

Article

Red Raspberry Maturity Detection Based on Multi-Module Optimized YOLOv11n and Its Application in Field and Greenhouse Environments

Rongxiang Luo ^{1,2}, Xue Ding ^{3,4,5,*} and Jinliang Wang ^{3,4,5}

¹ School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China; 2324100038@ynnu.edu.cn

² Southwest United Graduate School, Kunming 650500, China

³ Department of Geography, Yunnan Normal University, Kunming 650500, China; jlwang_ynnu@163.com

⁴ Key Laboratory of Resources and Environmental Remote Sensing for Universities in Yunnan Kunming, Kunming 650500, China

⁵ Center for Geospatial Information Engineering and Technology of Yunnan Province, Kunming 650500, China

* Correspondence: 4228@ynnu.edu.cn

Abstract: In order to achieve accurate and rapid identification of red raspberry fruits in the complex environments of fields and greenhouses, this study proposes a new red raspberry maturity detection model based on YOLOv11n. First, the proposed hybrid attention mechanism HCSA (halo attention with channel and spatial attention modules) is embedded in the neck of the YOLOv11n network. This mechanism integrates halo, channel, and spatial attention to enhance feature extraction and representation in fruit detection and improve attention to spatial and channel information. Secondly, dilation-wise residual (DWR) is fused with the C3k2 module of the network and applied to the entire network structure to enhance feature extraction, multi-scale perception, and computational efficiency in red raspberry detection. Concurrently, the DWR module optimizes the learning process through residual connections, thereby enhancing the accuracy and real-time performance of the model. Finally, a lightweight and efficient dynamic upsampling module (DySample) is introduced between the backbone and neck of the network. This module enhances the network's multi-scale feature extraction capabilities, reduces the interference of background noise, improves the recognition of structural details, and optimizes the spatial resolution of the image through the dynamic sampling mechanism. Reducing network parameters helps the model better capture the maturity characteristics of red raspberry fruits. Experiments were conducted on a custom-built 3167-image dataset of red raspberries, and the results demonstrated that the enhanced YOLOv11n model attained a precision of 0.922, mAP@0.5 of 0.925, and mAP@0.5 of 0.943, respectively, representing improvements of 0.7%, 4.4%, and 4.4%, respectively. At 3.4%, mAP@0.5-0.95 was 0.798, which was 2.0%, 9.8% and 3.7% higher than the original YOLOv11n model, respectively. The mAP@0.5 of unripe and ripe berries was 0.925 and 0.943, which was improved by 0.7% and 4.4%, respectively. The F1-score was enhanced to 0.89, while the computational complexity of the model was only 8.2 GFLOPs, thereby achieving a favorable balance between accuracy and efficiency. This research provides new technical support for precision agriculture and intelligent robotic harvesting.



Academic Editor: Maciej Zaborowicz

Received: 25 February 2025

Revised: 11 March 2025

Accepted: 16 April 2025

Published: 18 April 2025

Citation: Luo, R.; Ding, X.; Wang, J. Red Raspberry Maturity Detection Based on Multi-Module Optimized YOLOv11n and Its Application in Field and Greenhouse Environments. *Agriculture* **2025**, *15*, 881. <https://doi.org/10.3390/agriculture15080881>

Copyright: © 2025 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/](https://creativecommons.org/licenses/by/4.0/)

<https://creativecommons.org/licenses/by/4.0/>).

Keywords: red raspberries; ripeness detection; YOLOv11n; smart agriculture; hybrid attention mechanism

1. Introduction

Accurate detection of red raspberry maturity is crucial for timely harvesting and ensuring quality in agriculture. Traditional methods, relying on features like color histograms and shape descriptors, struggle with low robustness and accuracy in complex environments with lighting variations, background interference, and occlusion [1]. In recent years, deep learning, particularly convolutional neural networks (CNNs), has dramatically improved maturity detection, enhancing accuracy and stability [2]. Early research focused on image processing and computer vision techniques. For instance, Kienzle et al. (2012) utilized principal component and cluster analysis to analyze mango maturity [1]. Mohammadi et al. (2015) attained 0.9024 accuracy in persimmon maturity detection with LDA and QDA classifiers [2]. Khoshnam et al. (2016) employed acoustic testing to assess melon maturity, demonstrating that frequency changes indicate maturation [3]. Furthermore, Namdari Gharaghani et al. (2020) utilized finite element modal analysis to detect orange maturity, achieving over 0.91 consistency with experimental data [4]. Wakchaure et al. (2024) developed an image processing-based prototype for plantain maturity detection, automating manual classification [5], while Kumar Saha et al. (2024) estimated tomato maturity using dual-wavelength LiDAR data, achieving spatially resolved maturity classification [6].

The advent of deep learning, most notably convolutional neural networks (CNNs), has precipitated a paradigm shift in the domain of fruit maturity detection. A comprehensive review of fruit classification, maturity detection, and grading methodologies was conducted by Reshm and Sreekumar (2018) [7]. Subsequently, Surya Kiran and Niranjana (2019) provided a synopsis of the advancements in various maturity detection technologies [8]. Pardede et al. (2021) enhanced the efficiency of maturity classification by integrating VGG16 transfer learning with multi-layer perceptron (MLP) blocks [9]. Momeny et al. (2022) integrated a deep CNN with Bayesian optimization to enhance the robustness of detecting orange and black spot diseases and maturity [10]. Chen et al. (2022) proposed a method combining visual saliency maps and CNNs for citrus fruit maturity detection [11]. Azadnia et al. (2023) used the Inception-V3 model to detect high-precision hawthorn fruit maturity through machine vision and deep learning [12]. In another study, Olisah et al. (2024) addressed the challenge of inconspicuous features in Blackberry maturity detection by employing a multi-input CNN ensemble and optimizing the VGG16 model to enhance accuracy [13]. In a related study, Astuti et al. (2019) applied the K-nearest neighbor algorithm and image acquisition for oil palm fruit maturity detection [14]. Zhao and Chen (2021) utilized color information and an SVM model for wolfberry maturity detection [15]. Similarly, Kumar et al. (2022) developed a non-destructive tomato maturity model using reflectance data and chemometric analysis [16].

In recent years, YOLO (You Only Look Once) models have become increasingly prevalent in fruit maturity detection, owing to their efficiency and accuracy. Bonora et al. (2021) employed a YOLO-based convolutional neural network (CNN) to detect maturity and physiological disorders in “Abbé Fétel” pears, yielding favorable classification results [17]. Li et al. (2022) developed a real-time sweet cherry maturity detection algorithm using YOLOX, improving accuracy in complex environments [18]. Xiao et al. (2023) introduced a lightweight method for blueberry maturity detection based on an enhanced YOLOv5n algorithm, incorporating ShuffleNet and CBAM modules for better feature fusion and high recall [19]. Xu et al. (2023) proposed the YOLO Jujube method for detecting jujube fruit maturity in natural environments [20]. In addition, Xia Hongmei et al. (2021) utilized a Faster R-CNN model with an attention mechanism and multi-scale fusion to detect hydroponically grown broccoli buds with 0.965 accuracy [21]. Xingxu Li et al. (2023) designed a cascaded visual inspection system that improved cherry tomato picking efficiency through target detection and feature discrimination [22]. Fengjun Chen et al. (2024) enhanced

YOLOv7 to address occlusion in oil tea trees, boosting mAP to 0.946 [23]. Similarly, Ligang Wu et al. (2024) proposed YOLOv8-ABW, a system that integrates AIFI and Bi FPN to enhance the efficiency of yellow flower maturity detection [24]. Xu Tingting et al. (2024) introduced YOLO v7-RA, a system that combines ELAN_R3 and hybrid attention mechanisms to detect the maturity of dragon fruit [25]. Youwen Tian et al. (2024) optimized blueberry maturity detection with the MSC-YOLOv8 model, incorporating MobileNetV3 and CBAM modules [26]. Xuesong Jiang et al. (2024) reviewed profound learning advancements in non-destructive forest fruit quality detection [27]. Liu Zhigang et al. developed a machine vision-based method for apple maturity detection using an RGB model for high accuracy [28]. Runchi Zhang et al. (2024) enhanced YOLOv8 for tomato counting, achieving 0.938 accuracy [29]. Li Ying et al. (2024) proposed the YOLOv8s model for citrus fruit maturity, improving mAP with an adaptive fusion head [30]. Using multi-scale feature fusion, Liu et al. (2024) optimized YOLOv5ns for apple maturity detection [31]. Sun et al. (2024) introduced a lightweight YOLO-FHLD method for date maturity detection, improving accuracy and model expressiveness [32]. Jing et al. (2024) proposed YOLO-IA for peaches, achieving high-precision detection with a progressive feature pyramid network [33]. Zhu et al. (2024) used YOLO-LM to reduce camellia fruit occlusion with Criss-Cross Attention [34]. Ye et al. (2024) introduced CR-YOLOv9 for strawberry maturity detection, optimizing network design for efficiency [35]. Zhai et al. (2024) proposed an attention mechanism and bidirectional feature pyramid network for blueberry maturity detection, achieving 0.888 accuracy and 0.882 recall [36].

The findings of the aforementioned study demonstrate the efficacy of the deep learning model in recognizing the ripeness of various fruits. In the case of red raspberries, for instance, the fruits were arranged in clusters and spaced extensively in both greenhouse and outdoor field environments. The light in the greenhouse is more uniform, and there is less shade between the fruits. However, water vapor or haze may compromise the clarity of the fruits and branches due to the enclosed environment and higher humidity. Conversely, outdoor field environments are subject to light, wind, and temperature variations, and red raspberries frequently experience shading and overlap, which hinders accurate detection. While the deep learning model offers robust detection capability, its practical application is constrained by the device's computational capacity, preventing optimal real-time detection. This study proposes an enhanced version of a lightweight YOLO v11n red raspberry ripeness detection model to address these limitations. The model's primary strengths are its high accuracy, robustness, and computational efficiency. The specific strategies employed to achieve these objectives are as follows:

- (1) The HCSA attention mechanism: The HCSA module proposed in this study effectively combines halo attention, channel attention, and spatial attention, thereby enhancing the model's capacity to extract and represent features, particularly in recognizing spatial information and detailed features and demonstrating significant advantages.

- (2) The second component is the extended residual module (DWR), which enhances the feature extraction ability and multi-pixel perception accuracy while improving computational efficiency. The model's learning process is optimized by introducing residual connectivity to enhance its accuracy and real-time performance.

- (3) The lightweight dynamic upsampling module (DySample) is integrated into the network's backbone and neck, thereby enhancing the extraction capability of multi scale feature maps, reducing background noise with optimized spatial resolution, and improving the performance of detail changes.

