Article

Strawberry Detection and Ripeness Classification Using YOLOv8+ Model and Image Processing Method

Chenglin Wang [1], Haoming Wang [1], Qiyu Han [1], Zhaoguo Zhang [1,*], Dandan Kong [1] and Xiangjun Zou [2,3]

1. Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650504, China; wangcl86@kust.edu.cn (C.W.); haomw01@163.com (H.W.); qiyu.han2001@outlook.com (Q.H.); whyykongh@163.com (D.K.)
2. College of Intelligent Manufacturing and Modern Industry, Xinjiang University, Urumqi 830046, China; xjzou1@163.com
3. Foshan-Zhongke Innovation Research Institute of Intelligent Agriculture and Robotics, Foshan 528000, China
* Correspondence: zzg@kust.edu.cn

Abstract: As strawberries are a widely grown cash crop, the development of strawberry fruit-picking robots for an intelligent harvesting system should match the rapid development of strawberry cultivation technology. Ripeness identification is a key step to realizing selective harvesting by strawberry fruit-picking robots. Therefore, this study proposes combining deep learning and image processing for target detection and classification of ripe strawberries. First, the YOLOv8+ model is proposed for identifying ripe and unripe strawberries and extracting ripe strawberry targets in images. The ECA attention mechanism is added to the backbone network of YOLOv8+ to improve the performance of the model, and Focal-EIOU loss is used in loss function to solve the problem of imbalance between easy- and difficult-to-classify samples. Second, the centerline of the ripe strawberries is extracted, and the red pixels in the centerline of the ripe strawberries are counted according to the H-channel of their hue, saturation, and value (HSV). The percentage of red pixels in the centerline is calculated as a new parameter to quantify ripeness, and the ripe strawberries are classified as either fully ripe strawberries or not fully ripe strawberries. The results show that the improved YOLOv8+ model can accurately and comprehensively identify whether the strawberries are ripe or not, and the mAP50 curve steadily increases and converges to a relatively high value, with an accuracy of 97.81%, a recall of 96.36%, and an F1 score of 97.07. The accuracy of the image processing method for classifying ripe strawberries was 91.91%, FPR was 5.03%, and FNR was 14.28%. This study demonstrates the program's ability to quickly and accurately identify strawberries at different stages of ripeness in a facility environment, which can provide guidance for selective picking by subsequent fruit-picking robots.

Keywords: fruit detection; ripeness identification; deep learning; YOLOv8 model; computer vision

1. Introduction

Strawberries are widely grown around the world as a nutrient-rich cash crop [1]. The number of strawberries grown globally has increased 2.4-fold in the last 20 years [2]. This means that strawberry picking requires more sophisticated techniques to match the rapid development of strawberry-growing technology. Traditional manual harvesting, with low efficiency and high labor costs, cannot meet the demand for efficient harvesting [3]. The development of fruit-picking robots could improve fruit harvesting efficiency and save labor costs [4–6]. Ripeness identification is a key step for picking robots to realize selective harvesting and a prerequisite for counting crops and estimating yields [7,8]. Therefore, it is of great significance to study an efficient and accurate method for strawberry ripeness determination in a non-standardized environment for strawberry harvest management. However, when strawberries are in the transition from unripe to ripe (i.e., near-ripe),

the skin is characterized by both red and white features, making the distinction difficult [9]. Moreover, in the actual production process, it is necessary to separate the fully ripe strawberries from the ripe strawberries, taking into account the need for long-distance transportation of strawberries or local fresh sales [10]. This poses a great challenge in grading strawberry ripeness.

Deep learning has made significant progress in target detection and scene identification [11]. Current approaches for ripeness identification of fruit images are centered around deep learning. Deep learning target detection algorithms can be categorized into single-stage and dual-stage. Single-stage methods are faster and less complex than two-stage methods. Since most agricultural application scenarios require the deployment of network models into embedded devices, a lot of research has been carried out on single-level target detection algorithms. Typical single-stage methods include the YOLO ("you only look once") series [12], the SSD (single-shot multi-box detector) series [13], and others. Phan et al. [14] proposed four deep learning frameworks, Yolov5m, and models combining ResNet50, ResNet-101, and EfficientNet-B0, for classifying tomato fruits on the vine into ripe, unripe, and damaged categories. Azadnia et al. [15] classified hawthorn images into unripe, ripe, and overripe using Inception-V3, ResNet-50, and DL models. Yang et al. [16] proposed the LS-YOLOv8s model, which can accurately detect and grade strawberry ripeness by combining the YOLOv8s deep learning algorithm and the LW-Swin Transformer module. Chen et al. [17] first used YOLOv5 to detect citrus fruits and then used a 4-channel ResNet34 to detect citrus fruit ripeness. The accuracy reached 95.07%, which is better than the traditional RGB-based CNN and machine learning models. Zhang et al. [18] proposed a YOLOv5-based visual detection and pose classification algorithm to detect tomatoes and were able to identify the ripeness of tomatoes. The above methods and models have been successful in fruit target detection and ripeness classifying. However, more work needs to be carried out to improve the detection performance of the models in complex growing environments.

Image processing methods are also widely used in the field of fruit ripeness identification. Azarmdel et al. [19] segmented mulberry images using RGB color space and selected the B channel as the best channel to classify the fruit into three categories (unripe, ripe, and overripe). Alfatni et al. [20] used multivariate techniques to extract fruit image features and combine the information for oil palm species classifying and ripeness testing. By combining a chromatic aberration map of citrus fruits under normal conditions with a luminance map under light, Lu et al. [21] effectively solved the problem of the effect of light on the identification of citrus ripeness by using color features for threshold segmentation. Castro et al. [22] evaluated the ability of the combination of three color spaces (RGB, HSV, and L*a*b*) with machine learning for classifying potato endive fruits. Ropelewska et al. [23] developed a classification model using texture parameters of image color channels R, G, B, L, a, b, X, Y, and Z. The model was constructed using image texture parameters and traditional machine learning algorithm to quickly and accurately differentiate between different ripening stages of peaches.

There have been studies combining deep learning and image processing methods to classify strawberry ripeness. Wang et al. [24] proposed the Adaptive Strawberry Feature Augmentation Network (ASFA-net) for generating masks of strawberries. The red areas within the masks of ripening strawberries were segmented based on hue, saturation, and luminance (HSV) to calculate the proportion of the red area of individual strawberries, and the proportion of the red color was used as a new parameter for the quantification of strawberry ripeness. Tang et al. [25] used an improved Mask R-CNN backbone network and extracted the strawberry target in the image, divided the strawberry target into four sub-regions, and extracted the color eigenvalues of the B, G, L, a, and S channels of each sub-region, and classified the strawberry ripeness based on color eigenvalues. In this study, we combined deep learning methods and image processing methods to categorize strawberry ripeness into unripe, ripe, and fully ripe. Improving on the original YOLOv8 deep convolutional detection network structure, an attention module (efficient channel

attention, ECA) is introduced to accurately identify each strawberry without significantly increasing the memory of the network structure. In addition, the Focal-EIOU loss function, which is more suitable for strawberry maturity identification, was used in the loss function. After the strawberry bounding box is obtained using the improved model, the part of the bounding box that best represents the strawberry ripeness (i.e., the center line) is extracted. The number of pixels in the strawberry center line is much smaller compared to the whole strawberry image, which reduces the time to traverse each pixel in the image and improves the speed of ripeness classification to meet the needs of real-time detection in agricultural facilities. Strawberry ripeness is quantified by the ratio of the number of red pixels on the strawberry center line to the total number of pixels (i.e., red ratio).

