



## DMP-YOLO: Dense multi-scale perception for complex scenes YOLO algorithm *Prunus humilis* small target detection

Linyou Lv<sup>a,b,\*</sup>, Jiahui Li<sup>a</sup>, Yan Zhao<sup>b</sup>

<sup>a</sup> Liaoning Technical University, School of Environmental Science and Engineering, Fuxin 123000, China

<sup>b</sup> Liaoning Institute of Sand Land Management and Utilization, 55 Zhonghua Road, Fuxin, China

### ARTICLE INFO

**Keywords:**

Prunus humilis  
Deep learning YOLO  
Complex background  
Multi-scale perception  
Small Object

### ABSTRACT

To address the critical challenges of low detection accuracy and target loss in multi-scale *Prunus humilis* small fruit detection under severe occlusion by branches and leaves, this study proposes an improved DMP-YOLO model based on YOLO-v8. First, a multi-channel attention mechanism (C2f\_MSMA module) is integrated into the deep layers of the YOLO-v8 backbone network. This module enhances multi-scale fruit feature extraction capabilities, suppresses background noise interference, and improves small target recognition. Second, Spatial Pyramid Pooling (SPP) and TransformerBlock modules replace the original C2f modules in the final two backbone layers, establishing long-range dependencies and global contextual relationships to mitigate target loss. Third, a "Multi-Dimensional Adaptive Convolution" module (C2f\_MDAC) is embedded into the bottleneck, adaptively fusing multi-scale target information to optimize the detection efficiency of *Prunus humilis* fruits across varying sizes. Extensive experiments demonstrate that the improved model effectively enhances detection accuracy for small *Prunus humilis* targets in complex backgrounds with severe occlusions, both under normal lighting and dim conditions. Under normal lighting, the DMP-YOLO model achieves an average detection accuracy of 73.5 %, outperforming YOLO-v3, YOLO-v5, YOLO-v6, and YOLO-v8 by 4.7, 23, 19.3, and 14.3 percentage points respectively. Under dim lighting, the DMP-YOLO model achieves an average detection accuracy of 69.8 %, surpassing YOLO-v8, YOLO-v10, YOLO-v11, and YOLO-v12 by 12.9, 12.9, 16.9, and 16.5 percentage points respectively. Under normal lighting conditions, the DMP-YOLO model achieves an inference speed of 0.9 milliseconds per image, compared to 18.8, 10.9, 6.1, 2.1, 8.7 ms, 9.2 ms, and 9.0 ms respectively. Under dim lighting conditions, the detection inference speed is 7.1 ms, outperforming the other four models by 3.4 ms, 3.6 ms, 7.2 ms, and 5.5 ms respectively, thereby meeting the real-time detection requirements for precision agriculture. By conducting precise monitoring at the early stage of fruit development (especially the very small fruit stage), and based on the observed different morphological traits, this approach predicts the fruit set rate, maturity, and quality of *Prunus humilis* fruits. It provides a scientific basis and intelligent decision-making support for formulating precise and differentiated field water and fertilizer management strategies.

### 1. Introduction

*Prunus humilis*, a species within the genus *Prunus* (Rosaceae family), is an endemic shrub fruit tree native to China [1,2]. Its primary distribution encompasses northern regions, with documented occurrences in Hebei, Shanxi, Liaoning, and Inner Mongolia [3–5]. This deciduous shrub typically reaches 0.5–2 m in height and exhibits pronounced drought tolerance and cold resistance [6–8]. The fruits of *Prunus humilis* contain abundant vitamin C and exhibit a pleasantly sweet-tart flavor profile. Recognized as nutrient-dense agricultural products, they can be

processed into various value-added commodities including jam, dried fruits, juice, and fruit wine [9,10]. Consequently, owing to their exceptional nutritional profile and remarkable environmental adaptability, this species has been extensively cultivated across northern China. Nevertheless, the harvesting of fresh *Prunus humilis* fruits faces significant challenges, including labor-intensive operations and suboptimal efficiency. To address these limitations, the transition toward intelligent and mechanized harvesting solutions is critically needed. A fundamental prerequisite for such technological advancement is the precise detection of *Prunus humilis* fruits in agricultural settings. The

\* Corresponding author at: Liaoning Technical University, School of Environmental Science and Engineering, Fuxin 123000, China.

E-mail address: [llyou515@126.com](mailto:llyou515@126.com) (L. Lv).

low-growing stature of *Prunus humilis* shrubs, coupled with dense flowering and fruiting patterns where clusters predominantly develop on lower and middle branches, creates high target density and multiplicity during detection. This frequently results in missed detections of individual fruits. Compounding this challenge, immature fruits exhibit small dimensions and visual similarity to foliage, while pervasive leaf occlusion generates optically complex backgrounds that significantly impede reliable detection. Consequently, addressing target marginalization and ensuring real-time performance represent critical priorities in current detection methodologies. Substantial research efforts have been dedicated globally to applying deep learning for robotic vision and image-based detection [11,12], with several studies specifically targeting small object occlusion and complex background challenges.

Current methodologies addressing these challenges can be broadly categorized into two paradigms. The first encompasses region proposal-based approaches, including R-CNN, Fast R-CNN, and Faster R-CNN [13]. These frameworks employ a two-stage detection architecture involving initial region proposal generation followed by object classification. While achieving commendable detection accuracy, such methods suffer from feature misalignment due to coordinate quantization artifacts during region cropping. Furthermore, the region proposal step incurs substantial computational overhead, resulting in prolonged inference times incompatible with real-time applications. Consequently, these techniques remain better suited for precision-oriented detection scenarios. Harshana et al. [14] implemented a Deep Convolutional Neural Network (DCNN) for maturity classification of strawberries, though occlusion handling was not addressed. Fu L et al. [15] developed a LeNet-based convolutional neural network for detecting occluded kiwifruit (*Actinidia deliciosa*) clusters, demonstrating efficacy across occlusion types (occluded, overlapping, adjacent, and separated fruits). However, the average inference time reached 0.27 s per image. Liu et al. [16] developed an enhanced Faster R-CNN framework for multi-cluster green persimmon (*Diospyros kaki*) detection, achieving improved model accuracy while effectively addressing target detection challenges in cluttered environments with blurred foreground objects. Mao et al. [17] developed a modified Faster R-CNN architecture for wheat disease and pest identification, achieving superior recognition accuracy for pathological manifestations. Kaur P et al. [18] proposed an improved Mask R-CNN model applied to autonomous segmentation and detection of leaf diseases in tomato plants. The enhanced version significantly reduces storage requirements and computational costs while maintaining diagnostic accuracy.

The second category encompasses region proposal-free methods, primarily represented by the YOLO series [19–21], SSD algorithm [22], and CenterNet [23,24]. These approaches fundamentally directly identify targets in entire images through a single convolutional neural network pass, eliminating the region proposal generation stage. This paradigm shift significantly reduces computational redundancy while accelerating inference speeds, making them particularly suitable for real-time applications [25,26]. Xiong J et al. [27] developed a lightweight YOLO v5-Lite-based framework for papaya ripeness assessment under in-field conditions, demonstrating robust capability in rapid and effective maturity identification of papaya fruits within complex backgrounds. Fu L et al. [28] addressed challenges such as occlusion in kiwifruit detection by developing an enhanced YOLOv3-micro-based algorithm, achieving rapid and accurate automated identification of kiwifruit in orchard environments. Concurrently, Xu W et al. [29] developed an Adaptive Lighting-Angle Robust YOLOv5s (ALAD-YOLO) model that achieves precise apple leaf disease detection under complex environmental conditions—including variable imaging angles, diurnal illumination changes, and diverse weather patterns—demonstrating a 7.9 % absolute improvement in detection accuracy over the baseline YOLOv5. Miao R et al. [30] integrated MobileNetV3 as a lightweight backbone network into the YOLOv7 architecture and incorporated a global attention mechanism within the feature fusion module to achieve precise maturity detection in cherry tomatoes (*Solanum lycopersicum*

var. cerasiforme). Chen R et al. [31] developed an enhanced YOLO—CE architecture to address challenges of suboptimal illumination and light attenuation in underwater environments, significantly improving target recognition accuracy. The proposed solution demonstrates particular efficacy in boosting precision for both small objects and general detection tasks. To address the critical challenges of low accuracy, excessive model parameters, and slow processing speed in UAV-based infrared target detection, Cao et al. [32] developed YOLO-TSL, an enhanced lightweight infrared target detection algorithm based on an improved YOLOv8n architecture.

In the field of *Prunus* fruit trees, the detection technology for the harvest maturity of their small drupes has attracted sustained attention from some scholars. Khojastehnazhand et al. [33] utilized image processing techniques to classify apricots based on visual appearance and estimated their volume. Zhou C et al. [34] proposed a one-dimensional convolutional neural network (1D-CNN) model for classifying the maturity of green Chinese plums, which improves the intelligent sorting efficiency of these fruits.