2. Materials and Methods

2.1. The Dataset

2.1.1. Image Acquisition

The red raspberry image data utilized in this study were collected from the Goodberry Plantation, situated at No. 1 Xinhui Road, Chenggong District, Kunming City, Yunnan Province, China, which is managed by Dianong Group · Yunnan Goodberry Biotechnology Co., Ltd. To ensure the quality and representativeness of the data, the collection encompassed both greenhouse and outdoor environments and was conducted on both sunny and cloudy days. All images had a resolution of 3072×4090 pixels and were in JPG format. During the collection process, red raspberry fruits were divided into two categories based on their maturity: ripe and unripe, as shown in Figure 1. Unripe red raspberry fruits are green or light yellow, have a smooth and hard surface, lack noticeable color changes, have a sour taste, and are unsuitable for consumption. At this stage, the fruit is not yet fully expanded and has a low sugar and high moisture content, making it difficult to identify due to its strong light reflection and background interference. Ripe red raspberries are bright red, smooth and shiny, soft, and have accumulated sufficient sugar and moisture to be sweet. The fruit is loose at this time and often falls off quickly, but it has more apparent characteristics, making it easy to identify and pick accurately. The physical size of red raspberries ranges from 20 to 45 mm in diameter, corresponding to image sizes of 200 to 450 pixels after calibration. Fruits with a diameter of fewer than 15 pixels (i.e., a physical size of less than 1.5 mm) are defined as small targets, typically appearing when the fruit is covered or the shooting distance is far away.

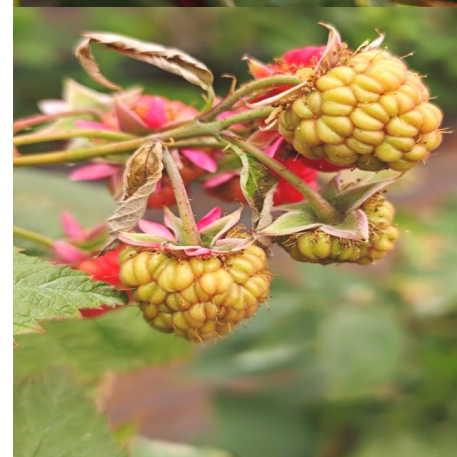
Greenhouse
environment



Outdoor field
environment



(a) Mature



(b) Immature

Figure 1. Ripening morphology of red raspberries in greenhouse greenhouses and outdoor fields.

As demonstrated in Figure 2, the process of detecting red raspberry fruits is hindered by several factors, including fruit overlap, occlusion, light variation, and complex backgrounds. Firstly, fruit overlap complicates the task of distinguishing individual fruits, particularly in areas where they are densely populated. This issue is further compounded by occlusion caused by branches and leaves, which impedes the visual system's ability to discern the individual fruits. Secondly, uneven lighting can cause changes in the appearance of the fruit, which can affect the recognition of color and detail, especially when immature red raspberries are close to the background color, which makes recognition more difficult. Thirdly, background complexity is also a key factor. A complex background contains many distracting objects, such as branches, leaves, soil, or other objects, which may confuse the fruit and affect the detection accuracy. Light uniformity is another influencing factor. The presence of uneven lighting can result in certain areas becoming overexposed or underexposed, thereby leading to the loss of fruit details and further reducing detection performance.



(a) Overlapped



(b) Branches occlusion



(c) Leaf interference



(d) Uneven illumination

Figure 2. Red raspberry images from different scenes.

2.1.2. Production of Datasets

This study meticulously annotated the dataset by employing the Roboflow online platform. A .txt file was generated using manual annotation, encompassing the bounding

box coordinates and category labels of the red raspberry fruits. The dataset comprised a total of 1022 original images. Following a series of pre-processing steps, including automatic orientation and resizing, and data enhancement methods such as flipping, scaling, panning, rotating, and saturation adjustment, the images were randomly combined to expand the sample set to 3167 images. The resulting dataset encompasses a diverse range of natural environments and is designed to accommodate complex scenes, such as fruit occlusions and changes in lighting.

The dataset was then divided into a training set of 2535 images, a validation set of 316 images, and a test set of 316 images in an 8:1:1 ratio. This division ensures the adequacy of the training data, while the independent validation and test set effectively test the model's generalization ability and reduce the risk of overfitting.

2.2. The Improved YOLOv11n Network Model

2.2.1. The YOLOv11n Network Architecture

YOLOv11n is an optimized version of the YOLO series, designed for real-time object detection to improve efficiency and accuracy [37]. Its architecture consists of an input layer, a backbone network, a neck network, and a head network. The backbone network uses the C3k2 module, an efficient cross-stage bottleneck structure. Compared to YOLOv8, the C3k2 module significantly reduces the amount of computation and speeds up processing by using smaller convolutional kernels. In the neck network, the C3k2 module replaces the C2f module in YOLOv8, improving the efficiency of feature fusion.

YOLOv11n also introduces fast spatial pyramid pooling (SPPF) and cross-stage partial with spatial attention (C2PSA) modules to further improve feature extraction capabilities. These modules enable the model to focus more effectively on key regions in the image, especially in the presence of multiple scales and complex backgrounds, further improving recognition accuracy. The C2PSA module helps the model focus more accurately on key regions by introducing a spatial attention mechanism that improves the detection of small and occluded objects.

The head network refines and optimizes the feature map through multiple C3k2 modules and Convolutional Batch Normalization-SiLU (CBS) modules to improve accuracy while maintaining fast detection speed. Combining these designs, YOLOv11n improves detection efficiency and provides a more accurate object detection solution while maintaining high efficiency.

Despite its outstanding efficiency, the detection accuracy of YOLOv11n is still affected to some extent when faced with overlapping fruits, branches, leaves, uneven lighting, and complex backgrounds. To solve these problems, we propose the following improvements: the HCSA attention mechanism improves the model's attention to spatial information and detailed features by fusing halo, channel, and spatial attention; the dilation-wise residual (DWR) [38] enhances feature extraction capabilities and improves pixel perception accuracy and computational efficiency; the dynamic upsampling module (DySample) [39] optimizes multi-scale feature extraction, improves spatial resolution, reduces background noise interference and enhances the ability to detect details. The improved YOLOv11n network structure is shown in Figure 3.

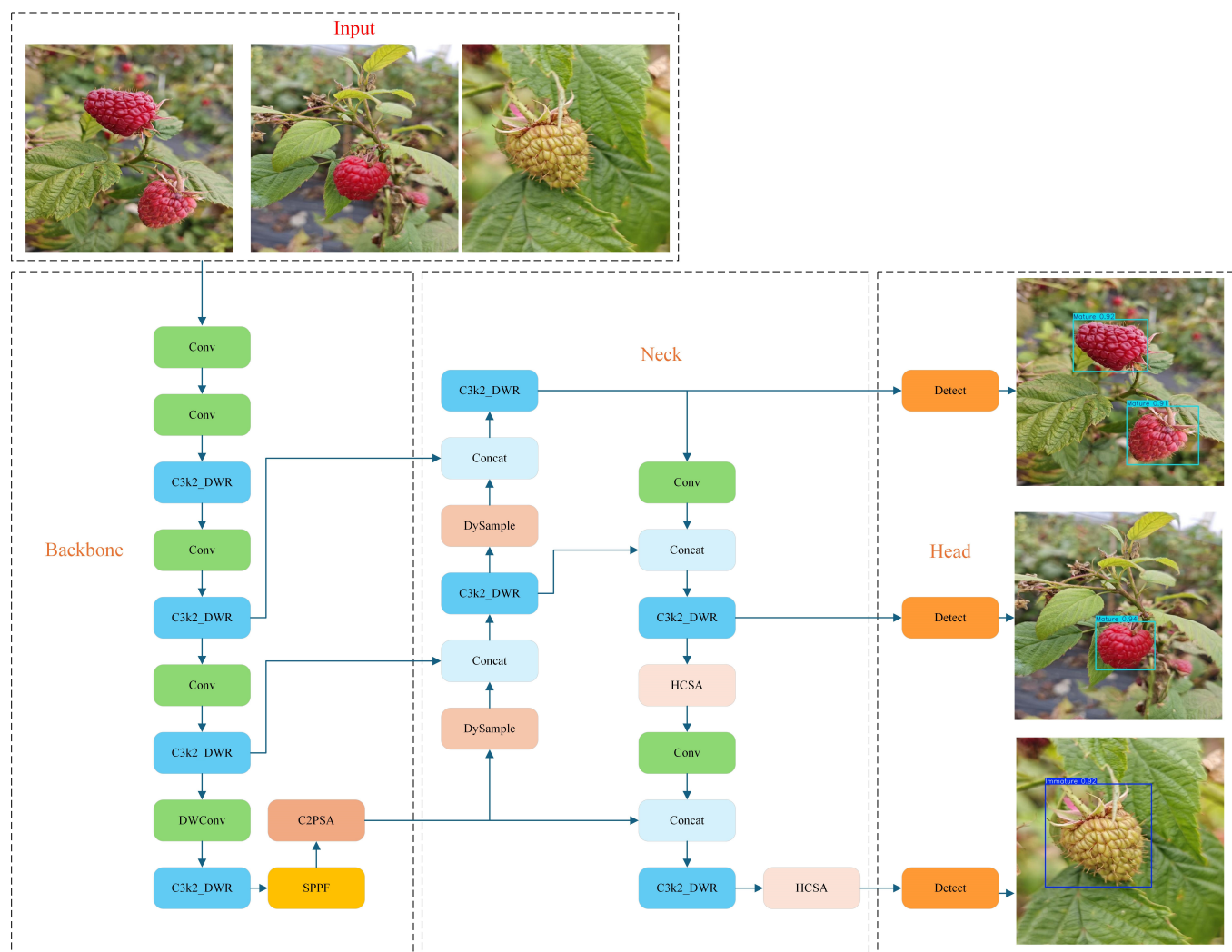


Figure 3. Improved YOLOv11n network structure diagram.

2.2.2. The Hybrid Attention Mechanism

In the context of red raspberry maturity detection, uneven lighting has been identified as a pivotal factor influencing the accuracy of the detection process. The possibility of lighting variations persists in the confines of a greenhouse environment, where lighting conditions may be relatively uniform. Conversely, in outdoor fields, lighting conditions are subject to more significant fluctuations, resulting in the alternation of intense light and shadows, which can hinder the precise recognition of the fruit. Enhancing the focus on local details by halo attention enables YOLOv11n to cope with lighting changes and capture fruit characteristics accurately and effectively. Furthermore, halo attention [40] assists the model in separating the target from overlapping fruits by strengthening the focus on local regions. The efficacy of halo attention in enhancing the accuracy of fruit recognition, even in areas with blurred details or low contrast, is noteworthy.

