scales are relatively small).

Therefore, in the interest of reducing the number of model parameters and enhancing the model detection speed, the YOLOX-S network, which possesses excellent multi-scale detection performance and takes into account both the detection speed and accuracy as its basis, was selected for this research. This work aims to improve on the original network model, in order to maintain the target detection accuracy while compressing the model, thus effectively achieving the detection of small kiwifruit targets.

2. Experimental Platform and Materials

2.1. Vision Platform System

In this paper, we primarily focus on the object detection task in the image processing field. The image recognition module was a Jetson Nano embedded development board, as presented in Figure 1. The improved model algorithm, which was trained in advance, was embedded in the board, and wireless communications, remote monitoring, and remote control were achieved through the 4G network module. The communication system is mainly divided into the picking-machine end, cloud server end, and client end, ensuring the transmission and storage of information. Remote wireless control of the picking robot can also be achieved. In addition, the left and right imagers of the depth camera capture video or image data, which are sent to a depth imaging processor. This processor correlates points in the left image with those in the right image, and calculates the depth value of each pixel in the image by shifting the points in the left image to match with the right image. Finally, it returns the result to the terminal in order to command the manipulator to act accordingly.

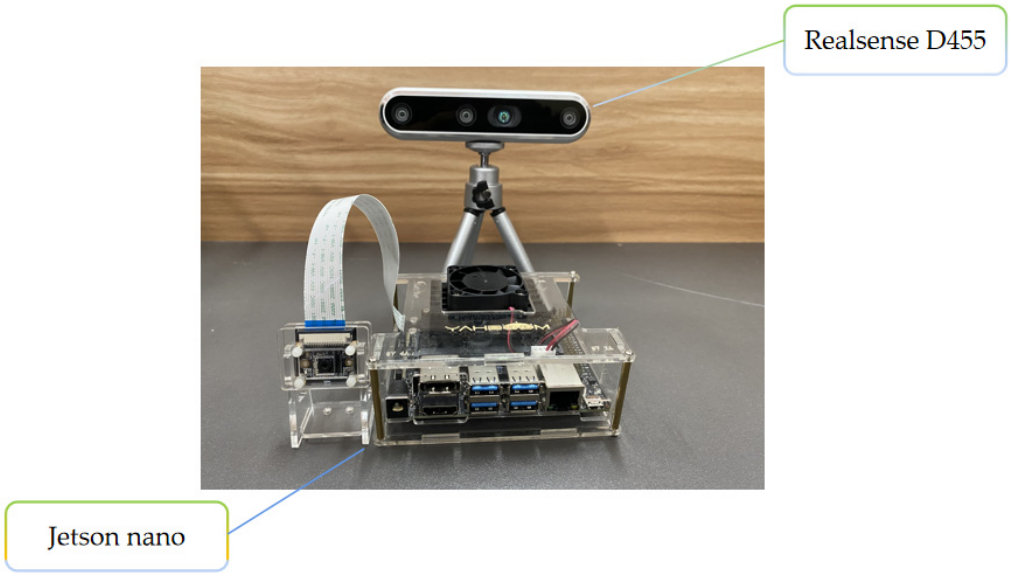


Figure 1. Image-recognition embedded module.

2.2. Hardware Platform

A test platform was independently developed by our team, which can be applied as a fruit picking and transferring platform in hilly and mountainous areas (Figure 2). The platform has a pure electric drive and a CAN interface for chassis speed regulation, steering, and attitude feedback. It is capable of stable driving and meets the hardware requirements of the platform positioning test for the chassis in hilly and mountainous areas.

Figure 2. Electric fruit picking platform used for experimental trials.

2.3. Experimental Configuration and Environment

The used graphics card was an NVIDIA GeForce GTX 3060, and the CPU was an AMD Ryzen 7 5800H with 16 GB memory. The experimental configuration was Windows 10, Python 3.8, PyTorch 1.8.1, and CUDA 10.1. The parameter settings are presented in Table 1.

Table 1. Training parameter settings.

Parameter	Value
Momentum	0.937
Weight_decay	0.0005
Batch_size	45
Learning_rate	0.0001
Epochs	500

2.4. Experimental Sample Dataset

The experimental data in this paper were collected from the Internet and from on-site filming. A total of 1500 images were collected. The photos taken on the spot are all taken from the orchard. Each picture contained a significantly large number of kiwifruit target fruits, and the total number of targets was 41,687. The targets in each image were labeled with fine granularity, in order to facilitate subsequent enhancements in the detection of small targets.

3. Principles and Methods

YOLOX [33] is a brand new high performance real-time target detection network, recently launched by Beijing Megvii Technology. It adopts cutting-edge technologies such as the anchor-free mechanism, decoupled heads, multi-positives, advanced label assignment strategy, and strong data augmentation. Hence, it has faster speed, higher recognition accuracy, smaller weight files, and can be easily mounted on mobile devices with lower configurations, thereby offering high research value. The structure of the YOLOX-S network selected for this paper is depicted in Figure 3.

