

Article

Apple Grading Method Design and Implementation for Automatic Grader Based on Improved YOLOv5

Bo Xu ¹, Xiang Cui ¹, Wei Ji ^{1,*}, Hao Yuan ² and Juncheng Wang ¹¹ School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China² School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, China

* Correspondence: jiwei@ujs.edu.cn

Abstract: Apple grading is an essential part of the apple marketing process to achieve high profits. In this paper, an improved YOLOv5 apple grading method is proposed to address the problems of low grading accuracy and slow grading speed in the apple grading process and is experimentally verified by the designed automatic apple grading machine. Firstly, the Mish activation function is used instead of the original YOLOv5 activation function, which allows the apple feature information to flow in the deep network and improves the generalization ability of the model. Secondly, the distance intersection over Union loss function (DIOU_Loss) is used to speed up the border regression rate and improve the model convergence speed. In order to refine the model to focus on apple feature information, a channel attention module (Squeeze Excitation) was added to the YOLOv5 backbone network to enhance information propagation between features and improve the model's ability to extract fruit features. The experimental results show that the improved YOLOv5 algorithm achieves an average accuracy of 90.6% for apple grading under the test set, which is 14.8%, 11.1%, and 3.7% better than the SSD, YOLOv4, and YOLOv5s models, respectively, with a real-time grading frame rate of 59.63 FPS. Finally, the improved YOLOv5 apple grading algorithm is experimentally validated on the developed apple auto-grader. The improved YOLOv5 apple grading algorithm was experimentally validated on the developed apple auto grader. The experimental results showed that the grading accuracy of the automatic apple grader reached 93%, and the grading speed was four apples/sec, indicating that this method has a high grading speed and accuracy for apples, which is of practical significance for advancing the development of automatic apple grading.



Citation: Xu, B.; Cui, X.; Ji, W.; Yuan, H.; Wang, J. Apple Grading Method Design and Implementation for Automatic Grader Based on Improved YOLOv5. *Agriculture* **2023**, *13*, 124. <https://doi.org/10.3390/agriculture13010124>

Academic Editor: Xiuliang Jin

Received: 22 December 2022

Revised: 29 December 2022

Accepted: 30 December 2022

Published: 2 January 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/>).

Keywords: apple grader; YOLOv5; attention mechanism SE; DIOU_Loss; mish

1. Introduction

Today, labour on farms and orchards relies heavily on manual labour by skilled farmers, which can lead to increased time and production costs. Smart farming has become a popular concept with the development of precision farming and information technology [1]. China is a major apple-producing country globally, and apple sorting has a high economic application value [2]. With increased economic development, people have higher requirements for fruit quality [3,4]. As a critical element in improving apple quality and liberating orchard labour, apple grading technology is of great significance in increasing the added value of products, improving market competitiveness, and alleviating labour shortages in orchards. Therefore, a high precision and speed grading method is needed for the effective and objective grading of apples.

In the research of fruit grading based on traditional machine learning, Abdullah et al. [5] detected the quality features of poppy peaches by machine learning, the features considered mainly included fruit surface color and fruit shape, and developed automatic machine vision detection software to detect the ripeness grade of poppy peaches by linear discriminant analysis and multilayer neural network. Marchant et al. [6] studied the method of automatic potato detection and grading based on a computer vision system. Moallem et al. [7]

proposed a computer vision-based grading algorithm for golden crown apples where texture and geometric features were extracted from the defective areas. Finally, a support vector machine (SVM), a multilayer perception (Muti-Layer Perception), and a K-Nearest Neighbor classifier were used to classify the apples into first-class, second-class, and out-of-class fruits. Gui et al. [8] proposed a wavelet rectangle-based apple classification method based on apple shape, which classified apples into normal fruit shape, mild deformity, and severe deformity with a classification accuracy of 86.2%, 85.8%, and 90.8%, respectively. In the above machine learning classification methods, preprocessing of images is often required, and the classification relies on single features, which has the problems of poor real-time performance and low robustness.

In the research of fruit grading based on deep learning, Fan et al. [9] used a convolutional neural network (CNN) architecture for apple quality recognition, trained a convolutional neural network, and achieved an accuracy of 96.5% in the test set, designed classification software for CNN-based convolutional neural networks, and used a computer vision module to sort at a rate of 5/s on a four-threaded fruit sorter. The classification accuracy reached 92%. However, the model was large, and the computational efficiency was relatively low. Raikar et al. [10] studied the quality grade of okra and used three deep learning models, AlexNet, GoogLeNet, and ResNet50, to classify okra into four types based on length: small, medium, large and extra large, where the accuracy of the ResNet deep learning model reached over 99%. Luna et al. [11] proposed a deep learning-based method for single tomato defect area detection, implemented through the OpenCV library and Python programming. He collected 1200 tomato images of different qualities using an image capture box and used the images for training VGG16, InceptionV3, and ResNet50 deep learning models, respectively, compared the experimental results and found that VGG16 was the best deep learning model for defect recognition. However, there are still problems of insufficient model optimization and poor real-time performance in the above deep learning model grading methods.

In terms of research on automatic fruit grading equipment, Cubero et al. [12] designed a computer vision-based automated citrus sorting device. The sorting device was deployed on a mobile platform, and the low-power industrial camera image acquisition and powerful lighting system enabled the device to work better in the field. Experiments showed that the sorting device could achieve a sorting speed of up to eight per second. Baigvand et al. [13] developed a machine learning-based fig sorting system, which first uses a feeder and a belt. The figs were first transported under a CCD camera by a feeder and belt conveyor. The figs were classified into five categories by extracting fig characteristics from the pictures taken by the CCD camera, including size, colour, segmentation size and fig centre position, etc. The experiments verified that the grading system was 95% accurate in recognizing the five categories of figs. However, the designed automatic fruit grader tends to be large and more suitable for large assembly line working modes and is not suitable for the needs of small and medium-sized farmers for detection and grading.

Although the above methods have achieved specific results in terms of fruit feature detection and equipment implementation, there are still problems, such as insufficient model optimization and equipment implementation. Based on this, this paper takes red Fuji apples as the research object. It provides an in-depth discussion on the grading detection of apple features and the implementation of automatic apple sorting equipment. An apple grading algorithm based on the improved YOLOv5 is proposed, using the Mish activation function instead of the original Relu activation function to improve the model generalization ability. A loss function (DIou_Loss) is introduced to speed up the rate of edge regression and improve localization accuracy. The attention mechanism squeeze excitation (SE) module is embedded into the backbone feature network to improve the feature extraction ability of the model. Experimental results show that the improved method can improve the model detection without increasing the model training cost. Finally, the automatic apple grader designed based on this paper was experimentally validated, and some conclusions were obtained.

2. Materials and Methods

2.1. Automatic Apple Grader Design

The structure of the automatic apple grader designed in this paper is shown in Figure 1. It consists of Feeding and material handling lifting mechanism, turnover detection conveyor, visual inspection and automatic grading control system, and graded actuators. The design is based on a two-level layout to reduce the space required.

