applied sciences

MDPI

Article

Non-destructive Internal Defect Detection of In-Shell Walnuts by X-ray Technology Based on Improved Faster R-CNN

Hui Zhang, Shuai Ji, Mingming Shao, Houxu Pu and Liping Zhang *

College of Mechanical Engineering, Xinjiang University, Urumqi 830017, China; hui@xju.edu.cn (H.Z.)
* Correspondence: zhlp@xju.edu.cn; Tel.: +86-177-9965-0112

Abstract: The purpose of this study was to achieve non-destructive detection of the internal defects of in-shell walnuts using X-ray radiography technology based on improved Faster R-CNN network model. First, the FPN structure was added to the feature-extraction layer to extract richer image information. Then, ROI Align was used instead of ROI Pooling for eliminating the localization bias problem caused by the quantization operation. Finally, the Softer-NMS module was introduced to the final regression layer with the predicted bounding box for improving the localization accuracy of the candidate boxes. The results of the study indicated that the proposed network model can effectively identify internal defects of in-shell walnuts. Specifically, the discrimination accuracies of the in-shell sound, shriveled, and empty-shell walnuts were 96.14%, 91.72%, and 94.80%, respectively, and the highest overall accuracy was 94.22%. Compared to the original Faster R-CNN network model, the improved Faster R-CNN model achieved an increase of 5.86% in *mAP* and 5.65% in F_1-value. Consequently, the proposed method can be applied for the in-shell walnuts with shriveled and empty-shell defects.

Keywords: walnuts; X-ray images; non-destructive detection; food-quality inspection; improved Faster R-CNN

check for updates

Citation: Zhang, H.; Ji, S.; Shao, M.; Pu, H.; Zhang, L. Non-destructive Internal Defect Detection of In-Shell Walnuts by X-ray Technology Based on Improved Faster R-CNN. *Appl. Sci.* **2023**, *13*, 7311. https://doi.org/10.3390/app13127311

Received: 27 April 2023
Revised: 12 June 2023
Accepted: 18 June 2023
Published: 20 June 2023

Copyright: © 2023 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

Walnut, as a characteristic dried fruit in Xinjiang, is favored by consumers worldwide because of its outstanding taste and high nutritional value [1]. Considering the rising consumer demand for processed walnut products, the internal quality of walnuts with shells plays a crucial role in purchasing decisions. After picking, internal defects, such as protein deterioration, flavor loss, shriveled seed kernel, and empty shell, and external mechanical damage in walnuts can occur during transportation, processing, and storage [2]. These seriously reduced the grade and commodity rate of walnuts and greatly weakened their market competitiveness [3]. At present, the detection methods used to identify the internal quality of walnuts are manual detection, physical means, or stoichiometry. However, the methods are time-consuming, labor-intensive, and inherently destructive. Therefore, a quick, effective, and non-destructive method for determining the internal defects of walnuts is highly desirable.

There are several common non-destructive detection-test methods in the field of industry and agriculture, such as machine vision, hyperspectral imaging, near-infrared spectroscopy, thermography, acoustic vibration method, magnetic resonance imaging (MRI), and X-ray imaging techniques. Machine-vision technology is limited to the detection of surface defects in objects and, thus, cannot effectively detect internal quality [4]. Hyperspectral imaging has disadvantages of more wavebands and large data volume, and the equipment is too expensive [5]. Near-infrared spectroscopy mainly provides a digital data analysis result that cannot visualize the internal defects, and requires high sample preparation and processing [6]. The acoustic vibration method has been highly susceptible to environmental impact, and it may be less applicable in batch-inspection

scenarios [7]. Thermography devices are mainly used to detect the temperature distribution on the surface of the object [8]. MRI is suitable for internal water-content detection, but not for the detection of agricultural products with extremely low water content, such as walnuts [9]. Among these, X-ray radiography is especially interesting in the field of internal quality inspection for agricultural produce, because it has good depths of X-rays and can be easily implemented inline [10]. Zehi et al. [11] also pointed out that X-ray technology is an appropriate alternative to the electron beam and gamma rays, and can be used in the food industry without harmful effects on human health or food quality and safety.

Recently, it has been widely used in research to detect internal disorders in fruit or nuts. Shahin et al. [12] utilized X-ray image technology to inspect the watercore disorder in apples with a higher accuracy of 88%. Van et al. [13] and Tim et al. [14] reported the internal-defects detection of apples and 'Conference' pears, achieving accuracy rates of 90% and 90.2%, respectively. Gao et al. [15] and Zhang et al. [16] also successfully applied it to detect whether hard-shelled walnuts had become hollow and also to detect the size of the walnut kernel. Such X-ray image technology has led to promising results in real-time, non-destructive testing of internal defects in intact walnuts with shells.

Over the past decade, deep-learning has played a very vital role in pattern recognition tasks, mainly because it does not require sophisticated image-processing pipelines and can automatically learn hierarchical features from the data [17]. Among various deep-learning networks, RNN often encounters the phenomenon of vanishing gradient and explosion gradient, and low computational efficiency [18]. The DBN model cannot specify the optimal classification surface between different classes, so the classification accuracy may not be as high as some discriminant model [19]. Nowadays, the application of convolutional neural networks (CNN) has yielded remarkable and significant results in pattern-recognition tasks related to various fields, because it performs excellently in the tasks of image recognition and object detection [20]. Based on the faster regions with CNN features (Faster R-CNN), the region proposal network (FPN), introduced by Ren et al. [21], is a two-stage detection network designed to enhance the speed and accuracy of real-time object detection. Zeng et al. [22] introduced FPN into the Fast R-CNN model to inspect the cotton-packaging defects with the *mAP* value increased by 9.08% compared with the original network. On the basis of Faster RCNN-FPN, Xia et al. [23] replaced ROI Pooling with ROI Align to detect polarizer surface defects, achieving an accuracy of up to 95%. Similarly, the suggested model has been successfully applied to aircraft-target detection [24]. The above results indicated that the improved Faster R-CNN model can be effectively applied in the industrial field. Notably, the improved Faster R-CNN algorithm has also been successfully proved in the pattern-recognition tasks of food and agriculture. Li et al. [25] optimized the anchor frame parameters in the Faster R-CNN model for apples in their natural environment with the average recognition rate of 97.6%. Chen et al. [26] utilized this model for the detection and recognition of Camellia oleifera fruit on trees to obtain superior results. Yan et al. [27] applied the improved Faster R-CNN model to classify 11 kinds of Rosa roxburghii fruits with 92.01% accuracy. However, there are no known studies that use the CNN-based Faster R-CNN for identifying pattern-recognition tasks in nuts, so far. Thus, it is reasonable to combine X-ray images with an improved Faster R-CNN network model for identifying internal defects of in-shell walnuts.