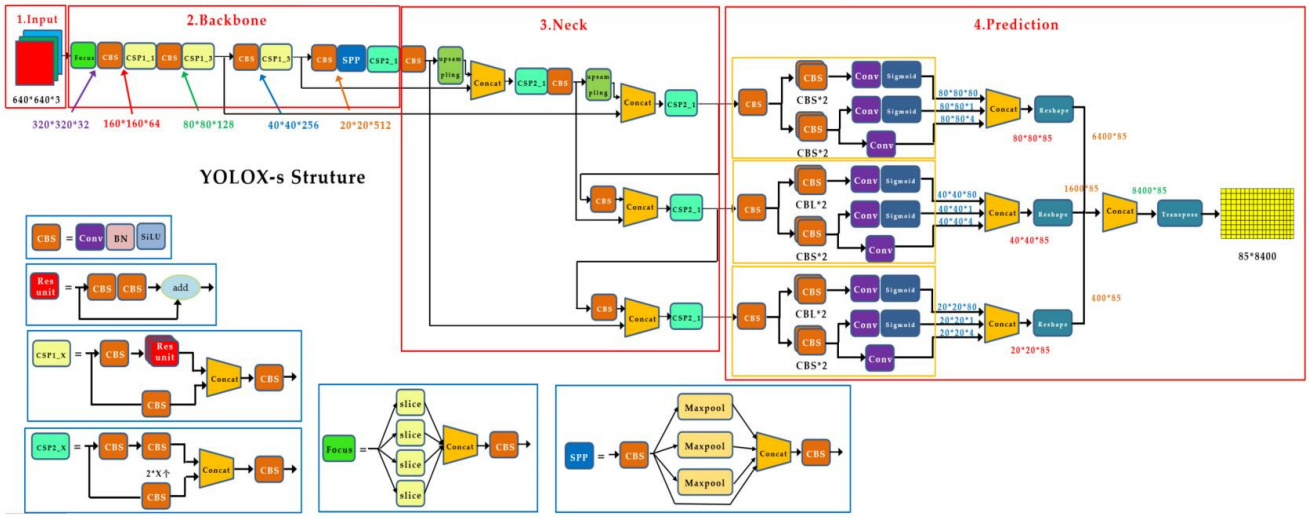


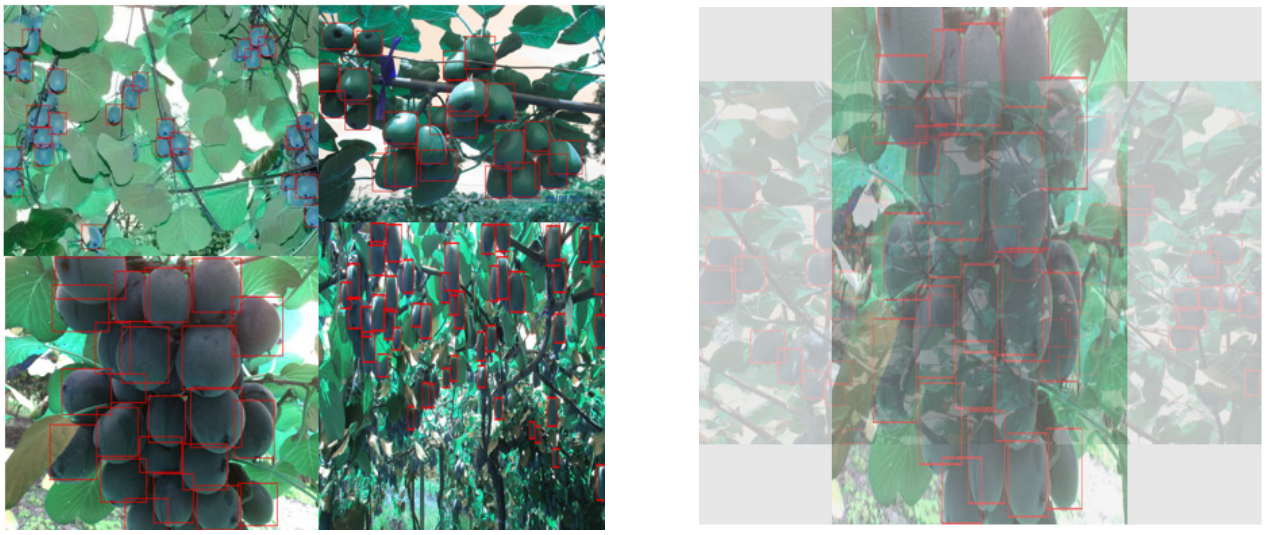
Figure 3. Structure of the YOLOX-S network model.

CSPDarknet is the backbone feature extraction network of the YOLOX algorithm, which is predominantly composed of three modules: Focus, CSPNet, and a spatial pyramid pooling network. The model first slices an input image for the operation. By sampling the complete image at equal intervals, multiple sampled images of appropriate size can be obtained. Subsequently, these images are combined in the channel dimension and the information in the image is transferred to the channel space, resulting in a down-sampled image with no information loss. The CSPNet module contains the backbone feature extraction and residual structure, which can effectively extract image features and significantly reduces the computational effort while maintaining high accuracy. The SPP network convolutes the output of the last CSPNet once, then utilizes three different scales of maximum pooled kernels to integrate the features of the feature image under different

receptive fields. FPN + PAN is a circular pyramid structure composed of convolution, sampling, and feature fusion operations, which repeatedly extracts the input image features, performs feature fusion at different scales, and finally outputs the three feature maps at different scales to the decoupled head for accurate prediction.

3.1. Pre-Processing of the Data Set

YOLOX utilizes mosaic and mix-up data augmentation methods to substantially enrich the detection dataset; in particular, random scaling is conducted to supplement the many small targets and make the network more robust. Mosaic augmentation involves performing a series of operations, such as flipping, scaling, and color shifting, on multiple different pictures, followed by cropping and splicing to recombine them into a new image. Hence, the generated images often contain more targets. Therefore, this kind of augmentation technique can significantly enrich the background and alleviate the imbalance of positive and negative samples in the detection process, to a certain extent. Mix-up augmentation refers to the fusion of two pictures, to some degree, in which the labels of the samples are also weighted. The prediction results are weighted using the weighted labels in order to calculate the loss; subsequently, the backpropagation update parameters can be enhanced. The effect is shown in Figure 4.



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

Figure 4. Data enhancement effects: (**a**) mosaic data enhancement and (**b**) mix-up data enhancement.