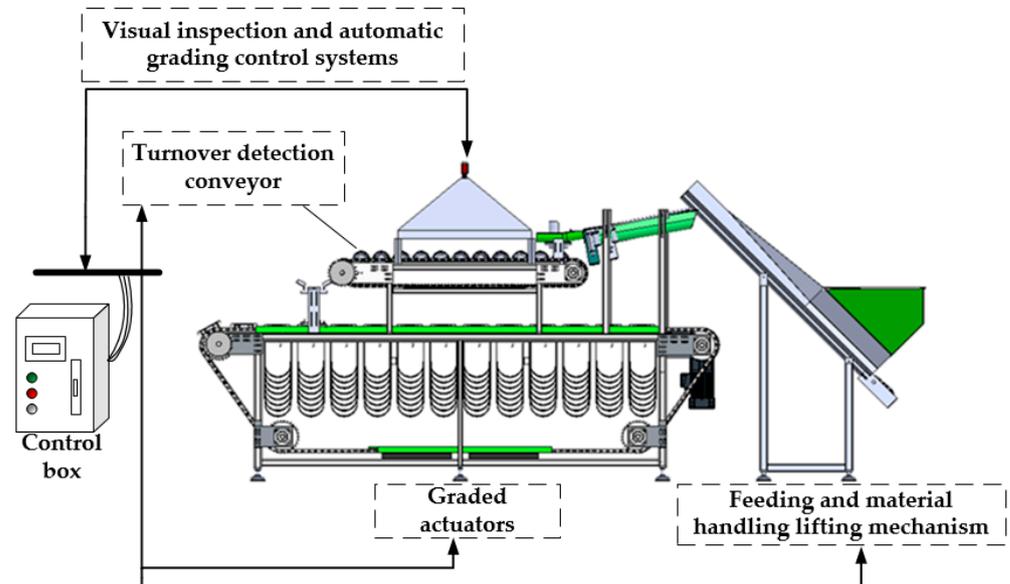


Figure 1. Structure of the automatic apple grader.

- (1) Feeding and material handling lifting mechanism. The Feeding and material handling lifting mechanism is a scraper elevator, as shown in Figure 2. The scraper elevator consists of a funnel-shaped storage tank and a vertical conveyor belt, where the funnel-shaped storage tank includes the back plate of the hopper and the support plate, the three-dimensional conveyor belt includes the guide plate and the curved scraper, and the whole mechanism is placed at an inclination of 45° . The scraper elevator moves the conveyor belt by means of an AC motor driven by a frequency converter, which organizes the disordered apples into an orderly quadruple queue, transporting them from the bottom upwards and conveying them into the Turnover detection conveyor.

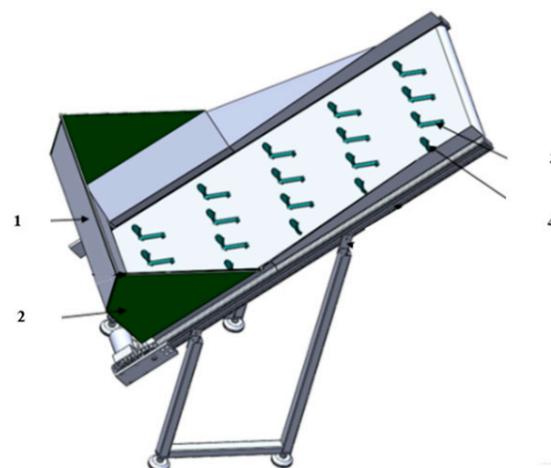


Figure 2. Scraper elevator. (1—hopper, 2—support plate, 3—guide plate, 4—curved scraper).

- (2) Turnover detection conveyor. The Turnover detection conveyor is shown in Figure 3 and consists of sprockets, chains, sponge rollers, and motors. The apples are lifted by the scraper elevator into the turnover detection conveyor. The turnover detection conveyor uses pairs of double-tapered rollers to turn the apples axially, and a CCD industrial camera mounted on top of the lampshade collects images of the tumbling apples several times to obtain complete surface information about the apples in a moving position.

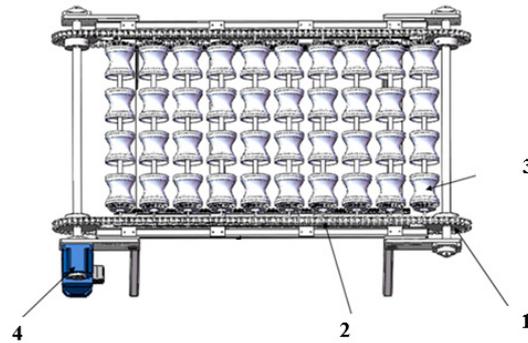


Figure 3. Turnover detection conveyor. (1—sprockets, 2—chains, 3—sponge rollers, 4—motor).

- (3) Visual inspection and automatic grading control system. The visual inspection and automatic grading control system are shown in Figure 4 and consist of a CCD industrial camera and automatic grading control system. The visual inspection and automatic grading control system determines the grading of apples according to the information collected by the CCD industrial camera on the whole surface of the apples and finally sends the grading results to the graded actuators.

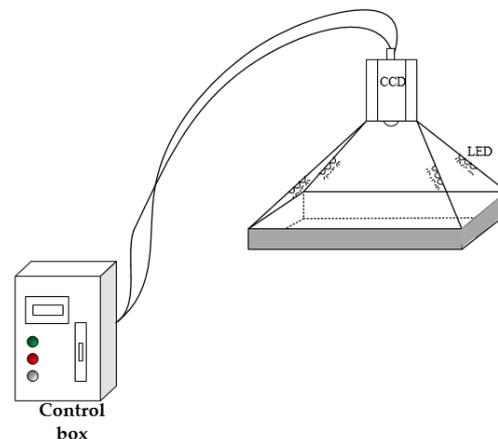


Figure 4. Visual inspection and automatic grading control system structure diagram.

- (4) Graded actuators. The graded actuator is shown in Figure 5 and consists of a Trigger grading mechanism, sprocket chain drive, grading fruit cup, grading channel, and a motor. The grading fruit cup is shown in Figure 5b and consists of a cup body, a drop door, and rollers. The grading actuator receives the grading results from the image detection and automatic grading system and allows the apples to reach the corresponding grade position and then open the cups and fall into the corresponding grade storage bin.

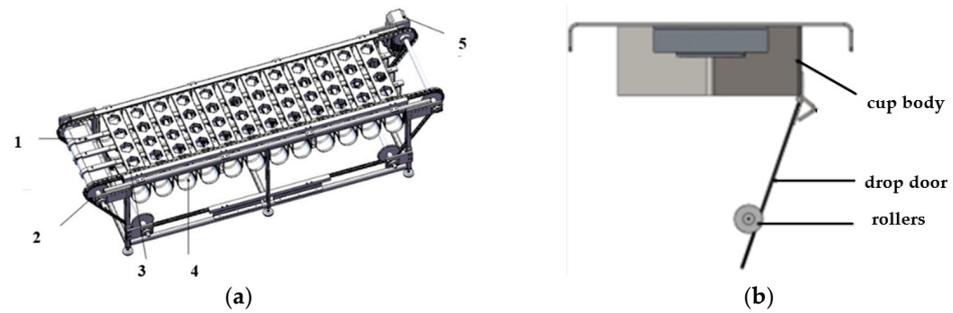


Figure 5. Graded actuators. (a) Grading actuator; (b) Detail of grading fruit cup. (1-Trigger grading mechanism; 2-Sprocket chain drive; 3-Grading fruit cup; 4-Grading channel; 5-Motor).

2.2. Apple Image Acquisition and Data Augmentation

2.2.1. Image Acquisition

The apples used in this dataset are “Fengxian apples” from Xuzhou City, Jiangsu Province, and “Yantai apples” from Yantai City, Shandong Province, which are among the representative brands of red Fuji. The sources of apples include purchases from apple markets and picking from orchards. The image acquisition equipment used in the experiments is a CCD industrial camera, MER2-G. The camera was mounted on a bracket above a lampshade with a 90° angle of view directly above the flip mechanism and a fixed height of 70 cm. The lampshade was illuminated with a LED strip as a fill light source to capture images of apples under diffuse lighting. A final dataset of 2000 apple images was obtained, including grade-1, grade-2, and grade-3 apples. The grade-1 and grade-2 apples were mainly purchased from the market (differentiated by price), while grade-3 apples were marketed in smaller quantities, mainly from orchard picking. The CCD industrial camera uses a GigE interface for data transmission and acquisition with an industrial computer. The image resolution is 1280×1024 , the pixel size is $4.5 \times 4.5 \mu\text{m}$, and the operating temperature range is 0° to 45° . The camera is installed according to the position of the flipping mechanism, the shooting angle is 90° directly above the flipping mechanism, the shooting height is a fixed value of 70 cm, and the image acquisition of the apple is carried out on the flipping mechanism, the image acquisition method and the image acquisition effect are shown in Figure 6.