This study notes that the fruit morphology of *Prunus humilis* exhibits multi-scale distribution characteristics, with a very small number of fruits scattered and most concentrated. Due to dense flowers and fruits, severe occlusion by branches and leaves, and varying light conditions, partial target loss occurs frequently. However, existing algorithms fail to comprehensively address these issues. Therefore, this study proposes an improved YOLO-v8-based network for detecting dense multi-scale small *Prunus humilis* targets in complex scenes. This network can resolve small target loss during accurate fruit detection under complex backgrounds and enhance detection precision for adjacent mature fruits and occluded fruits. Finally, experiments were conducted to verify the improved DMP-YOLO model's effectiveness in accurately detecting *Prunus humilis* fruits under complex backgrounds and to determine the distribution regions of mature fruits. The specific contributions of this study are as follows: a. To address the problem of small *Prunus humilis* target loss in complex backgrounds, we designed the C2f\_MSCA (Multi-Scale Context Aggregation Block) module for extracting information from small *Prunus humilis* fruit targets. This module enhances the feature extraction capability for small *Prunus humilis* targets, effectively suppresses background noise interference, and improves small target extraction efficiency; b. To resolve target loss caused by the dense arrangement of *Prunus humilis* fruits, SPP and TransformerBlock modules are introduced into the last two layers of the backbone network to establish long-range dependencies, thereby improving target extraction efficiency; c. To tackle target loss issues arising from overlapping areas between branches, leaves, and fruits, we propose the "Multi-Dimensional Adaptive Convolution" C2f\_MDAC (Multi-Dimensional Adaptive Convolution). This module adaptively fuses target information across different scales to enhance the fusion efficiency of multi-scale *Prunus humilis* fruits; d. Images of *Prunus humilis* fruits with varying scales and maturity levels were collected at different time points, followed by data cleaning and annotation. Additionally, extensive experiments were conducted using the model.

## 2. Materials and methods

### 2.1. Dataset establishment

Liaoning is a major cultivation region for *Prunus humilis*. The *Prunus humilis* images used in the experiment were mainly sourced from the Zhanggutai *Prunus humilis* Experimental Base in Zhangwu County, Fuxin City, Liaoning Province and collected between April and August 2025, with an image resolution of 3000 × 4000 pixels. Data are collected using mobile phone equipment and drone equipment. Image collection was conducted on average from 9:00 AM to 4:00 PM daily. Due to variable lighting conditions in the natural environment and complex external factors such as shading nets, the collection process included real *Prunus humilis* images with occlusions and complex

backgrounds. Additionally, different shooting angles were employed to enhance the diversity of target data [35,36]. Fig. 1 displays the collected real *Prunus humilis* images at different growth stages, including flowering periods, fruits obscured by leaves, extremely small fruits, and mature fruits.

Annotation software was used to annotate flowering, fruiting, immature fruits, and mature fruits in the images, respectively. The original dataset was split into a data augmentation set and a test set at an 8:2 ratio. *Prunus humilis* fruits are round or oval and often severely occluded. Additionally, in the early fruiting stage, the fruits are extremely small and easily blend into the background. Given these characteristics, data augmentation was applied to the *Prunus humilis* images in the data augmentation set. The dataset utilized mosaic augmentation and automatic augmentation strategies, with further augmentation achieved by adjusting hue, saturation, and brightness. Some images were also segmented and cropped to enrich the image dataset. A total of 1550 targets were collected, showing a normal distribution in their overall distribution (As can be seen from Fig. 2). Among these, medium immature fruits accounted for the largest proportion, while mature and young fruits were relatively fewer than medium-sized ones. However, the overall distribution remained uniform, confirming the feasibility of the dataset developed in this study.

## 2.2. Detection methods for the maturity of *cerasus humilis* fruits

To address the challenges associated with *Prunus humilis* fruits at different growth stages—such as their extremely small size during the early fruiting period and severe occlusion against complex backgrounds—this study proposes improvements based on the YOLO-v8 model. Specifically, a C2f\_MSCA (Multi-Scale Context Aggregation Block) module is designed and integrated into the backbone network. Additionally, the original C2f modules in the last two layers of the backbone network are replaced with a combination of TransformerBlock and SPP modules, enabling these components to partition input feature maps into regions of varying scales through multi-scale pooling operations. Furthermore, a novel convolution method, termed "Multi-Dimensional Adaptive Convolution," is developed based on an adaptive attention mechanism. Fig. 3 illustrates the network architecture of the improved DMP-YOLO model.

### 2.2.1. Multi-channel attention mechanism: C2f\_MSCA

*Prunus humilis* exhibits dense, clustered flowers and fruits, with short-statured plants, resulting in most fruits being shaded by leaves; therefore, robust resistance to complex backgrounds is crucial. In this study, the MSCA is integrated into the bottleneck and positioned under C2f convolution, termed C2f\_MSCA. Fig. 4 illustrates the principle structure of C2f\_MSCA, while its structural diagram is presented in Fig. 5. The C2f\_MSCA module was specifically designed and incorporated into the deep layers (Layer 6 and Layer 8) of the backbone network. Through the Multi-Scale Context Aggregation Block, the importance of features in each channel is evaluated, and a set of channel weights is learned to dynamically adjust channels, thereby modifying feature representations. Attention weights are generated via nonlinear mapping, and by assigning adaptive weights to each channel, target-relevant channel features are emphasized while irrelevant and redundant background information is suppressed. To address the challenge of detecting small *Prunus humilis* targets in complex backgrounds, the MSCA-Block (Multi-Scale Context Aggregation Block) is introduced into C2f. This module achieves feature optimization through a triple attention collaborative mechanism. Firstly, parallel  $3 \times 3 / 5 \times 5 / 7 \times 7$  convolution kernels are employed to construct hierarchical receptive fields, enabling the network to capture local details of *Prunus humilis* fruits at different scales, including their edge texture information, as well as global contextual relationships involving branch and leaf occlusion. This effectively resolves the issue of multi-scale representation of targets in complex backgrounds. Secondly, a Channel Attention Module is utilized to quantitatively evaluate the importance of each feature channel and generate channel weight vectors. This allows the network to adaptively enhance the contribution of fruit-related feature channels while suppressing background interference channels, such as those corresponding to trunk or branch and leaf shadow regions, thereby focusing feature expression on key target areas. Finally, a spatial attention module is embedded to apply spatial domain weighting to multi-scale features, thereby enhancing the positional responses of *Prunus humilis* fruits—such as their central regions or edge areas—and reducing the activation intensity of background noise.

Therefore, this study introduces a multi-channel attention mechanism and embeds the MSCA-Block into the deep layers of the backbone network. This enables the model to enhance small target features and



(a) Photo of *Prunus humilis* in bloom



(b) Physical diagram of *Prunus humilis* under leaf occlusion and complex background conditions