3.2. Improved YOLOX-S Network

3.2.1. Perturbing the Activation Function Gradient

The predominant function of the activation function is to provide non-linearity in the network structure. Considering that the difference between the gradient propagation effects of the SiLU and Mish loss functions utilized in the YOLOX model is slight, gradient perturbation was considered based on the SiLU activation function. As presented in Figure 5, the SiLU→SiLU-1 gradient change led to a smoother curve, while the SiLU→SiLU + 1 gradient change became steeper. Given that the Mish activation function worked relatively well in YOLOv4, it was considered to increase or decrease the gradient change based on SiLU. Introducing a gradient increase can enhance the generalization ability of the model more robustly, and as such, we found that the SiLU + 1 activation function enhanced the generalization ability of the model to a certain extent.

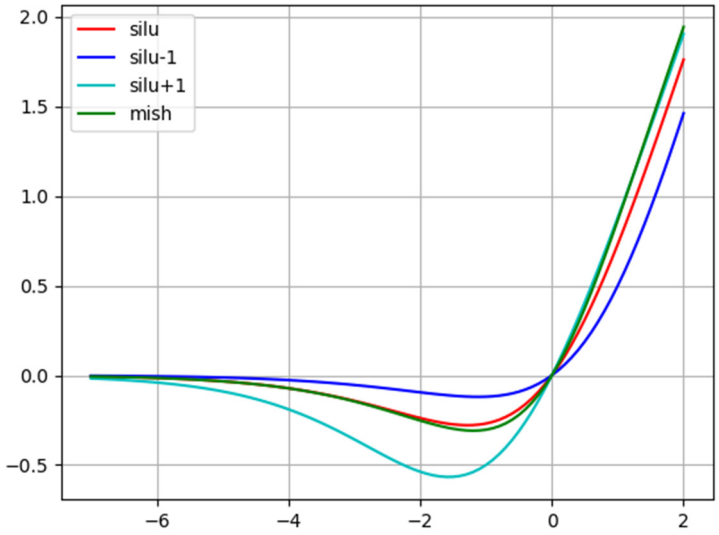


Figure 5. Variation between the perturbed gradient functions.

The dynamic positive and negative sample allocation algorithm utilized by YOLOX, SimOTA, is fast and effective. When determining the candidate areas for positive samples, the center point of a grid (20 × 20, 40 × 40, 80 × 80) was selected as the circle inside the ground truth (GT), with r being the radius centered on the center point of the GT. In Figure 6, the green box denotes the GT. It can be observed that there may be mismatches when using a small feature map. Subsequently, it can be observed that GTs are more likely to match smaller GTs in larger feature maps, but small feature maps can match a significantly small number of GTs.

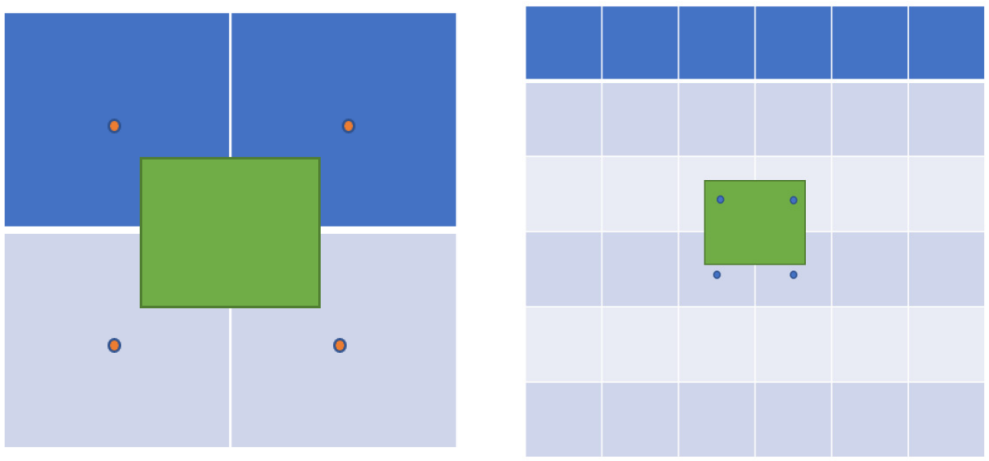
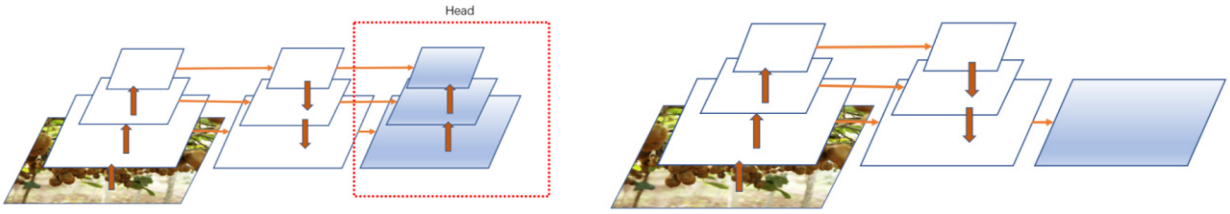


Figure 6. Feature matching candidate sample situation.

3.2.2. Nearest Neighbor Interpolation Up-Sampling of 80 × 80 Feature Map

Through in-depth research on the allocation strategy of positive and negative samples in the YOLOX model, the YOLOX model was found to reduce the number of predicted samples of the feature map in the confidence loss calculation, where almost all of the reduced samples were negative. Hence, the problem of imbalance in quantity caused by too many negative samples was alleviated, thereby suggesting that most targets in the kiwifruit detection task are small targets (i.e., both the absolute scale and relative scale are relatively small). Therefore, with the goal of reducing the number of model parameters and improving the model detection speed, the feature maps (20 × 20, 40 × 40) used for detecting large targets in the YOLOX model were eliminated. Subsequently, only the 80 × 80 feature map was retained, and a larger feature map size was introduced on this basis to match the

GT more effectively. When acquiring the final output from the 80 × 80 feature map, nearest-neighbor interpolation was utilized for up-sampling. This allows the model to provide more predictions and better match GTs, thereby extensively reducing the complexity of the model and the number of parameters. Figure 7 demonstrates the structure of the network before and after the improvement.