The specific objectives of this study were to: (1) examine the ability to inspect the internal defects in walnuts based on X-ray image technology associated with the improved Faster R-CNN model; (2) compare the performance of the object-detection algorithms based on deep-learning technology and select the best algorithm to build a model; and (3) visually demonstrate the effectiveness of the improved Faster R-CNN model for identifying the internal defects of the in-shell walnuts.

These results will help provide a theoretical support for developing non-destructive testing equipment for the internal quality of walnuts. Also, the method proposed by this study can provide a research strategy for non-destructive detection of the internal quality of other nuts.

2. Materials and Methods

2.1. Sample Preparation

A hand-harvested crop of "Wen 185" walnuts, a popular and distinctive type, was taken in September 2021 from an orchard in Ye City, Xinjiang, China (36°35′ N, 76°12′ E). The examples of walnut with no obvious exterior damage were randomly selected as experimental objects. Then, the remaining nuts were immediately kept in refrigerated storage at a temperature of 2–5 °C for further testing.

Referring to the national standard of "Quality Grade of Walnut", the samples were broken and classified based on visual assessment, by the same skilled fruit farmer. Then, a batch of walnuts was divided into three types, specifically, sound walnuts, shriveled walnuts, and empty-shell walnuts. As shown in Figure 1a, the proportion of walnut kernel area in sound walnut samples is large and the gap between the walnut shell and the kernel is small. The area of dried walnut kernel is relatively small and there is an obvious gap between the walnut shell and the kernel for the shriveled walnuts (Figure 1b). However, the difference between the empty-shell walnut and the other two groups of images is significant, and the nucleolar shape can hardly be seen in the X-ray diagram (Figure 1c).



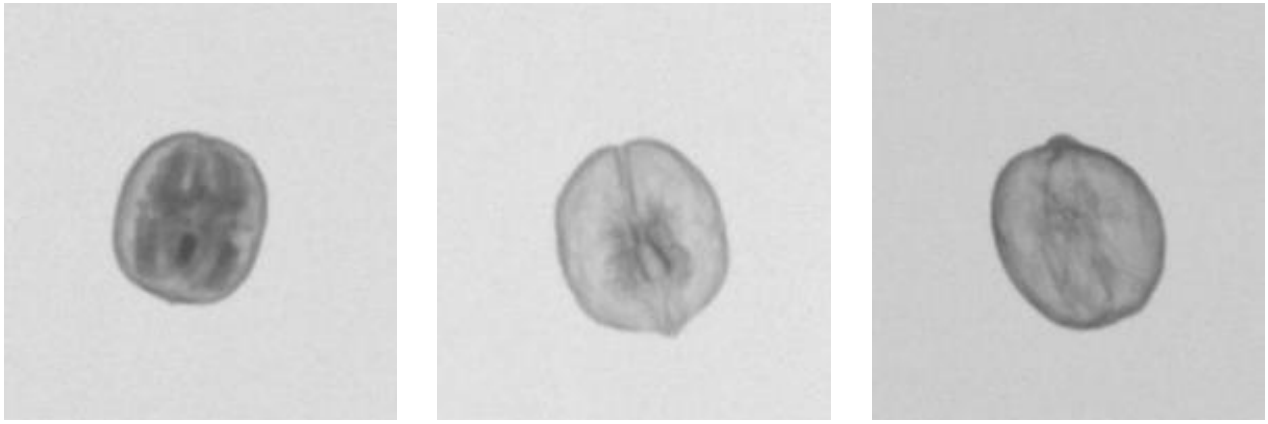
(**a**) (**b**) (**c**)

Figure 1. Examples of the (**a**) sound, (**b**) shriveled, and (**c**) empty-shell in walnut samples.

2.2. Walnut X-ray Images Acquisition

X-ray images of the walnut samples were obtained by using an X-ray radiography setup system (Techik Instrument Co., Ltd., Shanghai, China), mainly comprised of an HVC80804 X-ray source, a TK2-B-410-G04 linear-array detector, and a personal computer. In preliminary testing, it was determined that the tube voltage was to be set at 50 kV and the tube current was to be set at 6 mA, so that the best walnut X-ray images could be obtained. The number of walnuts acquired in each X-ray image was determined according to size and determined in partnership between the X-ray source and detector. There was a detection area of 410 mm^2, so the number of walnuts in each image varied from five to eight. When the walnut samples were transported to the X-ray inspection system via conveyor belt, the X-ray source generated X-rays. Due to the uneven thickness of walnuts and different degree of absorption of X-rays by each part, there were differences in the intensity of the X-rays after penetration. After receiving the X-rays with uneven strength and the figures subjected to A/D conversion, the information was transferred to the computer, which, in turn, generated X-ray images with different grayscale values. The X-ray images of the sound, shriveled, and empty-shell walnut samples are shown in Figure 2. In the experiment, a total of 3845 walnut samples were detected, in order to obtain 1000 X-ray images, which included 1327 sound walnuts, 1283 shriveled walnuts, and 1235 empty-shell walnuts. In order to enhance the generalization ability of the model using deep-learning algorithms, data-augmentation techniques are frequently employed. In this research, the operations such as flip (up–down and left–right), mirror, and brightness for walnut images were carried out to raise the proposed model's performance outcomes. Then, the number of X-ray images was increased to 4000 images. Of these, 2800 walnut X-ray images (70%) were selected as the training set for establishing the discrimination model, and the other

1200 images (30%) were set as the testing set for evaluating the discrimination effect of the constructed model.