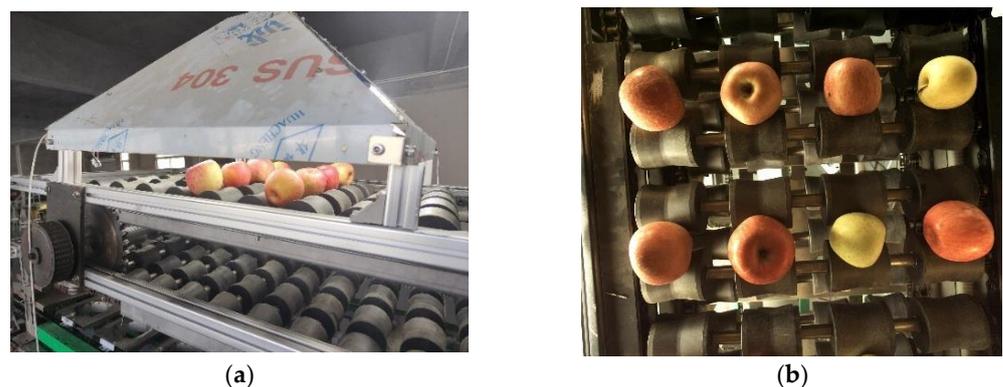


Figure 6. Data set production. (a) Image acquisition devices; (b) Image acquisition.

2.2.2. Apple Grading Criteria

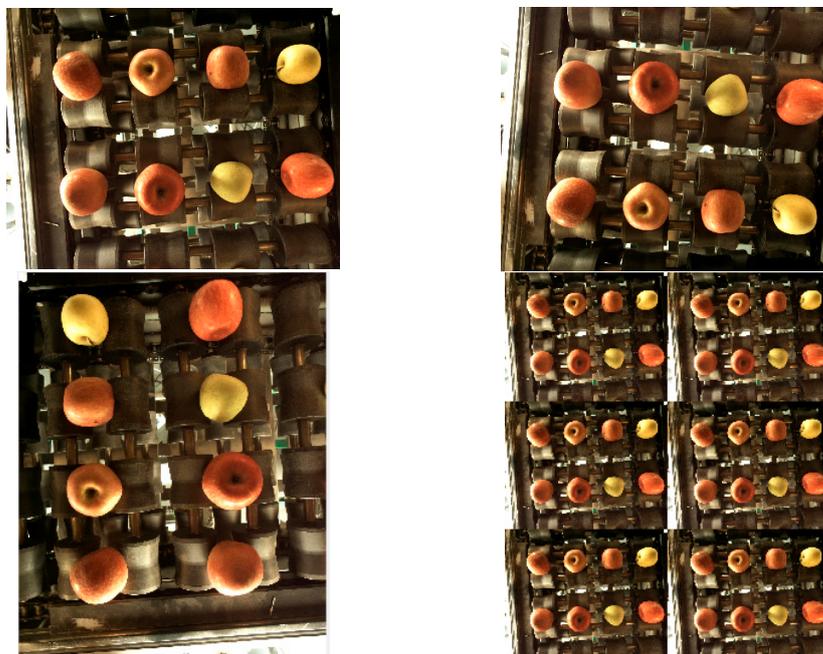
In this paper, based on the Red Fuji GB/T 10651-2008 national standard [14], as shown in Table 1, ripeness, fruit shape, defects, and fruit diameter were selected as grading features to classify Red Fuji apples into 3 grades for this dataset.

Table 1. Red Fuji GB/T 10651-2008 national standard.

Projects	Index	Quality Grade		
		Grade-1	Grade-2	Grade-3
Ripeness		Bright red or dark red	Greenish red	Greenish yellow
Fruit shape		No deformities	No deformities	deformities
defects		NO	NO	Area not exceeding 4 cm ²
diameter(Maximum cross-sectional diameter)/mm		≥70	≥70	≥65

2.2.3. Dataset Annotation and Expansion

LabelImg was used to annotate the apple images, saving the image categories and target rectangle boxes according to the PASCAL VOC dataset format, generating an annotation file in XML format. As the height of the industrial camera is a fixed value, the longest side of the rectangular box calibrated in the dataset is used as the criterion for fruit diameter; the ratio of the long side to the short side of the rectangular box is used as the criterion for fruit shape; apples with poor ripeness and defects are not carefully classified and are judged to be grade-3 apples. The collected apple images were expanded using MATLAB (2019) to make the training model more robust. The expansion methods included horizontal mirroring, vertical mirroring, multi-angle rotation (90.180.270), and image tiling. The expanded dataset is shown in Figure 7. The extended dataset has 6000 images with a uniform image size of 1280 × 1024, with a high number of grade-1 and grade-2 apples, each accounting for 40%, and a low number of grade-3 apples, accounting for 20%. The extended dataset was allocated to the training, test, and validation sets in a ratio of 7:2:1.

**Figure 7.** The expanded dataset.

2.3. Design of Apple Grading Method Based on Improved YOLOv5

YOLOv5 is an algorithm proposed by Glenn-Jocher with high real-time performance in terms of algorithmic efficiency [15,16]. The YOLOv5 network has four main components, which are the input side, the backbone network (backbone), the Neck network part, and the Output part. The YOLO family of algorithms has promising results on open-source datasets, but there is no comprehensive and mature method for grading different state fruits [17].

Therefore, this paper proposes an improved YOLOv5 model structure for apple grading based on the lightweight network YOLOv5s shown in Figure 8. Using the Mish activation function instead of Leaky-ReLU, the distance intersection overUnion (DIOU_Loss) loss function is used at the output of the model. Finally, a simple and efficient channel attention module (Squeeze Excitation) is introduced, which allows the model to focus on apple refinement features without increasing the computational effort of the model.

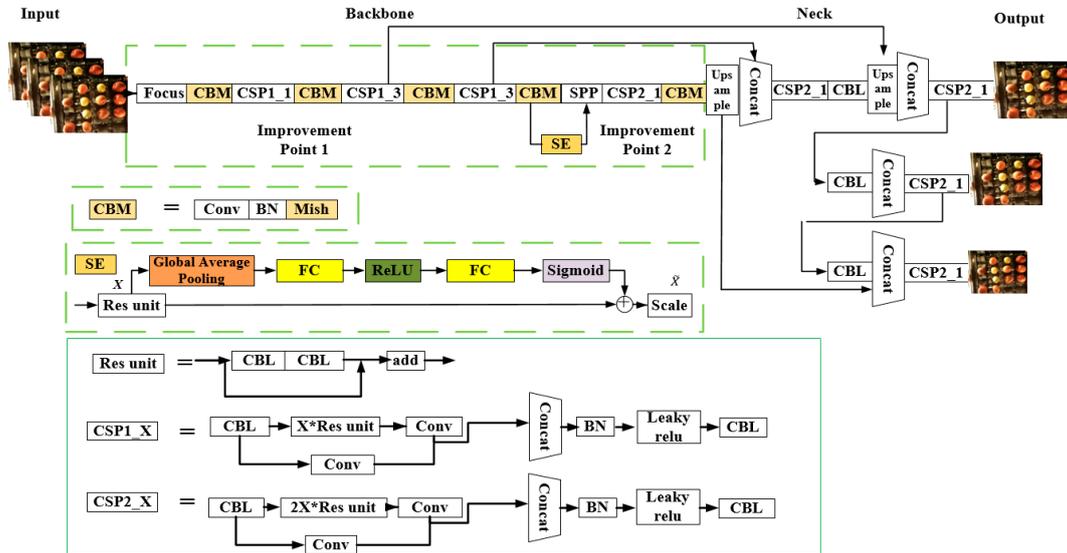


Figure 8. Diagram of the improved network structure of YOLOv5.