In Section 2, we present the construction of the dataset and the improved network in this study, describing how to extract the strawberry centerline and calculate the red ratio of the centerline. Experimental results are given in Section 3. Finally, discussions and conclusions are given in Sections 4 and 5.

2. Materials and Methods

2.1. Image Collection and Dataset Construction

2.1.1. Image Acquisition

The study was conducted on strawberries grown on raised beds, as shown in Figure 1a. The strawberry images in this study were taken from 24 October 2023 to 2 December 2023 and include strawberries at different stages of ripening. The site was located on a strawberry plantation in Chenggong District, Yunnan, China, and the photographed area consisted of 20 rows with 100 strawberries per row. Strawberry varieties include the Zhangji and Hongyan, both of which are red when ripe. Image acquisition was performed using the rear camera of a smartphone, with a picture resolution of 3024×4032, and a shooting imaging distance of 0.15~0.3 m. To improve the robustness of the model in various environments, we collected 1187 strawberry images in JPEG format, which contain images with different illumination conditions and different levels of occlusion, as shown in Figure 1b. Of these, 949 were used for the training set, 119 for the validation set, and 119 for the test set.
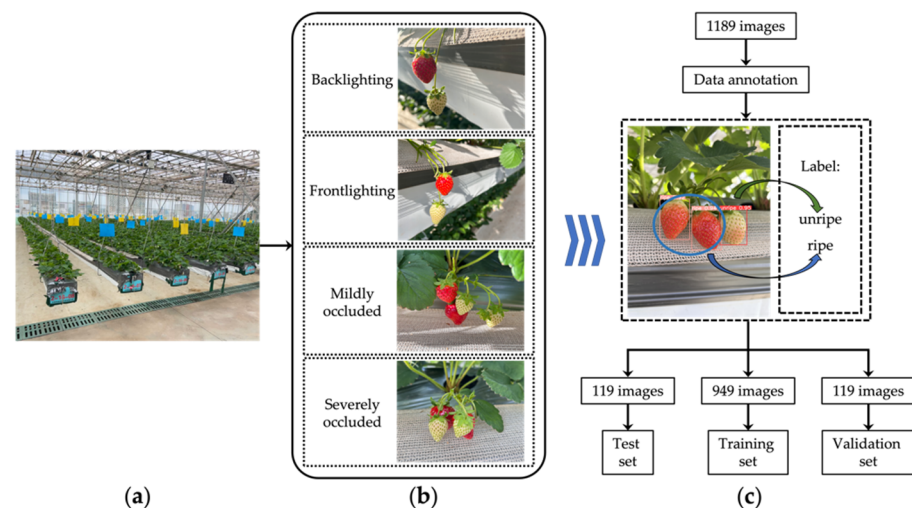

Figure 1. Image acquisition and processing: (**a**) Strawberry plantation. (**b**) Images of strawberries under different light and growth conditions. (**c**) Data annotation and dataset production.

2.1.2. Data Annotation and Dataset Production

The LabelImg labeling tool was used to label each strawberry fruit in 1187 images. Among them, 949 were used for the training set, 119 were used for the validation set, and 119 were used for the test set. The labeling situation and data set allocation are shown in Figure 1c. A ripe strawberry identified by the proposed deep learning model was used for

the data set of the image processing method for 1187 images, which was composed of 742 ripe strawberry images. The ripeness classifications of strawberries are shown in Figure 2.
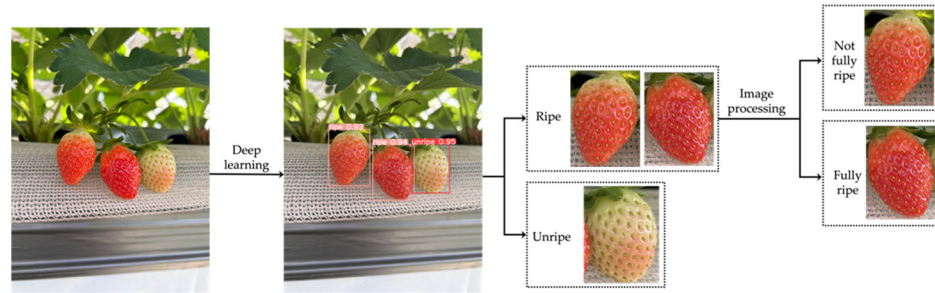

Figure 2. Ripeness classification of strawberries.

2.2. Construction of YOLOv8+ Model

2.2.1. Efficient Channel Attention Module

The efficient channel attention (ECA) module aims to extract inter-channel dependencies by using 1D convolution to promote local cross-channel interactions while avoiding dimensionality reduction [26]. In this study, ECA is added to the backbone network to enable cross-channel extraction of features from different regions of the strawberry, and its processing of the input content is shown in Figure 3.
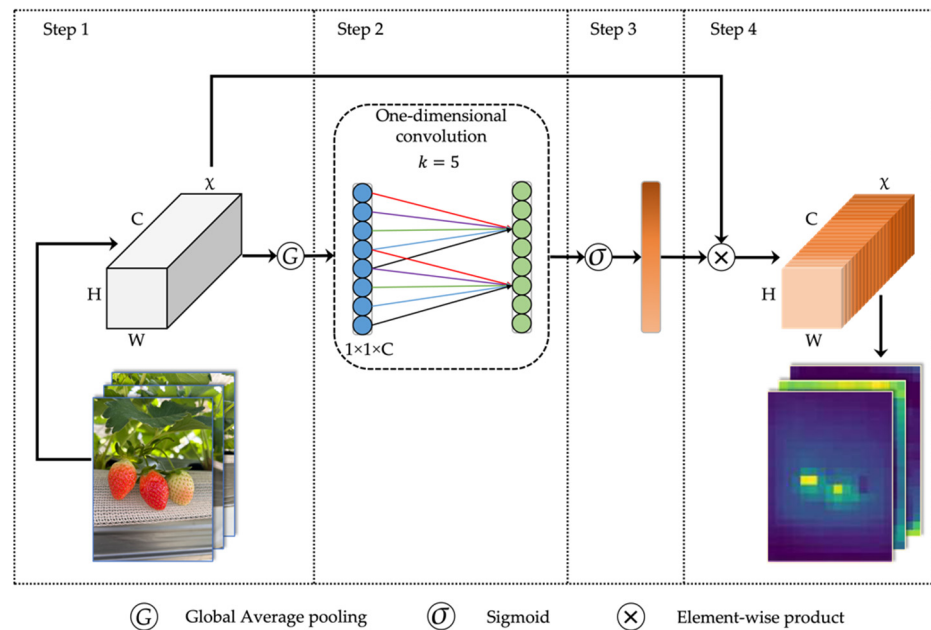

Figure 3. ECA module structure diagram.

In step 1, the input image is convolved with the original convolution to obtain the feature matrix $\chi^{W \times H \times C}$, which is globally average-pooled to capture the channel correlation, and the one-dimensional vector $L^{1 \times 1 \times C}$ is derived, where W, H, and C are the width, height, and channel dimensions of the convolution block, respectively.

In step 2, the approximate range of the channel interaction information (i.e., the size of the convolution kernel k for 1D convolution) needs to be determined before performance of the convolution operation for the input 1D vector. The kernel size, k, of a one-dimensional convolutional kernel is calculated through the mapping of $\psi(C)$ that exists between k and C. K is calculated through a one-dimensional convolutional operation, and k and K are calculated as follows:

$$k = \psi(C) = \left| \frac{\log_2(C)}{a} + \frac{b}{a} \right|_{odd}, \tag{1}$$

$$K = \text{Conv1D}_k(\mathbf{y}), \tag{2}$$

where Conv1D denotes a one-dimensional convolution, k is the kernel size of the one-dimensional convolution Conv1D, and $|t|_{odd}$ denotes the closest odd number to t. In this paper, a and b are set to 2 and 1, respectively.

