

Article

CES-YOLOv8: Strawberry Maturity Detection Based on the Improved YOLOv8

Yongkuai Chen ^{1,†}, Haobin Xu ^{1,2,†}, Pengyan Chang ¹, Yuyan Huang ¹, Fenglin Zhong ², Qi Jia ⁴, Lingxiao Chen ⁵, Huaiqin Zhong ^{3,*} and Shuang Liu ^{2,*}

¹ Institute of Digital Agriculture, Fujian Academy of Agricultural Sciences, Fuzhou 350003, China; stonecyk@126.com (Y.C.); 52303030056@fafu.edu.cn (H.X.); cpy155165@163.com (P.C.); 18759871027@163.com (Y.H.)

² College of Horticulture, Fujian Agriculture and Forestry University, Fuzhou 350002, China; zhong591@fafu.edu.cn

³ Crops Research Institute, Fujian Academy of Agricultural Sciences, Fuzhou 350003, China

⁴ Jiuquan Academy of Agriculture Sciences, Jiuquan 735099, China; jq17793520924@163.com

⁵ Fujian Agricultural Machinery Extension Station, Fuzhou 350002, China; snjztgk@163.com

* Correspondence: zhqeast@163.com (H.Z.); liush@fafu.edu.cn (S.L.)

† These authors contributed to the work equally and should be regarded as co-first authors.

Abstract: Automatic harvesting robots are crucial for enhancing agricultural productivity, and precise fruit maturity detection is a fundamental and core technology for efficient and accurate harvesting. Strawberries are distributed irregularly, and their images contain a wealth of characteristic information. This characteristic information includes both simple and intuitive features, as well as deeper abstract meanings. These complex features pose significant challenges to robots in determining fruit ripeness. To increase the precision, accuracy, and efficiency of robotic fruit maturity detection methods, a strawberry maturity detection algorithm based on an improved CES-YOLOv8 network structure from YOLOv8 was developed in this study. Initially, to reflect the characteristics of actual planting environments, the study collected image data under various lighting conditions, degrees of occlusion, and angles during the data collection phase. Subsequently, parts of the C2f module in the YOLOv8 model's backbone were replaced with the ConvNeXt V2 module to enhance the capture of features in strawberries of varying ripeness, and the ECA attention mechanism was introduced to further improve feature representation capability. Finally, the angle compensation and distance compensation of the SIoU loss function were employed to enhance the IoU, enabling the rapid localization of the model's prediction boxes. The experimental results show that the improved CES-YOLOv8 model achieves an accuracy, recall rate, mAP50, and F1 score of 88.20%, 89.80%, 92.10%, and 88.99%, respectively, in complex environments, indicating improvements of 4.8%, 2.9%, 2.05%, and 3.88%, respectively, over those of the original YOLOv8 network. This algorithm provides technical support for automated harvesting robots to achieve efficient and precise automated harvesting. Additionally, the algorithm is adaptable and can be extended to other fruit crops.

Keywords: automatic harvesting robots; CES-YOLOv8; strawberry maturity



Citation: Chen, Y.; Xu, H.; Chang, P.; Huang, Y.; Zhong, F.; Jia, Q.; Chen, L.; Zhong, H.; Liu, S. CES-YOLOv8: Strawberry Maturity Detection Based on the Improved YOLOv8. *Agronomy* **2024**, *14*, 1353. <https://doi.org/10.3390/agronomy14071353>

Academic Editors: Yanbo Huang and Alberto San Bautista

Received: 11 May 2024

Revised: 26 May 2024

Accepted: 20 June 2024

Published: 22 June 2024



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

1. Introduction

With the dual pressures of global population growth and a gradual reduction in arable land, increasing agricultural production has become an important societal challenge. The implementation of smart agriculture is a key solution to this challenge, in which the use of digital information technology and intelligent equipment is crucial for achieving efficient and sustainable agricultural development [1]. Among the many applications of smart agriculture, automated harvesting robot technology can replace manual labor, significantly increasing harvesting efficiency, which is especially important in regions with high labor

costs or labor shortages. Fruit maturity detection is a fundamental and critical technology for the efficient and accurate performance of automated harvesting robots.

Traditional automated harvesting systems mostly rely on simple color and size recognition for determining fruit maturity. Yamamoto and others proposed an algorithm based on color threshold segmentation to isolate strawberry targets [2]. Hayashi and others designed a strawberry-harvesting robot that also uses a color threshold segmentation algorithm to detect strawberries and estimate maturity [3]. Kaur and others utilized external quality features such as color, texture, and size to detect the maturity of plums [4]. Villaseñor-Aguilar and others proposed a new fuzzy classification framework based on the RGB color model to categorize the maturity of tomatoes [5]. Although these methods have resolved the maturity detection issue to some extent, they have stringent requirements for the detection environment and growth conditions. However, in actual production, the fruit maturation process is influenced by many factors, such as the fruit variety and growth conditions (such as light and humidity), which affect the color at maturity. Moreover, the color of the fruit may also change due to shading, pests, and diseases, among other reasons [6], which can affect color recognition accuracy.