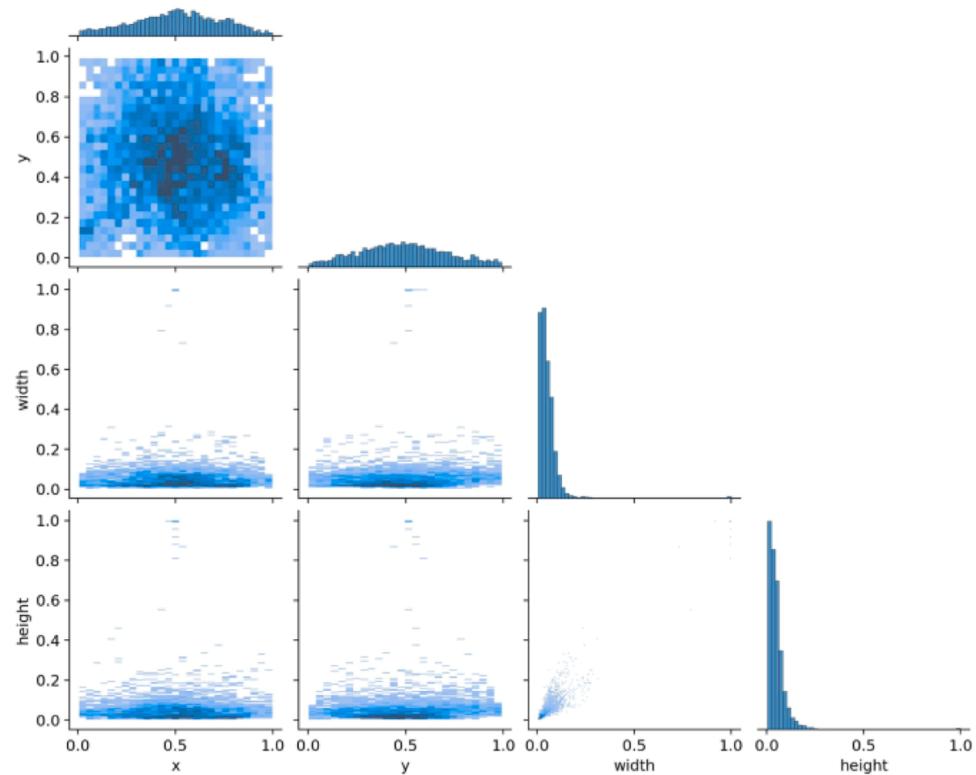


(c) Picture of the Actual Small Target Fruit

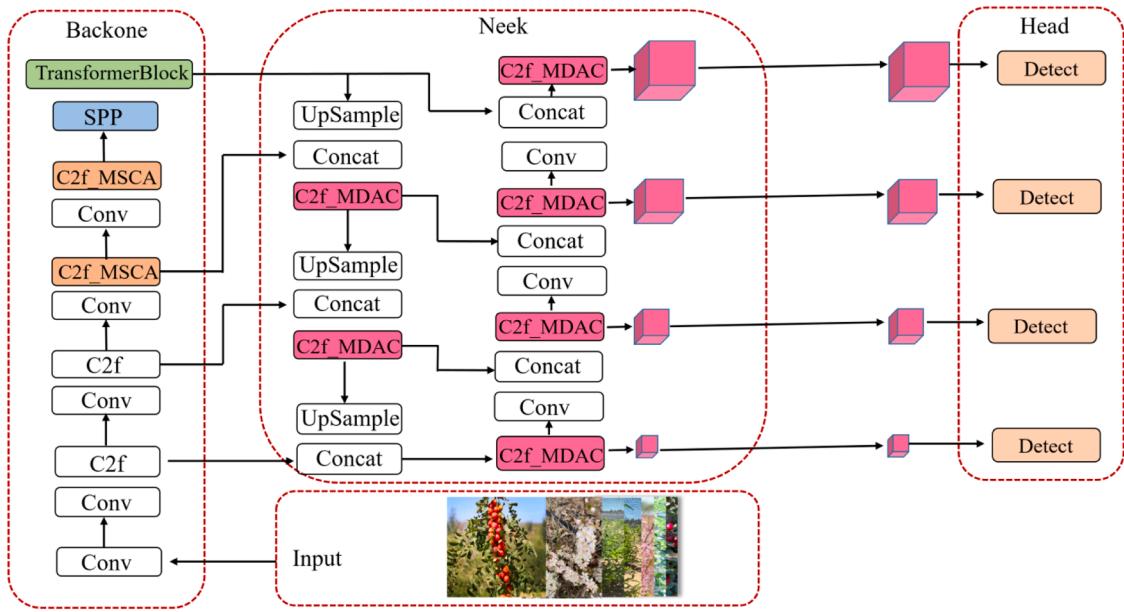


(d) Physical diagram of mature fruits obstructed by leaves

Fig. 1. Physical pictures of *Prunus humilis* at different periods.



**Fig. 2.** Distribution Map of the Dataset.



**Fig. 3.** DMP-YOLO Network Structure.

preserve the fine-grained features of *Prunus humilis* fruits within high-level semantic features. Through the attention mask, noise such as leaf occlusion and illumination variations is filtered out to suppress background interference. Additionally, this mechanism facilitates the detection head in more easily distinguishing dense fruits from similar features, thereby improving the discriminability of structural features.

#### 2.2.2. Introduce the TransformerBlock and SPP module

The SPP and TransformerBlock modules are introduced into the last

two layers of the backbone network to replace the original C2f module, allowing the TransformerBlock and SPP to partition the input feature map into regions of varying scales via multi-scale pooling operations. The SPP module constructs a feature pyramid using maximum pooling with four different kernel sizes (kernel\_size = 5, 9, 13, 1), enabling simultaneous capture of both microscopic details (such as young fruits with a diameter <15 pixels) and macroscopic contextual fruit cluster distribution patterns of *Prunus humilis* fruits within a single layer, thus resolving the resolution loss issue of small targets caused by

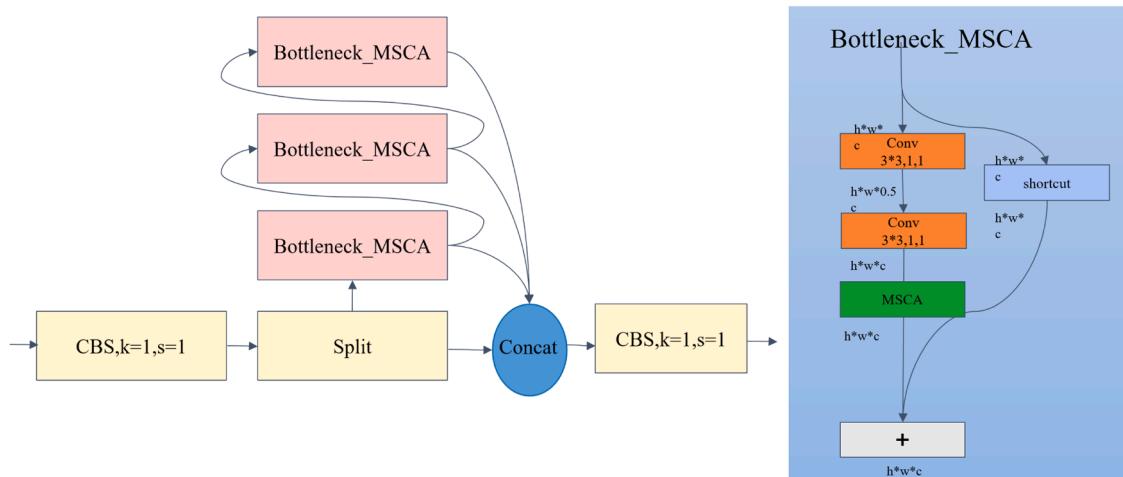


Fig. 4. Structure Diagram of C2f\_MSCA Module.

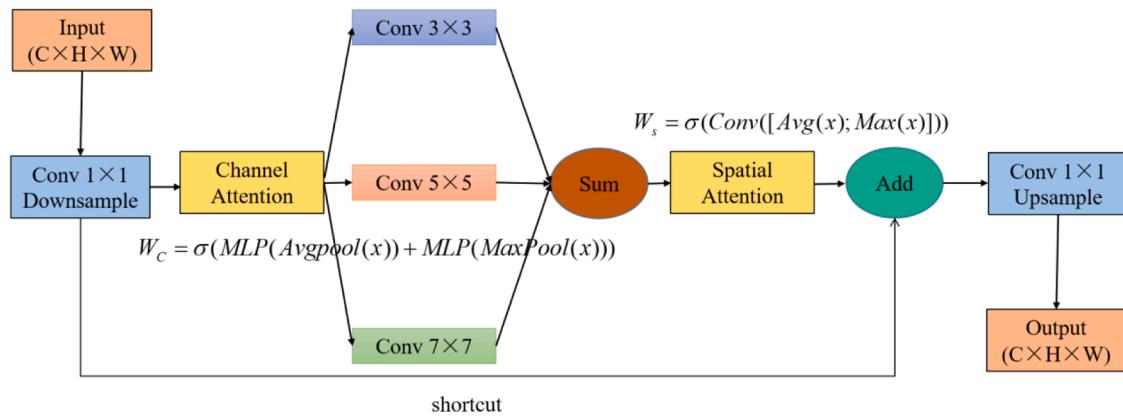


Fig. 5. Principle Structure of C2f\_MSCA.

downsampling. The extracted multi-scale detailed texture features are then fed into the TransformerBlock, which generates a weight matrix to perform weighted fusion of the input features. This process achieves long-range dependency modeling and efficient aggregation of contextual information, and capturing richer contextual information can enhance the expressive capability for multi-scale *Prunus humilis* target features. The two are combined in the final stage of feature extraction to achieve efficient fusion of multi-scale features and global information.

After establishing cross-layer feature associations with spatial attention and performing collaborative optimization following multi-layer feature concatenation, the problems of distant small fruit target loss and loss caused by foreground fruit-background fusion are solved. By fusing features with different receptive fields, the model can simultaneously recognize foreground single fruits and distant fruit clusters, focus on the spatial distribution hotspots of fruits, and enhance the model's generalization ability in complex backgrounds.

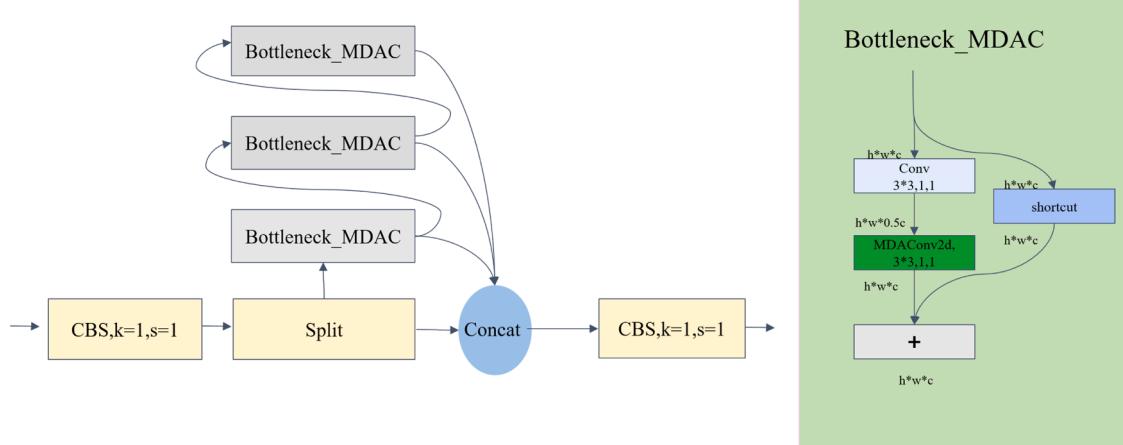


Fig. 6. C2f\_MDAC Principle Structure Diagram.