2.3.1. Improvement of the Activation Function

The role of the activation function in a convolutional neural network structure is to combine features thoroughly. The activation functions commonly used in YOLOv5 neural networks are Leaky ReLU, Sigmoid, etc. Leaky ReLU (see Equation (1)) can handle the gradient disappearance problem, but it suffers from neuron necrosis due to data sparsity, while Sigmoid (see Equation (2)) can map real numbers to a specified interval, and his curve is smooth and easy to find derivatives for, but it suffers from the problem of gradient disappearance. The Mish (see Equation (3)) activation function has outperformed the Leaky ReLU and other standard activation functions in many deep-learning models [18,19]. The depth of the model in this paper is deeper, and the apple features are more abstract, so this study uses the Mish activation function in the backbone of the YOLOv5 model to achieve better feature extraction results. The CBM module in the backbone network consists of a convolutional layer, a normalization layer, and the Mish activation function. The rest of the model still uses the Leaky ReLU activation function.

$$f_1(x) = \begin{cases} x & x > 0 \\ \alpha x & \text{others} \end{cases} \quad (1)$$

$$f_2(x) = \frac{1}{1+e^{-x}} \quad (2)$$

$$f_3(x) = \frac{x}{1+e^{-x}} \quad (3)$$

As can be seen from Figure 9, the Mish activation function can output arbitrarily large positive values while allowing slight negative gradient values, which avoids gradient saturation due to the gradient being close to zero. The Mish function is non-monotonic and continuously differentiable, which allows the deep neural network to achieve better accuracy and generalization, and facilitates the optimization of gradient updates [20,21].

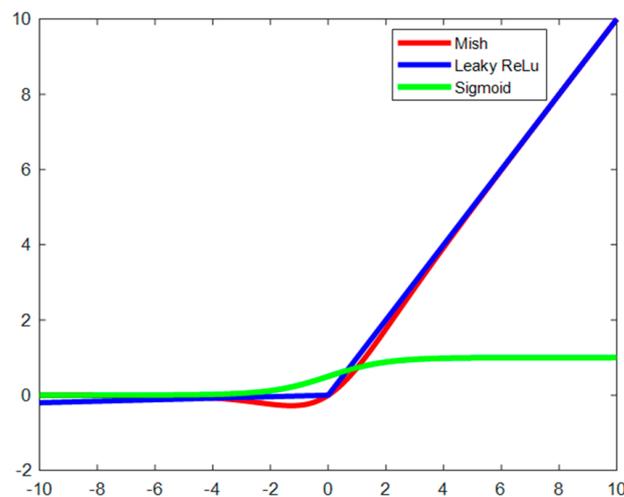


Figure 9. Comparison of Mish, Leaky ReLu, and Sigmoid function images.

2.3.2. Improvement of the Loss Function

Deep learning networks adjust the weights between the layers of the network during the training process through optimization algorithms, and they can reduce the loss so that the predicted frames and the actual frames overlap as much as possible. The loss function is the key to adjusting the weights [22,23]. GIoU has scale invariance. When the target is enlarged or reduced, the loss value can remain the same magnitude, and it considers both the overlapping and non-overlapping parts between the detection frame and the target frame. When $IoU = 0$, the distance of the bounding box does not affect the loss value, GIoU overcomes this shortcoming and can make the corresponding loss expression according to the distance of the two bounding boxes. GIoU expressions are as follows:

$$\begin{cases} GIoU = IoU - \frac{C - (A \cup B)}{C} \\ GIoU = -1 + \frac{A \cup B}{C} \quad (IoU = 0) \end{cases} \quad (4)$$

As shown in Equation (4), when there is an intersection between predicted frame A and actual frame B, convergence is slower in the horizontal and vertical directions. When there is an inclusion relation C between the predicted and actual frames (when C is the smallest closed frame containing A and B), the *GIoU* degrades to an *IoU* and does not work. In this paper, the apples in the flip turnover detection conveyor are relatively dense, and the apples rotate in all directions with the sponge rollers, which makes it impossible to accurately distinguish the actual region from the background region in the grading work of the prediction frame. Therefore, in this paper, *DIoU_Loss* is chosen as the boundary loss function in the output layer instead of *GIoU_Loss* to speed up the target grading accuracy and detection speed.

DIoU inherits the advantages of *GIoU* and adds the centroid distance geometric information [24,25]. As shown in Figure 10, which takes into account both the overlapping area and the distance between the two centroids, *DIoU* can provide the accurate gradient direction for the model when the prediction frame and the actual frame have crossed or overlapped. The introduction of the distance penalty makes *DIoU* converge faster than *GIoU*. The equation is shown in Equation (5).

$$DIoU = 1 - IoU + \frac{p^2(b, b^{st})}{c^2} \quad (5)$$

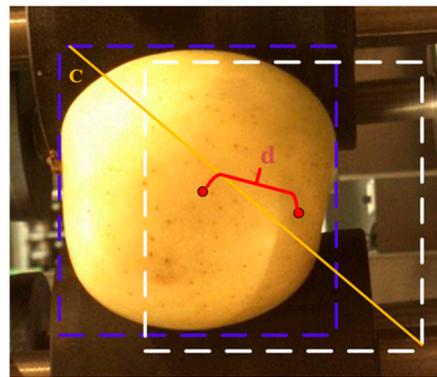


Figure 10. DIoU schematic.

In the above equation, b, b^{gt} represents the target and prediction box centroids, $p^{(*)}$ represents the Euclidean distance, and c is the diagonal length of the minimum enclosing box covering the target and prediction boxes.

2.3.3. Integration of Attentional Mechanisms

Attention is one of the most critical mechanisms in human perception. The human eye is adept at recognizing key image features from complex images and ignoring irrelevant information, which is where the attention mechanism excels. With the booming development of deep learning, the attention mechanism can be used for machine vision. Apples have characteristics such as many features and small sizes, which can easily lead to wrong and missed detection, thus making the grading accuracy of apple features low [26]. By introducing the attention mechanism in the convolutional layer, the learning representation can be enhanced autonomously, and the method is highly operational and effective [27,28]. The Backbone module in YOLOv5 adds the Focus structure, which improves the computational speed by slicing the feature map, but may have an impact on the features. In order to improve the target feature extraction effect of the Backbone module, this paper introduces the channel focus mechanism squeeze excitation (SE) module [29], which is embedded into the last layer of the Backbone module to improve the accuracy of apple grading without increasing the model size.

The SE module can effectively capture the channel and position information of the image, which in turn can improve the grading accuracy of the model. Figure 11 shows the working principle of the SE module, which consists of two main parts, Squeeze and Excitation. The SE module first obtains a global description of the input through Squeeze, which enables a wider perceptual field of view, and then obtains the weights of each channel in the Feature Map through Excitation’s two-layer fully connected bottleneck structure as input to the lower layer network.

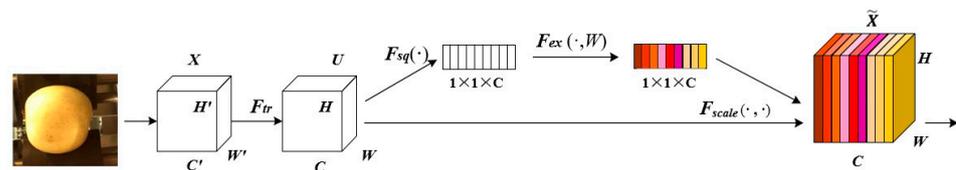


Figure 11. Squeeze and excitation.

In Figure 11, the squeeze operation first encodes the entire spatial feature on the channel as a local feature by global averaging pooling. Then the operation of the connected channel is performed through two fully connected layers and a non-linear activation function (see Equation (6)), followed by a Sigmoid activation function to obtain the weight of each channel, and finally, a multiplicative weighted multiplication to each channel to complete the recalibration of the attention mechanism. The calculation results are shown

in Equations (7) and (8). A correlation between channels was established through global average pooling, two fully connected layers, and a non-linear activation function.