By automatically learning the intrinsic connections and patterns within annotated datasets, deep learning technologies can be used to effectively extract deep features from images; they exhibit especially high accuracy and rapid identification in complex scene target detection and classification. In recent years, deep learning technologies have been rapidly integrated into various agricultural research fields, including fruit maturity detection in complex environments. Subramanian Parvathi and others have improved the region-based Faster R-CNN model for detecting the maturity of coconuts against complex backgrounds [7]. Wang Lingmin and colleagues utilized an enhanced AlexNet model to classify the maturity of bananas, achieving an accuracy of 96.67% [8]. Zan Wang and associates designed an improved Faster R-CNN model, MatDet, for the detection of tomato maturity. Experimental results indicate that, in complex scenes, the proposed model achieved optimal detection results under conditions of branch occlusion, fruit overlapping, and lighting effects, with a mean average precision (mAP) of 96.14% [9]. Chen Fengjun and colleagues proposed an improved method for detecting the maturity of olive fruits using EfficientDet, and the model's precision P, recall rate R, and mean average precision mAP in the test set were 92.89%, 93.59%, and 94.60%, respectively [10]. Wang Congyue and others introduced an enhanced object detection algorithm based on YOLOv5n for the real-time identification and maturity detection of cherry tomatoes, achieving an average accuracy of 95.2% [11]. Elizabeth Haruna Kazama and others used an enhanced YOLOv8 model modified through convolution blocks (RFCACnv) to classify the maturity stages of coffee fruits, with the model reaching an mAP@0.50 of 74.20% [12]. Megalingam, Rajesh Kannan, and colleagues proposed an integrated fuzzy deep learning model (IFDM) for classifying the maturity levels of coconuts. The study showed that the real-time learning model achieved an accuracy of 86.3% in classifying coconut maturity levels [13]. Currently, fruit maturity detection methods based on convolutional neural networks have rapidly developed, yet issues remain. Methods with high detection accuracy often have high computational complexity and slow detection speeds, while methods that are computationally simpler and faster tend to have lower accuracy [14].





To address the aforementioned issues, this study used strawberries as its research subject and basis for improvements on the YOLOv8 object detection network, proposing a novel strawberry ripeness detection algorithm named CES-YOLOv8 to enhance the accuracy of ripeness detection. The algorithm enhances the accuracy and robustness of strawberry ripeness recognition via automated harvesting robots under various environmental conditions without sacrificing real-time processing capabilities, providing technical support for efficient and precise automated harvesting. This research not only helps enhance the practicality and economic benefits of automated harvesting technology but also offers technical references for smart agriculture in precision agricultural management, harvesting, and sorting.

2. Materials and Methods

2.1. Classification of Strawberry Ripeness

In this study, strawberry ripeness was classified into four levels based on the growth and color changes of the fruit, as shown in Table 1.

Table 1. Classification of strawberry maturity levels.

Grade	Label	Description	Image
1	Immature_stage	Fruit remains green	
2	Mature_white_stage	Fruit begins to change from green to white; some varieties start to show light red spots	
3	Color_turning_stage	Red starts to spread and cover more of the fruit surface, but there are still areas that are white or light red	
4	Ripe_stage	The color of the strawberries uniformly turns bright red	

Level 1 is the unripe stage, characterized by hard flesh and a green surface on the skin. Level 2 is the white ripe stage, where the fruit begins to change from green to white, with pale red spots starting to appear in some areas, and Level 3 is the color-changing stage, where the color change of the fruit becomes more pronounced, with the red coloring starting to spread and cover more of the fruit surface, although some areas remain white or pale red.

2.2. Image Collection and Dataset Construction

The experimental data for this study were collected from the China Israel Demonstration Farm polytunnel greenhouse of the Fujian Academy of Agricultural Sciences. The strawberry varieties targeted in this experiment included Hongyan, Xiangye, and Yuexiu, among others.

2.2.1. Strawberry Image Acquisition

Strawberry image collection ensured that the dataset reflected the actual planting environment characteristics, such as irregular distribution, uneven lighting, and mutual occlusion between leaves and fruits. The collection times were set from 8:00 A.M. to 5:00 P.M., and a Sony (A550) camera was used to capture images of strawberries from various planting rack positions in the greenhouse without specifying the shooting angles. All strawberry images were captured under natural light conditions, and they totaled

546 original images with varying lighting conditions, degrees of occlusion, and angles, as shown in Figure 1.

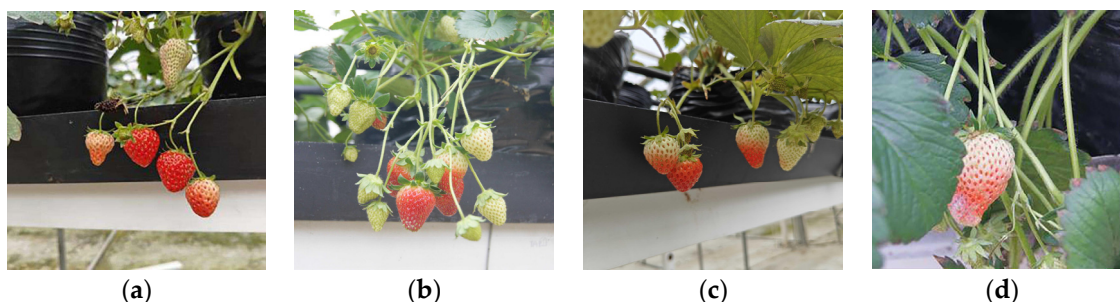


Figure 1. Captured images of strawberries. (a) Dispersed, (b) with light + overlapping fruits, (c) leaves blocking light, and (d) leaves obstructing.