The following section will provide a detailed overview of the function of halo attention, which enhances the model's capacity to capture long-distance dependencies through relative positional encoding. The correlation between the query (Q) and the key (K) is calculated by the dot product and combined with the relative positional encoding ($RelPos(Q, K)$) to obtain the attention matrix:

$$A = \frac{QK^T}{\sqrt{d}} + RelPos(Q, K) \quad (1)$$

where d is the dimension of the query and key, K^T is the transpose of the key matrix, and $RelPos(Q, K)$ denotes relative position encoding. The attention score is normalized by softmax:

$$A_{softmax} = softmax(A) \quad (2)$$

Finally, the weights and values (V) are weighted to obtain the output:

$$O = A_{softmax} V \quad (3)$$

Channel attention enhances the importance of feature channels by calculating the attention weight of each channel. First, the input feature map calculates channel statistics using global average pooling (Z_{avg}) and max pooling (Z_{max}):

$$Z_{avg} = AvgPool(X), Z_{max} = MaxPool(X) \quad (4)$$

These channel features are then used to learn channel attention weights through a shared multi-layer perceptron (MLP):

$$Z_c = \sigma(W_2 \cdot ReLU(W_1 \cdot [Z_{avg}; Z_{max}]) \quad (5)$$

W_1 and W_2 are the weights of the MLP, and σ is the sigmoid activation function. Finally, channel attention is applied to the input feature map:

$$X_c = X \times Z_c \quad (6)$$

Background leaf interference is a prevalent issue in field and greenhouse environments, where leaves and weeds in the background are analogous in color to red raspberry fruits, resulting in interference. The channel attention mechanism of HCSA intelligently adjusts attention to different channels, prioritizes feature channels related to fruits, and reduces the interference of the background to ensure the accurate recognition of red raspberry fruits. The specific steps involved are as follows: firstly, the spatial attention mechanism improves the response to key regions by focusing on the spatial dimension of the image; secondly, the spatial features are computed by average pooling and maximum pooling of the input feature map in the channel dimension:

$$Z_{avg} = Mean(X, \dim = 1) \quad (7)$$

$$Z_{max} = Max(X, \dim = 1) \quad (8)$$

These spatial features are then passed through a convolutional layer to generate a spatial attention map:

$$Z_s = \sigma(Conv2D([Z_{avg}; Z_{max}])) \quad (9)$$

HCSA combines the three attention mechanisms mentioned above to strengthen the feature extraction capability by complementing each other. Its final output is a fusion of the outputs of halo attention, channel attention and spatial attention:

$$X_s = X \times Z_s \quad (10)$$

The HCSA combines the three aforementioned attention mechanisms to enhance the feature extraction capability by complementing each other. The final output of the HCSA is a fusion of the outputs of halo attention, channel attention and spatial attention:

$$X_{HACB} = Halo \cdot Channel \cdot Spatial \quad (11)$$

The HCSA's design incorporates multiple attention mechanisms, enabling it to capture global and local contextual information and enhance the feature representation ability in image tasks. In the fruit ripeness target detection task context, the HCSA can recognize key information more accurately. This performance enhancement indicates the model's proficiency in capturing long-distance dependencies and focusing attention on key channels and spatial regions. Consequently, this integration of attentional mechanisms signifies the HCSA's efficacy in various visual tasks. The specific structure of The HCSA hybrid attention module is shown in Figure 4.

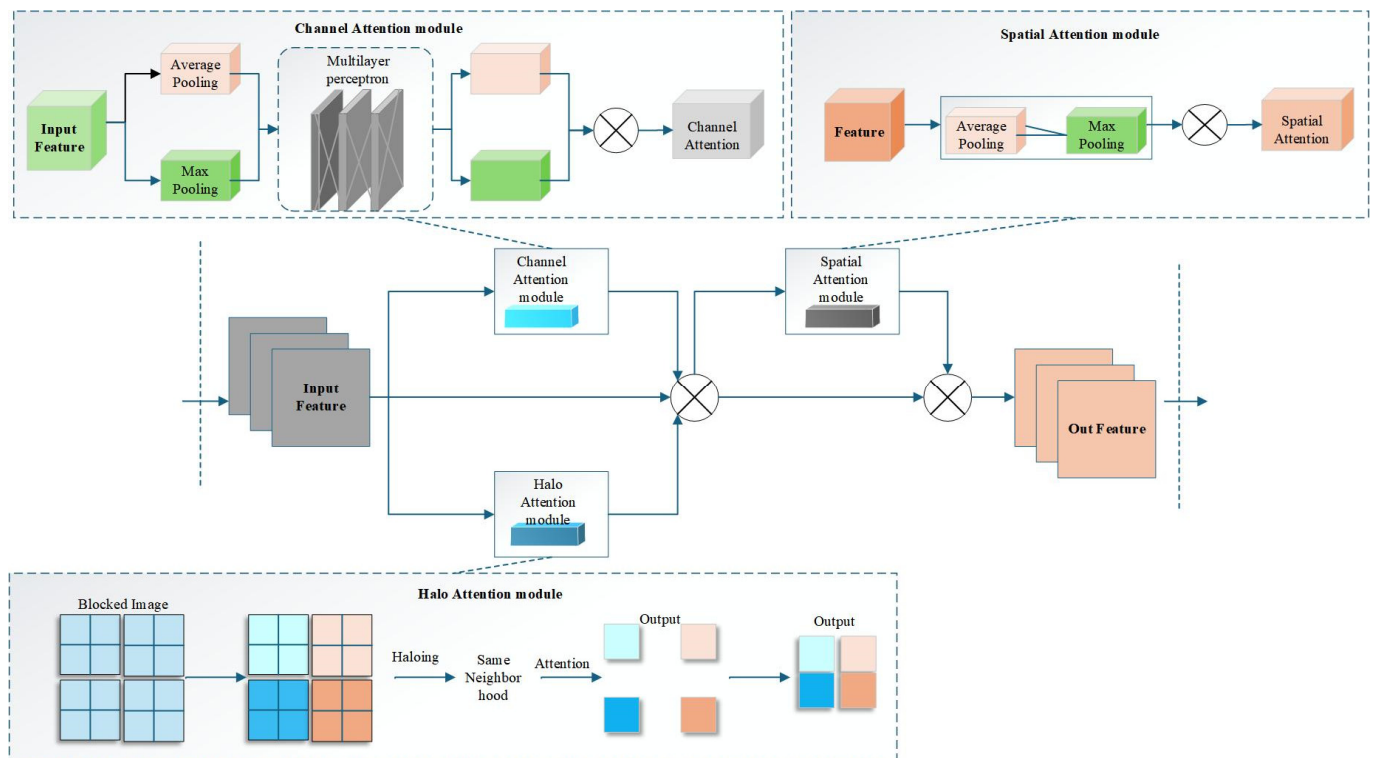


Figure 4. The HCSA hybrid attention module.

2.2.3. The Dilation-Wise Residual

The dilation-wise residual (DWR) module aims to effectively capture multi-scale contextual information through multi-rate dilated convolutions while maintaining low computational complexity [38]. Its structure is shown in Figure 5. In red raspberry ripeness detection, the C3k2_DWR module, as part of YOLOv11n, combines the dilation-wise residual (DWR) and CSP bottleneck structure to improve feature extraction and multi-scale perception significantly. The DWR module efficiently captures features at different scales by introducing dilated convolutions of different sizes ($d = 1$, $d = 3$, $d = 5$), especially in the case of fruit size variation or partial occlusion, which substantially improves the detection accuracy. The introduction of residual connection optimizes the information flow, improves the training efficiency, and effectively avoids the gradient vanishing problem.

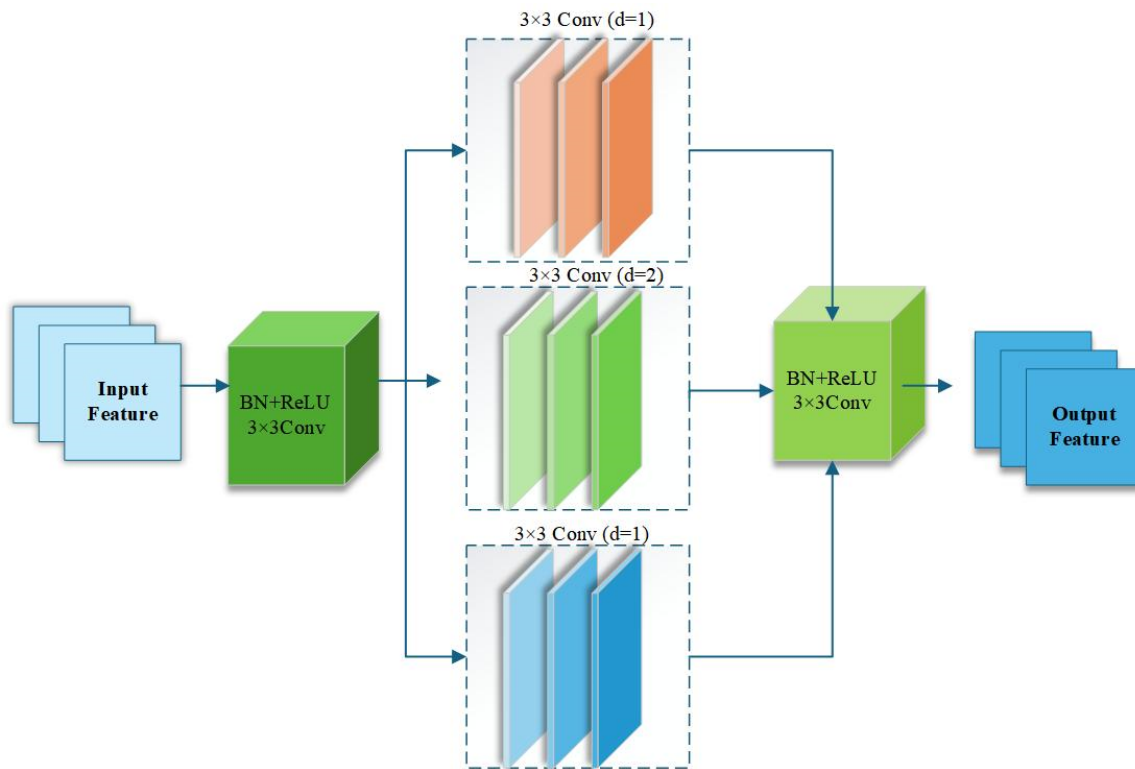


Figure 5. The dilation-wise residual module.

The C3k2_DWR module enhances computational efficiency by combining multiple convolution operations and reducing the number of channels using 1×1 convolution, ensuring efficient real-time detection. Meanwhile, optimizing the C3k module enhances the feature extraction capability, enabling the C3k2_DWR module to accurately identify red raspberries' maturity in complex environments. The following are the specific steps.

