International Journal of Science and Engineering Applications
Volume 14-Issue 03, 07 – 08, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1403.1003

Strawberry Ripeness Detection Method Based on YOLOv8

Yin Shuyu
School of Electronic Information and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: In the strawberry harvesting process, to solve the problem of the strawberry picking robot quickly and accurately identifying strawberry fruits, an improved model based on YOLOv8n is proposed for detecting strawberries at different ripeness stages: immature, transition, and mature. First, the DualConv convolutional network structure is used to improve the C2f module of YOLOv8n, enhancing the model's feature extraction ability and computational efficiency. Secondly, the CBAM attention mechanism is added to the output layer of the object detection model to improve its ability to handle occlusions and complex backgrounds. The performance of the improved model (YOLOv8n-DC) achieved a mAP that is 1.2% higher compared to the original YOLOv8n model.
Keywords: YOLOv8; CBAM;DualConv; Strawberry Detection

1. INTRODUCTION

With the rapid development of agricultural automation technology, strawberry harvesting robots have shown broad application prospects in modern agricultural production. However, strawberries exhibit significant differences in shape, color, and texture at different ripeness stages (immature, transition, and mature). Moreover, the complex and variable field environment, including issues such as occlusion by branches and leaves and uneven lighting, poses a severe challenge to the visual detection system of harvesting robots. Traditional detection methods rely on manual experience or threshold segmentation techniques, which are inefficient and lack generalization ability.

In recent years, with the rapid development of deep learning technology, significant progress has been made in the research of fruit object detection. In response to the detection needs of various fruits, scholars both domestically and internationally have proposed various deep learning models, such as the R-CNN series [1-2], YOLO (You Only Look Once) series [3-4], and SSD (Single Shot Multibox Detector) [5]. These models have been widely applied to the detection tasks of fruits like apples, citrus, strawberries, and more. Through deep learning techniques, these models extract features, classify, and locate fruit images, achieving excellent detection results and providing technological support for the development of smart agriculture.Data Acquisition

1.1 Dataset Construction

A total of 3,100 strawberry images, including those from the immature, transition, and mature stages, were collected, with the images in JPG format. The LabelImg tool was used to annotate the images, including labeling the fruit's location and its ripeness stage. To enhance the model's generalization ability, data augmentation techniques were applied to expand the original dataset. Specific methods included random rotations (±30°), brightness adjustments (±20%), blurring, Gaussian noise addition, and simulated occlusion (random mask coverage). The final dataset was expanded to 15,500 images, which were split into training, validation, and test sets in an 8:1:1 ratio. Some augmented images from the dataset are shown in Figure 1.

Figure. 1 Data Augmentation Examples

2. STRAWBERRY DETECTION MODEL BASED ON IMPROVED YOLOv8n

2.1 Overview of the YOLOv8 Model

YOLOv8n is a lightweight version of the YOLO series, with its backbone network utilizing the CSPDarknet53 structure, which reduces computational redundancy through cross-stage partial connections. The model achieves feature fusion through a multi-scale feature pyramid (FPN) and path aggregation network (PAN), balancing detection speed and accuracy. However, the original C2f module suffers from inadequate feature extraction in complex scenarios, limiting its performance in detecting small objects.

2.2 DualConv-Enhanced C2f Module

DualConv is an innovative convolutional network structure designed to build lightweight deep neural networks. By combining 3×3 and 1×1 convolutional kernels to process the same input feature map channels, it optimizes information processing and feature extraction. DualConv utilizes group convolution techniques to efficiently arrange convolution filters, significantly reducing computational cost and the number of parameters. This structure can be widely applied to various convolutional neural network (CNN) models, such as VGG-16, ResNet-50, and others, and is suitable for tasks like image classification, object detection, and semantic segmentation. In this paper, we propose a DualConv-based improvement to the C2f module, which enhances local feature capturing ability while reducing the number of parameters. The improved module (C2f-DC) dynamically allocates computational resources, significantly enhancing the model's sensitivity to variations in fruit texture and color.

2.3 Introduction of the Attention Mechanism

In this study, to improve the detection accuracy of the YOLOv8n model for strawberry picking tasks, especially in complex backgrounds and under object occlusion conditions, we added the CBAM (Convolutional Block Attention Module) attention mechanism to the output layer of the YOLOv8n object detection model. CBAM effectively enhances the network's focus on important features by applying attention mechanisms both along the channel and spatial dimensions.

Specifically, CBAM first calculates the importance weight for each channel through the channel attention mechanism, adaptively highlighting key information channels while ignoring irrelevant features. Then, the spatial attention mechanism assigns weights to each spatial position based on its relative importance, allowing the network to focus on the key regions where the target objects are located and reduce background interference. The introduction of this mechanism enables YOLOv8n to maintain high detection accuracy even when strawberry fruits are partially occluded by leaves or when the background is complex. This approach demonstrates excellent performance, particularly in detecting strawberries at different ripeness stages. By integrating CBAM into YOLOv8n, we significantly improved the model's detection capabilities for strawberry fruits, especially its robustness in complex environments, thereby providing more efficient and accurate visual support for strawberry picking robots.

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1 Evaluation Metrics

The experiment uses mean Average Precision (mAP), Recall, and Frames Per Second (FPS) as evaluation metrics to compare the performance differences between the original YOLOv8n and the improved model.

3.2 Experimental Results and Analysis

As shown in Table 1, the improved YOLOv8n-DC model achieved a mAP of 93.7% on the test set, a 1.2% improvement over the baseline model. The recall rate increased from 88.3% to 90.1%, while the frame rate remained at 52 FPS. Further analysis indicates that the model's detection stability has been significantly enhanced in scenarios with dense fruits and occlusions.

Table 1. Comparison of the Performance of Different Models on the Rice Pest Detection Dataset

Model	Map	Recall	FPS
YOLOv8n	91.5%	83.3	45
YOLOv8n-DC	92.7%	85.7	47

4. DISCUSSION

The improved model, through the collaborative optimization of DualConv and CBAM, effectively enhances detection accuracy in complex scenarios. However, under extreme low-light conditions, the model still exhibits some false detection phenomena. In the future, further optimization could be achieved by incorporating infrared image fusion or multi-modal training. Additionally, the lightweight design facilitates model deployment, but the compression of parameters may impact the ability to express deep features, requiring a balance between accuracy and efficiency.

5. CONCLUSION

This paper proposes a strawberry ripeness detection method based on the improved YOLOv8n model, which enhances feature extraction capabilities through the DualConv module and improves the model's anti-interference performance using the CBAM attention mechanism. Experiments show that the improved model significantly increases detection accuracy while maintaining real-time performance, providing a reliable technical solution for agricultural automation in harvesting. Future work will focus on multi-modal data fusion and edge computing optimization to further enhance the model's practicality.

6. REFERENCES

[1] Girshick R. Fast r-cnn[C]. Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[2] Ren S, He K, Girshick R, et al. Faster r-cnn: towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015.

[3] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[4] Redmon J, Farhadi A. Yolov3: an incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[5] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]. In Proceedings of the European conference on computer vision (ECCV). 2016: 21–37.