In step 3, the weight ω of each channel is obtained by the sigmoid activation function σ as shown in the following equation.

$$\omega = \sigma K, \tag{3}$$

In step 4, the weights ω are multiplied with the corresponding elements of the initial input feature map to obtain the final enhanced output feature map.

2.2.2. Focal–EIOU Loss

In target detection, bounding box regression (BBR) is a key step to determine the performance of target localization. Focal–EIOU loss combines EIOU loss and focal loss into a new BBR loss function [27]. The focal–EIOU loss effect is shown in Figure 4.
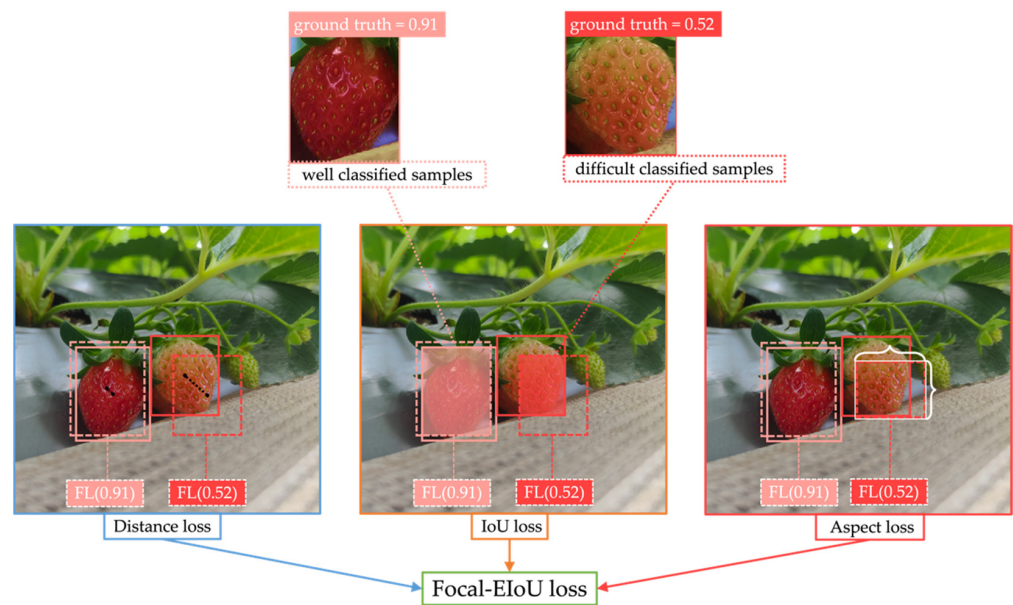

Figure 4. Focal–EIOU loss diagram.

The EIOU loss function consists of three parts: IOU loss L_{IOU}, distance loss L_{dis} and aspect loss L_{asp}. The calculation process of the three parts is shown in Formula (3). The calculation process of L_{EIOU} can be shown in Formula (4).

$$\begin{cases} L_{IOU} = 1 - IOU \\ L_{dis} = 1\text{-}DIOU = 1\text{-}IOU + \frac{\rho^2(\mathbf{b},\mathbf{b^{gt}})}{(w^c)^2 + (h^c)^2} \\ L_{asp} = IOU\text{-}1 + \frac{\rho^2(w,w^{gt})}{(w^c)^2} + \frac{\rho^2(h,h^{gt})}{(h^c)^2} \end{cases}, \tag{4}$$

$$\begin{aligned} L_{EIOU} &= L_{IOU} + L_{dis} + L_{asp} \\ &= 1\text{-}IOU + \frac{\rho^2(\mathbf{b},\mathbf{b^{gt}})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w,w^{gt})}{(w^c)^2} + \frac{\rho^2(h,h^{gt})}{(h^c)^2}, \end{aligned} \tag{5}$$

where $IOU = (A \cap B)/(A \cup B)$, $\mathbf{b}$ and $\mathbf{b^{gt}}$ denote the centroid of the target frame and the anchor frame respectively, w and w^{gt} denote the target box and the anchor box width

respectively, h and h^{gt} denote the target box and the anchor box width respectively, $\rho(\cdot) = \parallel \mathbf{b} - \mathbf{b}^{gt} \parallel_2$ denotes the Euclidean distance, and w^c and h^c are the widths and heights of the smallest outer rectangles of the target and anchor box.

Focal loss sets different weights for samples with different classifying difficulties by introducing the parameter γ. The parameter γ is calculated by the confidence level of the detection, samples with a higher confidence level have a smaller impact on the loss, and samples with a lower confidence level have a larger impact on the loss. The specific calculation process is shown as follows:

$$L_{\text{Focal-EIOU}} = \text{IOU}^{\gamma} L_{\text{EIOU}}, \tag{6}$$

where γ is the parameter used to regulate the sample imbalance problem.

2.2.3. Overall Structure of YOLOv8+

Yolov8 provides n, s, m, l, and x versions. Considering that most agricultural application scenarios require the deployment of network models into embedded devices, we chose the lightest version, Yolov8n, as the baseline model. In this study, we improve on the YOLOv8n model, and the improved YOLOv8+ model identifies strawberries as ripe or unripe, and the specific structure of the network is shown in Figure 5.
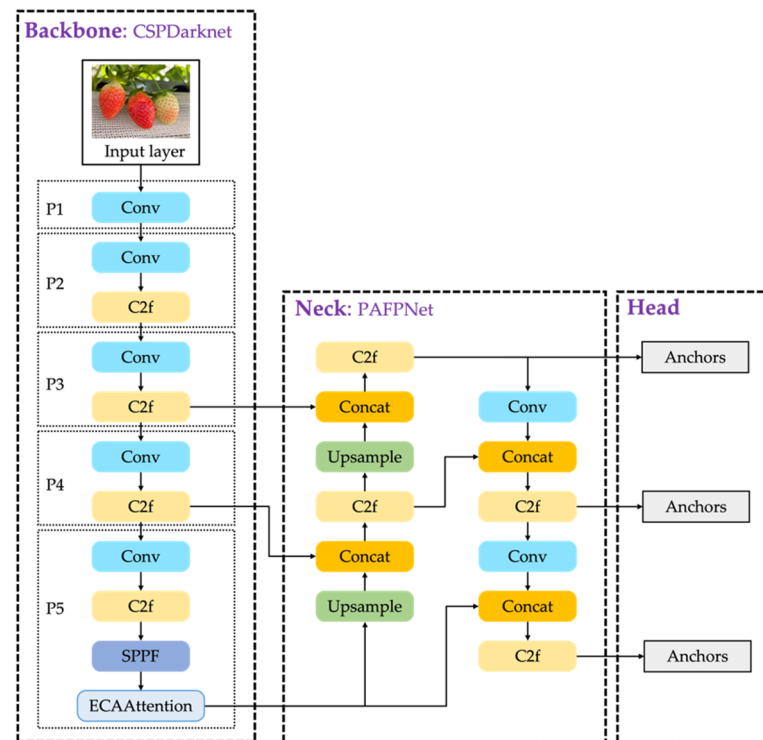

Figure 5. Improved YOLOv8+ structure.

The original strawberry image is input to the backbone network of the YOLO8+ model, and a series of convolutional layers are used to extract the features of strawberry size, shape, and color. The ECA mechanism is added to the backbone network of YOLOv8 to achieve cross-channel extraction of the features of different regions of the strawberries, and after the features are extracted, the preliminary feature map is generated. Then, the preliminary feature map is fed into the neck network of the YOLO8+ model to fuse the target features of strawberry fruit. Finally, the anchors of strawberry fruits with two different ripeness levels are output. At this point, the strawberry fruit target detection and ripeness classify based on YOLOv8+ model is completed.