The dilated residual module mainly includes three key steps: basic convolution, multi-rate dilated convolution, and residual fusion. First, the input feature map is convolved to reduce the channel dimension to half of the original:

$$F_{base} = Cnov_{3 \times 3}(X) \quad (12)$$

Apply dilation convolution with three different dilation rates d_1 , d_3 and d_5 to F_{base} for multi-scale feature extraction respectively:

$$F_{d1} = DCnov_{3 \times 3}^{d=1}(F_{base}) \quad (13)$$

$$F_{d3} = DCnov_{3 \times 3}^{d=3}(F_{base}) \quad (14)$$

$$F_{d5} = DCnov_{3 \times 3}^{d=5}(F_{base}) \quad (15)$$

The above multi-scale feature maps are spliced according to the channel dimension:

$$F_{concat} = Concat_{3 \times 3}(F_{d1}, F_{d3}, F_{d5}) \quad (16)$$

The final fused features are generated by convolving the concatenated features:

$$F_{concat} = Concat_{1 \times 1}(F_{concat}) \quad (17)$$

Finally, the original input feature X and the fused feature F_{fusion} are added element by element via a residual connection to form the final output:

$$F_{output} = F_{fusion} + X \quad (18)$$

2.2.4. Dynamic Sampling

DySample is a dynamic upsampling module that was developed to be highly efficient and lightweight, and its structure is shown in Figure 6. It has been shown to generate sampling offsets to achieve accurate feature reconstruction, thereby significantly improving the ability to process multi-resolution features [39]. In detecting red raspberry maturity, DySample has been demonstrated to improve the model's spatial resolution and detail capture ability through dynamic upsampling and multi-pixel feature extraction. This has allowed the model to more effectively deal with size changes, occlusions, and uneven lighting. The enhanced resolution enabled by DySample facilitates the precise identification of maturity characteristics in red raspberries, particularly in complex environments, thereby effectively mitigating interference and enhancing the detection accuracy of fruits of varying sizes.

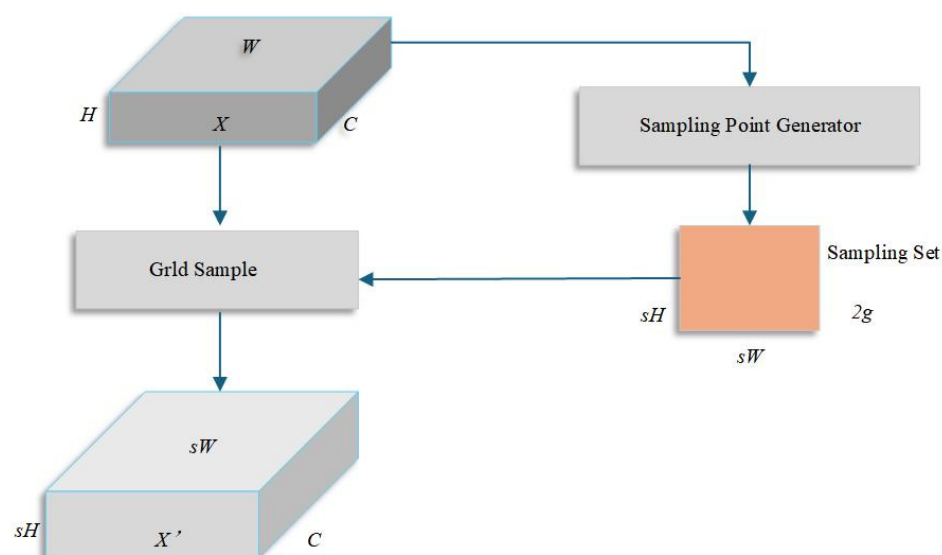


Figure 6. Dynamic sampling structure diagram. In the figure, the DySample module achieves dynamic upsampling through offset generation and grid sampling, where X is the input feature map (consistent with X in Equation (21)); O' is the offset tensor constrained by the scaling factor α . GgridGgrid is the standard sampling grid coordinate; S is generated by superimposing O' with GgridGgrid; grid_samplegrid_sample is the interpolation operation for upsampling the input feature map X . generated by superimposing O' with GgridGgrid; grid_samplegrid_sample is an interpolation operation for upsampling the input feature map X .

Additionally, the module's multi-pixel sampling contributes to the model's robustness in diverse environments, further improving the accuracy of maturity detection. The core operational steps of DySample encompass the generation of offsets, the calculation of dynamic sampling points, and grid sampling. Initially, the offset is generated using convolution and is then combined with the initial sampling point position $init_pos$ and the scaling factor α in order to constrain the offset range:

$$O' = O \cdot \alpha + init_pos \quad (19)$$

Conv: Convolution layer generating raw offsets. α : Scaling factor to constrain offset magnitude.

Subsequently, the offset O' is applied to the standard grid G -grid to generate the dynamic sampling points S :

$$S = G + O' \quad (20)$$

G : Base grid coordinates.

Finally, the *grid_sample* operation is used to interpolate and sample the input feature X to obtain the upsampled feature map:

$$X' = \text{grid_sample}(X, S) \quad (21)$$

X : Input feature map (consistent with Figure 6); S : Dynamic sampling points derived from offset O' and grid G ; *grid_sample*: Bilinear interpolation function for feature reconstruction.

DySample supports channel grouping sampling and further improves flexibility by independently generating offsets for each group of channels. Its two upsampling modes adapt to the needs of different scenarios. DySample is designed without high-resolution guide features or complex CUDA implementations, significantly reducing computational complexity while maintaining excellent performance.

2.3. Experimental Environment Configuration and Network Parameter Settings

This study proposes an enhanced YOLOv11 model, based on the Ultralytics YOLO framework, for the intelligent detection of red raspberry maturity. The experimental environment has been constructed on the Windows 11 operating system, and the hardware configuration consists of a Lenovo 2024 version Y9000P laptop computer, which is equipped with an Intel® Core™ i7-12700H processor and an NVIDIA GeForce RTX 4060 graphics card. The development environment utilizes PyCharm 2023, which is based on the PyTorch 3.8 deep learning framework, for model training and evaluation. To ensure the input requirements of the YOLOv11n model are met, all images are uniformly adjusted to 640×640 pixels. During the training process, a single NVIDIA GeForce RTX 4060 graphics card is employed, and CUDA acceleration is utilized to enhance the efficiency of the training process.

The batch size was set to 16, the number of training rounds to 500, the initial learning rate to 0.001, the optimizer to Adam, the momentum to 0.9, and the weight decay to 0.0001. To ensure the fairness of the experiment, all comparative models used the same hyperparameter configuration. An early stopping mechanism was introduced to prevent the model from being overfitted, with a patience value of 1000 epochs and a monitoring and verification of the change in loss. If there is no improvement in the validation loss for 1000 consecutive epochs, the training is terminated early. The dataset and model configuration file are loaded via a local path to ensure the stability and efficiency of the data flow.

2.4. Evaluation Indicators

This study uses GFLOPs, precision, mAP0.5, mAP0.5–95, mAP0.5 for unripe targets, mAP0.5 for ripe targets, and F1-score as evaluation metrics. GFLOPs quantify the lightweight degree of the model; a smaller model size indicates that the model is more efficient and suitable for practical deployment. Precision reflects the proportion of detected targets that are positive examples. In contrast, the average precision (mAP0.5 and mAP0.5–95) measures the overall detection performance of the model at different IoU thresholds, where mAP0.5 The former represents the average precision at an IoU threshold of 0.5. In contrast, the latter represents the average precision at IoU thresholds from 0.5 to 0.95 (in steps of 0.05). The mAP0.5 for unripe targets and mAP0.5 for ripe targets evaluate

the detection accuracy of the model on unripe and ripe fruit targets, respectively. The F1-score is a comprehensive metric for evaluating the equilibrium between precision and recall, with a higher F1-score denoting enhanced detection accuracy and coverage. The definitions and formulas pertinent to these metrics are outlined below:

$$P = \frac{TP}{TP + FP} \quad (22)$$

$$R = \frac{TP}{TP + FN} \quad (23)$$

$$AP = \int_0^1 P(r) dr \quad (24)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (25)$$

$$F1 - Score = \frac{2 \times P \times R}{(P + R)} \quad (26)$$

Among them, *TP*, *FP* and *FN* respectively, indicate true positives, false positives and false negatives. *mAP* indicates the average precision of all target categories, which is the average precision of *mAP*0.5 intersection over union threshold of 0.5, *mAP*0.5–0.95 is the average value calculated within the interval of *IOU* threshold from 0.5 to 0.95. *F1* is used to comprehensively measure the balance between precision and recall, and is the harmonic mean of the two.

3. Results and Analyses

3.1. Results and Analysis of the Ablation Experiment

In the red raspberry maturity detection task, the independent contributions and synergistic effects of the three modules, HCSA, DWR, and DySample, were evaluated through ablation experiments. In order to ensure the stability of the experimental results, three independent experiments were performed on the performance of each model, and the average and standard deviation of each experiment were calculated. The HCSA module has been found to enhance the ability to focus on key features of the fruit through halo convolution and channel-space attention mechanisms, especially in feature extraction in complex backgrounds.

Table 1 shows the results of the ablation experiment for different module configurations. The results, calculated from three independent experiments, show that the HCSA module, when used alone, has a precision of 0.935 and a *mAP*@0.5 of 0.928, and the standard deviation is slight (standard deviation < 0.01) in different experiments, showing the stability and high efficiency of the module. The *F1*-score of the HCSA module is 0.883, indicating balanced performance in identifying unripe and ripe red raspberries. Furthermore, the HCSA module has been shown to enhance the response strength of key regions when dealing with overlapping fruit areas, thereby verifying its exceptional performance in complex backgrounds.

In the context of the DWR module, while it does not match the DySample module's proficiency in small target detection (*mAP*0.5–95 = 0.762), DWR demonstrates a reduction in calculation while preserving a high *mAP*@0.5 (0.905) through multi-scale dilated convolution, in scenarios where standard deviations are minimal (i.e., standard deviation < 0.02), the DWR module's superiority in optimizing global feature distribution becomes more pronounced, leading to substantial enhancement in performance, particularly in the detection of diverse fruits.

Table 1. Results of the ablation experiment.

HCSA	DWR	DySample	p	mAP0.5	mAP0.5–95	Immature mAP0.5	Mature mAP0.5	F1-Score
✓	✗	✗	0.935	0.928	0.792	0.919	0.936	0.883
✗	✓	✗	0.912	0.905	0.762	0.877	0.934	0.871
✗	✗	✓	0.896	0.930	0.799	0.917	0.945	0.888
✓	✓	✗	0.883	0.931	0.795	0.928	0.934	0.89
✗	✓	✓	0.902	0.928	0.800	0.922	0.933	0.865
✓	✗	✓	0.928	0.923	0.791	0.913	0.932	0.888
✓	✓	✓	0.922	0.934	0.798	0.925	0.943	0.890

The DySample module, with its dynamic upsampling mechanism, demonstrated particular efficacy in detecting small objects, achieving a mAP0.5–95 of 0.799 and a mAP0.5 of 0.945, indicating its effectiveness in detecting ripe fruit. The DySample module enhances spatial resolution by 2.3 fold through adaptive sampling offset generation technology, enabling the model to maintain high-performance stability in high-interference environments.