2.2.2. Data Enhancement and Expansion

If the training sample size is insufficient during the model training process, it may cause overfitting [15], due to which the model performs well with the training data but poorly with the new test data. This is due to the insufficiency of training data causing the model to over-adapt to the characteristics of the training data and fail to generalize to other data. To increase feature diversity in order to prevent overfitting, achieve model convergence during training, and enhance the model's generalization ability and robustness, a series of data augmentation measures were implemented to expand the sample size. This study employed a random combination of data augmentation techniques such as mirroring, brightness adjustment, Gaussian blur, contrast adjustment, and random translation, with some of the augmented samples shown in Figure 2. This approach effectively expanded the dataset size, ultimately obtaining 2722 images that had undergone enhancement processing.

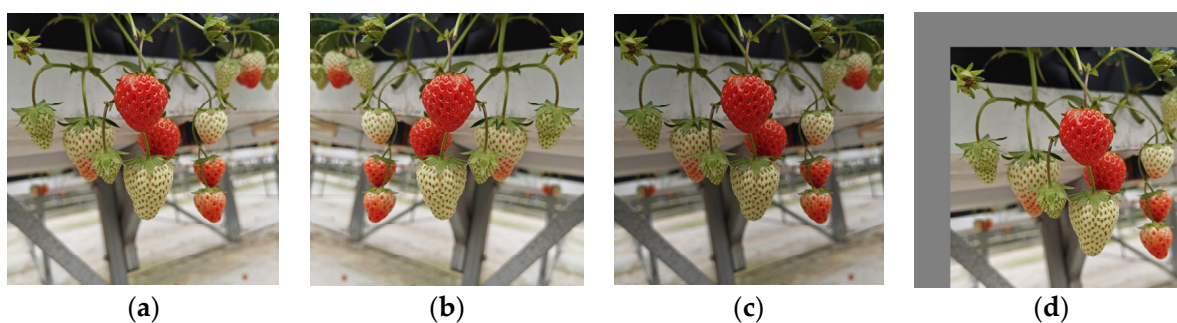


Figure 2. Data augmentation effects. (a) Original image, (b) mirrored flip, (c) contrast–brightness adjustment, and (d) random translation.

2.2.3. Data Labeling and Dataset Segmentation

Labeling software (labelImg v1.8.1) was used to annotate the strawberries with the following rules: (1) the smallest circumscribing rectangle was used as the annotation box, ensuring that the target was completely within the box and close to the boundary; (2) each target was marked via an independent box, with no sharing of boxes among multiple targets; (3) fruits that obscured each other in the image but did not affect the manual determination of ripeness levels were annotated separately; and (4) fruits that were heavily occluded (over 95% obscured) or too distant, causing the main body of the fruits to be at the edge and severely blurred, making it difficult to discern ripeness levels, were not labeled, with some labeled images shown in Figure 3. The annotation results were saved in a .txt file in YOLO format. The dataset was divided into training and test sets at a 4:1 ratio, resulting in 2177 images for training and 545 for testing.

sparse convolutions to process only the visible parts of the input, reducing the pretraining computational costs and allowing the model to use the remaining contextual information to predict missing parts, thus enhancing its ability to learn and understand visual data. Additionally, a global response normalization (GRN) layer is introduced in the convolutional network to enhance feature competition between the channels. The GRN enhances feature contrast, and through selective steps of global feature aggregation, normalization, and calibration, it helps prevent feature collapse and thereby improves the model’s expressive and generalization capabilities [19]. This module enhances the performance of pure convolutional neural networks in various downstream tasks, and the module structure is shown in Figure 6.

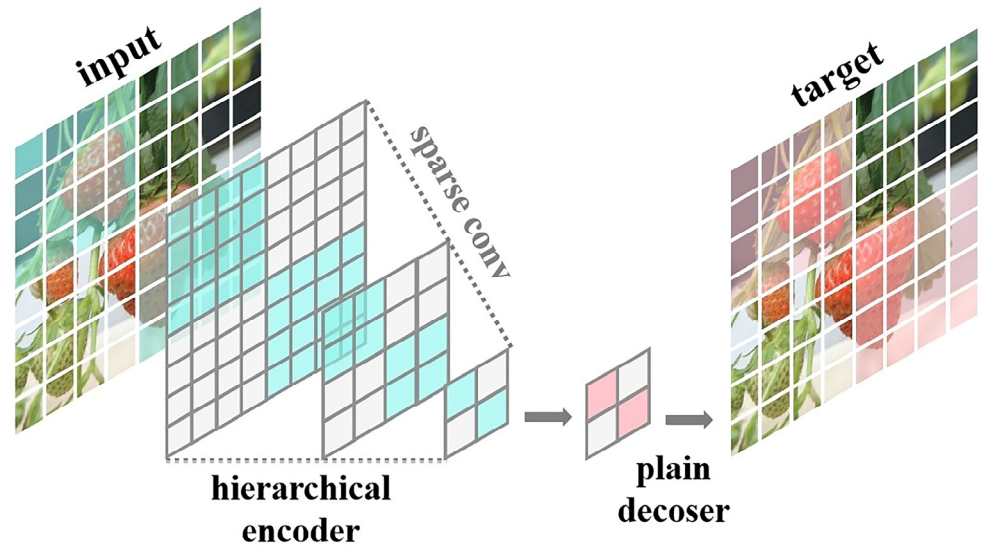


Figure 5. FCMAE full convolutional mask autoencoder.