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (6)$$

where Z_c represents the C th element in the statistic, H , W the space dimension, and the subscripts i, j the number of channels. After the Squeeze operation has obtained the channel information, it uses two fully connected layers to form the gate mechanism and activates it with Sigmoid. The calculation is as follows:

$$s = F_{ex}(z, W) = \sigma(g(z, W) = \sigma(W_2 \delta(W_1 z)) \quad (7)$$

where δ is the ReLu activation function, σ is the Sigmoid function, W_1 and W_2 are two fully connected layers equal to $C/r \times C$ and $C \times C/r$, respectively, r is the scaling parameter that limits the complexity of the model and increases its capability, and s represents the set of weights of the feature maps obtained through the fully connected and non-linear layers. Finally, the output weights are assigned to the original features. The calculation formula is as follows.

$$\widetilde{X}_c = s_c \times u_c \quad (8)$$

where \widetilde{X}_c is the feature map of the featured channel X , S_c is the weight, and U_c is a two-dimensional matrix.

3. Result and Discussion

3.1. Experimental Validation and Analysis of Results

3.1.1. Experimental Environment

The experimental models in this paper were constructed, trained, and the results were tested based on the Windows 10-x64-bit operating system. The experimental programming environment is Python 3.7, using Cudnn for GPU acceleration, and Apple hierarchical model training is implemented under the PyTorch 1.7 deep learning framework. The experimental environment configuration is shown in Table 2. The number of iterations of the training process was set to 150, the weight decay coefficient was 0.001, the learning rate was 0.917, and the maximum training batch was eight. An IOU threshold of 0.5 was taken as the standard.

Table 2. Experimental environment.

Computer Configuration	Specific Parameters
CPU	Intel i7-9750k
GPU	NVIDIA GTX1660Ti(16G)
Operating system	Windows 10-x64
Random Access Memory	DDR4 32G (8G*4)
CUDA	CUDA 10.3

In order to better calculate the classification accuracy and reliability of this model, this paper selects loss function curve (Loss), Precision, Average Precision (AP), Recall, Mean Average Precision(mAP), and frames per second (Fps) as the algorithm performance evaluation indexes [30]. The relevant evaluation indexes are calculated as shown in Equations (9)–(12).

$$Precision = \frac{TP(TruePositive)}{TP + FP(FalsePositive)} \quad (9)$$

$$Recall = \frac{TP(TruePositive)}{TP + FN(FalseNegative)} \quad (10)$$

$$\text{Average Precision} = \int_0^1 (P(R)dR) \quad (11)$$

$$\text{Mean Average Precision} = \frac{\sum \text{Average Precision}}{n(\text{Class})} \quad (12)$$

In the above equation, TP represents the number of apple samples correctly identified by the model, FP represents the number of apples incorrectly identified by the model, FN represents the number of apple samples not identified by the model, and n represents the number of categories.

3.1.2. Analysis of Experimental Results

(1) Experiments related to the improved algorithm

The loss function can visually reflect whether the model can converge stably or not [31]. In the process of network training, the following three models were selected for comparison, taking into account the comparative effects of different algorithmic models: the YOLOv5 algorithm using the Mish activation function to optimize the backbone network, denoted by YOLOv5-M; the YOLOv5 algorithm using the DIoU optimization loss function, denoted by YOLOv5-D; and the simultaneous use of the Mish activation function and DIoU optimization method, denoted by Im-YOLOv5. The resulting loss function curve after training is shown in Figure 12.

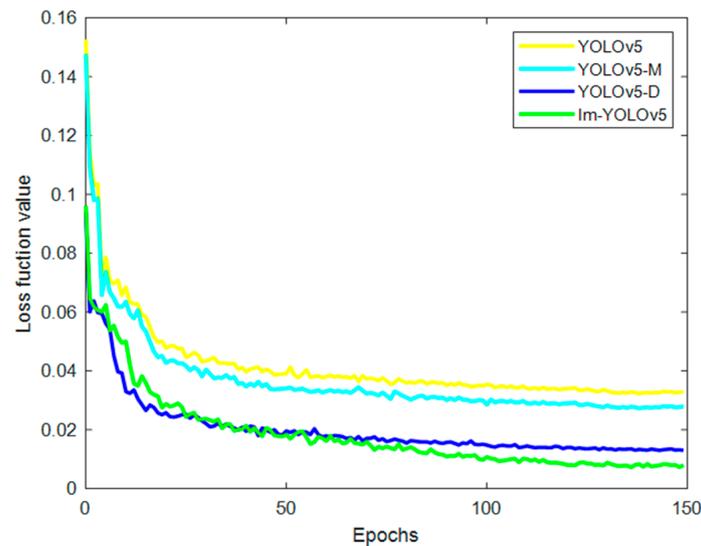


Figure 12. Loss value curve changes with epochs.

As shown in Figure 12, the overall trend of the loss values of the four models in training is the same. They decrease rapidly in training and eventually stabilize. The YOLOv5-M and YOLOv5-D loss values and convergence rates are significantly faster than the original YOLOv5 algorithm, and the degree of fluctuation is less, which proves that the localization accuracy and convergence rate of the models can be increased when using complete loss and activation functions [31]. From Im-YOLOv5, the loss value and convergence speed are slightly lower than YOLOv5-D for the first 50 iterations of the model, but after 50 iterations, the loss value and convergence speed are due to the rest of the models. This indicates that the Im-YOLOv5 algorithm can improve the convergence speed and localization accuracy of the model, which helps to obtain a more accurate resultant model, which proves the effectiveness of the improved model.

In order to verify the effectiveness of the improved method in this paper for apple grading, this study trained the YOLOv5 and the Im-YOLOv5 models under the same dataset and training set. The PR curve represents the relationship between accuracy and recall, which can measure the model's generalization ability. The PR curves of the two

models after the training was completed are shown in Figure 13. The area between the PR curve and the coordinate axes of the Im-YOLOv5 is larger than that of the original YOLOv5 model, which indicates that the improved model has better overall performance.

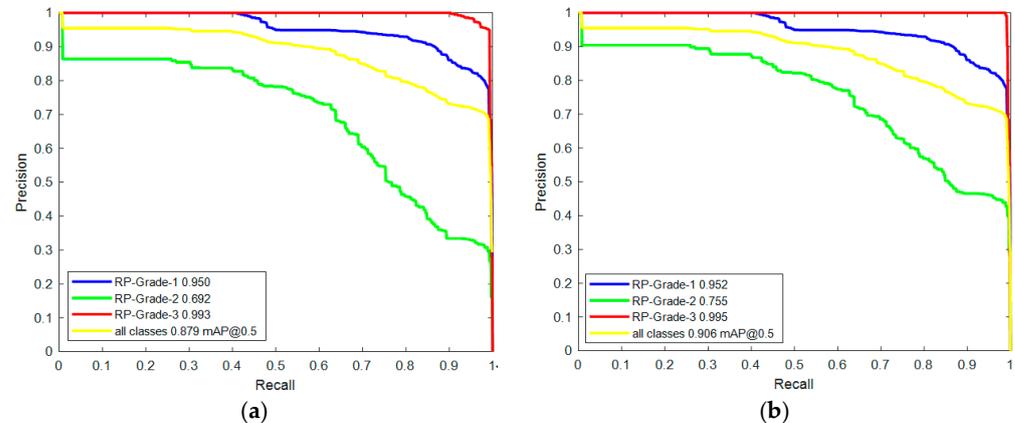


Figure 13. Comparison of YOLOv5 and Im-YOLOv5 PR curve. (a) YOLOv5; (b) Im-YOLOv5.