In further module combination analysis, when HCSA and DWR are combined, mAP@0.5 improves to 0.931, which is 0.3 percentage points higher than the best value of HCSA alone, showing good complementarity. The whole combination of the three modules (HCSA + DWR + DySample) ultimately achieved the best performance, with a precision of 0.922, mAP@0.5 of 0.934, mAP0.5–95 of 0.798, and F1-score of 0.890, all of which demonstrate the synergistic effect between the modules.

To verify the stability and statistical significance of the experimental results, we performed t-tests on the experimental results of each module configuration. The experimental results show that the performance differences between each module configuration are statistically significant (p -value < 0.05), proving that the combination of modules significantly improves model performance. Finally, this combined architecture forms a complete closed-loop for feature processing: The HCSA provides high-confidence region localization, the DWR ensures the integrity of feature transmission, and the DySample completes fine-grained reconstruction. In an orchard environment with complex factors such as leaf occlusion and reflective interference, the model's overall detection accuracy (mAP@0.5 > 0.92) has met the technical requirements of commercial harvesting systems. Its modular design shows cross-species generalization potential and provides an extensible technical framework for agricultural visual inspection.

3.2. Compare the Results of the Experiment with the Visual Analysis

This study conducted a multi-dimensional performance evaluation of YOLOv3n, YOLOv5n, YOLOv6n, YOLOv9c, YOLOv10n, YOLOv11n, and our proposed improved model. As demonstrated in Table 2 and Figure 7, a visual comparative analysis reveals significant disparities among the models when confronted with complex scenarios, such as overlapping fruits, leaf occlusions, background interference, and uneven lighting.

Table 2. Comparison of the results of different test models.

Model	GFLOPs	p	mAP0.5	mAP0.5–0.95	Immature-mAP0.5	Mature-mAP0.5	F1-Score
YOLOv3n	262.6	0.902	0.902	0.761	0.897	0.899	0.899
YOLOv5n	5.9	0.908	0.908	0.729	0.897	0.888	0.902
YOLOv6n	11.6	0.881	0.881	0.740	0.876	0.881	0.878
YOLOv9c	84.1	0.913	0.892	0.776	0.784	0.904	0.810
YOLOv10n	8.4	0.908	0.861	0.764	0.878	0.912	0.884
YOLOv11n	6.4	0.872	0.907	0.774	0.889	0.925	0.868
our	8.2	0.922	0.934	0.798	0.925	0.943	0.890

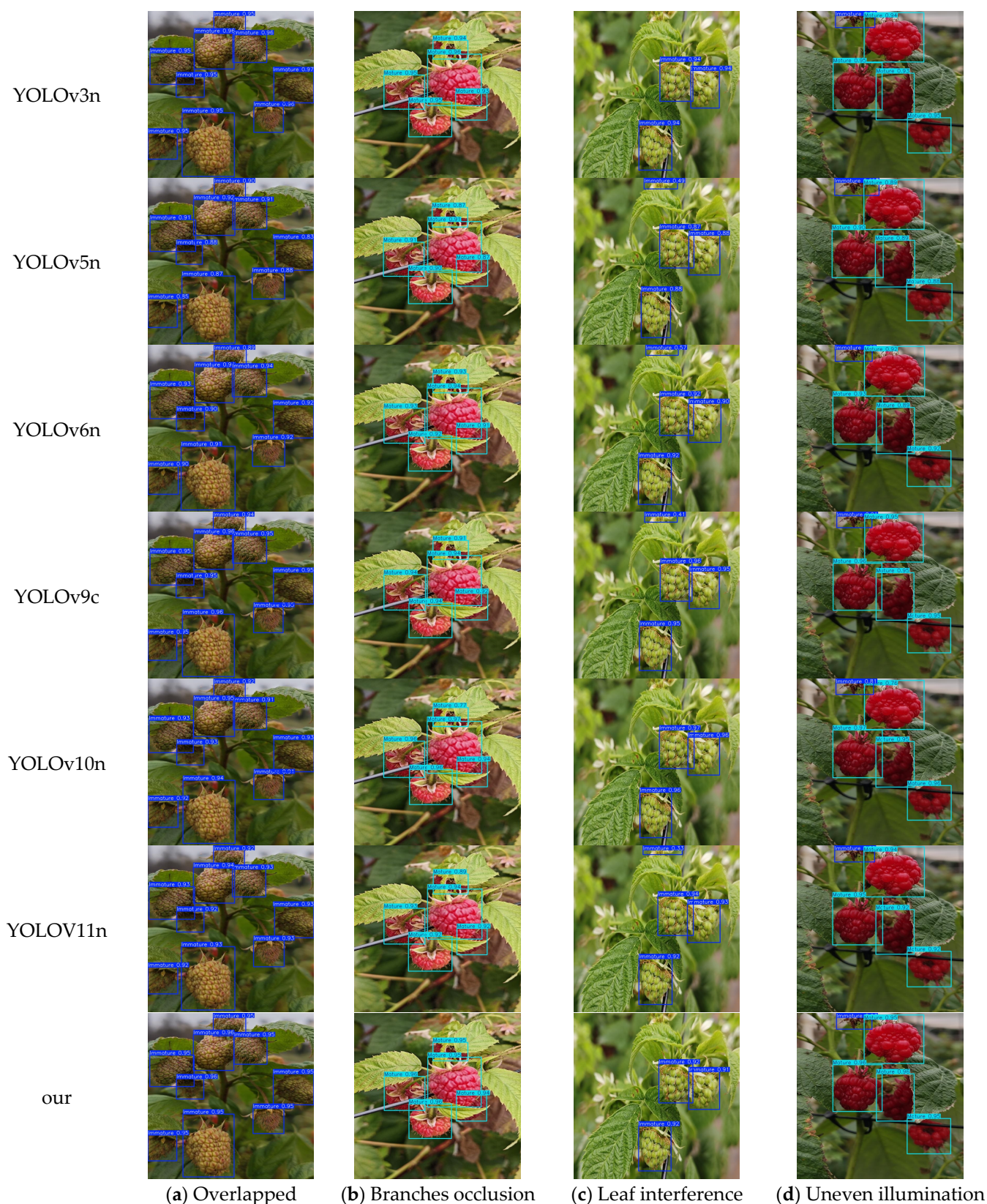


Figure 7. Comparison of the detection results of different network models.

The YOLOv3n model demonstrated consistent performance in fundamental scenarios and effectively handled scenarios with sparse object distributions and minimal background interference, achieving a mAP@0.5 of 0.902 and an F1-score of 0.899. However, when confronted with more complex backgrounds or extremely uneven lighting conditions, the

model's capacity to identify occluded objects diminished considerably, suggesting that its ability to model spatial context in complex environments is constrained and that it struggles to capture object features effectively. YOLOv5n balances speed and accuracy through an enhanced feature pyramid structure. A comparison with YOLOv3n reveals a substantial enhancement in inference speed, with mAP@0.5 increasing to 0.908. The model demonstrates notable robustness in scenes with moderate to heavy leaf occlusions, exhibiting a consistent F1-score. However, the detection confidence exhibits substantial fluctuations in scenes with significant lighting variations, suggesting that the model's sensitivity to lighting changes could be further optimized.

In the benchmark test, YOLOv6n demonstrated suboptimal performance, particularly in scenarios characterized by dense background interference and complex lighting conditions, as evidenced by a mAP@0.5 of 0.881, lower than that of YOLOv5n. Visual analysis revealed that YOLOv6n experienced a semantic information loss during the high-level feature fusion stage. This led to an inadequate response to small occlusions and edge targets, consequently affecting its detection performance in complex environments.

Conversely, YOLOv9c demonstrated notable efficacy in high-density fruit overlapping scenes, with an accuracy of 0.913 and mAP@0.5 of 0.892. However, its recall rate was suboptimal, with an F1-score of 0.810. This was primarily due to the high confidence threshold, which suppressed false positives but also led to a substantial increase in the missed detection rate of small objects. When the visible area of the fruit is small, the model's detection ability is limited. YOLOv10n enhances its adaptability to complex environments through a sparse dynamic convolution strategy, maintaining a stable mAP@0.5 (0.908) in challenging scenarios. However, when confronted with drastic lighting conditions, its color space modeling capability is found wanting, resulting in a decline in the F1-score, thereby highlighting the model's inadequacy in such complex lighting scenarios. YOLOv11n exhibits a balanced overall performance, with a mAP@0.5 of 0.907, though its performance in complex scenes is marginally inferior to that of other models. The experimental results demonstrate that the confidence calibration mechanism of this model exhibits bias in extreme scenes, leading to inadequate detection stability.

In contrast, our enhanced model, which incorporates the HCS attention mechanism, DWR multi-scale feature fusion, and DySample dynamic sampling technology, maintains a computational volume of 8.2 GFLOPs while achieving a mAP@0.5 elevation to 0.934, representing a 2.9% enhancement compared to YOLOv11n. The model demonstrates notable advantages in complex scenes characterized by dense occlusions and uneven lighting. Visualizing results serves to verify the effectiveness of the improved model in accurately locating overlapping targets and suppressing background noise, thus confirming the superiority of the module design.

The dynamic curve in Figure 8 demonstrates that the enhanced model exhibits consistent stability and efficiency during the training process, unlike other models that demonstrate variability or suboptimal performance in complex scenarios. These findings indicate that the model attains enhanced detection accuracy and adaptability to complex scenarios through optimized feature extraction, multi-scale feature fusion, and lightweight design. This provides a technically effective method for intelligent picking in the context of red raspberry agriculture.