### 2.2.3. Constructing "multi-dimensional adaptive convolution" C2f\_MDAC

Based on the adaptive attention mechanism, this study proposes a novel convolution method termed "Multi-Dimensional Adaptive Convolution" (MDAC). MDAC replaces the C2f module and is integrated into the Bottleneck structure, collectively named C2f\_MDAC as illustrated in Fig. 6, while Fig. 7 presents the structural diagram of the C2f\_MDAC module. By automatically focusing on critical damaged regions and adjusting the feature extraction capability of convolution operations, these modules enable the model to more accurately identify and locate fruit regions across different scales. MDAC enhances the multi-scale feature fusion capability of the Neck component through four-dimensional dynamic attention mechanisms: channel dimension, spatial dimension, convolution kernel dimension, and filter dimension. Under conditions of severe background interference, the collaborative optimization of feature extraction via multi-dimensional attention strengthens the perception of target features across multiple dimensions. This facilitates focusing on the central position of the *Prunus humilis* target, suppresses the activation of leaf-occluded areas and overlapping fruits, and reduces the weight of background-dominated channels.

The C2f\_MDAC module constructs an efficient bottleneck structure using Multi-Dimensional Adaptive Convolution. When applied to feature fusion networks, C2f\_MDAC can better extract information from feature maps of different scales, continuously adjust convolution morphology to adapt to multi-morphology target detection, thereby enhancing multi-scale features, especially those of small targets.

MDAC enhances the multi-scale feature fusion capability of the Neck component through a dynamic attention mechanism. Firstly, through the channel dimension, it dynamically strengthens feature channels related to *Prunus humilis* fruits, enhancing the response to red or yellow pigments in the fruits; secondly, through the spatial dimension, it focuses on the spatial positions of fruits and strengthens edge responses at the junctions of fruits and branches/leaves; thirdly, through the convolution kernel dimension, it adaptively selects optimal convolution kernel sizes to match fruits of different sizes; finally, through the filter dimension, it suppresses the activation of background filters, filters high-frequency noise, and mitigates pixel flicker caused by light reflection, etc., thereby effectively suppressing complex background features such as soil, leaves, and trunk textures. This mechanism enhances the dynamic temperature coefficient and adjusts the sharpness of the attention distribution via a learnable temperature parameter. The Dynamic Temperature Coefficient serves as an "attention regulator" in the C2f\_MDAC module, whose core function is to enable the model to adapt to the complexity of different input features by dynamically adjusting the characteristics of the attention weight distribution.

In the attention mechanism, the temperature coefficient ( $\tau$ ) controls the distribution characteristics of the softmax output:

$$\text{Attention} = \text{Softmax}\left(\frac{QK^T}{\tau}\right)V \quad (1)$$

When  $\tau > 1$ , the attention distribution is smoothed, enabling the exploration mode to focus on more positions with relatively balanced weights across all locations. This is suitable for complex backgrounds or scenarios with multiple coexisting targets, preventing over-focus on a single region; when  $\tau = 1$ , it functions as the standard softmax; when  $0 < \tau < 1$ , the attention distribution becomes sharpened, allowing the focusing mode to concentrate on key positions and strengthen their weights. This is applicable when target features are prominent, such as when fruits have high contrast against the background, to enhance discriminability. Therefore, when incorporating the dynamic temperature coefficient into the C2f\_MDAC module, the input captures channel-level statistical features through global average pooling. The output is then constrained to the range [0,1] by Sigmoid  $\tau \in [0.5, 1.5]$  and mapped to dynamically adjust the range. Therefore, when feature complexity is high—such as in multi-object and multi-scale scenarios—increasing  $\tau$  can help disperse attention; conversely, when features are simple—such as uniform backgrounds or single-color conditions—decreasing  $\tau$  can help concentrate attention. The C2f\_MDAC module resolves the conflicting requirements of multi-scale features, as convolution branches of different scales ( $3 \times 3 / 5 \times 5 / 7 \times 7$ ) have distinct  $\tau$  needs. A large convolution kernel ( $7 \times 7$ ) requires a larger  $\tau$  to cover broader contextual information, while a small convolution kernel ( $3 \times 3$ ) necessitates a smaller  $\tau$  to focus on local details. Through branch-independent  $\tau$  adjustments, the dynamic temperature coefficient enables the large-kernel branch to retain background context and the small-kernel branch to sharpen target edges. Additionally, it can suppress complex background interference. In *Prunus humilis* small target detection, dynamic  $\tau$  automatically adjusts to enhance robustness. In occlusion scenarios—where leaves and branches occlude parts of the fruits— $\tau$  increases automatically to distribute attention across multiple candidate regions, thereby improving recall rates.

The independent temperature coefficient is calculated for each scale branch:

$$\tau_i = \text{DynamicTemp}^{F_{mid}^{(i)}} \quad (2)$$

Dynamic Scaling When Calculating Attention:

$$A_i = \text{softmax}\left(\frac{QK^T}{\tau_i}\right)V \quad (3)$$

Weighted Feature Fusion of Temperature Coefficients:

$$X_{out} = \sum_{i=1}^n \text{Conv}(A_i) \quad (4)$$

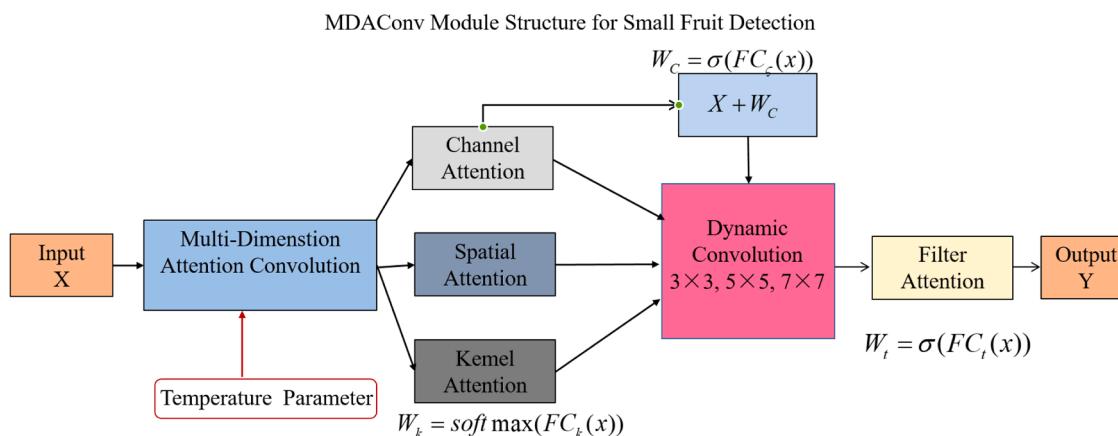


Fig. 7. Structure Diagram of C2f\_MDAC Model.

Therefore, when MDAC is embedded into the bottleneck of C2f and integrated into the Neek fusion network, for small fruit targets with a diameter <20 pixels,  $5 \times 5/7 \times 7$  large kernels are employed to capture both the fruit context information and the symbiotic relationship between fruits and adjacent foliage. For dense targets, spatial attention suppresses false activations in leaf-occluded regions, thereby enabling the detection of overlapping fruits. Finally, cross-layer feature correlation via spatial attention facilitates collaborative optimization after multi-layer feature concatenation, addressing target loss issues caused by detail degradation from downsampling for small targets (with diameters under 10 pixels) and color similarity between close-range fruits and their background. For instance, when mature red fruits are intermingled with withered leaves, this mechanism prevents target loss during precise detection.

### 3. Results and analysis

#### 3.1. Evaluation index

To evaluate the performance of the DMP-YOLO model in detecting dense and multi-scale *Prunus humilis* small targets, precision (P), recall (R), average precision (AP), and mean average precision (AP 0.5:0.95) were adopted as evaluation metrics for the validation dataset in this study. These metrics are calculated based on true positives (TP), false positives (FP), and false negatives (FN) [37,38]. TP refers to the number of targets where real fruits are correctly detected during the detection process, meaning both the prediction and the actual label are positive; FP refers to the number of targets that are actually non-fruits but mistakenly detected as fruits by the model, such as negative samples or interference from branches, leaves, and background that are predicted as positive; FN refers to the number of targets where fruits actually exist but are not detected by the model, i.e., positive samples that are predicted as negative.

Precision refers to the proportion of samples that are actually positive among the samples predicted as positive by the model, which reflects the degree to which the background or non-target is misjudged as *Prunus humilis* fruits. The calculation formula is as follows:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

Recall rate refers to the proportion of true positive samples that are correctly predicted as positive by the model, which reflects the degree to which the model misses detecting true *Prunus humilis* fruits. The calculation formula is as follows:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

Average Precision (AP) is the area under the precision-recall curve, with precision plotted on the y-axis and recall on the x-axis. A larger area under this curve indicates higher detection accuracy. AP(0.5) refers to the average precision at an IoU threshold of 0.5, while AP(0.5:0.95) represents the mean of average precisions across IoU thresholds ranging from 0.5 to 0.95. Their calculation formulas are shown in Eqs. (7) and (8):

$$\text{AP} = \int_0^1 p(r)dr, \quad (7)$$

$$\text{AP}(0.5 : 0.95) = \frac{\sum_{i=0.5}^{0.95} \text{AP}(i)}{10} \quad (8)$$

In the experiment, YOLO-v8 was employed as the baseline model. The model network was initially trained on a dataset in .txt format, then automatically optimized using either the SGD or Adam optimizer with a batch size of 8 for a total of 300 epochs. The backbone network was configured with an initial learning rate of 0.01, and the SGD momentum

parameter was set to 0.937. All experiments were performed on an NVIDIA GeForce RTX 3060 Ti GPU.