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

Figure 2. Examples of X-ray images of the (**a**) sound, (**b**) shriveled, and (**c**) empty-shell in walnuts.

2.3. The Basic Framework of the Faster R-CNN Network

The Faster R-CNN model consists of four parts: the main feature-extraction network (backbone), the region proposal network (RPN), the pooling layer, and the detector (classification and regression layer). First, the image features were extracted by backbone and then the extracted-feature maps were inputted to generate a series of candidate boxes using RPN. By combining feature maps and candidate boxes, the feature candidate boxes were extracted from the images. Finally, the category of candidate boxes was identified by the classification and regression layers, and the specific position of the prediction box was obtained.

Although Faster R-CNN has the advantages of high detection accuracy and strong robustness, it also has some shortcomings. The Faster R-CNN network only uses the last layer of the feature-extraction network for prediction, and when extracting features from the original walnut image requires multiple convolutions and pooling. These can easily cause the loss of target defect information in the image, resulting in missed and false detection. Additionally, two quantitative roundings of ROI Pooling will also lead to the loss of target information in the feature maps, which decreases the classification accuracy of in-shell walnuts with internal defects. In order to solve the problem of lower discrimination accuracy resulting from the loss of image information during the object-detection process, an improved Faster R-CNN network was proposed as follows.

2.4. Optimization Method of Fast R-CNN Model

2.4.1. Feature Fusion Based on FPN Structure

As shown in Figure 3, FPN structure was used to perform feature fusion in three forms, i.e., bottom–up, top–down, and lateral connection, in the feature-extraction layer. Among these, bottom–up represented the feed-forward process of the ResNet50 network in the backbone. It divided the extracted feature maps into five levels C1-C5 according to the specified size and channel numbers. In the top–down process, the spatial size of the deep feature maps was expanded by utilizing the bilinear interpolation up-sampling method to obtain a feature map with the same size as the previous level. The horizontal connection included two steps: (1) the feature maps of C2–C5 level adopted 1×1 convolution operation to alter the channel numbers of the feature maps and to increase the nonlinear features of the image, while the spatial size of the feature maps will not change; (2) two levels with the same dimension at the corresponding pixel positions were added to obtain the fused feature maps. 3×3 convolution operation was applied for removing the phenomenon of aliasing resulting from up-sampling and, then, the enhanced feature maps p2–p5 were

obtained. Considering that feature maps with high resolution can reduce detection speed, the C1 layer would not be fused.

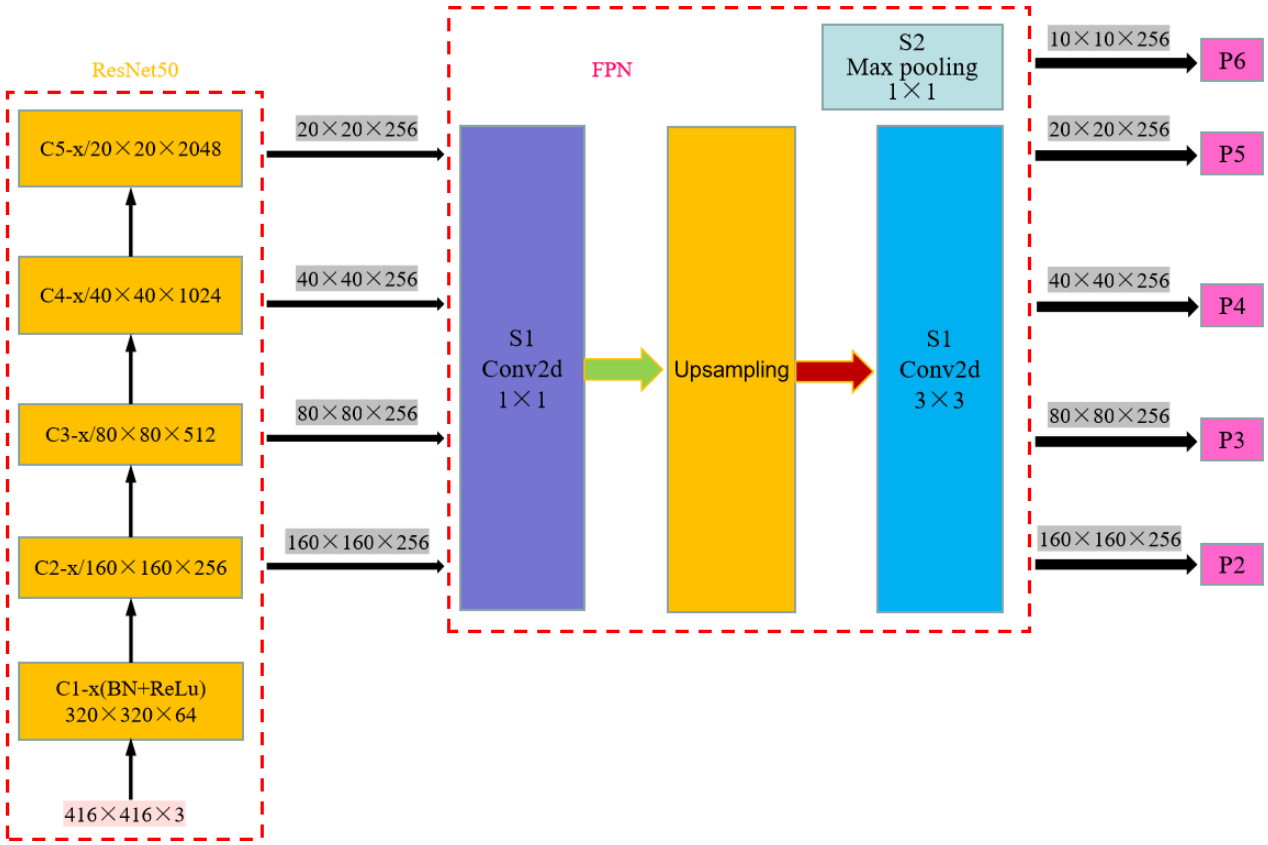


Figure 3. Diagram of feature-fusion structure based on FPN structure.