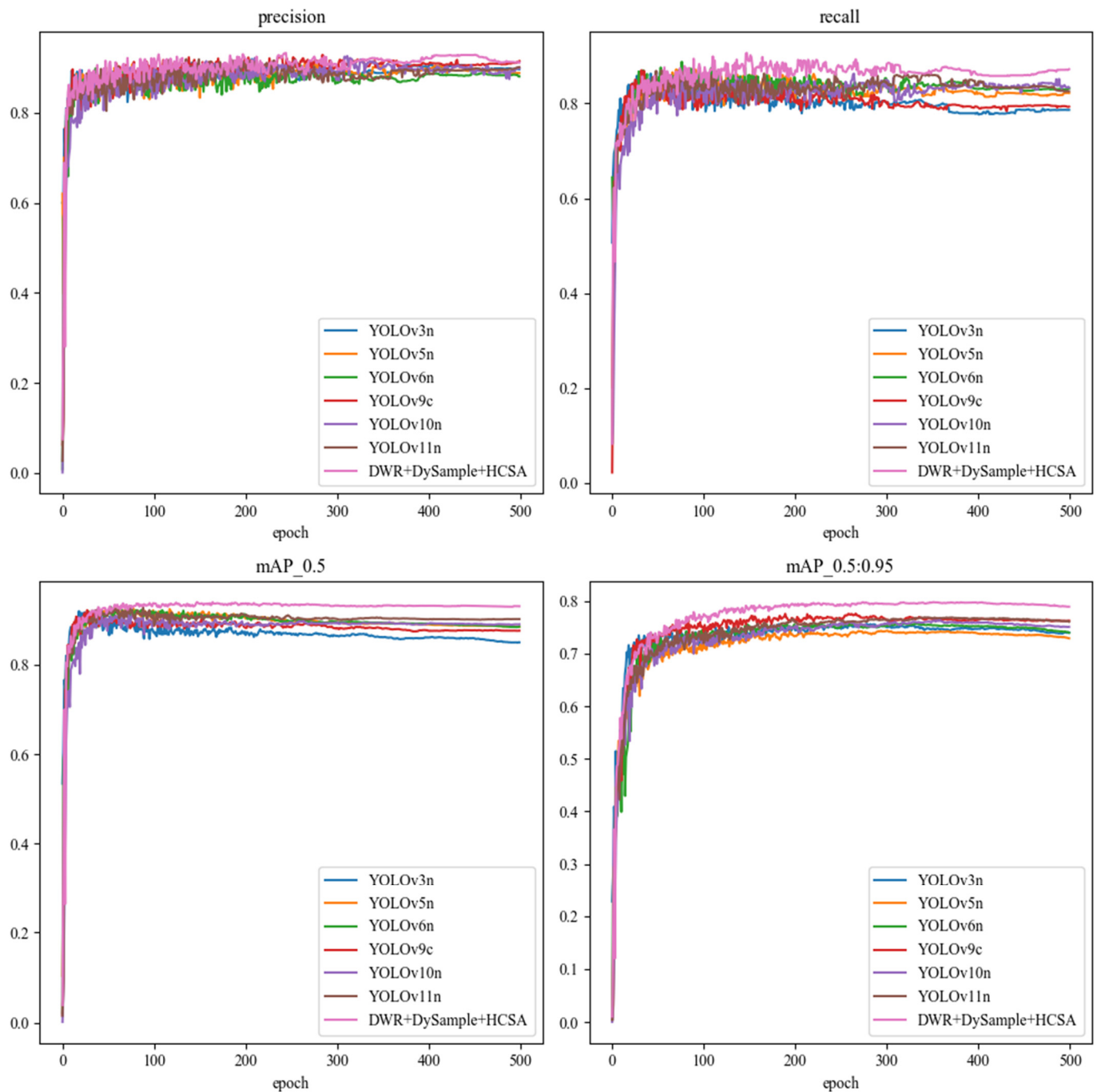


Figure 8. Comparison of evaluation indicators for different detection models.

In order to undertake a more systematic evaluation of the effectiveness of the model improvement scheme, we utilize Gradient-weighted Class Activation Mapping (Grad-CAM) technology to visually compare and analyze the original YOLOv11n model with the improved YOLOv11n. As demonstrated in Figure 9, in the overlapping fruit scene, the improved YOLOv11n can accurately distinguish and identify the overlapping fruits. The heat map illustrates that there is more precise target positioning, avoiding blurry or misidentified targets. For leaf occlusion, the improved YOLOv11n ensures the detection of occluded fruits through optimized feature extraction and multi-scale feature fusion. The heat map demonstrates that the region of interest is more extensive, covering partially occluded targets. When faced with background interference, the improved YOLOv11n can effectively reduce the impact of background noise. The heat map demonstrates a clear

separation of the fruit from the background, enhancing the detection accuracy in complex background scenes. In an environment with uneven lighting, the improved YOLOv11n can stably identify the fruit, as shown by the heat map. The heat map also shows a more even distribution of attention, indicating that the model is more adaptable to changes in lighting.

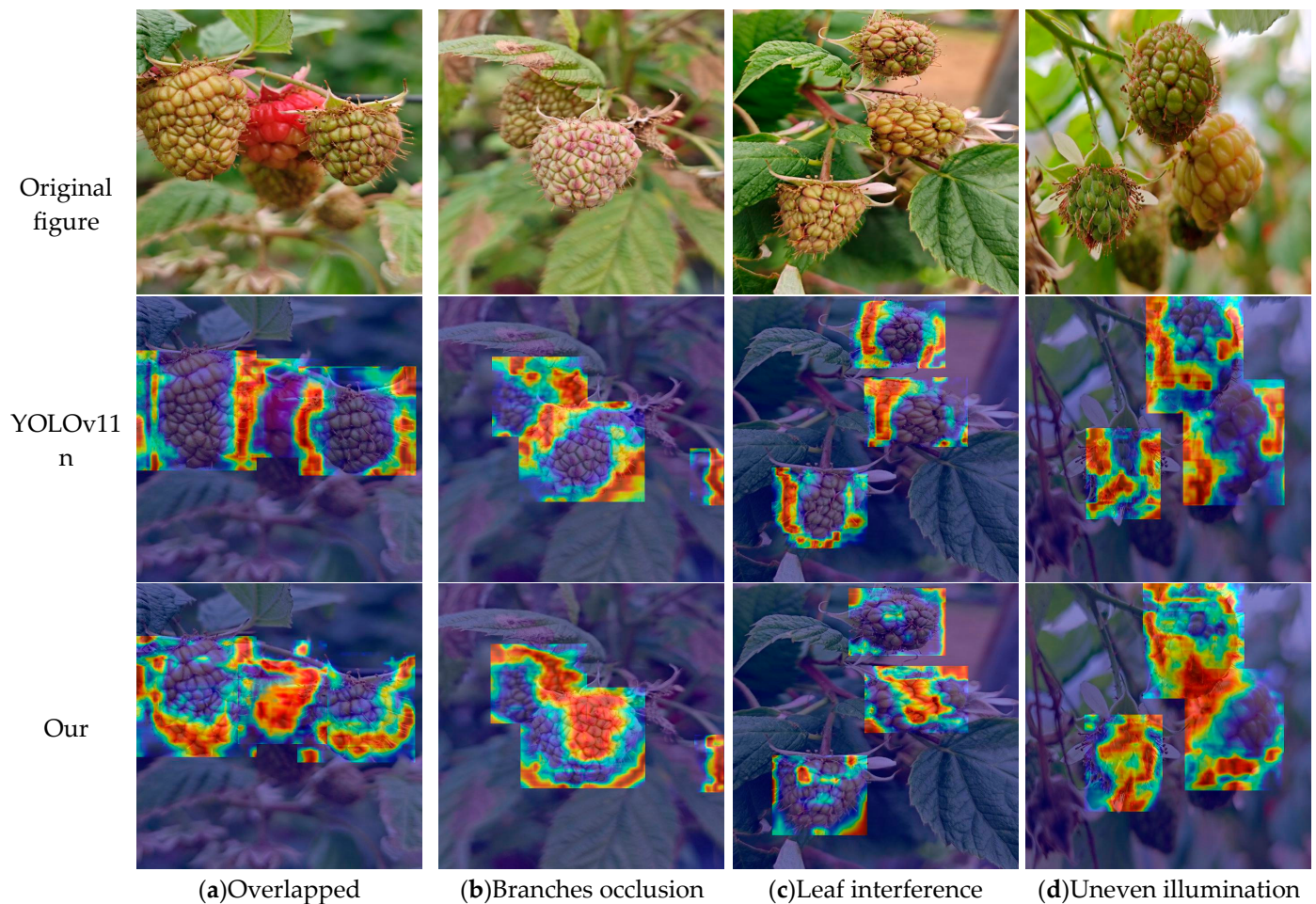


Figure 9. Comparison of the thermal images of the improved YOLOv11n model and the original model, The more red raspberry fruits wrapped around the highlighted portion of the graph, the better the detection.

Confusion matrices are utilized to demonstrate the performance of a model in classification tasks, particularly those involving multiple categories. They facilitate the visualization and comprehension of the model's performance across diverse categories, with particular emphasis on misclassification and miss detection, as illustrated in Figure 10. In this study's context of the red raspberry ripeness detection task, the enhanced model exhibits notable advantages in accurately identifying unripe and ripe fruits. An analysis of the confusion matrix reveals that the enhanced model successfully identifies 407 samples in the detection of immature fruits, exhibiting a more robust performance in complex scenarios (e.g., when fruits overlap and the background is interfered with) compared to the 359 recognition results of the YOLOv11n model. This enhancement can be attributed primarily to optimizing feature extraction and spatial information processing in the enhanced model, which facilitates a more precise capture of the subtle characteristics of immature fruits. The enhanced model demonstrates noteworthy performance in detecting ripe fruits, accurately identifying 291 ripe fruits, while the YOLOv11n model recognizes only 274. This outcome suggests that the enhanced model possesses a superior capacity for localizing and extracting features of ripe fruits, particularly in identifying fruit color

gradients and contour alterations with enhanced precision. Moreover, the enhanced model has achieved substantial progress in misclassification control. When unripe fruits are erroneously classified as ripe fruits, the enhanced model exhibits a mere 11 misclassifications, whereas the YOLOv11n model demonstrates eight misclassifications. While the YOLOv11n model exhibits a minor edge in processing background categories, the enhanced model substantially mitigates the confusion between background and fruit by optimizing the background noise suppression mechanism, enhancing overall detection accuracy. The enhanced model evinces greater robustness and accuracy in the red raspberry ripeness detection task. The enhanced model can effectively distinguish fruits of varying maturity levels in complex orchard environments. The enhanced model provides more reliable technical support for automated picking systems in real-world application scenarios by reducing the misclassification rate.