As can be seen from Figure 13, the Im-YOLOv5 model has improved the grading accuracy for different apple quality levels, with higher mAP of over 95% for Grade-1 and Grade-3 apples. The mAP for Grade-2 apples reached 0.755, an improvement of 9.1% over the original model. The average accuracy for all apple grades was 0.906, an increase of 3.1% compared to the original model.

The Im-YOLOv5 model and YOLOv5s model trained in this paper were used to grade apples of different qualities in an automatic apple grader.

Figure 14a shows the grading results before the improvement of the YOLOv5 model, and Figure 14b shows the results of the Im-YOLOv5 model grading. The accuracy of the apple grading in Figure 14a is low, where the apples in the first and second images appear to have duplicate detection frames, and the second image shows incorrect grading of the grade-1 and grade-2 apples, marking the grade-1 apples as grade-2 apples. The third panel shows no duplicate detection frames but incorrectly marks three grade-1 apples with low accuracy. In contrast, there is an improvement in grading accuracy for all grades of apples in Figure 14b, with no duplicate boxes. The improved model was able to pay more attention to apple feature information, which improved the robustness of the model while increasing the grading accuracy. Therefore, the Im-YOLOv5 model can satisfy apple grading in actual production environments.

In order to explore the effectiveness of visual attention mechanisms in convolutional networks and to enhance the interpretability of the apple grading model in this paper, a part of the improved YOLOv5 feature extraction layer in this paper was visualized [32]. The results of feature extraction from the convolutional layer of the backbone network are shown in Figure 15. As shown in Figure 15a, the initial feature size of the convolutional layer of the backbone network is large, the feature extraction is more fine-grained, and the apple features are extracted while containing complex background information; as the network deepens, the extracted features are gradually blurred and sparse and more semantic. As can be seen in Figure 15b, after the attention SE module, there are some highlighted areas in the figure, and the location of the apples is highlighted in the spatial pyramid pooling (SPP) output feature map, which indicates that after adding the SE module, the deep network layer of the Im-YOLOv5 model in this paper filtered the extracted features, which helped to highlight the target apples as well as filter the background information in the grading stage and improved the network model accuracy.



Figure 14. Grading results. (a) YOLOv5s model; (b) Im-YOLOv5 model.

(2) Comparison experiments between different models

In order to further verify the superiority of the proposed algorithm, the improved YOLOv5 algorithm was compared with several classical algorithms commonly used in the current deep learning field, including a single shot multibox detector (SSD) [33], a fast and superior generalization among One-Stage detectors, and YOLOv4 and YOLOv5s [34], which have better comprehensive performance. The comparison experiments selected accuracy, recall, mAP, and Fps as the evaluation metrics of each algorithm, and the models were trained and tested under the same initial conditions. The apple grading results are shown in Table 3 below.

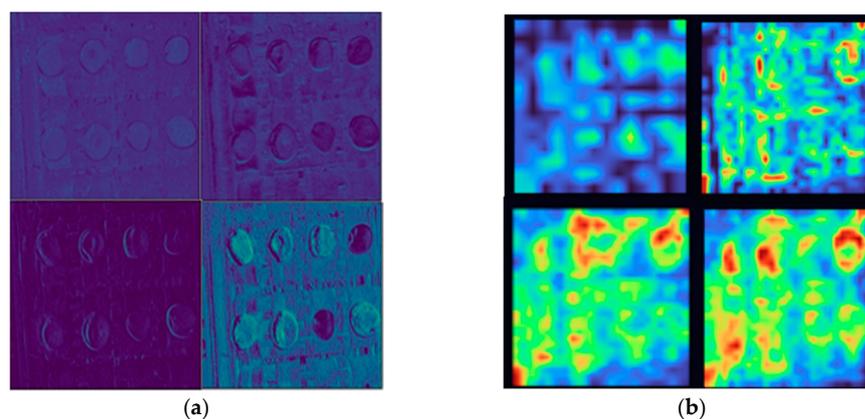


Figure 15. Feature map visualization of the Im-YOLOv5 model. (a) CSP network output 56×56 feature map; (b) SPP network output 14×14 feature map.

Table 3. Comparison of different models.

Index Models	Precision			Recall			mAP @0.5	FPS(f/s)
	Grade-1	Grade-2	Grade-3	Grade-1	Grade-2	Grade-3		
SSD	0.812	0.612	0.884	0.926	0.645	0.895	0.789	34.78
YOLOv4	0.821	0.656	0.892	0.862	0.609	0.923	0.815	50.42
YOLOv5s	0.938	0.692	0.991	0.950	0.655	0.993	0.879	56.64
Im-YOLOv5	0.951	0.806	0.992	0.952	0.751	0.995	0.906	59.63

As can be seen from the results in Table 3, the SSD model has lower accuracy and recall, with an average accuracy mAP of 0.789 and a real-time frame rate FPS of 34.78 for the Apple classification. As the model improves, its accuracy, recall, mAP, and FPS gradually increase, with the Im-YOLOv5 model having the highest mAP of 0.906, compared to the YOLOv5, YOLOv4, and SSD models by 14.8%, 11.1%, and 3.7%, respectively. The accuracy and recall of the grade-2 apple reached 0.806 and 0.751, respectively, which were 16.4% and 14.6% higher than the original YOLOv5 method. On the other hand, the real-time image frame rate of the Im-YOLOv5 method in this paper was improved, and the FPS of the improved model reached a maximum of 59.63, which has better real-time performance compared with the lightweight model YOLOv5s. The results show that the grading effect and real-time performance of the Im-YOLOv5 model proposed in this paper are better than those of the traditional deep learning model, proving the effectiveness of the proposed method.

3.2. System Solution Validation

3.2.1. Automatic Apple Grader Control System Set Up

The automatic apple grader designed and developed in this paper is shown in Figure 16, and its workflow is shown in Figure 17. When the automatic apple grader is started, the apples are lifted by a feeding and material handling lifting mechanism to the turnover detection conveyor. The turnover conveyor uses pairs of double conical sponge rollers to turn the apples. At this point, the automatic grading control system uses the improved YOLOv5 algorithm to grade the apples based on the surface information collected by the image acquisition device and sends the grading decision to the grading execution device [35,36]. The grading actuator automatically places the apples in the appropriate storage bin when they reach the appropriate grade based on the grading results assessed by the grading control system. The bins are equipped with cushioning material to reduce the impact of falling apples.

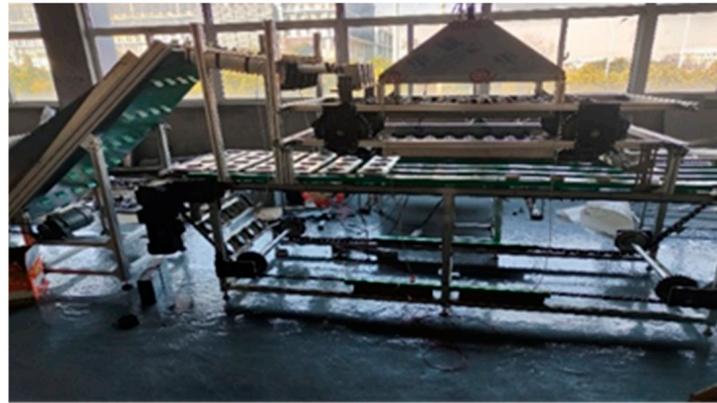


Figure 16. Physical view of the automatic apple grader.

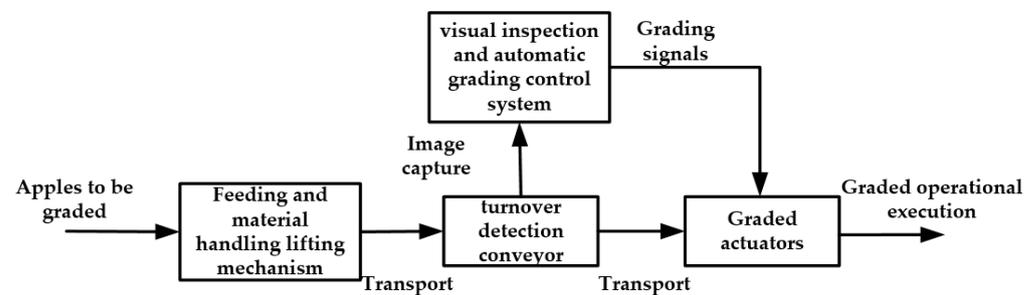


Figure 17. Workflow diagram for automatic apple grading.