In the original Faster R-CNN model, the input of RPN is the last layer in the backbone feature map, from which only a single-scale candidate frame can be obtained. In this study, FPN structure was used to input feature map p6 obtained from the enhanced feature map p2–p5 and feature map in the C5 level after maximum pooling into the RPN. Accordingly, in the RPN, the feature maps p2–p6 produced a sequence of anchor boxes with various sizes and aspect ratios. For p2–p6, the corresponding anchor areas were 32 × 32, 64 × 64, 128 × 128, 256 × 256, and 512 × 512, respectively. To cover the detection target of any size in the images, each feature map had three scales at each pixel position, that is 1:2, 1:1, and 2:1. Because the feature map inputted from the original RPN had only one scale, the method of combining shallow and deep features in the feature-extraction stage was proposed. This can more accurately obtain the information of internal defects in walnuts, improving the precision and accuracy of target recognition.

2.4.2. ROI Align

In the original Faster R-CNN model, ROI Pooling was employed to map the ROI area of the input image to the equivalent place of the feature maps. ROI was created through the selection and offset correction of area schemes with various sizes and proportions. The sizes are various and contain floating-point numbers. Hence, it is necessary to carry out a quantitative rounding operation to remove floating-point numbers. In addition, when inputting feature maps into the fully connected layer, it is necessary to adjust the feature maps to a uniform size. So, when the ROI is mapped to the matching spot on the feature maps, the quantified feature maps need to be scaled to a fixed size. In the course of the ROI pooling operation, there are two quantization rounding operations. However, there is

a certain loss of information in ROI after two quantization rounding operations, causing information to not match ROI and extracted features, thus, affecting detection accuracy.

In order to improve the recognition accuracy of the model, ROI Pooling in the original Faster R-CNN model was replaced by ROI Align. The principle of ROI Align was shown in Figure 4. Compared with ROI Pooling, ROI Align not only eliminated the quantization operation and kept all floating-point numbers, but also calculated the precise values of multiple sampling points by using the method of bilinear interpolation. By doing this, the final value can be obtained by gathering the highest or average values of multiple sampling points. In this process, the image information is not lost, and the image characteristics of the original area are preserved as much as possible, thus, improving the detection accuracy of the whole network.

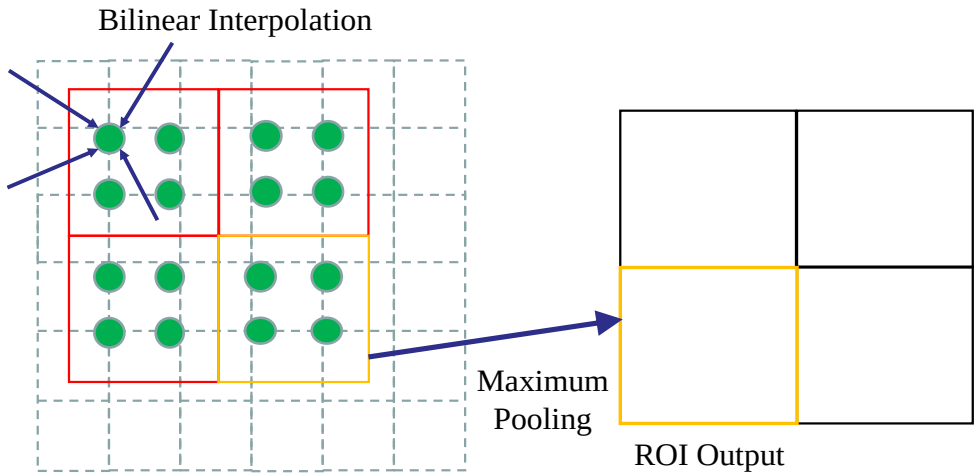


Figure 4. Principle of ROI Pooling.

2.4.3. Softer-NMS

In the target-detection process, the traditional non-maximum suppression algorithm (NMS) was used to extract candidate boxes with the highest confidence and suppress candidate boxes with low confidence. For the Faster R-CNN network, a large number of predicted candidate boxes were generated in the RPN network, many of which would be duplicated and located on the same target. In order to remove these duplicate candidate boxes, NMS was applied to obtain true candidate boxes. However, if an object appeared in the overlapping area of another object, that is, when two candidate boxes were close, then the candidate frame with the lower score would be deleted. That caused the failure of detecting the object and diminished the overall algorithm's average detection accuracy. Surprisingly, the application of Softer-NMS can select the candidate boxes more accurately by ranking it according to the confidence score of prediction candidates. In addition, Softer-NMS can perform a weighted average for the candidate boxes within the predicted labeled variance range, increasing the prediction confidence of the boundary boxes with high position reliability. Hereto, the architecture of the improved Faster R-CNN for in-shell walnuts with internal defects was shown in Figure 5.