2.3. Image Processing Method

Strawberries are classified into ripe and unripe by the above method, but in practical applications, ripe strawberries are classified into fully ripe strawberries and not fully ripe strawberries to be processed separately in consideration of transportation and other needs, so the proposed image processing method is used to solve the problem.

2.3.1. Strawberry Centerline Extraction

The strawberry image at the bounding box position in the output identification image is intercepted to obtain the image of all ripe strawberries. Let the width and height of the picture be w and h respectively, and take the three points and the two ends of the contour line segment at the top of the picture. Take three points and two endpoints of the contour line segment at the top of the picture, and the coordinates from left to right are a $(0, 0)$, b $(\frac{w}{4}, 0)$, c $(\frac{w}{2}, 0)$, d $(\frac{3w}{4}, 0)$, e $(w, 0)$. Then take the three points and two endpoints that divide the contour line segment at the bottom of the picture into four equal parts. From left to right, they are e′ $(0, h)$, d′ $(\frac{w}{4}, h)$, c′ $(\frac{w}{2}, h)$, b′ $(\frac{3w}{4}, h)$, a′ (w, h). Connecting a with a′, b with b′, c with c′, d with d′, and e with e′ are five lines, which are candidate line 1, candidate line 2, candidate line 3, candidate line 4, and candidate line 5 of the strawberry centerline, and the specific extraction of the candidate lines is shown in Figure 6a.
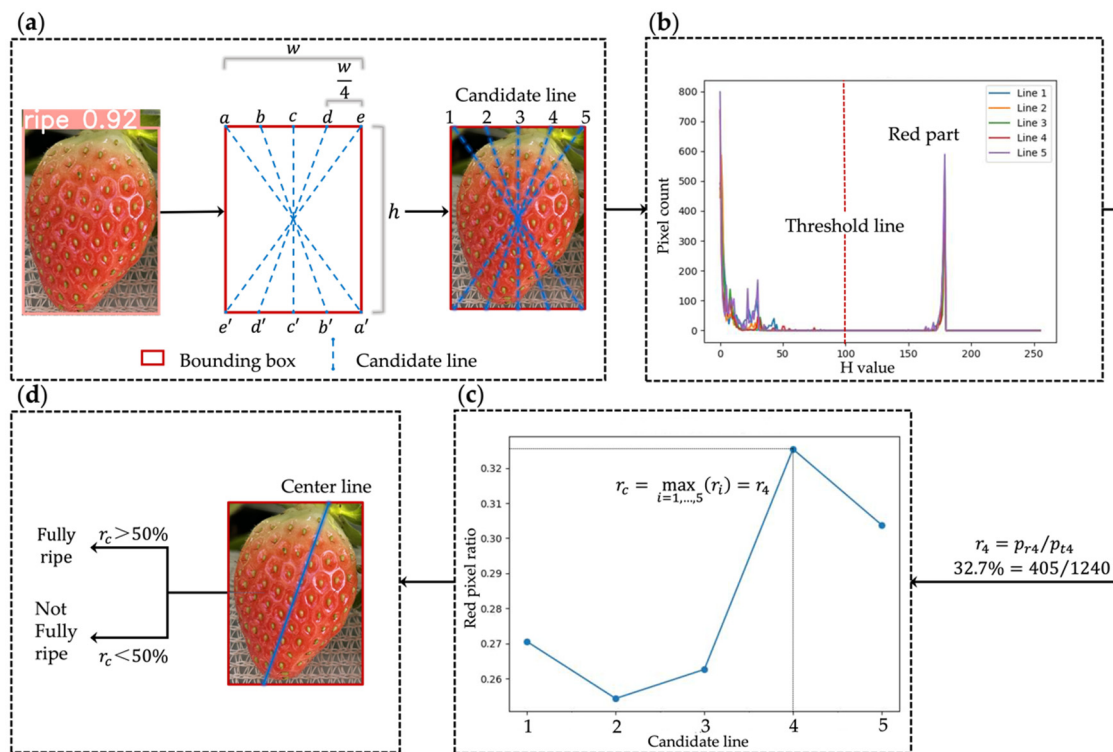

Figure 6. Ripe strawberry classification method: (**a**) Candidate line selection. (**b**) Color histogram of the H component of candidate line 1 to 5. (**c**) Red ratio of candidate line 1 to 5. (**d**) Classification based on the proportion of red pixels in the center line.

In an RGB image taken with a camera, all colors are composed of three color channels, and the percentage of red skin of strawberries cannot be reflected by the R-value alone. The HSV color space can effectively remove the effect of luminance on color by extracting the H color channel. Therefore, the HSV color space is chosen to make the histograms of the five candidate lines in the H component, as shown in Figure 6b. It can be seen that the pixels of the candidate line consist of the red pixels of the strawberry rind and the white-green pixels of the strawberry white rind and background. The number of red pixels p_{ri} in each candidate line can be obtained by setting the threshold: pixels with H > 100 in

the pixel belong to red pixels, and the percentage of red pixels in each candidate line r_i can be obtained.

$$r_i = p_{ri}/p_{ti}, i = 1, \ldots, 5, \tag{7}$$

where p_{ti} represents the number of all pixels in the image and i is the model of the candidate line i.

The line graph of the red pixel percentage r_i of the five candidate lines is shown in Figure 6c. Comparing the size of the red pixel ratio of the five candidate lines in each strawberry picture, the one with the largest ratio is the strawberry center line, and the red pixel ratio r_c of the strawberry center line can be obtained.

$$r_c = \max_{i=1,\ldots,5}(r_i), \tag{8}$$

2.3.2. Ripe Strawberry Classification Method

According to the proportion of red pixels in the center line of strawberry obtained above, the maximum value of the proportion of red pixels in the center line of a not fully ripe strawberry is 47.3%, and the minimum value of the proportion of red pixels in the center line of a fully ripe strawberry is 52.1%.

Analysis of the data indicates that the classification criteria for fully ripe strawberries and not fully ripe strawberries are as follows: strawberries with $r_c < 50\%$ can be classified as not fully ripe, while strawberries with $r_c \geq 50\%$ can be classified as fully ripe, as illustrated in Figure 6d.