The hardware of the automatic apple grading control system includes a CCD industrial camera, IPC-610L industrial computer, PLC-1212 controller, AC contactor, inverter, and AC motor. The CCD industrial camera uses a GigE interface for data transmission and acquisition with an IPC-610L industrial computer. The industrial computer and PLC-1212 controller use the snap7 library to transmit information via a network cable. PLC controls the AC motor to drive a grading actuator through the AC contactor. The processor CPU of the IPC-610L is the same as that of the training computer, an Intel i7-9750k, an Intel i7-9750k with two graphics cards GTX1660Ti (6G), the operating system is Windows 10-x64, and the software environment is Python3.7, CUDA10.3, TIA Portal V15.1.

In order to facilitate debugging and observe the improvement effect of the model algorithm, the PyQt-based apple automatic grading control system software developed in this study is shown in Figure 18, which implements local video detection and real-time grading functions to achieve fast and accurate apple grading. The software designed in this paper sends the processed apple grade and location information to the TIA Portal V15.1 software through the snap7 library. After the grading actuator receives the apples in order, the grading operation is finally completed by the PLC controller in the corresponding grading lane [36,37].

3.2.2. Results of the Grading Experiment

In order to verify the feasibility of the algorithm and the grading scheme of the apple automatic grading platform system in this paper, the designed and developed automatic apple grader was experimentally verified. One hundred apples of each quality grade were manually selected as samples, and the apple grades were determined based on the red Fuji GB/T 10651-2008 grading standard mentioned in Section 3.2. The experimental results are shown in Table 4.

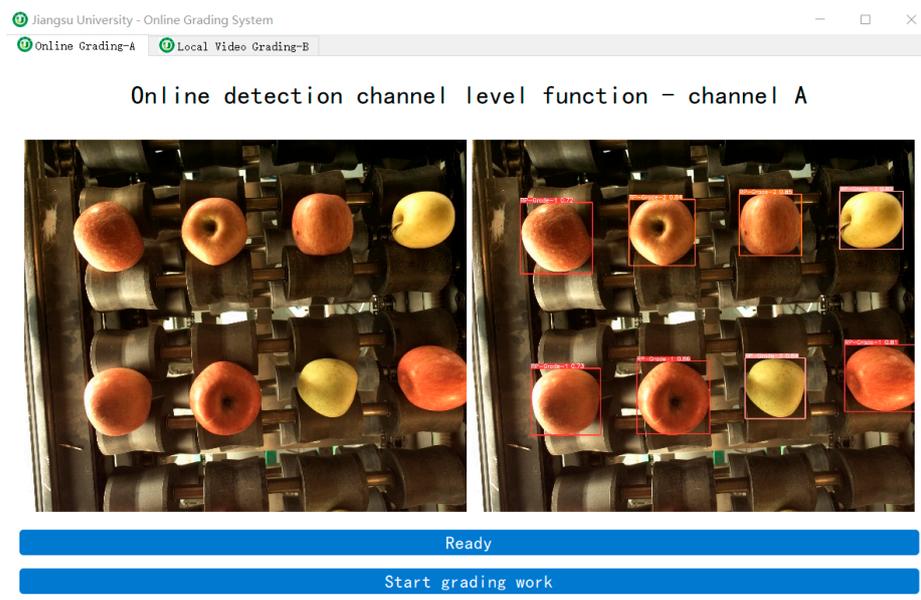


Figure 18. Apple auto-grading software.

Table 4. Grading experimental data.

Grade	Manual Grading Results	Equipment Grading Results	Consistency Rates	Completions Time/(s)
RP-grade-1	100	92	92%	27
RP-grade-2	100	88	88%	27
RP-grade-3	100	100	100%	27
Accuracy			93%	81

As can be seen from Table 4 of the experimental results, there was some grading error for Grade-1 and Grade-2 apples. Grade-1 apple sorting was 92% accurate, grade-2 apple sorting was 88% accurate, and grade-3 apple sorting was 100% accurate, with an average accuracy of 93%. The average classification accuracy for the three apple grades was 93% for the three apple grades, with an average classification speed of four apples/second. Both the real-time and accuracy rates are high enough to meet the grading requirements of small and medium-sized fruit farmers and to verify the effectiveness of the algorithm.

4. Conclusions

This paper proposes an improved apple grading model of YOLOv5, which better balances the grading accuracy and speed of apples, and also carries out experimental verification on the automatic apple grader designed and developed in this paper. The main conclusions of this study are as follows.

- (1) In order to achieve more accurate apple grading and better real-time performance, the DIoU loss function and Mish loss function were chosen to replace the GIoU function and Relu activation function of the original algorithm model in terms of algorithm optimization, which improved the feature extraction capability and convergence speed of the model. The attention SE module is embedded in the Backbone structure to discard unnecessary features, which improves the training accuracy of the model without burdening the model. The experimental results show that the improved YOLOv5 has improved the average accuracy rate mAP by 3.1% compared to YOLOv5, 11% compared to YOLOv4, and 15% compared to SSD, and the real-time grading speed has reached 59.63 FPS, which is a large improvement in both the apple-grade grading accuracy rate and real-time performance. A portion of the improved YOLOv5 feature extraction layer was visualized to show the features extracted by different

convolutional layers, enhancing the interpretability of the apple grading model in this paper.

- (2) An automatic apple grader was developed and designed, and the grading method in this paper was experimentally verified on an automatic apple grading machine platform. The experimental results showed that the grading accuracy of the grading method on the automatic apple grader reached 93%, with an average grading speed of four apples/sec. It has high accuracy and real-time performance, which can meet the grading needs of farmers and small and medium-sized enterprises in the field and has practical application in the apple grading industry.