#### 3.2. Introduction of the experiment

To verify the performance of the improved model DMP-YOLO, the *Prunus humilis* images in the test set were tested and evaluated. Table 1 presents the detection results of the proposed algorithm across different maturity levels. The model achieves an overall average detection precision of approximately 71.60 %. Since this study primarily focuses on *Prunus humilis* fruit detection, limited sample data was available for the flowering stage, leading to relatively coarse detection results. For the 513 instances of extremely small fruits, the precision stands at 85.6 %, recall at 59.8 %, and average precision at 73.2 %. For immature medium-sized green fruits, the proposed DMP-YOLO model achieves a detection precision of 87 %, a recall of 85.4 %, and an average precision as high as 90.6 %, indicating highly accurate detection of such fruits. In the detection of mature fruits, the precision reaches 92.70 %, recall is 81.20 %, and average detection precision is 92.90 %, similarly demonstrating excellent and accurate detection performance. The inference time per image is 0.9 ms, indicating a fast detection speed that meets the real-time requirements of agricultural detection. However, during the detection process, mAP50 and mAP50-95 show interpolation, which is caused by inaccurate bounding box localization or insufficient tightness of some detection boxes.

#### 3.3. Ablation experiment

To verify the effectiveness of the model improvements, ablation experiments were conducted using the test set, with the specific results presented in Table 2 below. This study implemented progressive modifications to the YOLO-v8 model to validate the efficacy of each improvement step. When the C2f\_MSCA module was designed and integrated into the base model, precision increased by 0.5 percentage points. This improvement is primarily attributed to the module's multi-channel attention mechanism, which enhances small target features and preserves fine-grained *Prunus humilis* fruit characteristics within high-level semantic features, thereby effectively boosting precision. Following the introduction of the TransformerBlock and SPP module, the model's precision improved by 2.1 percentage points and recall by approximately 2 percentage points. This module addresses target loss issues caused by downsampling of small fruits in dense clusters by establishing macroscopic connections with global fruit cluster information. Additionally, it enhances recall by creating long-range dependencies that provide contextual references for target detection. The integration of the multi-dimensional adaptive convolution C2f\_MDAC module resulted in a 5-percentage-point increase in recall and an approximate 1.7-percentage-point improvement in average precision. Added to the feature fusion stage, this module perceives target features through multiple dimensions, focuses on the central position of *Prunus humilis* targets, and adaptively adjusts the convolution module structure to match fruits of varying sizes, thereby significantly enhancing both recall and precision.

When both the C2f\_MSCA and TransformerBlock+SPP modules were embedded into the base model, precision increased by 4.6 percentage points and recall by 5 percentage points. This combined enhancement

**Table 1**  
Network detection results based on the DMP-YOLO Model.

| Class  | Images | Instances | Box:p | R     | mAP50 | mAP50-95 |
|--------|--------|-----------|-------|-------|-------|----------|
| All    | 45     | 1228      | 0.773 | 0.686 | 0.735 | 0.505    |
| Hua    | 5      | 10        | 0.441 | 0.481 | 0.372 | 0.289    |
| Guo    | 27     | 513       | 0.856 | 0.598 | 0.732 | 0.512    |
| Unripe | 9      | 486       | 0.87  | 0.854 | 0.906 | 0.588    |
| ripe   | 10     | 219       | 0.927 | 0.812 | 0.929 | 0.629    |

**Table 2**

Comparison table of ablation experiment models.

| Model    | P                    | R        | mAP50(%) | mAP50–95 | Parameter | GFLOPS |
|----------|----------------------|----------|----------|----------|-----------|--------|
| C2f_MSCA | TransformerBlock+SPP | C2f_MDAC |          |          |           |        |
| —        | —                    | —        | 0.924    | 0.72     | 59.58     | 32.69  |
| ✓        | —                    | —        | 0.929    | 0.70     | 54.74     | 26.04  |
| —        | ✓                    | —        | 0.945    | 0.74     | 57.27     | 26.97  |
| —        | —                    | ✓        | 0.891    | 0.77     | 61.20     | 32.54  |
| ✓        | ✓                    | —        | 0.970    | 0.77     | 60.82     | 30.33  |
| —        | ✓                    | ✓        | 0.937    | 0.76     | 60.12     | 31.16  |
| ✓        | ✓                    | ✓        | 0.953    | 0.80     | 73.50     | 46.60  |
|          |                      |          |          |          |           | 3.88M  |
|          |                      |          |          |          |           | 11.70  |

addresses small target loss in complex backgrounds and target occlusion in dense fruit arrangements through backbone network improvements and long-range dependency modeling, leading to substantial gains in detection precision and recall. Finally, simultaneous integration of the C2f\_MDAC module and TransformerBlock+SPP into the base model yielded a 1.3-percentage-point improvement in precision.

Therefore, the above experiments demonstrate that the C2f\_MSCA and TransformerBlock+SPP modules significantly improve the model's precision and recall, while the C2f\_MDAC module better enhances the overall average precision metrics mAP50 and mAP50–95 for model detection. Consequently, when all three modules are simultaneously integrated into the DMP-YOLO model, all evaluation metrics are significantly improved. Specifically, the accuracy of the DMP-YOLO model increases by 2.9 percentage points compared to the baseline model, with an 8-percentage-point improvement in recall. Notably, mAP50 achieves a breakthrough increase of approximately 13.92 percentage points, and mAP50–95 rises by approximately 13.91 percentage points. The final parameter of the DMP-YOLO model is 8.88 M, and the GFLOPS is 11.70.

#### 3.4. Comparative experiment

##### 3.4.1. Experiment under normal light conditions

To evaluate the detection capability of various models for dense multi-scale *Prunus humilis* fruits, this study selected mainstream YOLO detection models including YOLOv3, YOLOv5, YOLOv6, YOLOv8, YOLOv10, YOLOv11, and YOLOv12 for comparative experiments [37], using the same evaluation metrics for comparison. The detailed results are shown in Table 3. The DMP-YOLO model achieved the best performance across all metrics, with an mAP50 of 73.5 %, representing improvements of 4.7, 23, 19.3, 14.3, 17.8, 16.5, and 17.5 percentage points compared to YOLOv3, YOLOv5 [38], YOLOv6, YOLOv8 [39,40], YOLOv10, YOLOv11, and YOLOv12, respectively. Furthermore, the DMP-YOLO model achieved a detection accuracy of 73.2 % for small immature fruits, outperforming the other models by 3, 36.4, 30.3, 21.4, 32.9, 43.5, and 37.8 percentage points, respectively, demonstrating its strong performance in detecting dense and extremely small immature fruits. For medium-sized green immature fruits and mature fruits, the detection accuracies reached 90.6 % and 92.9 %, respectively. The DMP-YOLO model also achieved higher recall rates compared to other models in detecting flowering, small immature fruits, medium immature fruits, and mature fruits, with values of 48.1 %, 59.8 %, 85.4 %, and 81.2

%, respectively. Additionally, the model has 3.88 M parameters and 11.70 GFLOPS, making it more lightweight compared to YOLOv3, which also delivered relatively good detection results. Fig. 8 illustrates the distribution of metrics across different models, highlighting the superior performance of the DMP-YOLO model in all aspects.

The inference speed of the DMP-YOLO model is 0.9 ms per image in Fig. 9, compared to 18.8 ms, 10.9 ms, 6.1 ms, 2.1 ms, 8.7 ms, 9.2 ms, and 9.0 ms for the other seven models, respectively. This breakthrough in inference time enables ultra-low latency, providing a solid technical foundation for real-time field detection in precision agriculture and significantly enhancing the timeliness and responsiveness of agricultural intelligent operations.

Fig. 10 presents the visualization results of dataset detection, where (a), (b), (c), and (d) represent the original data image, label image, YOLO-v8-based detection results, and DMP-YOLO model-based detection results, respectively. By comparing the second and fourth images of each category, it is evident that the YOLOv8 detection results exhibit issues of missed small *Prunus humilis* fruit targets. In the fifth image of Fig. 10(c) and (d), it is clearly observed that the baseline model misclassifies flowers as fruits, whereas the DMP-YOLO detection results demonstrate high accuracy. The DMP-YOLO model effectively addresses this problem, primarily due to its multi-module integration mechanism, which establishes long-range dependencies and significantly enhances the feature extraction capability for small *Prunus humilis* fruit targets. Additionally, through multi-scale perception, it adaptively strengthens the contribution of fruit-related feature channels, thereby resolving the issue of missed detection of small *Prunus humilis* fruit targets. This approach notably improves the model's detection precision, recall, and average precision metrics.

Therefore, based on the performance comparison of different detection models, the DMP-YOLO model exhibits higher precision, a higher recall rate, and faster inference detection speed. Thus, it is demonstrated that DMP-YOLO possesses a remarkable advantage in the detection of dense multi-scale *Prunus humilis* small target fruits.