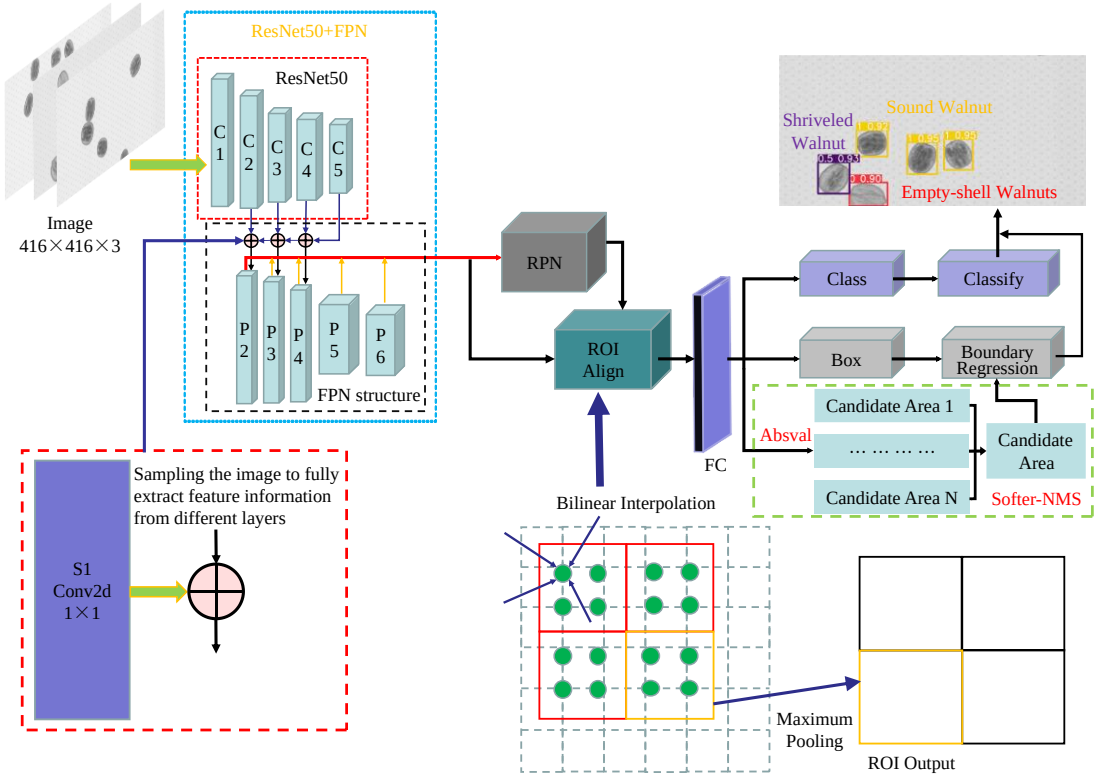


Figure 5. Architecture of the improved Faster R-CNN network.

2.5. Training Platform

The process was based around a personal computer with some specific features, namely, an Intel Core i5-8500 CPU, 3.5 GHz, 16 GB video memory and running memory, and an ASUS RTX2060 GPU. PyTorch, a deep-learning framework, was used to provide both training and testing environments under the Windows 10 operating system. Python3.8, cuda11.1, cudnn10.2, and other required libraries were utilized to train and test the target-recognition model for walnut samples.

In the training stage, the hyperparameters were set to the values recommended by Ren et al. [21]. The hyperparameters were set as follows: the batch size was configured as 4, while the momentum factor was assigned a value of 0.9 for avoiding the memory limitation of GPU. The total amount of training epochs was set as 200. With the stochastic gradient descent (SGD) adopted, the learning rate was set to 0.01 and the weight of the model was updated every 4 epochs with the attenuation coefficient of 0.0001. The confidence threshold and intersection over union (IOU) threshold were all set to 0.5. After training, the weight file of the constructed model was saved, and the testing set was utilized to evaluate the discrimination effect of the model. The ultimate output of the network consists of prediction boxes of the three classes of walnut samples location and the probability of belonging to a particular category.

2.6. Evaluation Indicators of the Model

In this study, the discrimination performance of the constructed model was evaluated by the confusion matrix, including precision, recall, F_1-value, and mean average precision (*mAP*), which were calculated as follows [28]:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F_1 = \frac{2PR}{P+R} \tag{3}$$

$$mAP = \frac{\sum_{k=1}^{n} PR}{N} \tag{4}$$

where *TP* and *FN* are the number of positive samples that are classified as positive and negative, respectively; *TN* and *FP* are the number of negative samples that are classified as negative and positive, respectively; *N* represents the number of walnut sample categories; *n* represents the IOU thresholds for the quantity, and *k* represents the IOU threshold. If one kind of sample was determined as positive, the other two kinds of samples were symbolized as negative. For example, when the empty-shell walnut sample was positive, the shriveled and sound walnut samples were defined as negative. In order to comprehensively evaluate the stability and accuracy of the model, the 10-fold cross-validation method was applied [29]. That is, the data were split into 10 equal portions and one part was used for validation in each iteration, while the remaining 9 parts were used as the training model. In this way, to represent the discrimination performance of the created model, we computed the average value of 10 recordings for confusion matrix results.

3. Results

3.1. Construction of the Fast R-CNN Model

Currently, deep-learning-based object-detection algorithms mainly include single-stage object-detection algorithms represented by YOLOv3 and YOLOv5, as well as two-stage object-detection algorithms represented by Faster R-CNN. In this study, three discrimination models based on the YOLOv3, YOLOv5, and Faster R-CNN algorithms were, respectively, established to analyze and compare the recognition performance for walnut samples with different internal defects.

The *mAP* curves of three models are pictured together in Figure 6, among which, the YOLOV5 model fitted rapidly in the first 20 epochs, but there were still small fluctuations in the model fitting before 60 epochs and, basically, tending to be stable after 140 epochs of training. The YOLOv3 model had the worst fitting performance result in that it was gradually stable after 140 epochs and there were significant fluctuations within 140 epochs. In contrast, the Faster R-CNN model quickly tended to fit within 10 epochs and, subsequently, remained stable always, indicating that the Faster R-CNN model had a fast fitting speed and strong robustness.

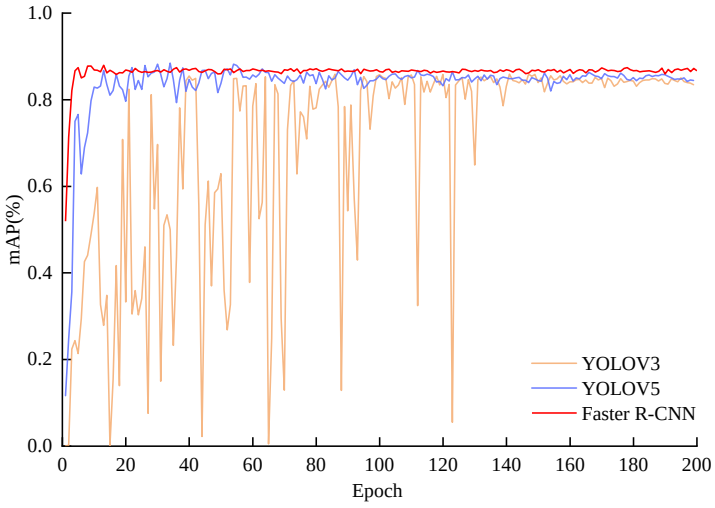


Figure 6. The *mAP* curves of the YOLOv3 model, the YOLOv5 model, and the Faster R-CNN model.