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

Figure 7. (**a**) Original YOLOX feature fusion structure and (**b**) improved structure, in which only the 80 × 80 feature map structure is preserved.

3.2.3. Transfer of Shallow Features

The performance when using a single output feature map may be unstable under specific conditions. Considering that the low-level feature semantic information is relatively small but the target position is accurate, the final feature map and the feature map in the shallow network were concatenated, in order to better integrate the semantic and representation information to a certain extent, such that the accuracy of the regression box could be significantly enhanced (see Figure 8).

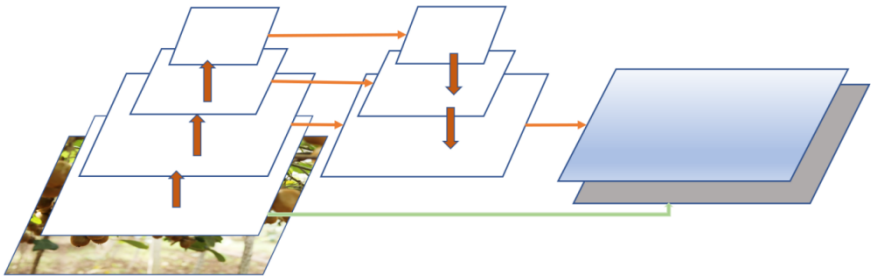


Figure 8. Structure of the final feature map.

3.2.4. Enhancing the Loss Function

Equations (1)–(5) are the loss functions of the YOLOX-S algorithm. The bounding box loss functions GIOU_loss and IOU_loss for predicting Reg have certain limitations, resulting in an inability to effectively optimize the overlap between the detection box and the real box when one is included in the other. Subsequently, for the confidence degree and category loss, the original algorithm adopts a binary cross-entropy loss function, which is not conducive to the classification of positive and negative samples.

$$Loss = GIOU\_Loss + Loss_{conf} + Loss_{class}, \tag{1}$$

$$GIOU\_Loss = 1 - GIOU = 1 - (IOU - \frac{|Q|}{C}), \tag{2}$$

$$\begin{aligned} Loss_{conf} = & \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - \hat{C}_i^j)] - \\ & \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{noobj} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - \hat{C}_i^j)] \end{aligned}, \tag{3}$$

$$Loss_{class} = \sum_{i=0}^{S^2} I_{ij}^{obj} \sum_{c \in classes} [\hat{P}_i^j \log(P_i^j) + (1 - \hat{P}_i^j) \log(1 - \hat{P}_i^j)]. \quad (4)$$

In Equation (2), C represents the minimum circumscribed rectangle of the detection frame and the priori frame, and Q represents the difference between the minimum circumscribed rectangle and the concatenation of the two frames.

In Equations (3) and (4), I_{ij}^{obj} and I_{ij}^{noobj} indicate whether the target falls into detection frame j of grid i, and λ_{noobj} represents the loss weight of the localization error. Subsequently, C_j^i and P_j^i refer to the training values, and $\hat{C}_i^j$ and $\hat{P}_i^j$ refer to the prediction values.

Therefore, we adopted CIOU_loss as the Reg bounding box loss function and increased the aspect-ratio restriction mechanism, compared with the previous one, such that the prediction box was more in line with the real box, as demonstrated in Equation (5). Equation (6) was used to measure the consistency of the aspect ratio, and the confidence degree and category loss function utilized the PolyLoss function based on the Taylor expansion approximation of the focal loss [34]. Thus, it not only took into account the superior binary classification performance of the focal function, but also achieved enhancement of the accuracy and performance on this basis. The convergence speed was also effectively accelerated.

$$CIOU\_Loss = 1 - (IOU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v), \quad (5)$$

$$\begin{gathered} v = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w^p}{h^p})^2 \\ \alpha = \frac{v}{(1-IOU)+V} \end{gathered} \quad (6)$$

where $\rho()$ is the Euclidean distance between the center points of the two boxes, c is the diagonal length of the smallest circumscribed rectangle of the two, α is the weight coefficient, and v is the aspect ratio distance between the two frames.

$$L_{Ploy-1} = -\log(P_t) + \varepsilon_1(1 - P_t), \quad (7)$$

where P_t represents the probability of target label prediction.

4. Results and Discussion

4.1. Evaluation of Model Performance

In order to evaluate the effectiveness of the proposed method for kiwifruit detection in different aspects, the mean average accuracy, the number of model parameters utilized, and the *FPS*, along with the detection time per sheet, were selected as evaluation metrics. *mAP* refers to a comprehensive consideration of precision and recall, which is used to evaluate the effectiveness of the model, while *FPS* refers to the number of frames per second, which can be utilized to measure the real-time performance of the model. Finally, the number of model parameters reflects the lightness of the model.

$$\begin{gathered} recall = \frac{TP}{TP + FN} \\ precision = \frac{TP}{TP + FP}, \\ mAP = \frac{\sum_{i=0}^{N-1} \int_0^1 P(R)dR}{N} \end{gathered} \quad (8)$$

where TP represents the number of correctly identified images, FP represents the number of misidentified images, and FN represents the number of missed images. When there is only one category, mAP is equal to the AP.

4.2. Analysis and Comparison of the Enhanced Model

In order to efficiently verify the effectiveness of the proposed model, a comparative experiment was conducted on the enhanced YOLOX-S network using the same training parameters. Table 2 provides the detailed scores of each evaluation index before and after the enhancement. Figure 9 presents the *AP* diagram of the model before and after enhancement.