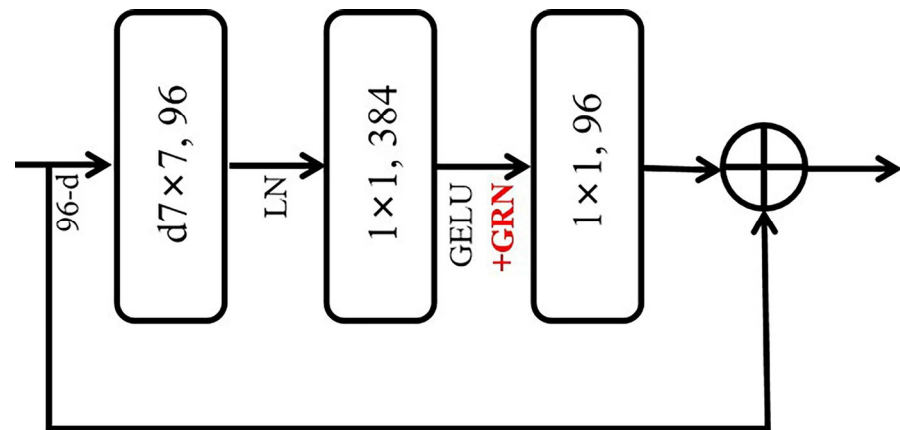


Figure 6. ConvNeXt V2 module.

In the detection of strawberry ripeness, ConvNeXt V2 randomly masks parts of the strawberry image. Through processing with sparse convolution, it predicts the masked areas to capture details within the strawberry image, accurately capturing features while reducing computational costs without sacrificing performance. Concurrently, the GRN layer enhances the competition among feature channels, helping the model better distinguish subtle differences between strawberries of different maturities, thus improving recognition accuracy.

2.3.2. ECA Attention Mechanism

Attention mechanisms dynamically adjust the weights of the input features within a network [20], enabling better perception of the distinctive features in images and facilitating rapid target localization. This mechanism has been widely adopted in computer vision. The efficient channel attention (ECA) module (Figure 7) avoids the dimension reduction found in the squeeze-and-excitation (SE) module. It learns channel attention directly after global average pooling using a one-dimensional convolution, maintaining the dimensionality of the channels [21]. A key feature of the ECA module is its adaptive method for determining the size (k) of the one-dimensional convolutional kernel, which aligns the local cross-channel interaction range with the channel dimensions, facilitating efficient learning without manual adjustments. Due to its light weight and minimal additional parameters, the ECA module significantly reduces model complexity while maintaining performance.

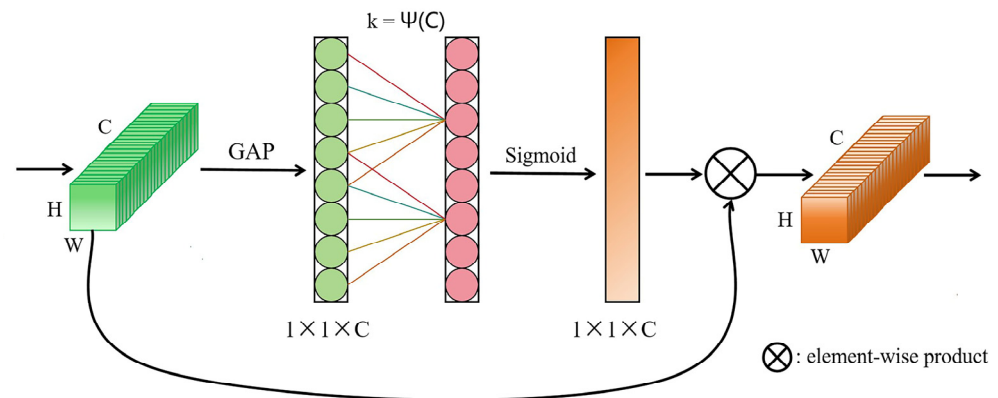


Figure 7. ECA attention mechanism structure.

In this study, an ECA attention mechanism was added above the SPPF layer of the backbone network. The ECA attention mechanism avoids dimensionality reduction, preserving more original feature information of strawberries at different maturity levels, thereby enhancing feature-representation capabilities. Local interactions of one-dimensional convolution enable the model to focus more on key feature areas related to maturity and automatically adjust the range of the receptive field based on different feature layers, allowing the model to flexibly handle changes in the strawberry-ripening process.

2.3.3. SIoU Loss Function

The CIoU loss function primarily relies on the aggregation of bounding box regression, but it overlooks the misalignment issue between the expected true boxes and predicted boxes [22], a flaw that can slow convergence and reduce training efficiency.

To address the issue of IoU calculations when ground truth boxes and predicted boxes overlap, this study employed the SIoU loss function proposed by Gevorgyan [23]. As shown in Figure 8, in addition to considering the distance σ , overlap area, and aspect ratio between the predicted boxes, the SIoU loss function also takes into account the vector angles α and β between the true box, B^{GT} , and the predicted box, Box B. It incorporates an angular penalty term, redefining the related loss function [24]. In the detection of strawberry ripeness, the angle cost introduced via SIoU can reduce the distance between prediction and ground truth boxes by optimizing angles, thereby indirectly improving IoU. Additionally, the distance cost, aside from considering straight-line distances, also adjusts angles, increasing the intersection between prediction and ground truth boxes over a wider range. Finally, SIoU dynamically adjusts the weights of distance and angle to accommodate different overlaps of strawberries, preventing gradient-vanishing issues.

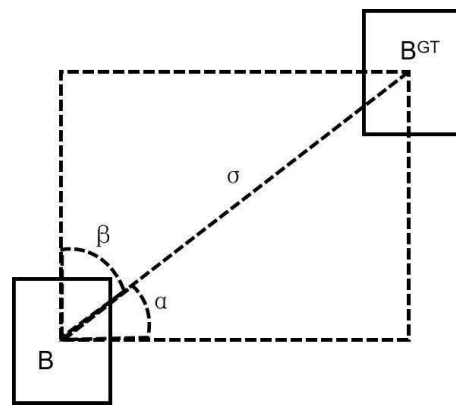


Figure 8. Schematic diagram of SIoU loss function.