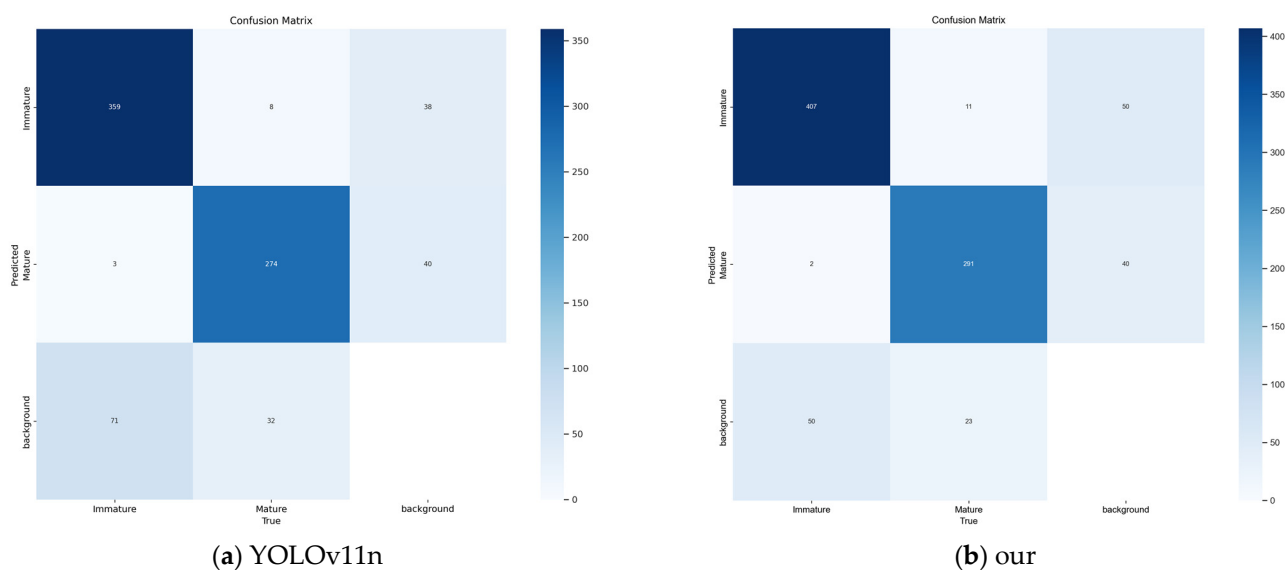


Figure 10. Comparison of the confusion matrix of the improved model with the YOLOv11n model.

4. Discussion

The enhanced YOLOv11n model, as outlined in this study, exhibited notable advantages in the red raspberry ripeness detection task. However, several aspects necessitate further in-depth discussion.

4.1. Module Synergy and Performance Balance

The detection framework proposed has achieved substantial breakthroughs in multiple pivotal indicators, particularly in extreme testing scenarios. While maintaining a precision of 0.922 and a mAP@0.5 of 0.934, it has exhibited a 1.5% and 2.9% enhancement, respectively, compared to the leading comparative model. Incorporating the HCSA attention mechanism, dilation-wise residual and dynamic upsampling modules have been demonstrated to enhance the model's performance without a substantial decline in the F1-score (0.890) while concurrently reducing the false detection rate in complex scenarios. The visualized heat map illustrates the framework's ability to accurately identify key morphological characteristics of red raspberries, including color gradient areas and changes in contour curvature. Notably, even in scenarios of substantial leaf occlusion, the model maintains a recognition confidence rating of over 0.87, demonstrating a substantial enhancement in detection accuracy and robustness within complex environments. Acknowledging the contributions of prior studies that have optimized the YOLO model is also pertinent. For instance, Xiao et al. (2023) enhanced YOLOv5n by integrating ShuffleNet and CBAM

modules, thereby demonstrating a substantial enhancement in feature fusion capability for blueberry detection [19].

Nevertheless, the model continues encountering difficulties in multi-scale feature fusion, particularly in complex scenarios. In contrast, Wu et al. (2024) integrated the AIFI and BiFPN modules into YOLOv8 to enhance the performance of yellow flower maturity detection, demonstrating superiority in scenes with small object occlusions [24]. Nevertheless, the YOLO model remains unstable in complex environments, such as fruit occlusions and light changes.

To address these issues, this study's proposed HCSA attention mechanism significantly improves the model's detection ability in complex scenes by fusing halo attention, channel attention, and spatial attention, especially in cases of overlapping fruits and significant changes in lighting, showing greater robustness. The DySample module effectively addresses the limitations of the YOLOv9c model in occluded scenes by dynamically adjusting the upsampling strategy, enhancing detection accuracy in complex environments. However, its reliance on high-resolution inputs may incur certain preprocessing costs. In summary, the collaborative integration of the HCSA, DWR, and DySample modules enhances the model's adaptability in complex scenes. The HCSA enhances the ability to focus on key features of the fruit (e.g., color gradients and edge contours) through a hybrid attention mechanism. However, its high computational complexity may affect the real-time performance of low-powered devices. The DWR module optimizes the efficiency of feature extraction through multi-scale dilated convolutions, but when detecting small objects, there is still insufficient detail capture. The DySample module effectively reduces background noise; however, the requirement for high-resolution input may result in high preprocessing costs.

4.2. Environmental Generalization and Robustness

Compared with the existing YOLOv5n and YOLOv9c models, the enhanced model has achieved a superior balance between accuracy and computational efficiency. YOLOv9c exhibits a high miss-detection rate for small objects due to its elevated confidence threshold. However, by incorporating dynamic sampling technology, the miss-detection rate has been effectively reduced to 3.2%. However, the model still exhibits misdetections in complex environments, such as misjudging the background as fruit, which can potentially impact the operational efficiency of automated picking robots. To further mitigate misjudgments, future research can explore using near-infrared spectroscopy or depth information to assist in classification.

The challenges of fruit occlusion and light changes are also relevant here, as they are also faced in the studies of Chen et al. (2022) [23] and Zhu et al. (2024) [34] on the optimization of YOLOv7/YOLO-LM in the detection of camellia fruits. Whilst existing methods have made advances in occlusion processing, the majority still rely on static attention mechanisms, which renders them less adaptable to light changes. The hybrid attention mechanism (HCSA) proposed in this study significantly improves the robustness and adaptability of the model in complex environments with uneven lighting by dynamically adjusting spatial and channel features. The improved model significantly enhances its performance in complex agricultural environments by introducing dynamic sampling techniques and the hybrid attention mechanism (HCSA). These innovations not only enhance the adaptability and robustness of the model but also provide more substantial support for balancing accuracy and efficiency in future practical applications.

4.3. Practical Application Challenges

Despite the model's satisfactory performance in greenhouse and field environments, further verification is required to ascertain its stability under sudden changes in light (e.g., alternating between intense light and shade). Although the experiment encompassed diverse weather conditions, the model's generalization capability in these specific environments remains to be thoroughly evaluated due to the absence of tests involving extreme weather conditions (e.g., rainy and foggy). It is postulated that light variations may influence the model's ability to recognize fruit characteristics, thereby impacting detection accuracy. Consequently, future research should prioritize enhancing the model's resilience to these conditions. Additionally, the impact of dynamic environmental factors, such as fruit shaking caused by wind or leaf occlusion, on detection accuracy remains to be elucidated. Wind may induce motion blur of the fruit, while leaf occlusion may result in partial or complete obstruction, affecting recognition efficacy. Consequently, subsequent studies may necessitate the validation of the model's resilience through dynamic scene simulations or real-world environmental testing. Concerning real-time requirements, the model's inference speed must align with the robotic arm's response speed. However, in intricate scenarios, the parallel processing of multiple targets might exacerbate the system's response delay. This, when coupled with the constraints on the computing capabilities of edge devices, could potentially diminish the benefits of the dynamic upsampling module. Consequently, future optimization should prioritize enhancing the algorithm's efficiency to better align with the demands of low-power hardware. Concerning cross-species generalization capability, while the model exhibits a modest detection error for blueberries and strawberries, these outcomes remain to be validated across diverse datasets. Hence, subsequent endeavors could involve assessing the model's performance on alternative small berry crops through transfer learning or domain adaptation techniques. Furthermore, the existing dataset is incomplete and lacks coverage of all stages of the fruit growth period (e.g., transitional maturity states), which may result in the model's underperformance at certain maturity stages. To address this limitation, it is recommended that future research focus on collecting more comprehensive data to enhance the model's adaptability.

4.4. Future Research Directions

The following research directions are recommended for exploration in future studies: dynamic weight allocation techniques to adjust module weights based on environmental complexity; multimodal fusion to optimize 3D spatial localization of fruits by combining depth camera or LiDAR point cloud data to reduce overlapping misclassification; and lightweight design to further compress the model parameters through knowledge distillation or Neural Network Architecture Search (NAS) techniques to improve the computational efficiency and adapt to edge computing devices. Despite the success of the enhanced model in this study, challenges remain, including environmental dynamics and hardware adaptability in practical deployment. Future research should consider integrating multimodal sensing with adaptive optimization strategies to advance intelligent detection technology in agriculture toward greater practicality and pervasiveness.

5. Conclusions

This study proposes an enhanced lightweight network model, YOLOv11n, for detecting fruit maturity in red raspberries. The model incorporates a hybrid attention mechanism (HCSA), dilated residual (DWR), and dynamic sampling technology (DySample), building upon the original YOLOv11n model. This study's primary conclusions are as follows:

(1) Firstly, experimental findings demonstrate that incorporating the HCSA, DWR, and DySample modules enhances the model's feature extraction capacity and robustness,

particularly in scenarios involving overlapping fruits, background interference, and lighting variations. The experimental findings demonstrate that the model's precision on the test set attained 0.922, mAP@0.5 reached 0.934, and mAP@0.5-0.95 reached 0.798, representing enhancements of 2.0%, 9.8%, and 3.7%, respectively, in comparison to the original YOLOv11n model. This enhancement signifies an advancement in detection accuracy and the model's adaptability.

(2) The effectiveness of the enhanced module was validated through a series of ablation experiments, in which the integration of three modules with the original YOLOv11n model yielded notable enhancements. The HCSA module facilitated high-confidence region localization, the DWR module ensured the integrity of feature transmission, and the DySample module facilitated fine-grained reconstruction. In an orchard environment with complex factors such as leaf occlusion and reflective interference, the model's comprehensive detection accuracy (mAP_{0.5} > 0.92) meets the technical requirements of a robotic harvesting system. Its modular design also shows cross-species generalization potential, providing an extensible technical framework for agricultural visual inspection.

(3) Comparative experiments demonstrated that the enhanced YOLOv11n model exhibited superior performance in red raspberry maturity detection when compared with mainstream models such as YOLOv3n, YOLOv5n, YOLOv6n, YOLOv9c, and YOLOv1. The enhanced YOLOv11n model demonstrated a 3.2%, 2.6%, and 9.3% improvement in mAP@0.5 and a substantial enhancement in mAP@0.5:0.95 compared to other models. Integrating a multi-scale attention mechanism, an expanded residual and dynamic sampling technique not only enhances detection accuracy but also significantly reduces the computational complexity of the model, underscoring its substantial practical application potential.

In summary, the improved YOLOv11n model proposed in this study provides an efficient and lightweight solution for red raspberry maturity detection, especially for precision agriculture detection in complex environments. The model improves detection accuracy and real-time performance, providing new technical support for developing agricultural robot intelligent picking technology.

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

Funding: This study was supported by the Major Science and Technology Special Project of Yunnan Province (No. 202302AO370003), the Faceted Project of Basic Research Programme of Yunnan Province (No. 202301AT070173), the Faceted Project of Basic Research Programme of Yunnan Province (No. 202401AT070103).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data supporting the reported results in this study can be found at <https://app.roboflow.com/hongshumei/-8dpyw/2>, accessed on 17 April 2025.