Table 2. The comparison of model scores before and after enhancement.

Model	*AP*	Param	Time/ms	*FPS*
IMPROVED	82.62	5,483,590	15.6	101
ORI	76.10	9,937,682	43.2	88

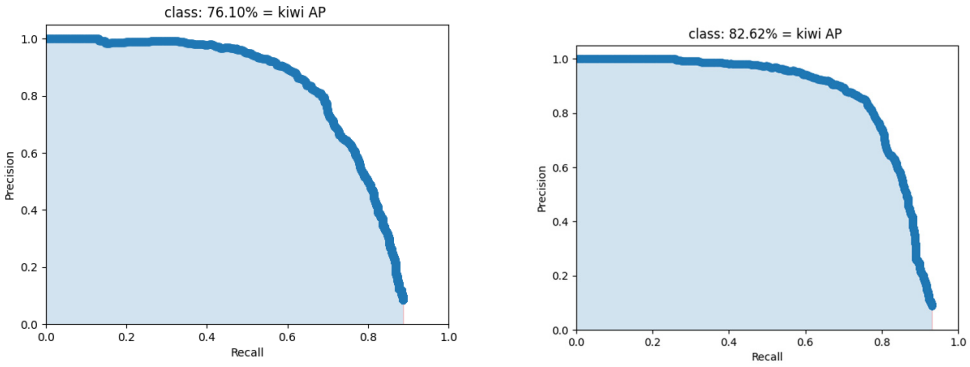


Figure 9. The YOLOX-S training curve.

From the perspective of model lightness, the improved model parameters were reduced by 44.8% and the model detection speed was increased by 63.9%, verifying the feature expression ability of the model. Feature map up-sampling and nearest-neighbor interpolation reduced the computational complexity by omitting unnecessary computations, thus achieving the effect of making the network lightweight.

So as to more intuitively depict the improvement in various aspects for the considered models, we created a performance comparison diagram with respect to the model improvement strategies, as shown in Figure 10.

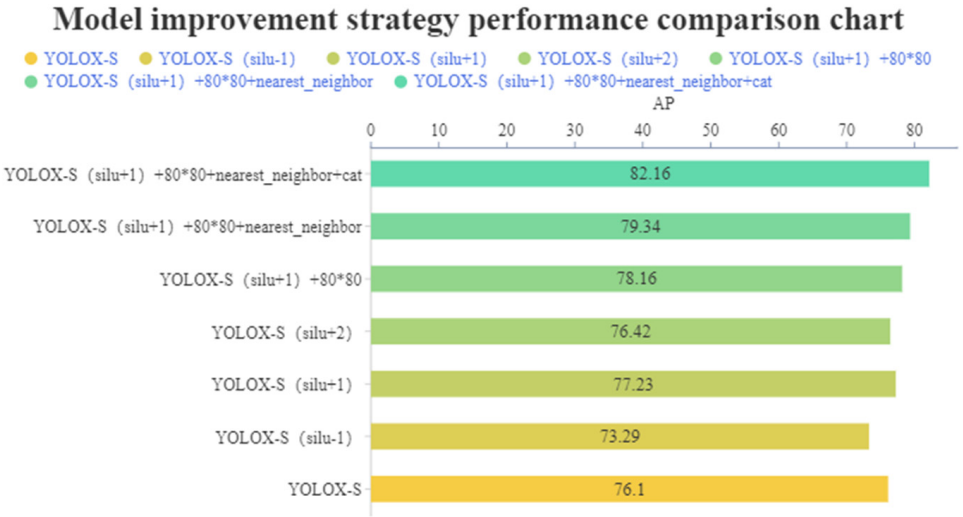


Figure 10. Model improvement strategy performance comparison chart.

As shown by Figure 10, in terms of model effectiveness and accuracy, the expressiveness growth of the model is mainly divided into three stages. The first stage is that the

perturbation of the activation function enhances the generalization ability of the model by selecting the SILU+1 function, which increases the *AP* value by 1.13%. The improvement in the second stage is due to the cancellation of the feature map of the redundant large target in this detection task, so that the network detection is all concentrated on the small target, which reduces the calculation of negative samples and the misjudgment of positive samples, so that the *AP* value continues to increase by 2.01%. The improvement in the last stage comes from the design of the new network fusion structure. By splicing the final output feature map and shallow-level features, the semantic information of the two is combined, and the loss function is improved in the prediction segment. Compared with the original model, the enhanced model significantly improved the *AP* value on the kiwifruit images by 6.52%, which is a substantial increase.

Figure 11 presents the before and after images for comparison. By comparing the groups of images, it can be seen that in the (Figure 11a) group of experiments, the fruit could be effectively identified by the enhanced model, even when there were tree trunks, branches, and leaves in the way.