2.4. Overall Process of Strawberry Ripeness Identification

The overall process of grading strawberry ripeness by combining deep learning and image processing is shown in Figure 7. The input raw strawberry color image is in in the backbone network to obtain the preliminary feature map. The preliminary feature map is processed in the neck network to get the feature pyramid map. After processing the feature pyramid map in the backbone network, the original image with anchor box is output. The ripe strawberries in the original image are extracted for image processing according to the output anchor boxes, and after the image processing process, the percentage of red pixels in the centerline of strawberries is used as an index to judge the strawberries as fully ripe and not fully ripe. So far, strawberry ripeness is classified as unripe, not fully ripe, or fully ripe in the original strawberry image.
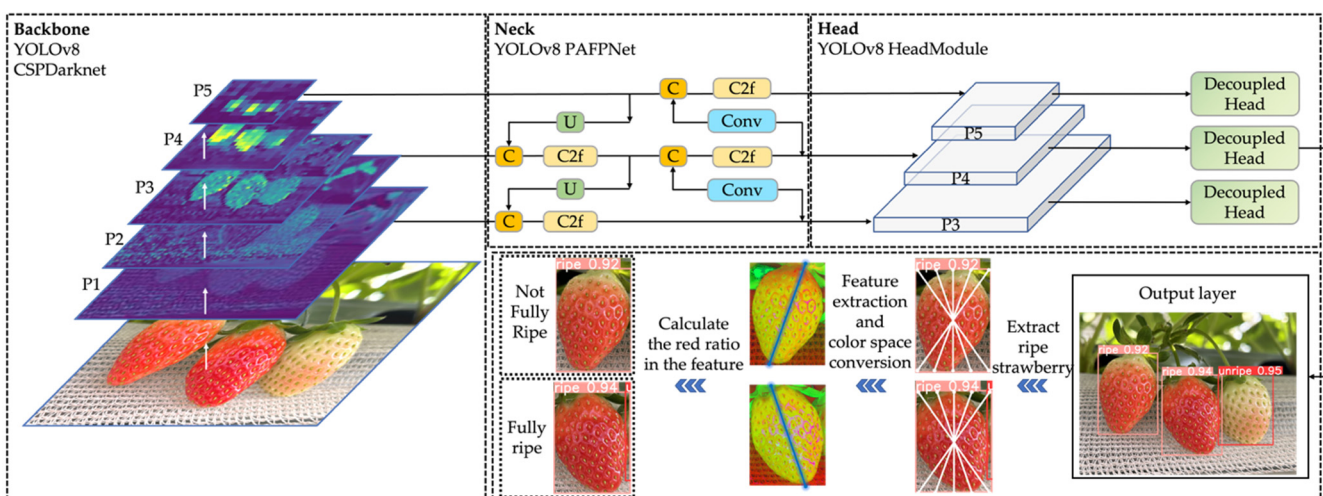

Figure 7. Overall workflow of the proposed method.

2.5. Experiments

In terms of hardware configuration, a computer equipped with an Intel i5-136000kf processor, 32 GB RAM, and a GeForce GTX 4080 GPU was utilized. The computer employed

CUDA 11.2 parallel computing architecture and NVIDIA (Santa Clara, CA, USA) cuDNN 8.0.5 GPU acceleration library. The software simulation was conducted using the Pytorch deep learning framework (Python 3.11 version).

In the first set of experiments, three classic attention mechanisms, efficient channel attention (ECA), squeeze-and-excitation attention (SEA), and shuffle attention (SA), were added to the backbone and head network of the YOLOv8 model. The YOLOv8n model as well as the YOLOv8n model with different attention mechanisms added were trained on the training set, and the performance parameters of the models were recorded. The aim was to compare the impact of these different attention mechanisms on the model's detection performance.

In the second set of experiments, the dataset was trained and identifications were made using the YOLOv8+ model proposed in this study. The proposed model was compared with several common deep learning network models including YOLOv3, YOLOv4, YOLOv5, YOLOv8n, SSD, and Faster-RCNN. This evaluation is performed using the training set.

In the third experimental group, the image processing algorithm introduced in this study was utilized to further categorize the ripe strawberries into partially ripe and fully ripe ones.

2.5.1. Model Performance Evaluation Metrics

To test the performance of the model, F_1 score, mean average precision (mAP), and frames per second (FPS) were selected as the indexes for evaluating the performance of the model, defined as follows:

$$\text{precision} = \frac{TP_1}{TP_1 + FP_1}, \tag{9}$$

$$\text{recall} = \frac{TP_1}{TP_1 + FN_1}, \tag{10}$$

$$F_1\text{-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \tag{11}$$

$$AP = \frac{\sum \text{precision}}{Q}, \tag{12}$$

$$mAP = \frac{\sum_{i=1}^{Q} AP_i}{Q}, \tag{13}$$

where true positive (TP_1) is the number of samples that correctly identified strawberries as unripe and ripe, false positive (FP_1) is the number of samples where labeling box were generated but in the wrong location of the box or the wrong classification of the box, and false negative (FN_1) is the number of samples where no labeling frames were generated in the strawberry labeling region. AP is equal to the area under the precision-recall curve, mAP is the average of the AP, and Q is the number of categories in the training set. There are 2 categories for the detection of strawberry ripeness with deep learning in this study, and Q is 2.

2.5.2. Image Processing Evaluation Metrics

The image processing algorithm classifies the strawberries into fully ripe and not fully ripe strawberries. False positive rate (FPR), false negative rate (FNR), and accuracy are used as the metrics to evaluate the performance of the image processing algorithm. speed(s) is used as the metric to evaluate the speed.

$$FPR = \frac{FP_2}{FP_2 + TN_2}, \tag{14}$$

$$FNR = \frac{FN_2}{FN_2 + TP_2}, \tag{15}$$

$$\text{Accuracy} = \frac{TP_2 + TN_2}{TP_2 + FP_2 + FN_2 + TN_2}, \tag{16}$$

$$\text{speed(s)} = \frac{t}{N}, \tag{17}$$

where true positive (TP_2) is the number of samples that correctly classify strawberries as fully ripe, false positive (FP_2) is the number of samples that are not fully ripe classified as fully ripe, true negative (TN_2) is the number of samples that correctly classify not fully ripe strawberries, and false negative (FN_2) is the number of samples that are fully ripe classified as not fully ripe. N is the number of all samples, i.e., 742 images of ripe strawberries obtained by model identification, N = 742. t is the time taken to classify and process all the images, t = 4.99 s.

3. Results

3.1. Improved Module Performance Comparison

3.1.1. Performance Comparison of Attentional Mechanism

To verify the effectiveness of the attention mechanisms used in this study for model performance enhancement. ECA, SEA, and SA were added to the backbone and head network of the yolov8 model, respectively, to compare the effectiveness of the different attentional mechanisms on the model's detection performance improvement of different attention mechanisms. Figure 8 shows mAP50 changes during training after adding the three classical attention mechanisms at different network of the YOLOv8 model.
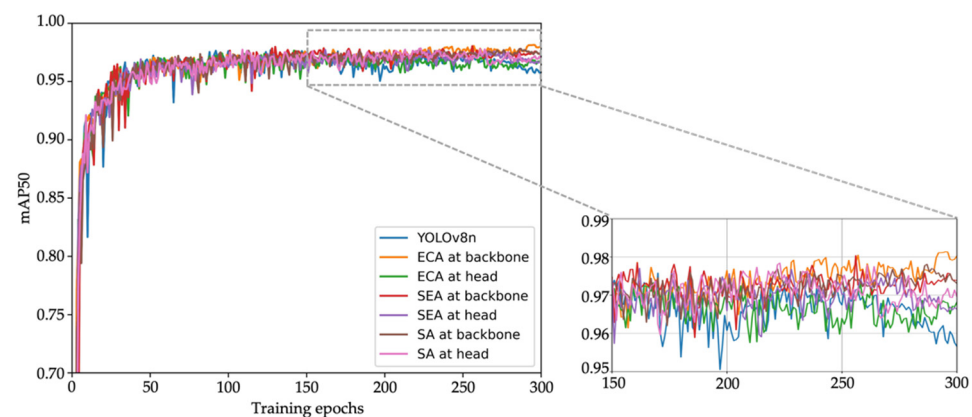

Figure 8. Changes in mAP50 curves for different attention mechanisms added to YOLOv8n model training.

As can be seen in Figure 8, all curves rise in a similar trend. When the attention mechanisms were added to the backbone or head network of YOLOv8, the values of mAP50 that finally converge during the model training process were all higher than those of the YOLOv8n model. Adding the attention mechanisms to the backbone network of the YOLOv8 model instead of the head network during the training process resulted in higher values of mAP50 convergence after 250 epochs. Among them, the highest value of final convergence of the mAP50 curve was obtained when ECA was added to the backbone position of YOLOv8. The comparison results show that the addition of ECA could effectively improve the model to learn the characteristics of strawberry at different ripeness stages.