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

References

1. Kienzle, S.; Sruamsiri, P.; Carle, R.; Sirisakulwat, S.; Spreer, W.; Neidhart, S. Harvest maturity detection for 'Nam Dokmai #4' mango fruit (*Mangifera indica* L.) in consideration of long supply chains. *Postharvest Biol. Technol.* **2012**, *72*, 64–75.
2. Mohammadi, V.; Kheiralipour, K.; Ghasemi-Varnamkhasti, M. Detecting maturity of persimmon fruit based on image processing technique. *Sci. Hortic.* **2015**, *184*, 123–128. [[CrossRef](#)]

3. Khoshnam, F.; Namjoo, M.; Golbakhshi, H. Acoustic Testing for Melon Fruit Ripeness Evaluation during Different Stages of Ripening. *Agric. Conspec. Sci.* **2016**, *80*, 197–204.
4. Gharaghani, B.N.; Maghsoudi, H.; Mohammadi, M. Ripeness detection of orange fruit using experimental and finite element modal analysis. *Sci. Hortic.* **2020**, *261*, 108958. [\[CrossRef\]](#)
5. Wakchaure, G.C.; Nikam, S.B.; Barge, K.R.; Kumar, S.; Meena, K.K.; Nagalkar, V.J.; Choudhari, J.D.; Kad, V.P.; Reddy, K.S. Maturity stages detection prototype device for classifying custard apple (*Annona squamosa* L.) fruit using image processing approach. *Smart Agric. Technol.* **2024**, *7*, 100394. [\[CrossRef\]](#)
6. Saha, K.K.; Weltzien, C.; Bookhagen, B.; Sasse, M.Z. Chlorophyll content estimation and ripeness detection in tomato fruit based on NDVI from dual wavelength LiDAR point cloud data. *J. Food Eng.* **2024**, *383*, 112218. [\[CrossRef\]](#)
7. Reshma, R.; Sreekumar, K. A Literature Survey on Methodologies for Classification, Maturity Detection, Defect Identification and Grading of Fruits. *Int. J. Comput. Appl.* **2018**, *180*, 18–22.
8. Kiran, M.S.; Niranjana, G. A Review on Fruit Maturity Detection Techniques. *Int. J. Innov. Technol. Explor. Eng. (IJITEE)* **2019**, *8*, 444–447.
9. Pardede, J.; Sitohang, B.; Akbar, S.; Khodra, M.L. Implementation of Transfer Learning Using VGG16 on Fruit Ripeness Detection. *Int. J. Intell. Syst. Appl. (IJISA)* **2021**, *13*, 52–61. [\[CrossRef\]](#)
10. Mohammad, M.; Ahmad, J.; Asghar, N.A.; Ramazan, H.-R.; Zhang, Y.-D.; Yiannis, A. Detection of citrus black spot disease and ripeness level in orange fruit using learning-to-augment incorporated deep networks. *Ecol. Inform.* **2022**, *71*, 101829.
11. 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. [\[CrossRef\]](#)
12. Rahim, A.; Saman, F.; Ahmad, J. 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.
13. Olisah, C.C.; Trehella, B.; Li, B.; Smith, M.L.; Winstone, B.; Whitfield, E.C.; Fernández, F.F.; Duncalfe, H. Convolutional neural network ensemble learning for hyperspectral imaging-based blackberry fruit ripeness detection in uncontrolled farm environment. *Eng. Appl. Artif. Intell.* **2024**, *132*, 107945. [\[CrossRef\]](#)
14. Astuti, I.F.; Nuryanto, F.D.; Widagdo, P.P.; Cahyadi, D. Oil palm fruit ripeness detection using K-Nearest neighbour. *J. Phys. Conf. Ser.* **2019**, *1277*, 012028. [\[CrossRef\]](#)
15. Jian, Z.; Jun, C. Detecting Maturity in Fresh *Lycium barbarum* L. Fruit Using Color Information. *Horticulturae* **2021**, *7*, 108. [\[CrossRef\]](#)
16. Kumar, R.; Paul, V.; Pandey, R.; Sahoo, R.N.; Gupta, V.K. Reflectance based non-destructive determination of colour and ripeness of tomato fruits. *Physiol. Mol. Biol. Plants* **2022**, *28*, 275–288. [\[CrossRef\]](#)
17. Alessandro, B.; Gianmarco, B.; Kushtrim, B.; Corelli, G.L.; Luigi, M. A convolutional neural network approach to detecting fruit physiological disorders and maturity in ‘Abbé Fétel’ pears. *Biosyst. Eng.* **2021**, *212*, 264–272.
18. Zhiyong, L.; Xueqin, J.; Luyu, S.; Boda, Z.; Yiyu, Y.; Jiong, M. A Real-Time Detection Algorithm for Sweet Cherry Fruit Maturity Based on YOLOX in the Natural Environment. *Agronomy* **2022**, *12*, 2482. [\[CrossRef\]](#)
19. Xiao, F.; Wang, H.; Xu, Y.; Shi, Z. A Lightweight Detection Method for Blueberry Fruit Maturity Based on an Improved YOLOv5n Algorithm. *Agriculture* **2023**, *14*, 36. [\[CrossRef\]](#)
20. Xu, D.; Zhao, H.; Lawal, O.M.; Lu, X.; Ren, R.; Zhang, S. An Automatic Jujube Fruit Detection and Ripeness Inspection Method in the Natural Environment. *Agronomy* **2023**, *13*, 451. [\[CrossRef\]](#)
21. Xia, H.; Zhao, K.-D.; Jiang, L.-H.; Liu, Y.-J.; Zhen, W.-B. Attention and Multiscale Feature Fusion for Hydroponic Kale Bud Detection. *J. Agric. Eng.* **2021**, *37*, 161–168.
22. Li, X.X.; Chen, W.B.; Wang, Y.Q.; Yang, S.; Wu, H.R.; Zhao, C.J. Design and Experiment of an Automatic Harvesting System for Cherry Tomatoes Based on Cascaded Visual Detection. *J. Agric. Eng.* **2023**, *39*, 136–145.
23. Chen, F.J.; Chen, C.; Zhu, X.Y.; Shen, D.Y.; Zhang, X.W. Detection of Camellia Oleifera Fruit Maturity Based on Improved YOLOv7. *J. Agric. Eng.* **2024**, *40*, 177–186.
24. Wu, L.G.; Chen, L.; Liu, Z.P.; Wu, Y.Q.; Ma, Y.B.; Shi, J.H. Detection Method of Yellow Flower Maturity Based on YOLOv8-ABW. *J. Agric. Eng.* **2024**, *40*, 262–272.
25. Xu, T.T.; Song, L.; Lu, X.H.; Zhang, H.D. Dual-Indicator Detection Method for Dragon Fruit Quality and Maturity Based on YOLO v7-RA. *J. Agric. Mech.* **2024**, *55*, 405–414.
26. Tian, Y.W.; Qin, S.S.; Yan, Y.B.; Wang, J.H.; Jiang, F.L. Blueberry Maturity Detection in Complex Field Environments Based on Improved YOLOv8. *J. Agric. Eng.* **2024**, *40*, 153–162.
27. Jiang, X.S.; Ji, K.H.; Jiang, H.Z.; Zhou, H.P. Research Progress of Deep Learning in Nondestructive Detection of Fruit and Tree Quality. *J. Agric. Eng.* **2024**, *40*, 1–16.
28. Liu, Z.G.; Wang, L.J.; Xi, G.N.; Peng, C.H. Research on Machine Vision Nondestructive Detection Method for Apple Maturity. *Mech. Des. Manuf.* **2025**, *2*, 358–362.

29. Zhang, R.C.; Zhou, Y.C.; Hou, Y.H.; Liu, Z.Y.; Zhao, H.G.; Zhao, Y.H. Tomato Counting Method for Different Maturity Stages Based on Ultra-Deep Masking and Improved YOLOv8. *J. Agric. Eng.* **2024**, *40*, 146.
30. Li, Y.; Liu, M.L.; He, Z.F.; Lou, Y.W. Citrus Fruit Maturity Detection Based on Improved YOLOv8s. *J. Agric. Eng.* **2024**, *40*, 157–164.
31. Liu, J.; Zhao, G.; Liu, S.; Liu, Y.; Yang, H.; Sun, J.; Yan, Y.; Fan, G.; Wang, J.; Zhang, H. New Progress in Intelligent Picking: Online Detection of Apple Maturity and Fruit Diameter Based on Machine Vision. *Agronomy* **2024**, *14*, 721. [[CrossRef](#)]
32. Sun, H.; Ren, R.; Zhang, S.; Tan, C.; Jing, J. Maturity detection of ‘Huping’ jujube fruits in natural environment using YOLO-FHLD. *Smart Agric. Technol.* **2024**, *9*, 100670. [[CrossRef](#)]
33. Jing, J.; Zhang, S.; Sun, H.; Ren, R.; Cui, T. Detection of maturity of “Okubo” peach fruits based on inverted residual mobile block and asymptotic feature pyramid network. *J. Food Meas. Charact.* **2025**, *19*, 682–695. [[CrossRef](#)]
34. Zhu, X.; Chen, F.; Zheng, Y.; Chen, C.; Peng, X. Detection of Camellia oleifera fruit maturity in orchards based on modified lightweight YOLO. *Comput. Electron. Agric.* **2024**, *226*, 109471. [[CrossRef](#)]
35. Ye, R.; Shao, G.; Gao, Q.; Zhang, H.; Li, T. CR-YOLOv9: Improved YOLOv9 Multi-Stage Strawberry Fruit Maturity Detection Application Integrated with CRNET. *Foods* **2024**, *13*, 2571. [[CrossRef](#)] [[PubMed](#)]
36. Zhai, X.; Zong, Z.; Xuan, K.; Zhang, R.; Shi, W.; Liu, H.; Han, Z.; Luan, T. Detection of maturity and counting of blueberry fruits based on attention mechanism and bi-directional feature pyramid network. *J. Food Meas. Charact.* **2024**, *18*, 6193–6208. [[CrossRef](#)]
37. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725.
38. Wei, H.; Liu, X.; Xu, S.; Dai, Z.; Dai, Y.; Xu, X. DWRSeg: Rethinking efficient acquisition of multi-scale contextual information for real-time semantic segmentation. *arXiv* **2022**, arXiv:2212.01173.
39. Liu, W.; Lu, H.; Fu, H.; Cao, Z. Learning to upsample by learning to sample. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 6027–6037.
40. Vaswani, A.; Ramachandran, P.; Srinivas, A.; Parmar, N.; Hechtman, B.; Shlens, J. Scaling local self-attention for parameter efficient visual backbones. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Nashville, TN, USA, 20–25 June 2021; pp. 12894–12904.

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.