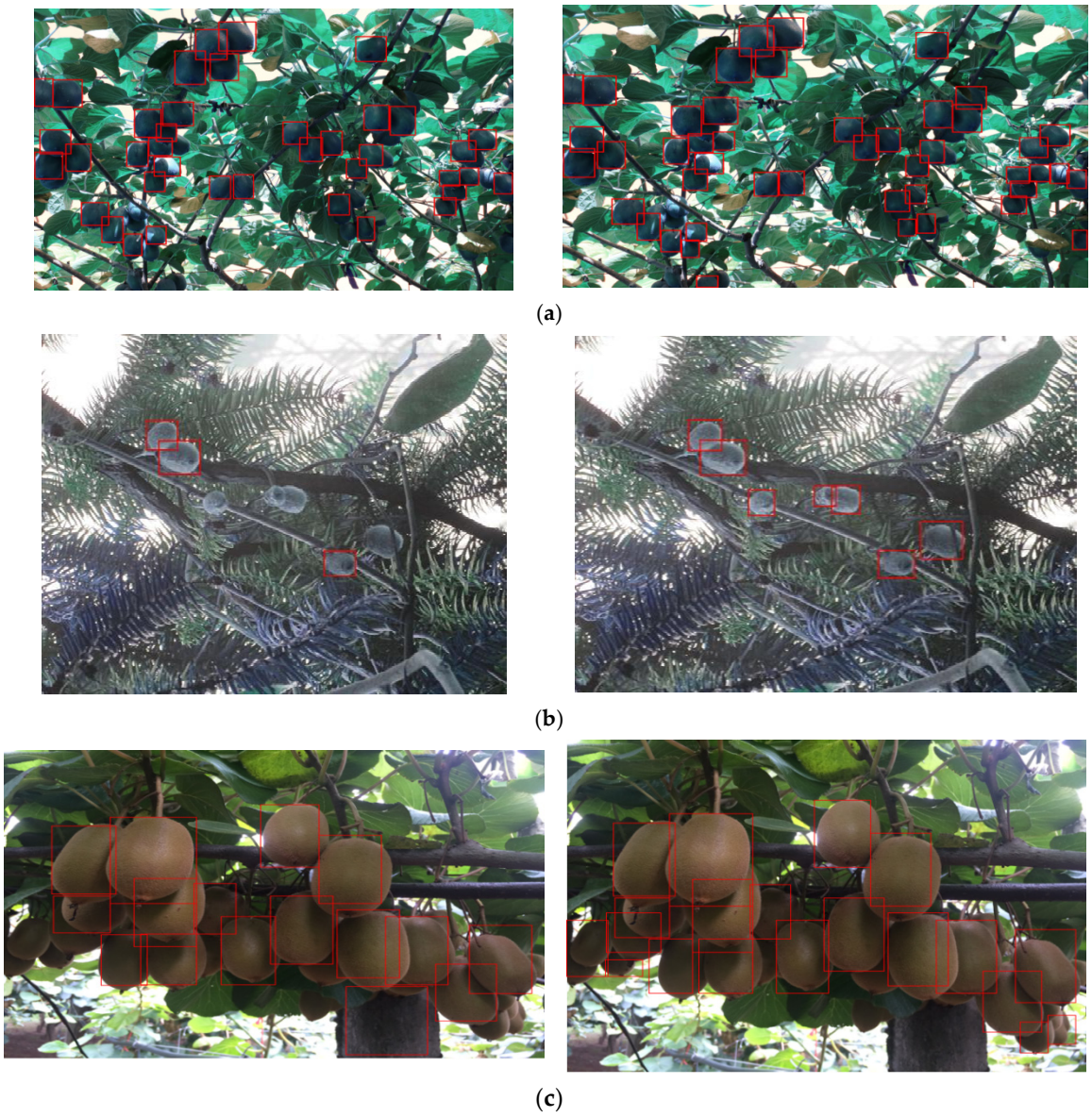


Figure 11. (**a**) Enhancement of fruit recognition under tree trunk and leaf occlusion. (**b**) Enhancement of low-density fruit missed recognition. (**c**) Enhancement of non-target fruit misrecognition.

However, for the (Figure 11b) group of experiments, the original algorithm was not able to recognize the low-density fruits effectively. Additionally, for the (Figure 11c) group of experiments, the original algorithm misidentified the tree trunk as a fruit, but the enhanced model corrected it, and was able to accurately identify more fruit.

In light of the above, the enhanced algorithm significantly improved the ability to detect small-scale target fruit and reduced the false recognition and misrecognition rates. In addition, we compared several state-of-the-art algorithms and conducted training tests under the same conditions. The proposed enhanced model provided improved results in all aspects. The performance comparison is given in Table 3. In addition, Table 4 compares our findings with those of various scholars around the world, and details the advantages and disadvantages of their techniques.

Table 3. Comparison of mainstream models.

Model	mAP@0.5/%	*FPS*
Ours	82.62	101
YOLOv5s	74.12	83
YOLOv3	69.46	68
YOLOv2	67.83	63
Fast R-CNN	80.15	52

Table 4. Comparison and analysis of advantages and disadvantages of methods.

Reference	Description	Advantages	Disadvantages
[24]	Utilized the L*a*b* color space a* channel for kiwifruit image segmentation	Accurate segmentation of a single fruit	susceptible to external changes
[25]	Combines least circumscribed rectangle method and elliptic Hough transform	Accurate segmentation of a single fruit	Not ideal for fruit cluster identification
[35]	Improved K-means algorithm	Multi-target detection possible	Easily disturbed by shape and texture
[36]	Built a color classifier	Low hardware requirements	slow detection
[37]	Construction of the positional relationship between fruit and calyx in linear clusters	high speed	Does not meet the multi-robot collaborative operation
[31]	Merge Region Proposal Network (RPN) and Fast R-CNN Network	high speed	Poor adaptability in complex situations
[32]	Convolutional Neural Network Based on LeNet	Multiple fruit clusters can be recognized	High equipment requirements and large amount of parameters
Ours	Cancel the large object detection layer Concatenate shallow features with final features	high speed few parameters Suitable for embedded mobile devices	AP can be further improved

It can be seen from the table that the algorithm proposed in this paper can solve the problem of poor recognition of multiple fruit clusters compared with the traditional image processing method used in past research [24,25,35–37]. Compared with studies based on deep learning methods [31,32], the improved accuracy of our algorithm alleviates the problems of the network model being too large and the equipment requirements being too high. The recognition in the case of fruit occlusion and misjudgment is improved, the recognition accuracy and speed of the fruit are further improved, and the parameter

amount of the model is reduced. It can effectively complete the detection task of kiwifruit in agricultural production, and has a positive impact on future picking of kiwifruit.