After training, the specific discrimination results of YOLOv3, YOLOv5, and Faster R-CNN models for walnut samples were shown in Table 1. It can be observed that the Faster R-CNN model exhibits a better discriminative result and surpasses the other two models. The overall identification precision, recall, *mAP*, and F_1-value were 89.47%, 86.47%, 87.94%, and 89.71%, respectively. In the aspect of training time, the Faster R-CNN model required a slightly longer time because it was affected by the size of the model framework. Nevertheless, the Faster R-CNN model had the advantages of a fast fitting speed, strong robustness, and a good classification performance. Given that the discrimination accuracy for walnut quality was less than 90%, further improvement of Faster R-CNN is desirable.

Table 1. The training results of YOLOv3, YOLOv5, and Faster R-CNN models for identifying internal defects in walnuts.

Model	Accuracy (%)	Recall (%)	F_1-Value (%)	*mAP* (%)	Total Training Time (h)	Training Time of Single Image (ms)
YOLOv3	86.14	79.02	82.43	85.87	10.27	14
YOLOv5	87.32	83.25	85.24	88.43	9.73	8
Faster R-CNN	89.47	86.47	87.94	89.71	11.38	10

3.2. The Training Results of the Improved Faster R-CNN Model

The training loss curves of original and improved Faster R-CNN model were shown in Figure 7. In the case of the improved Faster R-CNN model, the training loss experienced a rapid decline to approximately 0.0075 within the initial 20 epochs. Subsequently, the loss continued to decrease gradually and, eventually, reached a stable value of 0.002. While the original Faster R-CNN model still kept an increase in loss values in the first 20 epochs, there was a clear increase during the later iteration process. The result indicated that the improved Faster R-CNN has a faster fitting speed and better robustness.

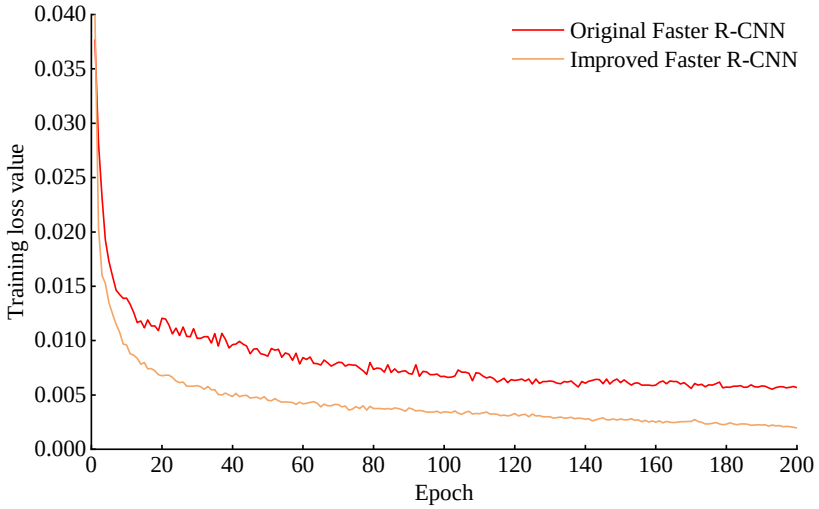


Figure 7. Training loss curve of the Faster R-CNN model showing improvement by comparing before and after.

Based on the method of Gao et al. [3], the impact of each improvement point on the discrimination performance of the improved Faster R-CNN model has been compared. As shown in Table 2, the *mAP* of the improved Faster R-CNN model was increased by 1.62% after adding the FPN structure for image-feature fusion compared with the original model. Subsequently, by replacing ROI pooling with ROI Align to eliminate the quantitative rounding operation, the *mAP* was increased by 1.73%. On this basis, Softer-NMS was conducted to perform weighted averaging on candidate regions in the predicted layer, and the *mAP* was further improved by 2.51%. By doing this, the *mAP* and F_1-value

of the improved Faster R-CNN model reached 95.57% and 93.59%, respectively. The results indicate that compared to the original Faster R-CNN model, the improved Faster R-CNN model exhibits a significantly enhanced ability to discriminate in-shell walnuts with different internal defects.

Table 2. Performance comparison of different improvement points.

Model Framework	*mAP* (%)	F_1-Value (%)
Faster R-CNN	89.71	87.94
Faster R-CNN + FPN	91.33	89.07
Faster R-CNN +FPN + ROI Align	93.06	91.43
Faster R-CNN + FPN + ROI Align + Softer-NMS	95.57	93.59

3.3. Performance Analysis of the Improved Faster R-CNN Model

In order to evaluate the recognition capability of the improved Faster R-CNN model for internal defects in walnuts more intuitively, the discrimination effect of the model on 155 images of the testing set were further analyzed with a confusion matrix. There are a total of 525 walnut samples in the test image set, among which there are 525 sound walnuts is 207, 145 shriveled walnuts, and 173 empty-shell walnuts. As shown in Table 3, the discrimination accuracy of the constructed model for shriveled walnuts is the lowest (91.72%). There are 12 misjudgments among 145 shriveled walnut samples, of which 8 shriveled walnuts are misjudged as empty-shell samples, and 4 shriveled walnuts are misjudged as sound samples. For the empty-shell walnuts, 9 samples are wrongly discriminated with an error rate of 5.2%, obtaining a slightly higher discrimination accuracy of 94.8%. This may be because the feature information of sample images is not complete, caused by various shooting angles using X-ray radiography. Additionally, different shriveled degrees occurred in in-shell walnuts, and it is likely that this is also the reason why the shriveled walnuts and empty-shell walnuts are wrongly discriminated against each other. Significantly, it is found that the improved Faster R-CNN model obtains a discrimination accuracy of 96.14% for sound walnut samples, and only 8 sound walnuts are wrongly discriminated with an error rate of 3.9%. In general, 496 walnuts of 525 discrimination samples are correctly recognized and the overall discrimination accuracy reaches 94.22%. It is believed that the improved Faster R-CNN model proposed by this work can effectively discriminate the internal defects in intact walnuts with shells.