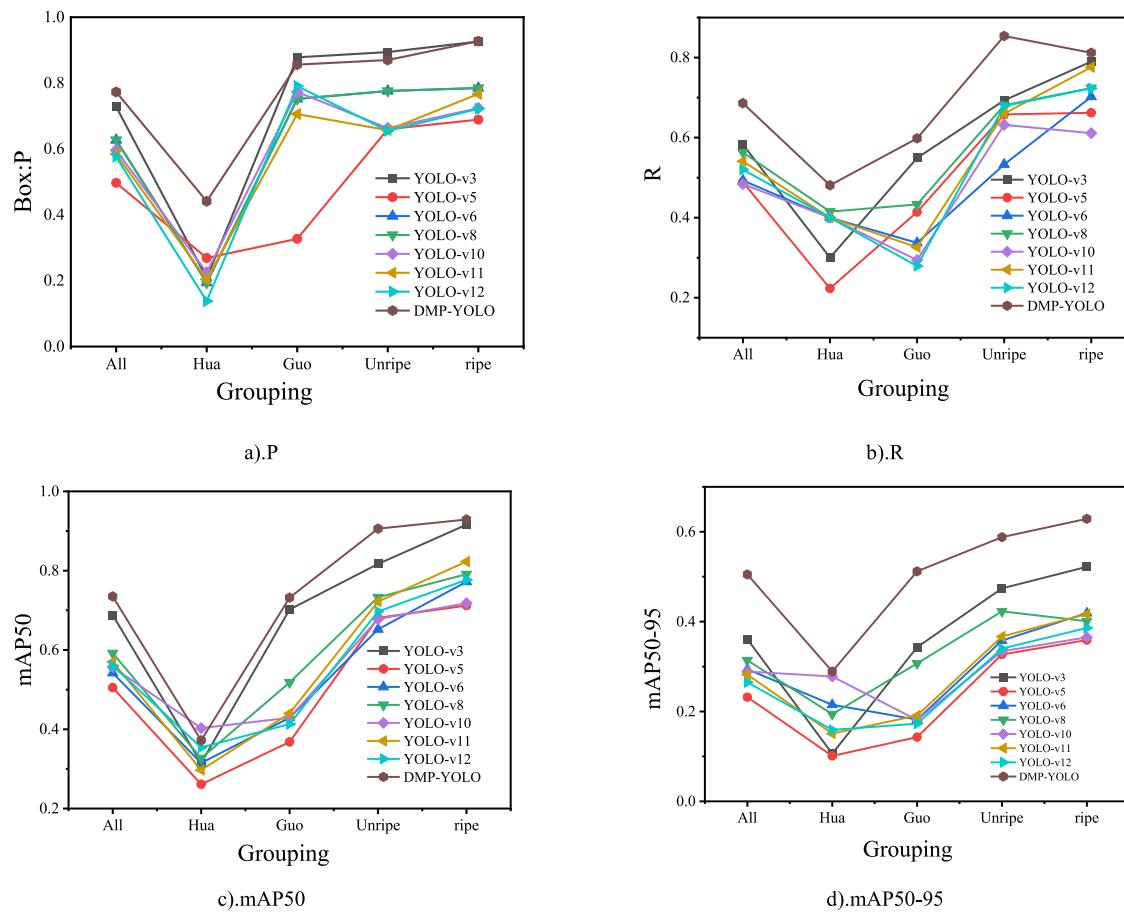
##### 3.4.2. Experiment under dim lighting conditions

The original dataset underwent simulated low-light processing to generate an adversarial low-light scene dataset. On this dataset, we conducted multi-model comparisons of multi-scale *Prunus humilis* small object detection, simultaneously validating that the improved DMP-YOLO model demonstrates greater generalisability in detection tasks. The detection results based on the DMP-YOLO model are presented in

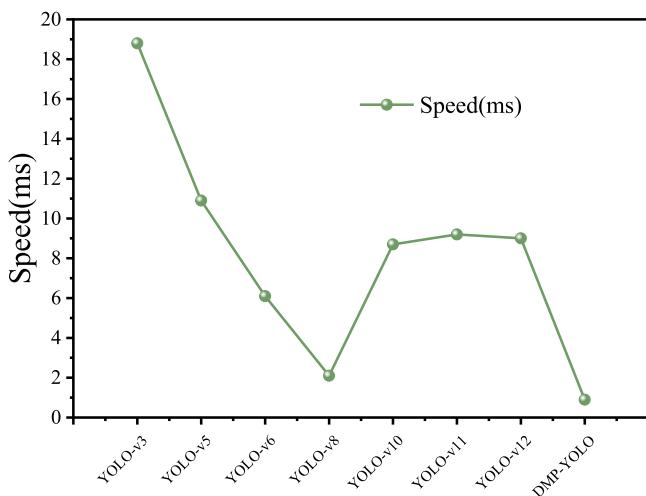
**Table 3**

Comparison test results of different models.

| Model    | Images | Instances | Box:p | R     | mAP50 | mAP50–95 | Speed  | Parameter | GFLOPS |
|----------|--------|-----------|-------|-------|-------|----------|--------|-----------|--------|
| YOLO-v3  | 45     | 1228      | 0.728 | 0.583 | 0.688 | 0.361    | 18.8ms | 103.69M   | 283.00 |
| YOLO-v5  | 45     | 1228      | 0.497 | 0.489 | 0.505 | 0.232    | 10.9ms | 2.51M     | 7.20   |
| YOLO-v6  | 45     | 1228      | 0.627 | 0.493 | 0.542 | 0.294    | 6.1ms  | 4.24M     | 11.90  |
| YOLO-v8  | 45     | 1228      | 0.627 | 0.563 | 0.592 | 0.314    | 2.1ms  | 3.01M     | 8.20   |
| YOLO-v10 | 45     | 1228      | 0.596 | 0.484 | 0.557 | 0.289    | 8.7ms  | 2.71M     | 8.40   |
| YOLO-v11 | 45     | 1228      | 0.584 | 0.541 | 0.57  | 0.292    | 9.2ms  | 2.59M     | 6.40   |
| YOLO-v12 | 45     | 1228      | 0.576 | 0.52  | 0.56  | 0.265    | 9.0ms  | 2.57M     | 6.50   |
| DMP-YOLO | 45     | 1228      | 0.773 | 0.686 | 0.735 | 0.505    | 0.9ms  | 3.88M     | 11.70  |



**Fig. 8.** Comparison chart of multi-model evaluation indicators.



**Fig. 9.** Reasoning speed of different models.

**Table 4.** The model achieves an overall detection accuracy of 69.8 % with a recall rate of 61.7 %. Under dim lighting conditions, the model achieved 70 % detection accuracy for extremely small fruits. For immature and mature fruits, detection accuracies were 87.5 % and 89.5 % respectively. The model's high detection accuracy across different categories demonstrates the robust performance of the DMP-YOLO model even in low-light conditions, while also proving its strong generalisation capability.

In Fig. 11, (a) shows the labelled images from the dataset, (b)

presents the detection results from the YOLO-v8 model, and (c) displays the detection results from the DMP-YOLO model. Through visual comparison, the YOLO-v8 model exhibits instances of missed detections, whereas the DMP-YOLO model performs well, significantly reducing the occurrence of missed detections for small target fruits[41,42].

Validation experiments for detecting dense, heavily occluded *Prunus humilis* targets in complex backgrounds under dim lighting conditions were conducted. The DMP-YOLO model was compared with YOLO-v8, YOLO-v10, YOLO-v11, and YOLO-v12. The evaluation metrics employed were identical to those used for normal illumination conditions. As detailed in Table 5, the DMP-YOLO model demonstrated superior performance across all metrics, achieving an mAP50 of 69.8 %. Compared to YOLO-v8, YOLO-v10, YOLO-v11, and YOLO-v12 by 12.9, 12.9, 16.9, and 16.5 percentage points respectively. The DMP-YOLO model achieved a recall rate of 61.7 %, surpassing other models by 12.4, 16.0, 13.7, and 16.6 percentage points respectively. The DMP-YOLO model achieved an mAP50-95 of 43.7 %, surpassing other models by 13.5, 16.1, 16.4, and 16.9 percentage points respectively. Detection inference speed was the fastest at 7.1 ms, outperforming others by 3.4 ms, 3.6 ms, 7.2 ms, and 5.5 ms respectively. Fig. 12 illustrates the distribution of metrics across different models in low-light conditions. Fig. 13 illustrates the detection speeds of various models for *Prunus humilis* under dim lighting conditions. Comparative analysis reveals that the DMP-YOLO model retains a significant advantage. While its detection speed is slower than under normal illumination, it substantially reduces detection time compared to other detection models while maintaining high detection accuracy. This effectively meets the real-time requirements of precision agriculture. The comparative experiments in low-light conditions further demonstrate the enhanced robustness of the modified DMP-YOLO model, with all evaluation



**Fig. 10.** visualization of data results.

**Table 4**  
Detection results of the DMP-YOLO model in low-light conditions.

| Class  | Images | Instances | Box:p | R     | mAP50 | mAP50-95 |
|--------|--------|-----------|-------|-------|-------|----------|
| All    | 45     | 1228      | 0.744 | 0.617 | 0.698 | 0.437    |
| Hua    | 5      | 10        | 0.328 | 0.40  | 0.322 | 0.244    |
| Guo    | 27     | 513       | 0.878 | 0.481 | 0.70  | 0.455    |
| Unripe | 9      | 486       | 0.824 | 0.826 | 0.875 | 0.508    |
| ripe   | 10     | 219       | 0.892 | 0.762 | 0.895 | 0.542    |

metrics outperforming other mainstream detection models.