5. Conclusions

The research ultimately proposed an enhanced YOLOX-S target detection algorithm for kiwifruit picking robots. In order to effectively improve the detection of small-scale targets, the YOLOX-S algorithm was enhanced through fine-grained annotation of the target frame of the data set, as well as mosaic and mix-up data augmentation methods. Through up-sampling of the nearest-neighbor value in the small target feature map and the splicing of superficial features with the final features, in addition to the optimization of the loss function, the number of parameters of the enhanced YOLOX-S were significantly reduced while the *AP* values was increased. We demonstrated that the proposed enhancement method is applicable to actual fruit-picking environments, and is beneficial for embedment in mobile devices.

This research predominantly focused on the detection of kiwifruit. Simultaneously, the critical key to effective picking is to locate the target and return its three-dimensional coordinate points.

In further research, we intend to focus on:

(a) At present, the *AP* of the model has not reached the ideal state. Next, the data set will be enriched to further improve the performance and accuracy of the model.
(b) We will use pre-processing of the depth image data and color image data by utilizing the camera's external and internal parameters, triangulation principles, and the conversion of pixel coordinates to 3D spatial coordinates to carry out fruit localization.
(c) The proposed algorithm effectively met the basic requirements for fruit picking using a large-end actuator. However, due to the large number of kiwifruit that need to be picked, in order to further enhance the efficiency of the manipulator, it is necessary to further research the picking sequence allocation for kiwifruit.
(d) We will analyze the correlation between data, identify a variety of other types of fruit through transfer learning, and design a multi-classification general picking model for orchards.

Author Contributions: Conceptualization, J.Z. and W.H.; methodology, P.J.; software, J.Z. and T.L.; validation, J.Z., A.Z. and S.Z.; formal analysis, J.Z.; investigation, J.Z.; resources, W.H. and P.J.; data curation, J.Z. and W.Y.; writing—original draft preparation, J.Z.; writing—review and editing, J.Z., P.J. and W.H.; visualization, J.Z.; supervision, P.J., T.L. and W.H.; project administration, P.J. and W.H.; funding acquisition, P.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Excellent Youth Project of the Hunan Education Department (No. 20B292), the research and application of key technologies for remote active monitoring of wild animals under national key protection in forests, the Hunan Agricultural Machinery Equipment and Technology Innovation R&D Project (Xiang Cai Nong Zhi [2020] No.107), and the Hunan Province Science and Technology Achievement Transformation and Industrialization Plan Project (2020GK4075).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Junaid, M.; Shaikh, A.; Hassan, M.U.; Alghamdi, A.; Rajab, K.; Al Reshan, M.S.; Alkinani, M. Smart Agriculture Cloud Using AI Based Techniques. *Energies* **2021**, *14*, 5129. [CrossRef]
2. Liu, C.; Feng, Q.; Tang, Z.; Wang, X.; Geng, J.; Xu, L. Motion Planning of the Citrus-Picking Manipulator Based on the TO-RRT Algorithm. *Agriculture* **2022**, *12*, 581. [CrossRef]