Table 3. The discrimination results of the improved Faster R-CNN model.

Actual Class	Predicted Class			Discrimination Accuracy (%)	Overall Accuracy (%)
	Empty-Shell Walnut	Shriveled Walnut	Sound Walnut		
Empty-shell walnut (173)	164	8	1	94.80%	94.22%
Shriveled walnut (145)	8	133	4	91.72%	
Sound walnut (207)	2	6	199	96.14%	

Examples of the recognition results of the improved Faster R-CNN model for sound, shriveled, and empty-shell walnut samples are depicted in Figure 8. The red, purple, and yellow boxes were used for labeling the empty-shell, shriveled, and sound walnuts. As shown, the confidence levels for individual walnuts with complete image information are between 92% and 95%, and the overall confidence level is relatively close to the *mAP* value from the testing set. Although the confidence level of a single empty-shell walnut sample with incomplete information is lower than that of the other two types of walnut samples, there are not wrong detection or missed detection caused by incomplete information of the walnut images.

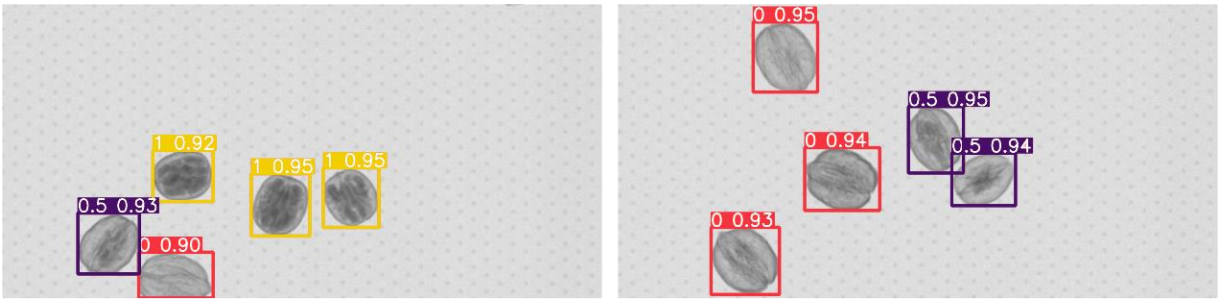


Figure 8. Examples of visualization results of the improved Faster RCNN model.

4. Discussion

In this paper, the following improvements are conducted to address the shortcomings of the Faster R-CNN model. Firstly, in the feature-extraction part the ResNet50 network is used to replace the VGG16 network, and is combined with the FPN structure to enrich the information of the feature map and improve the detection performance of the model. In addition, ROI Align is used to replace ROI Pooling to solve the problems of false and missed detection caused by the quantization operation. Finally, the NMS part applies the Softer-NMS algorithm to improve the detection accuracy of the model by increasing the classification confidence of the bounding candidate frames. By doing this, the improved Faster R-CNN model can meet the expected expectations and provide a theoretical basis for the nondestructive detection equipment for internal quality of in-shell walnuts.

Walnuts were randomly placed on the conveyor belt without considering the influence of posture on image quality. In the future research, we will study X-ray images of walnuts with different angles and design a transfer device to improve inspection accuracy. Only the thin-skinned walnuts from Yecheng were selected as the object in this paper, the applicability to the detection for internal defects of thick-skinned walnuts using X-ray technology needs to be further verified. Finally, whether other internal defects, such as browning and mildew, within in-shell walnuts can effectively be detected with the proposed method above, further researches may be needed.

5. Conclusions

In this study, X-ray radiography technology was employed for the non-destructive detection of in-shell walnuts with shriveled and empty-shell defects. After comparison of three target detection algorithms, the Faster R-CNN model was found to be more appropriate than the YOLOv3 model and the YOLOv5 model for the identification of the internal defects in hard-shell walnuts. In the improved Faster R-CNN network architecture, the FPN structure used for feature fusion was firstly introduced to enrich the feature map information of internal defects in walnuts. To solve the problems of false and missed detection caused by quantization operation, ROI Pooling module was replaced with ROI Align module. To increase the prediction confidence of the bounding boxes and, thereby, improve the discrimination accuracy of the network, the Softer-NMS structure was inputted into the final regression layer with the predicted bounding box. Contrasted with original Faster R-CNN network model, the *mAP* and F_1-value of the improved Faster R-CNN model were increased by 5.86% and 5.65%, respectively. The detection results of the testing set indicated that the proposed improved Faster R-CNN model can effectively realize the recognition of the internal defects of in-shell walnuts. The discrimination accuracy of the in-shell sound, shriveled, and empty-shell walnuts were 96.14%, 91.72%, and 94.80%, respectively, and the highest overall accuracy can reach 94.22%. These results indicated that X-ray image technology associated with the improved Faster R-CNN model can be effectively applied to the recognition tasks of the in-shell walnuts with shriveled and empty-shell defects.

Author Contributions: Conceptualization, methodology, data curation, writing—review and editing, and funding acquisition, H.Z.; conceptualization, supervision, and writing—review and editing, S.J.; investigation, and validation, M.S.; software, and visualization, H.P.; and methodology, data curation, and funding acquisition, L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Start Fund of Scientific and Research in Xinjiang University (grant number 620320039) and Key R&D Special Project of Xinjiang Uygur Autonomous Region (grant number 2022B02028-4).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are very grateful for all constructive comments that helped us improve the original version of the manuscript.

Conflicts of Interest: The authors declare that we have no known competing financial interest or personal relationship that influenced, or could have appeared to influence, the work reported in this paper.