2.4. Model Evaluation Metrics

The process used to determine the ripeness of strawberry fruits must consider both detection accuracy and performance. For model detection accuracy, precision, recall, and the F1 score are used as evaluation metrics. For model performance, the mAP50, which is the mean average precision at a threshold of 50%, was selected as the evaluation metric. The formula is as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (2)$$

$$\text{F1} = \frac{2 * \text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})} \times 100\% \quad (3)$$

$$\text{AP} = \frac{\sum_1^k P \times R}{K} \times 1 \quad (4)$$

$$\text{mAP} = \frac{\sum_1^k \text{AP}}{k} \times 1 \quad (5)$$

TP (true positives) represents the number of actual positive samples predicted as positive. FP (false positives) represents the number of actual negative samples predicted as positive. FN (false negatives) represents the number of actual positive samples predicted as negative. TN (true negatives) represents the number of actual negative samples predicted as negative.

3. Experiments and Result Analysis

3.1. Experimental Environment Configuration and Network Parameters

The training and testing of this study's model were performed on a computer equipped with an Intel Core i7-13700K CPU at 3.4 GHz, 32 GB of RAM, and a Windows 10 (64-bit) operating system accelerated by a GeForce RTX 4070 Ti GPU with 12 GB of VRAM. The programming language used was Python 3.8.10, the deep learning framework was PyTorch 1.2.0, and the OpenCV version was 4.8. The initial learning rate was set to 0.001 to balance the model's convergence speed and learning efficiency, preventing instability due to rapid convergence. Additionally, a momentum decay strategy was employed, with a value set to 0.937, to speed up the learning process and avoid local minima. Finally, to enhance the model's generalization capability, a weight decay of 0.0005 was set, which helped reduce the risk of overfitting.

3.2. CES-YOLOv8 Model Experiments

To validate the performance of the CES-YOLOv8 model, 545 strawberry images from the test set were evaluated. Table 2 presents the detection results of the algorithm at

different maturity levels. According to Table 2, the algorithm achieved a precision of 88.2%, a recall of 89.8%, an mAP50 of 92.10%, and an F1 score of 88.99%.

Table 2. Detection results for different maturity levels.

Maturity Level	Precision	Recall	mAP50	F1 Score
Immature_stage	80.80%	75.80%	82.70%	78.22%
Mature_white_stage	88.10%	91.80%	93%	89.91%
Color_turning_stage	90.10%	94.30%	94.60%	92.15%
Ripe_stage	93.60%	97.10%	98.10%	95.32%
Average	88.20%	89.80%	92.10%	88.99%

The improved algorithm further integrates the positional and semantic information of occluded fruits, enabling the extraction of the fine-grained features of the fruit phenotypes for the accurate detection of fruits at different maturity levels. Figure 9 clearly shows that the algorithm can accurately detect strawberries of different maturity levels in images with single and multiple targets and overlapping occlusions. Figure 9d shows that the algorithm can accurately identify small targets and severely occluded strawberries that have fallen onto planting racks. In summary, the improved CES-YOLOv8 model can accurately detect the maturity of fruits, exhibiting good detection performance for small targets, multiple targets, foliage occlusion, heavy fruiting, and varying lighting conditions.

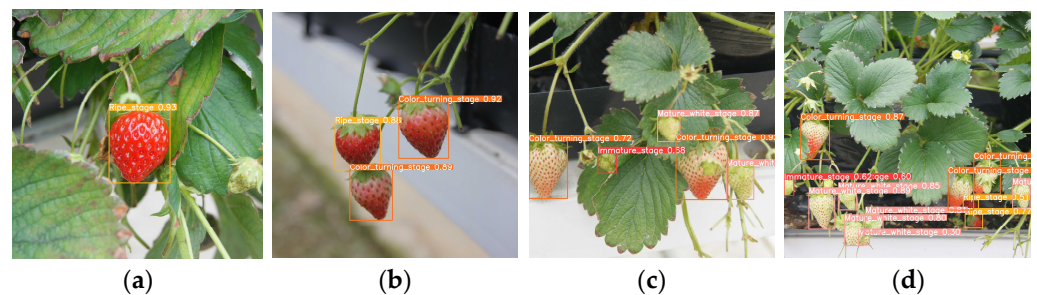


Figure 9. Model detection example image. (a) Single-object detection, (b) discrete multi-object detection, (c) minor obstruction detection, and (d) severe obstruction detection.

3.3. Ablation Study of the Improved CES-YOLOv8

To further validate the impact of the improved CES-YOLOv8 on model performance in strawberry ripeness detection, the modified algorithm was systematically compared with the initial algorithm in order to assess the impact of each improvement. The specific experimental results are shown in Table 3, where “-” indicates no change to the original structure.

Table 3. Ablation experiment results.