3. Kong, J.; Wang, H.; Yang, C.; Jin, X.; Zuo, M.; Zhang, X. A Spatial Feature-Enhanced Attention Neural Network with High-Order Pooling Representation for Application in Pest and Disease Recognition. *Agriculture* **2022**, *12*, 500. [CrossRef]
4. Jiang, H.; Zhang, C.; Qiao, Y.; Zhang, Z.; Zhang, W.; Song, C. CNN feature based graph convolutional network for weed and crop recognition in smart farming. *Comput. Electron. Agric.* **2020**, *174*, 105450. [CrossRef]
5. Mesa, A.R.; Chiang, J.Y. Multi-Input Deep Learning Model with RGB and Hyperspectral Imaging for Banana Grading. *Agriculture* **2021**, *11*, 687. [CrossRef]
6. Jia, W.; Tian, Y.; Luo, R.; Zhang, Z.; Lian, J.; Zheng, Y. Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot. *Comput. Electron. Agric.* **2020**, *172*, 105380. [CrossRef]
7. Fu, L.; Feng, Y.; Wu, J.; Liu, Z.; Gao, F.; Majeed, Y.; AI-Mallahi, A.; Zhang, Q.; Li, R.; Cui, Y. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model. *Precis. Agric.* **2020**, *13*, 754–776. [CrossRef]
8. Jia, W.; Zhang, Y.; Lian, J.; Zheng, Y.; Zhao, D.; Li, C. Apple harvesting robot under information technology: A review. *Int. J. Adv. Robot. Syst.* **2020**, *17*, 1729881420925310. [CrossRef]
9. Kang, H.; Chen, C. Fruit detection and segmentation for apple harvesting using visual sensor in orchards. *Sensors* **2019**, *19*, 4599. [CrossRef] [PubMed]
10. Yang, L.; Luo, J.; Song, X.; Li, M.; Wen, P.; Xiong, Z. Robust Vehicle Speed Measurement Based on Feature Information Fusion for Vehicle Multi-Characteristic Detection. *Entropy* **2021**, *23*, 910. [CrossRef]
11. Zhou, J.; Jiang, P.; Zou, A.; Chen, X.; Hu, W. Ship Target Detection Algorithm Based on Improved YOLOv5. *J. Mar. Sci. Eng.* **2021**, *9*, 908. [CrossRef]
12. Liu, T.; Ma, Y.J.; Yang, W.; Li, W.; Wang, R.; Jiang, P. Spatial-temporal interaction learning based two-stream network for action recognition. *Inform. Sci.* **2022**, *606*, 864–876. [CrossRef]
13. Liu, C.; Su, J.; Wang, L.; Lu, S.; Li, L. LA-DeepLab V3+: A Novel Counting Network for Pigs. *Agriculture* **2022**, *12*, 284. [CrossRef]
14. Arefi, A.; Motlagh, A.; Mollazade, K.; Teimourlou, R. Recognition and Localization of Ripen Tomato Basedon Machine Vision. *Aust. J. Crop. Sci.* **2011**, *5*, 1144–1149.
15. Xiang, R.; Ying, Y.; Jiang, H.; Rao, X.; Peng, Y. Recognition of Overlapping Tomatoes Based on Edge Curvature Analysis. *Trans. Chin. Soc. Agric. Mach.* **2012**, *43*, 157–162.
16. Si, Y.; Liu, G.; Gao, R. Segmentation Algorithm for Green Apples Recognition Based on K-means Algorithm. In Proceedings of the 3rd Asian Conference on Precision Agriculture, Beijing, China, 14–17 October 2009; pp. 100–105.
17. Zulkifley, M.A.; Moubark, A.M.; Saputro, A.H.; Abdani, S.R. Automated Apple Recognition System Using Semantic Segmentation Networks with Group and Shuffle Operators. *Agriculture* **2022**, *12*, 756. [CrossRef]
18. Jing, W.; Leqi, W.; Yanling, H.; Yun, Z.; Ruyan, Z. On Combining DeepSnake and Global Saliency for Detection of Orchard Apples. *Appl. Sci.* **2021**, *11*, 6269. [CrossRef]
19. Henten, E.; Tuijl, B.; Hoogakker, G.J.; Weerd, M.; Hemming, J.; Kornet, J.G.; Bontsema, J. An autonomous robot for de-leafing cucumber plants in a high-wire cultivation system. *Biosyst. Eng.* **2006**, *94*, 317–323. [CrossRef]
20. Liu, C.; Zhao, C.; Wu, H.; Han, X.; Li, S. ADDLight: An Energy-Saving Adder Neural Network for Cucumber Disease Classification. *Agriculture* **2022**, *12*, 452. [CrossRef]
21. Xie, Z.; Zhang, T.; Zhao, J. Ripened Strawberry Recognition Based on Hough Transform. *Trans. Chin. Soc. Agric. Mach.* **2007**, *38*, 106–109.
22. Lu, S.; Wen, Y.; Ge, W.; Peng, H. Recogntition and Features Extraction of Suagrcane Nodes Based on Machine Vision. *Trans. Chin. Soc. Agric. Mach.* **2010**, *41*, 190–194.
23. Li, B.; Wang, M.; Li, L. In-field pineapple recognition based on monocular vision. *Trans. Chin. Soc. Agric. Eng.* **2010**, *26*, 345–349.
24. Cui, Y.; Su, S.; Lyu, Z.; Li, P.; Ding, X. A Method for Separation of Kiwifruit Adjacent Fruits Based on Hough Transformation. *J. Agric. Mech. Res.* **2012**, *34*, 166–169.
25. Fu, L.; Wang, B.; Cui, Y.; Gejima, Y.; Taiichi, K. Kiwifruit recognition at nighttime using artificial lighting based on machine vision. *Int. J. Agric. Biol. Eng.* **2015**, *8*, 52–59.
26. Lowe, D.G. Object Recognition from Local Scale-Invariant Features. In Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV), Kerkyra, Greece, 20–27 September 1999; pp. 1150–1157.
27. Dalal, N.; Tniggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 July 2005; pp. 886–893.
28. Lin, C.; Xu, G.L.; Cao, Y.J.; Liang, C.H.; Li, Y. Improved contour detection model with spatial summation properties based on nonclassical receptive field. *J. Electron. Imaging* **2016**, *25*, 043018. [CrossRef]
29. Suh, H.K.; Hofstee, J.W.; De Ijsselmui, N.J.; Henten, E.V. Sugar beet and volunteer potato classification using Bag-of-Visual Words model, Scale-Invariant Feature Transform, or Speeded Up Robust Feature descriptors and crop row information. *Biosyst. Eng.* **2018**, *166*, 210–226. [CrossRef]
30. Mukherjee, P.; Lall, B. Saliency and KAZE features assisted object segmentation. *Image Vis. Comput.* **2017**, *65*, 82–97. [CrossRef]
31. Song, Z.; Fu, L.; Wu, J.; Liu, Z.; Li, R.; Cui, Y. Kiwifruit detection in field images using Faster R-CNN with VGG16. *IFAC-PapersOnLine* **2019**, *52*, 76–81. [CrossRef]
32. Fu, L.; Feng, Y.; Elkamil, T.; Liu, Z.; Li, R.; Cui, Y. Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks. *Trans. Chin. Soc. Agric. Eng.* **2018**, *34*, 205–211.
33. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding Yolo Series in 2021. *arXiv* **2021**, arXiv:2107.08430.

34. Leng, H.; Tan, M.; Liu, C.; Cubuk, E.; Shi, X.; Cheng, S.; Anguelov, D. PolyLoss: A Polynomial Expansion Perspective of Classification Loss Functions. In Proceedings of the Tenth International Conference on Learning Representations (ICLR), Virtual, 25–29 April 2022.
35. Chen, L. The Multi-Objectrecognition Method of Cluster Kiwifruits Based on Machinevision. Master's Thesis, Northwest A&F University, Yangling, China, 2018.
36. Cui, Y.; Su, S.; Wang, X.; Tian, Y.; Li, P.; Zhang, F. Recognition and Feature Extraction of Kiwifruit in Natural Environment Based on Machine Vision. *Trans. Chin. Soc. Agric. Mach.* **2013**, *44*, 247–252.
37. Fu, L.; Tola, E.; Al-Mallahi, A.; Li, R.; Cui, Y. A novel image proo cessing algorithm to separate linearly clustered kiwifruits. *Biosyst. Eng.* **2019**, *183*, 184–195. [CrossRef]