#### 4. Conclusion

This paper proposes an improved DMP-YOLO model based on YOLO-

v8 to detect the loss of small *Prunus humilis* fruit targets and address the issue of low detection accuracy. A multi-channel attention mechanism, the C2f\_MSCA module, is designed and integrated into the deep layers of the backbone network to mitigate interference from complex background noise and enhance the feature extraction capability for small *Prunus humilis* targets. Subsequently, the original C2f modules in the last two layers of the backbone network are replaced with TransformerBlock and SPP modules. This modification focuses on the spatial distribution hotspots of fruits, establishes long-range dependencies, and improves the model's generalization ability in complex backgrounds. Furthermore, based on an adaptive attention mechanism, a novel "multi-dimensional adaptive convolution" module, C2f\_MDAC, is developed. This module automatically adjusts convolution kernel sizes to accommodate fruits of varying scales, enabling the model to more accurately



**Fig. 11.** Visualization of data results.

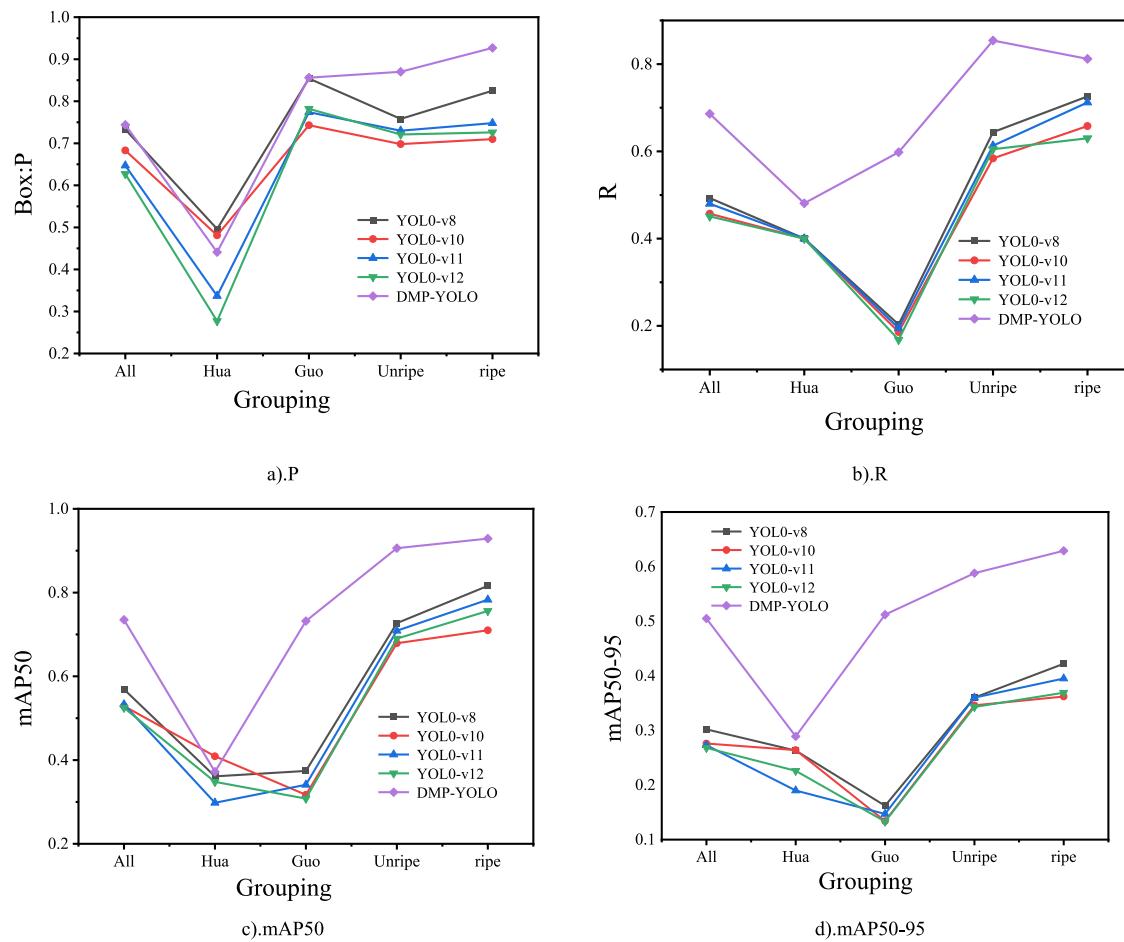
**Table 5**  
Detection results of DMP-YOLO model in dim environment.

| Model    | Images | Instances | Box:p | R     | mAP50 | mAP50–95 | Speed  | Parameter | GFLOPS |
|----------|--------|-----------|-------|-------|-------|----------|--------|-----------|--------|
| YOLO-v8  | 45     | 1228      | 0.733 | 0.493 | 0.569 | 0.302    | 10.5ms | 3.06M     | 8.20   |
| YOLO-v10 | 45     | 1228      | 0.683 | 0.457 | 0.529 | 0.276    | 10.7ms | 2.71M     | 8.40   |
| YOLO-v11 | 45     | 1228      | 0.647 | 0.48  | 0.533 | 0.273    | 14.3ms | 2.59M     | 6.40   |
| YOLO-v12 | 45     | 1228      | 0.627 | 0.451 | 0.525 | 0.268    | 12.6ms | 2.57M     | 6.50   |
| DMP-YOLO | 45     | 1228      | 0.744 | 0.617 | 0.698 | 0.437    | 7.1ms  | 3.88M     | 11.60  |

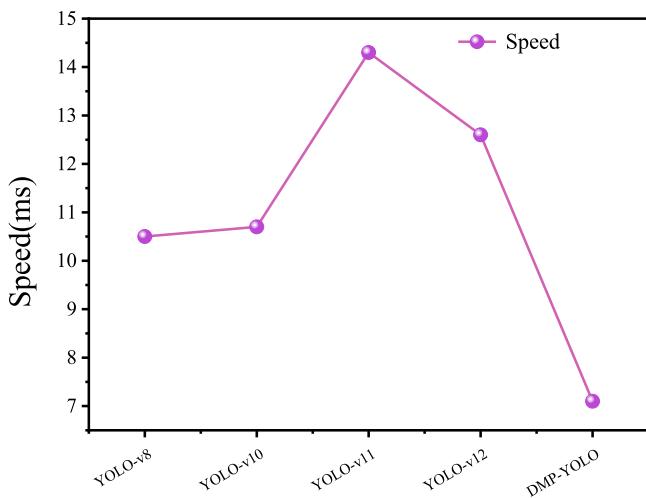
identify and localize fruit regions across different scales.

To verify the detection capability of the DMP-YOLO model, five groups of ablation experiments were conducted to evaluate the contribution of different modules to detection performance improvement. The experimental results indicate that embedding the MSCA into the C2f module of the backbone network enhances the model's detection accuracy, while incorporating the TransformerBlock and SPP modules improves both precision and recall. Replacing the C2f module with the MDAC module yields a more significant improvement in average detection precision. The optimal performance is achieved when all these modules are combined. Through a large number of experiments, it has been proven that the DMP-YOLO mock-up has boosts in all metrics.

Under healthy lighting conditions, compared to the baseline, the detect accuracy is boosted by 14.6 percentage points, and the recall is boosted by 12.3 percentage points; in dim environments, compared to the baseline, the detect accuracy is boosted by 12.9 percentage points, and the recall is boosted by 12.4 percentage points. These experiments fully confirm that the proposed improved DMP-YOLO model achieves significant enhancements in all evaluation indicators. This study provides a scientific basis and intelligent decision support for formulating precise and differentiated field water and fertilizer management strategies, based on distinct morphological traits identified through accurate detection during the early stages of fruit development (particularly the minimal fruit stage). Additionally, it offers theoretical support and



**Fig. 12.** Comparison chart of multi-model evaluation metrics in dim environment.



**Fig. 13.** Reasoning speed of different models in dim environments.

technical guidance for the intelligent production of *Prunus humilis* fruits.

The primary limitation of this study is the proposed model's high parameter count and computational complexity (high GFLOPS), potentially posing deployment challenges on resource-constrained hardware platforms.

## 5. Prospect

This study primarily focuses on the accurate detection of small multi-scale *Prunus humilis* fruit targets in complex backgrounds, providing important support for intelligent fruit detection. However, to meet the demands of precision agriculture in the future, further research is needed to explore the detection of *Prunus humilis* fruits under extreme weather conditions (e.g., cloudy, rainy, and foggy weather). This includes addressing scientific challenges such as improving robustness in extreme lighting environments and ensuring detection adaptability across varying canopy coverage.

## Ethical statement

### Ethical Approval

The main research object of this article is *Prunus humilis* plants, and deep learning is used to detect fruit maturity. No ethical approval was required as it did not involve the collection or analysis of data involving human or animal subjects.

## CRediT authorship contribution statement

**Linyou Lv:** Project administration, Funding acquisition. **Jiahui Li:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yan Zhao:** Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The authors gratefully acknowledge the financial support from the National Key Research and Development Program of China (Grant No 2024YFD1501405) and the Science and Technology Program of Liaoning Province (Grant No 2023JH1/10400001).