3.1.2. Performance Comparison with Classic Network Models

To verify the effectiveness of the proposed improved method on the proposed model, the overall performance of the model was compared by training the dataset using different deep learning models. Figure 9 illustrates the changes in mAP50 curves of YOLOv8+, YOLOv8n, YOLOv8s, YOLOv8m, YOLOv5, YOLOv4, YOLOv3, SSD and, Faster RCNN models during training.
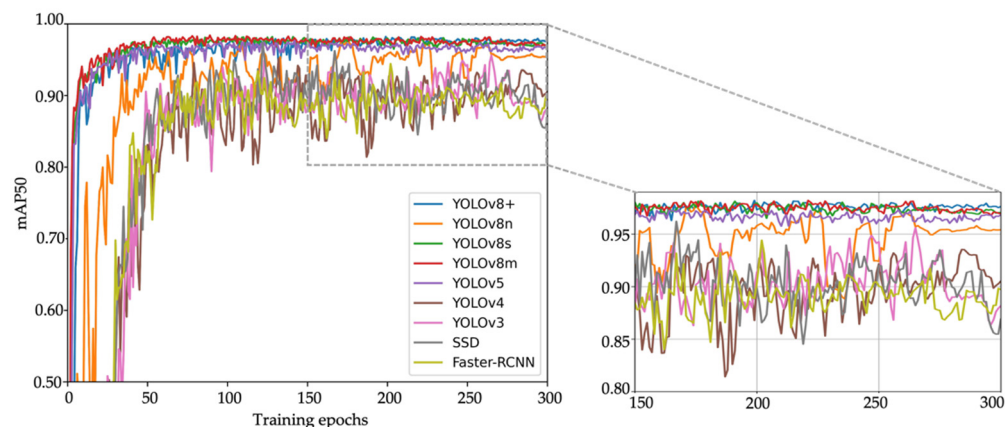
Figure 9. Changes in mAP50 curves during different model training.

As can be seen in Figure 9, the YOLOv8+ model has higher values of final convergence of mAP50 curves during training compared to the YOLOv8 models of different sizes. This indicates that the model could effectively learn the characteristics of strawberry fruits at different ripeness stages, thus showing a relatively stable improvement in strawberry ripeness identification accuracy. The YOLOv5 model eventually converged on a relatively high value but with slight ups and downs in the overall trend. The YOLOv3, YOLOv4, SSD, and Faster RCNN models had large fluctuations in the mAP50 curves during the training process, their performance was not stable enough, and the final mAP50 values were low. The Precision, Recall, and F_1 score obtained by training these models are shown in Table 1.

Table 1. Results of comparison detection between different models.

Methods	Precision (%)	Recall (%)	F_1-Score
YOLOv8+	97.81	96.36	97.07
YOLOv8n	96.54	95.57	96.05
YOLOv8s	96.68	96.02	96.35
YOLOv8m	97.63	98.02	97.82
YOLOv5n	95.84	95.66	95.75
YOLOv4	95.14	92.33	93.71
YOLOv3	95.62	93.21	94.41
SSD	94.47	95.59	95.03
Faster-RCNN	88.12	94.57	91.23

By observing the results in Tables 1 and 2, it could be found that the YOLOv8+ model had a precision value of 97.81%, which is 0.18% to 9.69% higher than the other models. Except for the YOLOv8m model, the recall values of the YOLOv8+ model were higher than those of the other models by 0.34% to 4.03%, while the F_1 score was higher than the other models by 0.72% to 5.84%. The YOLOv8+ model has a slightly lower Recall and F1 score than the YOLOv8m model, but its FPS was much higher on the GPU and CPU than that of the YOLOv8m model.

To verify the ripeness identification ability of the YOLOv8+ model in a realistic environment, the trained model was used to identify strawberry fruits at different ripeness stages. The detection results are shown in Figure 10. (For simplicity, we only show the detection results of YOLOv8+ and YOLOv8n, which are the best performers among the models). A comparison of (a) and (d) shows that YOLOv8+ could accurately identify difficult-to-classify samples with an identification confidence of 0.92 or higher. Comparisons (b) and (e) show that YOLOv8+ could accurately identify the ripening stages of strawberry fruits under different light conditions with an identification confidence level of 0.92 or higher. Comparisons (c) and (f) show that the confidence level for the identification of ripe strawberries in the case of overlapping fruits ranged from 0.43 to 0.83. In conclusion, the improved YOLOv8+ model can effectively localize and identify strawberries as ripe or not under various environmental conditions.

Table 2. Comparison of detection speed between different models.

Methods	Device	Inference Time (s/per Image)	FPS
YOLOv8+	GPU	0.012	83.33
	CPU	0.039	25.64
YOLOv8n	GPU	0.013	76.92
	CPU	0.035	28.57
YOLOv8s	GPU	0.022	45.45
	CPU	0.103	9.70
YOLOv8m	GPU	0.024	41.67
	CPU	0.199	5.02
YOLOv5n	GPU	0.014	71.42
	CPU	0.046	21.73
YOLOv4	GPU	0.018	55.56
	CPU	0.053	18.86
YOLOv3	GPU	0.017	58.82
	CPU	0.049	20.41
SSD	GPU	0.015	66.67
	CPU	0.047	21.27
Faster-RCNN	GPU	0.028	35.71
	CPU	0.205	4.87

3.2. Ripe Strawberry Classification Experiment

The proposed image processing algorithm is utilized to classify the ripe strawberries into fully ripe and not fully ripe, and the results are shown in Table 3. Accuracy, FPR, and FNR were 91.91%, 5.03%, and 14.28%, respectively, and the average processing time per image was 6.7ms. This shows that image processing can classify ripe strawberries effectively and with speed. Analysis of the data shows that misclassification is mainly classifying fully ripe strawberries as not fully ripe strawberries.

The visualization results of image processing are shown in Figure 11. Due to the high FNR value, we selected a part of FN samples for analysis. (a–c) show that lightly occluded ripe strawberries can be classified as fully ripe better. (d–f) show that strawberries obscured by calyx, stems, and leaves, or other fruits are not favorable for judging their ripeness based on their percentage of red color.

Figure 10. Comparison YOLOv8+ with YOLOv8n: (**a**–**c**) YOLOv8+; (**d**–**f**) YOLOv8n.

Table 3. Results for image processing classify ripeness.