Model	Backbone Network	Attention Mechanism	Loss Function	Precision	Recall	mAP50	F1 Score	FPS
YOLOv8	-	-	-	83.40%	86.90%	90.10%	85.11%	220.92
YOLOv8	-	-	SIoU	85.90%	86.60%	90.30%	86.25%	224.74
YOLOv8	-	ECA	-	83.50%	87.50%	90.40%	85.45%	225.30
YOLOv8	ConvNeXt V2	-	-	86.60%	87.80%	90.50%	87.20%	192.82
YOLOv8	-	ECA	SIoU	86.30%	87.60%	90.70%	86.95%	230.39
YOLOv8	ConvNeXt V2	-	SIoU	87.70%	89.70%	91.70%	88.69%	191.16
YOLOv8	ConvNeXt V2	ECA	-	86.80%	86.70%	91.10%	86.75%	186.06
YOLOv8	ConvNeXt V2	ECA	SIoU	88.20%	89.80%	92.10%	88.99%	184.52

After the SIoU loss function was adopted, there were slight increases in the mAP50 and F1 values, with the greatest improvement in precision being 2.5%. This indicates that the SIoU loss function, through more refined bounding box regression optimization, helped the model achieve better performance. The application of the ECA attention mechanism

also improved the model's precision, recall, and mAP50, demonstrating that the enhancement of the attention mechanism strengthened the feature expression of specific channels, significantly improving the accurate positioning and recognition of targets. Replacing the backbone network with ConvNeXt V2 increased the precision, recall, mAP50, and F1 values by 3.2%, 0.9%, 0.4%, and 2.09%, respectively. This significant improvement indicates that ConvNeXt V2, compared to the original YOLOv8 structure, better captures image features and enhances model performance. Combining ConvNeXt V2, ECA, and SIoU allowed the model's performance to reach the highest values among all the metrics, with precision, recall, mAP50, and F1 values of 88.20%, 89.80%, 92.10%, and 88.99%, respectively, indicating increases of 4.8%, 2.9%, 2.05%, and 3.88%.

This study also calculated the frames per second (FPS) with the test dataset. The results show that the original YOLOv8 had an FPS of 220.92. The introduction of the ECA attention mechanism and SIoU loss function improved the processing speed, but the inclusion of ConvNeXt V2 in the backbone network increased the image features and computational time, leading to a noticeable decrease in FPS. Despite this, the FPS of the improved model still reached 184.52, far exceeding the requirements for real-time processing applications (over 30 FPS).

In summary, when all improvements are combined, their effects are further amplified, significantly enhancing the overall performance of strawberry ripeness detection without sacrificing real-time processing capabilities.

3.4. Comparative Analysis of Different Target Detection Networks

To qualitatively evaluate the detection results of the improved CES-YOLOv8 model, this model was compared with Faster R-CNN, RetinaNet, YOLOv5, YOLOv7, and the original YOLOv8 models in detecting strawberry images from a test set. The results, as shown in Table 4, indicate that the improved CES-YOLOv8 model excels in terms of precision, recall, and F1 score by 88.20%, 89.80%, and 88.99%, respectively, demonstrating that CES-YOLOv8 has high accuracy in detecting strawberry ripening.

Table 4. Comparative experiment results of different models.

Model	Precision	Recall	mAP50	Model Size /M
YOLOv5	86.50%	89.40%	80.26%	3.74
YOLOv7	77.10%	83.70%	80.26%	71.3
Retinanet	64.42%	92.59%	80.26%	139
Faster RCNN	65.26%	89.22%	80.26%	108
YOLOv8	85.90%	86.60%	86.25%	5.91
CES-YOLOv8	88.20%	89.80%	88.99%	41.3

Some of the inference results from the different models are shown in Figure 10, where the red arrows indicate missed detections, the blue arrows indicate false detections, and the yellow arrows indicate duplicate detections. The results show that the CES-YOLOv8 model surpasses other models in accurately identifying occluded targets and reducing false positives and false negatives. YOLOv8 and YOLOv5 have significant issues with missed detections in occluded strawberries. Although YOLOv7 has a lower miss rate, it suffers from multiple duplicate detections, which could affect accuracy and increase the detection time. Faster R-CNN and RetinaNet perform poorly in terms of accuracy and false detections, especially RetinaNet, which has numerous duplicate detection issues, making it impractical for real-world applications. Overall, CES-YOLOv8 is advantageous for reducing common issues in agricultural production applications, such as fruit and leaf occlusions, and it has significantly improved accuracy in fruit identification and positioning.



Figure 10. Strawberry maturity detection images using different models.

4. Discussion

As the global population increases and land resources become increasingly scarce, improving agricultural production efficiency has become particularly important. Automated harvesting robot technology, as a key advancement in smart agriculture, holds significant value in increasing harvesting efficiency and reducing labor costs [25]. Traditional methods for determining fruit ripeness are limited, primarily relying on a simple recognition of color and size, making it difficult to adapt to variable growing conditions and the significant color differences in fruits during the ripening process. Existing studies have proposed multiple solutions, but there are significant environmental dependency issues and difficulties in balancing model accuracy and efficiency [26]. Therefore, developing an efficient and accurate algorithm for detecting strawberry ripeness is crucial.

In response to these shortcomings, in this study, a CES-YOLOv8 network model based on improvements to YOLOv8 was proposed; it enhances the accuracy and robustness of strawberry ripeness recognition while balancing the model's use of computational resources. During the data collection phase, the impacts of different lighting conditions, varying levels of occlusion, and different angles on image acquisition were comprehensively considered, greatly enhancing the model's adaptability and robustness in real agricultural production environments. Additionally, by replacing some C2f modules of the backbone layer with ConvNeXt V2 modules and introducing ECA attention in the layer above the SPFF, the model's feature diversity and generalization capability were effectively enhanced, improving performance while reducing memory resource usage and increasing the accuracy of fruit detection in complex environments. The experimental results show that the improved network achieved significant performance enhancements in strawberry ripeness recognition tasks, with an accuracy of 88.2%, a recall of 89.8%, an mAP50 of 92.10%, and an F1 score of 88.99%, representing improvements of 4.8%, 2.9%, 2.05%, and 3.88%, respectively, over the corresponding values of the original YOLOv8 network.