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

Funding: This research was funded by the National Natural Science Foundation of China (No. 61973141 and 62173162), the Jiangsu agriculture science and technology innovation fund (No. CX(20)3059), and A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (No. PAPD-2018-87).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Tian, Y.N.; Yang, G.D.; Wang, Z.; Wang, H.; Li, E.; Liang, Z.Z. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* **2019**, *157*, 417–426. [\[CrossRef\]](#)
2. Zhu, D.S.; Ren, X.J.; Wei, L.; Cao, X.H.; Ge, Y.; Liu, H.; Li, J.R. Collaborative analysis on difference of apple fruits flavour using electronic nose and electronic tongue. *Sci. Hort.* **2020**, *260*, 108879. [\[CrossRef\]](#)
3. Musacchi, S.; Serra, S. Apple fruit quality: Overview on pre-harvest factors. *Sci. Hort.* **2018**, *234*, 409–430. [\[CrossRef\]](#)
4. Courtemanche, B.; Montpetit, B.; Royer, A.; Roy, A. Creation of a Lambertian microwave surface for retrieving the downwelling contribution in ground-based radiometric measurements. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 462–466. [\[CrossRef\]](#)
5. Abdullah, M.Z.; Mohamad-Saleh, J.; Fathinul-Syahir, A.S.; Mohd-Azemi, B.M.N. Discrimination and classification of fresh-cut starfruits using automated machine vision system. *J. Food Eng.* **2006**, *76*, 506–523. [\[CrossRef\]](#)
6. Marchant, J.A.; Onyango, C.M.; Street, M.J. High speed sorting of potatoes using computer vision. *ASAE Pap.* **1988**, *88*, 3540.
7. Moallem, P.; Serajoddin, A.; Pourghassem, H. Computer vision-based apple grading for golden delicious apples based on surface features. *Inf. Processing Agric.* **2017**, *4*, 33–40. [\[CrossRef\]](#)
8. Gui, J.S.; Zhang, Q.; Hao, L. Apple shape classification method based on wavelet moment. *Sens. Transducers* **2014**, *178*, 182–187.
9. Fan, S.X.; Li, J.B.; Zhang, Y.H.; Tian, X.; Wang, Q.Y.; He, X.; Zhang, C.; Huang, W.Q. On line detection of defective apples using computer vision system combined with deep learning methods. *J. Food Eng.* **2020**, *286*, 110102. [\[CrossRef\]](#)
10. Raikar, M.M.; Meena, S.M.; Kuchanur, C.; Girraddi, S.; Benagi, P. Classification and grading of Okra-ladies finger using deep learning. *Procedia Comput. Sci.* **2020**, *171*, 2380–2389. [\[CrossRef\]](#)
11. Luna, R.; Dadios, E.P.; Bandala, A.A.; Vicer, R.R.P. Tomato fruit image dataset for deep transfer learning-based defect detection. In Proceedings of the 2019 IEEE International Conference on Cybernetics and Intelligent Systems and IEEE Conference on Robotics, Automation and Mechatronics, Bangkok, Thailand, 18–20 November 2019; pp. 356–361.
12. Cubero, S.; Aleixos, N.; Albert, F.; Torregrosa, A.; Ortiz, C.; Garcia-Navarrete, O.; Blasco, J. Optimised computer vision system for automatic pre-grading of citrus fruit in the field using a mobile platform. *Precis. Agric.* **2014**, *15*, 80–94. [\[CrossRef\]](#)
13. Baigvand, M.; Banakar, A.; Minaei, S.; Khodaei, J.; Behroozi-Khazaei, N. Machine vision system for grading of dried figs. *Comput. Electron. Agric.* **2015**, *119*, 158–165. [\[CrossRef\]](#)
14. GB/T 10651—2008; Fresh apple General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. Standardization Administration of PRC (SAC): Beijing, China, 2008.
15. Chen, Z.Y.; Wu, R.H.; Lin, Y.Y.; Li, C.Y.; Chen, S.Y.; Yuan, Z.E.; Chen, S.W.; Zou, X.J. Plant disease recognition model based on improved YOLOv5. *Agronomy* **2022**, *12*, 365. [\[CrossRef\]](#)
16. Xu, X.W.; Zhang, X.L.; Zhang, T.W. Lite-YOLOv5: A lightweight deep learning detector for on-board ship detection in large-scene sentinel-1 SAR images. *Remote Sens.* **2022**, *14*, 1018. [\[CrossRef\]](#)

17. Lv, J.D.; Xu, H.; Han, Y.; Lu, W.B.; Xu, L.M.; Rong, H.L.; Yang, B.A.; Zou, L.; Ma, Z.H. A visual identification method for the apple growth forms in the orchard. *Comput. Electron. Agric.* **2022**, *197*, 106954. [[CrossRef](#)]
18. Jagtap, A.D.; Kawaguchi, K.; Karniadakis, G.E. Adaptive activation functions accelerate convergence in deep and physics-informed neural networks. *J. Comput. Phys.* **2020**, *404*, 109136. [[CrossRef](#)]
19. Li, S.; Wang, Y.H.; Feng, C.Y.; Zhang, D.; Li, H.Z.; Huang, W.; Shi, L. A thermal imaging flame-detection model for firefighting robot based on YOLOv4-F model. *Fire* **2022**, *5*, 172. [[CrossRef](#)]
20. Li, Y.M.; Zhang, J.; Hu, Y.; Zhao, Y.N.; Cao, Y. Real-time safety helmet-wearing detection based on improved YOLOv5. *Comput. Syst. Sci. Eng.* **2022**, *43*, 1219–1230. [[CrossRef](#)]
21. Wang, Y.N.; Fu, G.Q. A novel object recognition algorithm based on improved YOLOv5 model for patient care robots. *Int. J. Hum. Robot.* **2022**, *19*, 2250010. [[CrossRef](#)]
22. Lv, H.F.; Lu, H.C. Research on traffic sign recognition technology based on YOLOv5 algorithm. *J. Electron. Meas. Instrum.* **2021**, *35*, 137–144.
23. Ji, W.; Gao, X.X.; Xu, B.; Pan, Y.; Zhang, Z.; Zhao, D. Apple target recognition method in complex environment based on improved YOLOv4. *J. Food Process Eng.* **2021**, *44*, e13866. [[CrossRef](#)]
24. Ji, W.; Peng, J.Q.; Xu, B.o.; Zhang, T. Real-time detection of underwater river crab based on multi-scale pyramid fusion image enhancement and MobileCenterNet model. *Comput. Electron. Agric.* **2023**, *204*, 107522. [[CrossRef](#)]
25. Zheng, Z.; Wang, P.; Liu, W.; Li, J.Z.; Ye, R.G.; Ren, D.W. Distance-IoU Loss: Faster and better learning for bounding box regression. In Proceedings of the 34th AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 December 2020.
26. Chen, W.; Zhang, J.F.; Guo, B.Y.; Wei, Q.Y.; Zhu, Z.Y. An apple detection method based on des-YOLO v4 algorithm for harvesting robots in complex environment. *Math. Probl. Eng.* **2021**, *2021*, 7351470. [[CrossRef](#)]
27. Guo, S.Y.; Li, L.L.; Guo, T.Y.; Cao, Y.Y.; Li, Y.L. Research on mask-wearing detection algorithm based on improved YOLOv5. *Sensors* **2022**, *22*, 4933. [[CrossRef](#)] [[PubMed](#)]
28. Ji, W.; Pan, Y.; Xu, B.; Wang, J.C. A real-time apple targets detection method for picking robot based on ShufflenetV2-YOLOX. *Agriculture* **2022**, *12*, 856. [[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. Jung, H.K.; Choi, G.S. Improved YOLOv5: Efficient object detection using drone images under various conditions. *Appl. Sci.* **2022**, *12*, 7255. [[CrossRef](#)]
31. Wang, C.; Zhang, X.F.; Liu, C.; Zhang, W.; Tang, Y. Detection method of wheel hub weld defects based on the improved YOLOv3. *Opt. Precis. Eng.* **2021**, *29*, 1942–1954. [[CrossRef](#)]
32. Shin, H.C.; Roth, H.R.; Gao, M.C.; Lu, L.; Xu, Z.Y.; Nogues, I.; Yao, J.H.; Mollura, D.; Summers, R.M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [[CrossRef](#)]
33. Zhou, T.; Yu, Z.T.; Cao, Y.; Bai, H.Y.; Su, Y. Study on an infrared multi-target detection method based on the pseudo-two-stage model. *Infrared Phys. Technol.* **2021**, *118*, 103883. [[CrossRef](#)]
34. Chang, Y.H.; Zhang, Y.Y. Deep learning for clothing style recognition using YOLOv5. *Micromachines* **2022**, *13*, 1678. [[CrossRef](#)]
35. Ji, W.; Cui, X.; Xu, B.; Ding, S.H.; Ding, Y.; Peng, J.Q. Cross-coupled control for contour tracking error of free-form curve based on fuzzy PID optimized by improved PSO algorithm. *Meas. Control.* **2022**, *55*, 807–820. [[CrossRef](#)]
36. Raza, T.Z.; Lang, W.; Jedermann, R. Integration of wireless sensor networks into industrial control systems. In *Dynamics in Logistics*; Springer: Cham, Switzerland, 2017; pp. 209–218.
37. Cai, W.Z.; Tian, J.Y.; Wang, S.Y.; Li, J.F.; Yang, S.Q. Research of joint virtual commissioning of robotic grinding based on NX MCD and TIA. *Mod. Manuf. Eng.* **2022**, *7*, 37–42.

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.