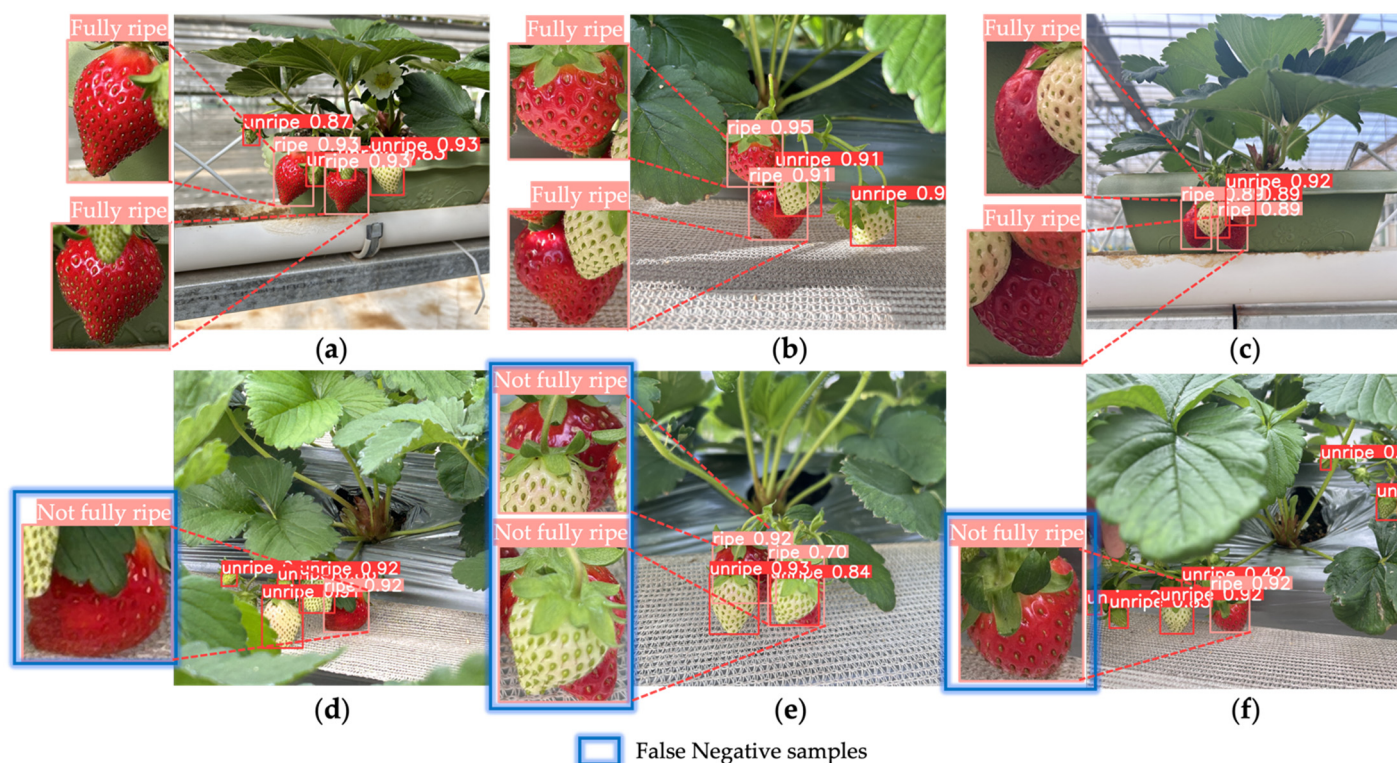| Accuracy (%)	FPR (%)	FNR (%)	Speed (ms)
91.91	5.03	14.28	6.7

Figure 11. Visualization results of image processing: (**a**–**c**) Lightly occluded strawberries; (**d**–**f**) Heavily occluded strawberries.

4. Discussion

In this study, the YOLOv8 model was improved using ECA and focal–EIOU loss, which were used to improve the performance of the model in identifying ripe strawberries. Since the shape characteristics of ripe strawberry fruits do not differ much and it is challenging to identify the ripeness of classified strawberries using the model only, this study also investigated color characteristics.

In Experiment 1, the model performance was compared by comparing ECA, SEA, and SA added to the backbone or head network of the YOLOv8 model. ECA added to the backbone network of the model; the model showed more significant advantages in ripe strawberry identification than the other attentional mechanisms. The reasons why the ECA mechanism was able to work well in identifying ripe strawberries may be as follows:

First, the reddening of strawberries is random, and there is no way to determine where the redness will come first. Ripeness is judged mainly on the basis of the total area of red color, and a strawberry is considered ripe when it reaches a certain level. Therefore, the characteristic of red coloration may exist at a distance rather than continuously. Second, the backbone network is used for feature extraction, and ECA is added to the backbone of the YOLOv8 model to capture long-distance dependency, which better captures strawberry features and improves the accuracy of the YOLOv8 model in identifying ripe and unripe strawberries. Adding an appropriate attention mechanism based on the characteristics of the detection object can enable the network to better capture the characteristics of the detection object and improve the accuracy of the model in detecting the target object. Similar conclusions were obtained in a related study [28–30].

In Experiment 2, focal–EIOU loss was invoked to improve YOLOv8 based on Experiment 1. In a comparison of the YOLOv8+ model with other state-of-the-art detection models, the YOLOv8+ model had the highest values of precision and mAP50, which indicated that the model was able to accurately identify strawberry ripening or not. The model also exhibits the best recall, which suggests that it has a low tendency to miss detections.

The reasons why the model was able to fully detect the ripeness of all strawberries in the image using focal–EIOU loss may be as follows:

Near-ripe strawberries are difficult to identify and classify, and the training process may incorrectly identify ripe as unripe, while ripe strawberries are less difficult to identify. Therefore, the samples are characterized by uneven difficulty of classification. Focal–EIOU loss adds an adjustment factor by modifying the cross-entropy loss. This factor reduces the value of the loss for those samples that have been correctly classified, allowing the training of the model to focus more on samples that are difficult to classify. The use of focal–EIOU loss in the loss function can boost the weight of high-quality bounding boxes, suppress the weight of low-quality bounding boxes, and solve the problem of imbalance between difficult and easy samples. Similar conclusions have been reached in related studies [31,32].

In Experiment 3, strawberries labeled "ripe" were further classified as "not fully ripe" and "fully ripe". The experimental results show that the accuracy and detection time of using image processing to further classify ripe strawberries can meet the needs of real-time detection. Analysis of the data showed that most of the strawberries that were misclassified were fully ripe strawberries that could easily be mistaken for not fully ripe strawberries. Due to the lighter red color of the fully ripe strawberry part of the skin, its H-value may be lower than the set threshold, leading to an error in judgment. By extracting the color channels of the image, the threshold values corresponding to different colors in the image are divided. By obtaining the ratio of pixels corresponding to a specific color to classify the different ripening stages of strawberries. Similar conclusions were obtained in related studies [20].

5. Conclusions

In this paper, we first propose an improved YOLOv8+ model based on YOLOv8, which can accurately and comprehensively identify the ripening stages of strawberry fruits in complex environments. This paper also proposes a method to further classify strawberry ripeness using image processing. The specific conclusions are as follows:

1. Add the ECA mechanism to the backbone of the YOLOv8 model to capture long-distance dependencies and better capture strawberry features for this growth characteristic of strawberries.
2. The use of focal–EIOU loss in the loss function can enhance the weight of high-quality bounding boxes and suppress the weight of low-quality bounding boxes, solving the problem of imbalance between difficult and easy samples.
3. The trained YOLOv8+ model has an accuracy of 97.81%, a recall of 96.36%, and an F1 score of 97.07. It demonstrates comprehensive identification of strawberries at the ripeness stage in complex environments, where complex environments including frontlight, backlight, and occlusion.

These results validate that the proposed YOLOv8+ model combined with image processing can comprehensively and accurately identify strawberries at different ripening stages in complex environments. By using the insights gained from accurately classifying the ripening stages, customized path planning methods can be designed for fruit-picking robots to harvest ripe fruits and optimize the process of efficiently harvesting ripe fruits.

Author Contributions: Conceptualization, C.W., H.W. and Q.H.; methodology, C.W., D.K. and Z.Z.; investigation, Q.H., D.K. and Z.Z.; resources, C.W., D.K. and X.Z.; writing—original draft preparation, H.W.; writing—review and editing, C.W. and Q.H.; project administration, Z.Z. and C.W.; funding acquisition, Z.Z. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in https://github.com/haomw01/strawberry_ripeness_classification_based_on_YOLOv8.

Conflicts of Interest: We declare that we do not have any commercial or associative interests that represent any conflicts of interest in connection with the work submitted.