The improvements in precision and recall mean that the model is more accurate in detecting the ripeness of strawberries, reducing misidentification. In practical applications, this can prevent fruits from being harvested at the wrong time, ensuring product quality and market value. A higher mAP50 indicates that the model maintains high performance in real-time dynamic environments. A higher F1 score ensures that all ripe strawberries are correctly classified, which is important for actual production. As the model is improved, the increase in feature computation also leads to a reduction in FPS, but it still far surpasses

the need for real-time processing. An FPS of 184.52 supports the rapid scanning of fruits using robots without slowing down production due to the image processing speed. Overall, the improved model enhances various performance metrics without sacrificing real-time processing capabilities. These results not only validate the effectiveness of the improvements but also demonstrate that the methods proposed in this study can accurately identify strawberry ripeness in complex environments, significantly advancing the development of automated harvesting technologies.

However, this study has certain limitations. First, although the model in this study performs well in detecting the ripeness of strawberries, its generality and applicability to other types of fruits or crops need further verification. Second, considering the complexity of agricultural production, such as the impact of climate conditions and soil types in different regions on fruit ripeness, it is necessary to explore the adaptability and robustness of the model under more diverse conditions [27]. Additionally, although the experiment considered various issues such as different lighting conditions, degrees of occlusion, and angles, it overlooked more physiological details; future work can further investigate the specific mechanisms through which these factors affect model performance and how to further optimize the model in order to address these challenges.

To address the aforementioned shortcomings, future research will further explore the generalization capabilities of the model, especially for different types of fruits, and for ripeness detection at various growth stages. Moreover, given the complexities of actual agricultural production, future research should focus more on the adaptability and robustness of the model under real field conditions, including its response to different climatic conditions and pest impacts. Through continuous optimization and improvement, more technical support for the development of smart agriculture and automated harvesting technologies can be provided, contributing to the enhancement of agricultural production efficiency and sustainable development.

5. Conclusions

Addressing the current difficulty of balancing model accuracy and performance in ripeness detection using automated harvesting robots, this study focused on strawberries and proposed an improved CES-YOLOv8 network structure. During the data collection phase, the effects of different lighting conditions, degrees of occlusion, and angles were considered, and image data covering these scenarios were collected, effectively enhancing the model's applicability and robustness in real agricultural environments. Targeted improvements were made to the YOLOv8 object detection network, including the replacement of some C2f modules in the backbone layer with ConvNeXt V2 modules and the introduction of ECA attention in the layer above the SPFF. The improvements enhanced the model's feature diversity and generalization ability, boosting its performance. The model's accuracy, recall, mAP50, and F1 score reached 88.20%, 89.80%, 92.10%, and 88.99%, respectively, showing increases of 4.8%, 2.9%, 2.05%, and 3.88%, respectively, compared to the corresponding values of the initial YOLOv8 structure. While improving the accuracy and precision of strawberry ripeness detection, the enhancements also effectively reduced the problems of missed and duplicate detections. This study provides an efficient and precise ripeness detection technology for automated harvesting robots in the field of smart agriculture, which advances the field of smart agriculture, enhances agricultural production efficiency, and supports sustainable agricultural development.

Author Contributions: Conceptualization, methodology, investigation, formal analysis, data curation, validation, writing—original draft, and writing—review and editing Y.C.; conceptualization, methodology, software, investigation, formal analysis, data curation, validation, writing—original draft, and validation, H.X.; methodology, writing—original draft, visualization, investigation, validation, and writing—review and editing, P.C.; methodology, investigation, writing—review and editing, and validation, Y.H. and F.Z.; investigation, formal analysis, writing—review and editing, and validation, L.C. and Q.J.; conceptualization, resources, supervision, and writing—review and editing, H.Z. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded through the following grant: Key Technology for Digitization of Characteristic Agricultural Industries in Fujian Province (XTCXGC2021015).