References

1. Xi, J.; Jiang, Z. Analysis of the current situation of walnut industry development in Xinjiang region. *Mod. Hortic.* **2023**, *46*, 38–40. [CrossRef]
2. Dong, C.L. Thoughts on high-quality development of walnut industry in Chuxiong. *Green Sci. Technol.* **2021**, *23*, 110–112. [CrossRef]
3. Gao, T.; Zhang, S.; Sun, H.; Ren, R. Mass detection of walnut based on X-ray imaging technology. *J. Food Process Eng.* **2022**, *45*, e14034. [CrossRef]
4. Ren, Z.; Fang, F.; Yan, N.; Wu, Y. State of the art in defect detection based on machine vision. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2022**, *9*, 661–691. [CrossRef]
5. Ghamisi, P.; Plaza, J.; Chen, Y.; Li, J.; Plaza, A.J. Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–32. [CrossRef]
6. Larson, J.E.; Perkins-Veazie, P.; Ma, G.; Kon, T.M. Quantification and prediction with near infrared spectroscopy of carbohydrates throughout apple fruit development. *Horticulturae* **2023**, *9*, 279. [CrossRef]
7. Zhang, W.; Lv, Z.; Xiong, S. Nondestructive quality evaluation of agro-products using acoustic vibration methods—A review. *Crit. Rev. Food Sci. Nutr.* **2018**, *58*, 2386–2397. [CrossRef]
8. Baranowski, P.; Lipecki, J.; Mazurek, W.; Walczak, R.T. Detection of watercore in 'Gloster'apples using thermography. *Postharvest Biol. Technol.* **2008**, *47*, 358–366. [CrossRef]
9. Razavi, M.S.; Asghari, A.; Azadbakh, M.; Shamsabadi, H.-A. Analyzing the pear bruised volume after static loading by Magnetic Resonance Imaging (MRI). *Sci. Hortic.* **2018**, *229*, 33–39. [CrossRef]
10. Kotwaliwale, N.; Singh, K.; Kalne, A.; Jha, S.N.; Seth, N.; Kar, A. X-ray imaging methods for internal quality evaluation of agricultural produce. *J. Food Sci. Technol* **2014**, *51*, 1–15. [CrossRef]
11. Zehi, Z.B.; Afshari, A.; Noori, S.; Jannat, B.; Hashemi, M. The effects of X-ray irradiation on safety and nutritional value of food: A systematic review article. *Curr. Pharm. Biotechnol.* **2020**, *21*, 919–926. [CrossRef]
12. Shahin, M.A.; Tollner, E.W.; McClendon, R.W. AE—Automation and Emerging Technologies. *J. Agric. Eng. Res.* **2001**, *79*, 265–274. [CrossRef]
13. Van Dael, M.; Verboven, P.; Zanella, A.; Sijbers, J.; Nicolai, B. Combination of shape and X-ray inspection for apple internal quality control: In silico analysis of the methodology based on X-ray computed tomography. *Postharvest Biol. Technol.* **2019**, *148*, 218–227. [CrossRef]
14. Van de Looverbosch, T.; Raeymaekers, E.; Verboven, P.; Sijbers, J.; Nicolai, B. Non-destructive internal disorder detection of Conference pears by semantic segmentation of X-ray CT scans using deep learning. *Expert Syst. Appl.* **2021**, *176*, 114925. [CrossRef]
15. Gao, T.Y.; Zhang, S.J.; Sun, P.; Zhao, H.M.; Sun, H.X.; Niu, R.M. Variety Classification of walnut based on X-ray image. *Food Sci. Technol.* **2020**, *45*, 284–288. [CrossRef]
16. Zhang, S.; Gao, T.; Ren, R.; Sun, H. Detection of Walnut Internal Quality Based on X-ray Imaging Technology and Convolution Neural Network. *Trans. Chin. Soc. Agric. Mach* **2022**, *53*, 383–388.
17. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
18. Wang, B.; Kong, W.; Zhao, P. An air quality forecasting model based on improved convnet and RNN. *Soft Comput.* **2021**, *25*, 9209–9218. [CrossRef]

19. Lee, H.; Grosse, R.; Ranganath, R.; Ng, A.Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 609–616.
20. Jogin, M.; Madhulika, M.; Divya, G.; Meghana, R.; Apoorva, S. Feature extraction using convolution neural networks (CNN) and deep learning. In Proceedings of the 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 18–19 May 2018; pp. 2319–2323.
21. Ren, S.Q.; He, K.M.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal.* **2017**, *39*, 1137–1149. [CrossRef]
22. Zeng, Y.X.; Lu, H.C.; Lv, H.F. Research on cotton packaging defect detection method based on improved Faster R-CNN. *Electron. Meas. Instrum.* **2022**, *36*, 179–186. [CrossRef]
23. Xia, Y.; Xiao, J.Q.; Weng, Y.S. Surface defect detection of polarizer based on improved Faster R-CNN. *Opt. Tech.* **2021**, *47*, 695–702. [CrossRef]
24. Zhu, W.T.; Lan, X.C.; Luo, H.L.; Yue, B.; Wang, Y. Remote sensing aircraft target detection based on improved Fasterr-CNN. *Comput. Sci.* **2022**, *49*, 378–383.
25. Li, L.S.; Zeng, P.P. Apple target detection based on improved Faster-RCNN framework of deep learning. *Mach. Des. Res.* **2019**, *35*, 24–27. [CrossRef]
26. Chen, B.; Rao, H.H.; Wang, Y.L.; Li, Q.S.; Wang, B.Y.; Liu, M.H. Study on detection of camellia fruit in natural environment based on Faster-RCNN. *Acta Agric. Jiangxi* **2021**, *33*, 67–70. [CrossRef]
27. Yan, J.W.; Zhao, Y.; Zhang, L.W.; Su, X.D.; Liu, H.Y.; Zhang, F.G.; Fan, W.G.; He, L. Recognition of Rosa roxbunghii in natural environment based on improved Faster-RCNN. *Trans. Chin. Soc. Agric. Eng.* **2019**, *35*, 143–150.
28. Wei, R.; Pei, Y.K.; Jiang, Y.C.; Zhou, P.Z.; Zhang, Y.F. Detection of cherry defects based on improved Faster R-CNN model. *Food Mach.* **2021**, *37*, 98–105+201. [CrossRef]
29. Saidi, L.; Ben Ali, J.; Fnaiech, F. Application of higher order spectral features and support vector machines for bearing faults classification. *ISA Trans.* **2015**, *54*, 193–206. [CrossRef]

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.