References

1. Zhu, H.; Liu, X.; Zheng, H.; Yang, L.; Li, X.; Han, Z. Identifying strawberry appearance quality based on unsupervised deep learning. *Precis. Agric.* **2024**, *25*, 614–632. [CrossRef]
2. Soode-Schimonsky, E.; Richter, K.; Weber-Blaschke, G. Product environmental footprint of strawberries: Case studies in Estonia and Germany. *J. Environ. Manag.* **2017**, *203*, 564–577. [CrossRef] [PubMed]
3. Anjom, F.K.; Vougioukas, S.G.; Slaughter, D.C. Development and application of a strawberry yield-monitoring picking cart. *Comput. Electron. Agric.* **2018**, *155*, 400–411. [CrossRef]
4. Wang, C.; Li, C.; Han, Q.; Wu, F.; Zou, X. A Performance Analysis of a Litchi Picking Robot System for Actively Removing Obstructions, Using an Artificial Intelligence Algorithm. *Agronomy* **2023**, *13*, 2795. [CrossRef]
5. Tang, Y.; Qi, S.; Zhu, L.; Zhuo, X.; Zhang, Y.; Meng, F. Obstacle Avoidance Motion in Mobile Robotics. *J. Syst. Simul.* **2024**, *36*, 1–26.
6. Ye, L.; Wu, F.; Zou, X.; Li, J. Path planning for mobile robots in unstructured orchard environments: An improved kinematically constrained bi-directional RRT approach. *Comput. Electron. Agric.* **2023**, *215*, 108453. [CrossRef]
7. Du, X.; Cheng, H.; Ma, Z.; Lu, W.; Wang, M.; Meng, Z.; Jiang, C.; Hong, F. DSW-YOLO: A detection method for ground-planted strawberry fruits under different occlusion levels. *Comput. Electron. Agric.* **2023**, *214*, 108304. [CrossRef]
8. Wu, F.; Yang, Z.; Mo, X.; Wu, Z.; Tang, W.; Duan, J.; Zou, X. Detection and counting of banana bunches by integrating deep learning and classic image-processing algorithms. *Comput. Electron. Agric.* **2023**, *209*, 107827. [CrossRef]
9. Zhou, X.; Lee, W.S.; Ampatzidis, Y.; Chen, Y.; Peres, N.; Fraisse, C. Strawberry Maturity Classification from UAV and Near-Ground Imaging Using Deep Learning. *Smart Agric. Technol.* **2021**, *1*, 100001. [CrossRef]
10. Yu, Y.; Zhang, K.; Yang, L.; Zhang, D. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Comput. Electron. Agric.* **2019**, *163*, 104846. [CrossRef]
11. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
12. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
13. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
14. Phan, Q.-H.; Nguyen, V.-T.; Lien, C.-H.; Duong, T.-P.; Hou, M.T.-K.; Le, N.-B. Classification of Tomato Fruit Using Yolov5 and Convolutional Neural Network Models. *Plants* **2023**, *12*, 790. [CrossRef] [PubMed]
15. Azadnia, R.; Fouladi, S.; Ahmad Jahanbakhshi, A. Intelligent detection and waste control of hawthorn fruit based on ripening level using machine vision system and deep learning techniques. *Results Eng.* **2023**, *17*, 100891. [CrossRef]
16. Yang, S.; Wang, W.; Gao, S.; Deng, Z. Strawberry ripeness detection based on YOLOv8 algorithm fused with LW-Swin Transformer. *Comput. Electron. Agric.* **2023**, *215*, 108360. [CrossRef]
17. Chen, S.; Xiong, J.; Jiao, J.; Xie, Z.; Huo, Z.; Hu, W. Citrus fruits maturity detection in natural environments based on convolutional neural networks and visual saliency map. *Precis. Agric.* **2022**, *23*, 1515–1531. [CrossRef]
18. Zhang, J.; Xie, J.; Zhang, F.; Gao, J.; Yang, C.; Song, C.; Rao, W.; Zhang, Y. Greenhouse tomato detection and pose classification algorithm based on improved YOLOv5. *Comput. Electron. Agric.* **2024**, *216*, 108519. [CrossRef]
19. Azarmdel, H.; Jahanbakhshi, A.; Mohtasebi, S.S.; Muñoz, A.R. Evaluation of image processing technique as an expert system in mulberry fruit grading based on ripeness level using artificial neural networks (ANNs) and support vector machine (SVM). *Postharvest Biol. Technol.* **2020**, *166*, 111201. [CrossRef]
20. Alfatni, M.S.M.; Khairunniza-Bejo, S.; Marhaban, M.H.B.; Saaed, O.M.B.; Mustapha, A.; Shariff, A.R.M. Towards a Real-Time Oil Palm Fruit Maturity System Using Supervised Classifiers Based on Feature Analysis. *Agriculture* **2022**, *12*, 1461. [CrossRef]
21. Lu, J.; Sang, N.; Hu, Y.; Fu, H. Detecting citrus fruits with highlight on tree based on fusion of multi-map. *J. Light-Electronoptic* **2014**, *125*, 1903–1907. [CrossRef]
22. Castro, W.; Oblitas, J.; De-La-Torre, M.; Cotrina, C.; Bazán, K.; Avila-George, H. Classification of Cape Gooseberry Fruit According to its Level of Ripeness Using Machine Learning Techniques and Different Color Spaces. *IEEE Access* **2019**, *7*, 27389–27400. [CrossRef]
23. Ropelewska, E.; Rutkowski, K.P. The Classification of Peaches at Different Ripening Stages Using Machine Learning Models Based on Texture Parameters of Flesh Images. *Agriculture* **2023**, *13*, 498. [CrossRef]
24. Wang, D.; Wang, X.; Chen, Y.; Wu, Y.; Zhang, X. Strawberry ripeness classification method in facility environment based on red color ratio of fruit rind. *Comput. Electron. Agric.* **2023**, *214*, 108313. [CrossRef]
25. Tang, C.; Chen, D.; Wang, X.; Ni, X.; Liu, Y.; Liu, Y.; Mao, X.; Wang, S. A fine recognition method of strawberry ripeness combining Mask R-CNN and region segmentation. *Front. Plant Sci.* **2023**, *14*, 1211830. [CrossRef] [PubMed]

26. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.

27. Zhang, Y.-F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146–157. [CrossRef]

28. Zhang, Y.; Ma, B.; Hu, Y.; Li, C.; Li, Y. Accurate cotton diseases and pests detection in complex background based on an improved YOLOX model. *Comput. Electron. Agric.* **2022**, *203*, 107484. [CrossRef]

29. Yang, G.; Wang, J.; Nie, Z.; Yang, H.; Yu, S. A Lightweight YOLOv8 Tomato Detection Algorithm Combining Feature Enhancement and Attention. *Agronomy* **2023**, *13*, 1824. [CrossRef]

30. Solimani, F.; Cardellicchio, A.; Dimauro, G.; Petrozza, A.; Summerer, S.; Cellini, F.; Renò, V. Optimizing tomato plant phenotyping detection: Boosting YOLOv8 architecture to tackle data complexity. *Comput. Electron. Agric.* **2024**, *218*, 108728. [CrossRef]

31. Gong, X.; Zhang, X.; Zhang, R.; Wu, Q.; Wang, H.; Guo, R.; Chen, Z. U3-YOLOXs: An improved YOLOXs for Uncommon Unregular Unbalance detection of the rape subhealth regions. *Comput. Electron. Agric.* **2022**, *203*, 107461. [CrossRef]

32. Shi, T.; Ding, Y.; Zhu, W. YOLOv5s_2E: Improved YOLOv5s for Aerial Small Target Detection. *IEEE Access* **2023**, *11*, 80479–80490. [CrossRef]