Data Availability Statement: Since the project presented in this research has not yet concluded, the experimental data will not be disclosed for the time being. Should readers require any supporting information, they may contact the corresponding author via email.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Rehman, A.; Saba, T.; Kashif, M.; Fati, S.M.; Bahaj, S.A.; Chaudhry, H. A revisit of internet of things technologies for monitoring and control strategies in smart agriculture. *Agronomy* **2022**, *12*, 127. [CrossRef]
2. Yamamoto, S.; Hayashi, S.; Yoshida, H.; Kobayashi, K. Development of a stationary robotic strawberry harvester with a picking mechanism that approaches the target fruit from below. *Jpn. Agric. Res. Q.* **2014**, *48*, 261–269. [CrossRef]
3. Hayashi, S.; Yamamoto, S.; Saito, S.; Ochiai, Y.; Kamata, J.; Kurita, M.; Yamamoto, K. Field operation of a movable strawberry-harvesting robot using a travel platform. *Jpn. Agric. Res. Q.* **2014**, *48*, 307–316. [CrossRef]
4. Kaur, H.; Sawhney, B.K.; Jawandha, S.K. Evaluation of plum fruit maturity by image processing techniques. *J. Food Sci. Technol.* **2018**, *55*, 3008–3015. [CrossRef] [PubMed]
5. Villaseñor-Aguilar, M.J.; Botello-Álvarez, J.E.; Pérez-Pinal, F.J.; Cano-Lara, M.; León-Galván, M.F.; Bravo-Sánchez, M.-G.; Barranco-Gutierrez, A.I. Fuzzy classification of the maturity of the tomato using a vision system. *J. Sens.* **2019**, *2019*, 3175848. [CrossRef]
6. Yin, Y.; Guo, C.; Shi, H.; Zhao, J.; Ma, F.; An, W.; He, X.; Luo, Q.; Cao, Y.; Zhan, X. Genome-wide comparative analysis of the R2R3-MYB gene family in five solanaceae species and identification of members regulating carotenoid biosynthesis in wolfberry. *Int. J. Mol. Sci.* **2022**, *23*, 2259. [CrossRef] [PubMed]
7. Parvathi, S.; Selvi, S.T. Detection of maturity stages of coconuts in complex background using Faster R-CNN model. *Biosyst. Eng.* **2021**, *202*, 119–132. [CrossRef]
8. Wang, L.M.; Jiang, Y. Automatic grading of banana ripeness based on deep learning. *Food Mach.* **2022**, *38*, 149–154. [CrossRef]
9. Wang, Z.; Ling, Y.; Wang, X.; Meng, D.; Nie, L.; An, G.; Wang, X. An improved Faster R-CNN model for multi-object tomato maturity detection in complex scenarios. *Ecol. Inform.* **2022**, *72*, 101886. [CrossRef]
10. Chen, F.; Zhang, X.; Zhu, X.; Li, Z.; Lin, J. Detection of olive fruit maturity based on improved EfficientDet. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 158–166.
11. Wang, C.; Wang, C.; Wang, L.; Wang, J.; Liao, J.; Li, Y.; Lan, Y. A lightweight cherry tomato maturity real-time detection algorithm based on improved YOLOV5n. *Agronomy* **2023**, *13*, 2106. [CrossRef]
12. Kazama, E.H.; Tedesco, D.; Carreira, V.d.S.; Júnior, M.B.; de Oliveira, M.F.; Ferreira, F.M.; Junior, W.M.; da Silva, R.P. Monitoring coffee fruit maturity using an enhanced convolutional neural network under different image acquisition settings. *Sci. Hortic.* **2024**, *328*, 112957. [CrossRef]
13. Megalingam, R.K.; Manoharan, S.K.; Maruthababu, R.B. Integrated fuzzy and deep learning model for identification of coconut maturity without human intervention. *Neural Comput. Appl.* **2024**, 1–13. [CrossRef]
14. Zhang, W.; Liu, Y.; Chen, K.; Li, H.; Duan, Y.; Wu, W.; Shi, Y.; Guo, W. Lightweight fruit-detection algorithm for edge computing applications. *Front. Plant Sci.* **2021**, *12*, 740936. [CrossRef] [PubMed]
15. Xiao, Z.Q.; He, J.X.; Chen, D.B.; Zhan, Y.; Lu, Y.L. Automatic classification method of rock spectra based on twin network model. *Spectrosc. Spectr. Anal.* **2024**, *44*, 558–562.
16. Wang, Y.T.; Zhou, H.Q.; Yan, J.X.; He, C.; Huang, L.L. Progress in computational optics research based on deep learning algorithms. *Chin. J. Lasers* **2021**, *48*, 1918004.
17. Zhao, J.D.; Zhen, G.Y.; Chu, C.Q. Drone image target detection algorithm based on YOLOv8. *Comput. Eng.* **2024**, *50*, 113–120.
18. Woo, S.; Debnath, S.; Hu, R.; Chen, X.; Liu, Z.; Kweon, I.S.; Xie, S. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 16133–16142.
19. Li, Y.; He, Z.; Ma, J.; Zhang, Z.; Zhang, W.; Chatterjee, P.; Pamucar, D. A Novel Feature Aggregation Approach for Image Retrieval Using Local and Global Features. *CMES-Comput. Model. Eng. Sci.* **2022**, *131*, 239–262. [CrossRef]
20. Zhu, M.L.; Ren, Y.Z. Screw surface defect detection based on neural networks. *J. Ordnance Equip. Eng.* **2024**, *45*, 224–231.
21. 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.
22. Li, G.M.; Gong, H.B.; Yuan, K. Research on Sichuan pepper cluster detection based on lightweight YOLOv5s. *Chin. J. Agric. Mech.* **2023**, *44*, 153.
23. Gevorgyan, Z. Siou Loss: More Powerful Learning for Bounding Box Regression, 23 May 2022. Available online: <https://arxiv.org/abs/2205.12740> (accessed on 16 April 2024).

24. Gu, Z.; Zhu, K.; You, S. YOLO-SSFS: A Method Combining SPD-Conv/STDL/IM-FPN/SIoU for Outdoor Small Target Vehicle Detection. *Electronics* **2023**, *12*, 3744. [[CrossRef](#)]
25. Raja, V.; Bhaskaran, B.; Nagaraj, K.; Sampathkumar, J.; Senthilkumar, S. Agricultural harvesting using integrated robot system. *Indones. J. Electr. Eng. Comput. Sci.* **2022**, *25*, 152. [[CrossRef](#)]
26. Yoshida, T.; Onishi, Y.; Kawahara, T.; Fukao, T. Automated harvesting by a dual-arm fruit harvesting robot. *Robomech J.* **2022**, *9*, 19. [[CrossRef](#)]
27. Vincent, D.R.; Deepa, N.; Elavarasan, D.; Srinivasan, K.; Chauhdary, S.H.; Iwendi, C. Sensors driven AI-based agriculture recommendation model for assessing land suitability. *Sensors* **2019**, *19*, 3667. [[CrossRef](#)]

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