## Data availability

Data will be made available on request.

## References

- [1] L. Wu, Z. Li, J. Chai, et al., Application value and cultivation techniques of *Prunus humilis*, *Mod. Agric. Sci. Technol.* (20) (2022) 59–62.
- [2] Q. Hua, T. Wei, W. Han, et al., Research progress on the components, volatile flavors and biological activities of *Prunus humilis*, *Food Mach.* 40 (12) (2024) 186–192.
- [3] L. Fan, Y. Zhang, X. Li, et al., Study On Drought Resistance of Wild *Prunus Humilis* with Different Leaf Shapes under the Same Site Conditions, *Chinese Fruits*, 2024, pp. 55–59.
- [4] Y. Yang, J. Wei, J. Guo, et al., Breeding of a new variety of *Prunus humilis*, *Luo'ou No.1. Chin. Fruits* (08) (2024) 152–153.
- [5] Y. Liu, L. Zhang, H. Ye, et al., Genetic variation analysis of fruit traits in the hybrid F1 generation of *Prunus humilis* 'Nongda No.7' and 'DS-1', *J. Shanxi Agric. Univ. (Nat. Sci. Ed.)* 45 (01) (2025) 41–51.
- [6] H. Zhang, L. Lv, Q. Zhao, et al., Cultivation and management techniques for Nongda No.7 *Prunus humilis*, *J. Fruit Resour.* 5 (04) (2024) 73–74.
- [7] C. Wu, X. Han, J. Wu, et al., Research Progress on the Nutritional Value and Development and Utilization of *Prunus humilis*, *Chinese Fruits*, 2024, pp. 1–6.
- [8] Y. He, K. Lu, Y. Qin, et al., Adaptive evaluation of the introduction and cultivation of different *Prunus humilis* varieties, *Fruits South. China* 52 (04) (2023) 126–130.
- [9] L. Yu, L. Ren, W. Xin, et al., Cultivation techniques for *Prunus humilis* excellent varieties in Sandy Land, *Guangdong Seric.* 55 (05) (2021) 97–98.
- [10] L. Song, New variety of Chinese dwarf cherry-Yanshan No.2, *Inf. China's Fruit Ind.* 40 (06) (2023) 65.
- [11] Z. Zhao, Y. Hicks, X. Sun, et al., FruitQuery: a lightweight query-based instance segmentation model for in-field fruit ripeness determination, *Smart Agric. Technol.* 12 (2025) 101068.
- [12] Z. Liu, L. Zhuo, C. Dong, et al., TBD-Y: automatic tea bud detection with synergistic object-spatial attention and global-local attention guided feature fusion, *Smart Agric. Technol.* 12 (2025) 101066.
- [13] K. Zhang, Q. Wu, Q. Chen, Detecting soybean leaf disease from synthetic image using multi-feature fusion faster R-CNN, *Comput. Electron. Agric.* (2021) 183.
- [14] H. Harshana, O. Yuichi, S. Tetsuhito, et al., Detecting greenhouse strawberries (mature and immature), using deep convolutional neural network, *Eng. Agric. Environ. Food* 11 (3) (2018) 127–138.
- [15] L. Fu, K. Feng, Elkamil Tola, et al., Field multi-cluster Kiwifruit image recognition method based on convolutional neural network, *Trans. Chin. Soc. Agric. Eng.* 34 (02) (2018) 205–211.
- [16] Y. Liu, H. Ren, Z. Zhang, et al., Research on multi-cluster green persimmon detection method based on improved faster RCNN, *Front. Plant Sci.* 14 (2023) 1177114.
- [17] R. Mao, Y. Zhang, Z. Wang, et al., Identifying wheat stripe rust and yellow dwarf disease via improved faster-RCNN, *Trans. Chin. Soc. Agric. Eng.* 38 (17) (2022) 176–185.
- [18] P. Kaur, S. Harnal, V. Gautam, et al., An approach for characterization of infected area in tomato leaf disease based on deep learning and object detection technique, *Eng. Appl. Artif. Intell.* 115 (2022) 105210.
- [19] Z. Liu, H. Yao, Y. Lyu, et al., A deep learning method for log diameter measurement using wood images based on Yolov3 and DeepLabv3+, *Forests* 15 (5) (2024).
- [20] F. Wen, H. Wu, X. Zhang, et al., Accurate recognition and segmentation of Northern corn leaf blight in drone RGB images: a CycleGAN-augmented YOLOv5-mobile-seg lightweight network approach, *Comput. Electron. Agric.* 236 (2025) 110433.
- [21] Z. Qi, J. Wang, G. Yang, et al., Lightweight YOLOv8-based model for weed detection in Dryland spring wheat fields, *Sustainability* 17 (13) (2025), 6150-6150.
- [22] H. Li, E. Yan, J. Jiang, et al., Monitoring of key *Camellia Oleifera* phenology features using field cameras and Deep learning, *Comput. Electron. Agric.* 219 (2024) 108748.
- [23] A. Paul, R. Machavaram, et al., Smart Solutions for capsicum harvesting: unleashing the power of YOLO for detection, segmentation, in: *Growth Stage Classification, Counting, and Real-time Mobile Identification*. *Computers and Electronics in Agriculture*, 219, 2024 108832.
- [24] G. Zheng, Y. Jiang, J. Shen, Rice pest recognition method based on improved YOLOv7, *J. Huazhong Agric. Univ.* 42 (03) (2023) 143–151.
- [25] H. Li, J. Yang, J. Song, et al., Stem base detection in sugarcane plants using improved YOLOv5m model, *Sugar Tech.* (2025) 1–11 (prepublish).
- [26] X. Ji, J. Li, X. Cai, et al., Driving by a publicly available RGB image dataset for rice planthopper detection and counting by fusing swin transformer and YOLOv8-p2 architectures in field landscapes, *Agriculture* 15 (13) (2025), 1366–1366.
- [27] J. Xiong, Y. Han, X. Wang, et al., Maturity detection of Papaya in natural environment based on YOLO v5 - lite, *Trans. Chin. Soc. Agric. Mach.* 54 (06) (2023) 243–252.
- [28] L. Fu, Y. Feng, J. Wu, et al., Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model, *Precis. Agric.* (2020) 1–23 (prepublish).
- [29] W. Xu, R. Wang, ALAD-YOLO: an lightweight and accurate detector for Apple leaf diseases, *Front. Plant Sci.* 14 (2023) 1204569.
- [30] R. Miao, Z. Li, J. Wu, Lightweight cherry tomato ripeness detection method based on improved YOLO v7, *Trans. Chin. Soc. Agric. Mach.* 54 (10) (2023) 225–233.
- [31] R. Chen, H. Zhou, H. Xie, et al., YOLO-CE: an underwater low-visibility environment target detection algorithm based on YOLO11, *J. Supercomput.* 81 (5) (2025), 723–723.
- [32] L. Cao, Q. Wang, Y. Luo, et al., YOLO-TSL: a lightweight target detection algorithm for UAV infrared images based on triplet attention and slim-neck, *Infrared Phys. Technol.* 141 (2024) 105487.
- [33] M. Khojastehnazhand, V. Mohammadi, S. Minaei, Maturity detection and volume estimation of apricot using image processing technique, *Sci. Hortic.* 251 (2019) 247–251.
- [34] C. Zhou, Y. Liu, H. Xia, et al., A method for classifying the maturity of green Chinese plums based on hyperspectral data, *J. For. Eng.* 10 (03) (2025) 146–153.
- [35] L. Zhang, T. Yu, G. Zheng, et al., Using machine learning to predict selenium content in crops: implications for soil health and agricultural land utilization in longevity regions, *Sci. Total Environ.* (2025), 964178520–178520.
- [36] L. Chen, B. Hu, J. Sun, et al., Using remote sensing and machine learning to generate 100-cm soil moisture at 30-m resolution for the black soil region of China: implication for agricultural water management, *Agric. Water Manag.* 309 (2025) 109353.
- [37] Z. Yang, K. Hu, W. Kou, et al., Plant recognition and counting of amorphophallus Konjac based on UAV RGB imagery and Deep learning, *Comput. Electron. Agric.* 235 (2025) 110352.
- [38] F. Wang, H. Jiang, J. Wu, et al., Efficient detection and counting method for maize seedling plots, *Smart Agric. Technol.* 11 (2025) 100914.
- [39] D. Liu, J. Fang, Y. Zhao, DMSF-YOLO: a dynamic multi-scale fusion method for maize tassel detection in UAV low-altitude remote sensing images, *Agriculture* 15 (12) (2025) 1259.
- [40] Y. Yue, W. Zhang, Detection and counting model of soybean at the flowering and podding stage in the field based on improved YOLOv5, *Agriculture* 15 (5) (2025) 528.
- [41] F. Qu, H. Li, P. Wang, et al., Rice spike identification and number prediction in different periods based on UAV imagery and improved YOLOv8, *Comput. Mater. Contin.* 84 (2) (2025) 3911–3925.
- [42] S. Sood, H. Singh, B. Khan, et al., Leveraging the WFD2020 dataset for multi-class detection of wheat fungal diseases with YOLOv8 and faster R-CNN, *Comput. Mater. Contin.* 84 (2) (2025) 2751–2787.