Tomato Ripeness Detection and Harvesting Decision Making in a Greenhouse Using BIIE-YOLOv10n

Fanjia Meng[1], Ming Lu[1], Lihong Huang[2], Xin Wang[2,*]
[1]College of Information and Electrical Engineering, China Agricultural University (East Campus), Beijing, China
[2]Beijing Key Laboratory of Optimized Design for Modern Agricultural Equipment, College of Engineering, China Agricultural University (East Campus), Beijing, China
*meng@cau.edu.cn; lming@cau.edu.cn; 2024307150326@cau.edu.cn; *wangxin117@cau.edu.cn*

***Abstract*—Accurate detection of tomato ripeness is crucial to improving harvesting efficiency and supporting precise picking decisions. However, occlusion and overlap of fruits, along with complex background interference, can lead to loss of local information, unstable color distribution, and increased difficulty in ripeness discrimination, thereby undermining the stability and reliability of detection. To address these challenges, this study proposes a real-time tomato ripeness detection model based on an improved YOLOv10n framework, named BIIE-YOLOv10n. The model employs an improved bidirectional feature pyramid network (IBiFPN) to achieve adaptive multi-scale feature fusion and enhanced contextual information exchange, integrates an improved iterative channel-spatial attentional feature fusion (ICSAFF) mechanism into the C2f module for effective global-local feature aggregation, and introduces the Inner-EIoU loss function to balance positive and negative samples, thereby improving bounding box regression accuracy under complex environments. Experimental results on a self-constructed tomato ripeness dataset show that the proposed model achieves an accuracy of 82.6 %, a recall of 80.5 %, an F1-score of 82.0 %, and an mAP50 of 85.4 %, representing improvements of 1.3 %, 5.2 %, 4.0 %, and 5.0 % over the baseline model, respectively. Based on the detection results, a visual-driven strategy is developed for the assessment of cluster-level ripeness and picking decision-making, providing support for automated harvesting systems in greenhouse environments. In summary, BIIE-YOLOv10n significantly enhances tomato ripeness detection performance and provides reliable decision-making support for automated harvesting and intelligent grading in greenhouse settings.**

***Index Terms*—Tomato ripeness detection; YOLOv10n; Greenhouse scenarios; Harvesting decision.**

I. INTRODUCTION

Tomato is an economically important crop that is widely cultivated in greenhouse environments [1], and its ripeness directly affects harvest timing, fruit quality [2], and subsequent supply management. Currently, tomato harvesting is largely based on manual experience for assessing ripeness, which suffers from strong subjectivity [3], low efficiency, and insufficient consistency in evaluation standards [4], [5]. With the gradual adoption of harvesting robots and intelligent agricultural equipment, the automatic, objective, and reliable visual recognition of fruit ripeness has become a critical prerequisite to realize intelligent harvesting systems [6]. Moreover, accurate detection of ripeness contributes to optimized inventory management and reduced post-harvest losses [7], thus promoting the intelligent and large-scale and development of the tomato industry.

Although greenhouse cultivation alleviates external environmental disturbances to some extent, visual recognition of tomato ripeness remains a challenge. Tomato fruits exhibit continuous and gradual changes across different ripeness stages, especially during the color-transition phase from pink to red, where high similarity in color distribution and texture appearance significantly increases the difficulty of ripeness discrimination. In the early growth stage, immature green fruits also show a strong visual resemblance to surrounding leaves, which can easily lead to misclassification. Moreover, under dense greenhouse planting conditions, tomatoes typically grow in clusters, resulting in frequent inter-fruit occlusion and overlap. Such occlusion causes incomplete object boundaries and loss of discriminative local features, ultimately degrading the accuracy and robustness of ripeness detection models.

During the transitional ripeness stages, tomatoes exhibit a high similarity in both color distribution and texture appearance [8], causing discriminative cues between different ripeness classes to reside primarily in subtle local variations and cross-scale semantic differences. Although the original C2f module enhances feature flow through multi-branch convolutional structures, it treats all channels and spatial locations uniformly, lacking explicit modeling of fine-grained color and texture variations [9]. SENet emphasizes responses to color channels related to rippeness through channel recalibration and suppresses redundant feature channels [10], [11]; however, by ignoring spatial and contextual information, it struggles to capture the relationship

Manuscript received 17 October, 2025; accepted 28 January, 2026.
This work was financially supported by the 2025 Integrated Project for Agricultural Machinery R&D, Manufacturing, Promotion, and Application and the National Key Research and Development Program of China under Grant No. 2023YFD2000902.

between local textures and global semantics. In contrast, the iAFF module iteratively fuses local and global characteristics to improve feature consistency [12], but without adaptive channel-wise weighting, it remains insufficient to effectively distinguishing subtle ripeness-related differences during transitional stages.

In tomato ripeness recognition, single-fruit identification is based primarily on detailed cues such as edges, color, and local textures. Although these characteristics can describe individual fruits, they do not form stable holistic representations under occlusion or dense distribution conditions [13]. When shallow and deep features are inadequately integrated, low-level boundary information cannot be effectively utilized by higher-level semantics, and high-level features cannot compensate for missing fine details. This imbalance increases the risk of missed detections and reduces the reliability of the detection. To mitigate this problem, feature pyramid networks (FPN) introduce a top-down pathway to propagate high-level semantic information to lower layers [14], thus enhancing the representation of small and occluded targets. Building upon this, BiFPN further incorporates bidirectional information flow and employs learnable weights for adaptive multi-scale feature fusion [15], [16], demonstrating improved stability and discriminative capacity in dense fruit scenarios. Nevertheless, its feature weighting mechanism relies primarily on local feature responses and lacks explicit modeling of global contextual relationships. When low-level features are incomplete or severely occluded, uncertainty remains in evaluating cross-scale feature importance, which partially weakens multi-scale integration and target discrimination.

Under dense greenhouse planting conditions, severe fruit overlap is common, and traditional CIoU-based loss functions impose relatively weak constraints on geometric differences between the bounding boxes [17], which can lead to unstable regression. To improve regression performance, EIoU [18] decouples width-height discrepancies from enclosing box constraints to accelerate convergence and enhance localization accuracy [19], while Inner-IoU strengthens the gradient responses for high-IoU samples by introducing auxiliary bounding boxes [20]. However, these methods still struggle to simultaneously balance geometric constraints and sample discrimination. Therefore, improving the accuracy and stability of bounding box regression under dense occlusion remains a critical challenge.

Motivated by the above challenges, this study proposes an improved YOLOv10-based detection model, termed BIIE-YOLOv10n, which enables fast and accurate detection of tomatoes at three ripeness stages: green, pink, and red. The main contributions of this work are summarized as follows. First, a color shift estimation-and-correction (CSEC) module is employed to perform illumination correction on input images, enhancing color consistency, and improving discrimination among different ripeness stages. Second, an iterative channel-spatial attentional feature fusion (ICSAFF) module is embedded in the C2f structure to effectively capture both global and local information, thus strengthening feature extraction for tomatoes at different ripeness stages. Additionally, an optimized BiFPN structure is introduced to replace the original concatenation modules, enabling dynamic adjustment of multi-scale feature contributions and enhanced contextual integration, which improves detection stability and grading performance in densely overlapped scenarios. Finally, the Inner-EIoU loss is adopted in place of CIoU by incorporating auxiliary bounding boxes to emphasize high-quality samples and combining width-height discrepancy losses to reinforce geometric constraints, resulting in more stable regression and improved localization accuracy under complex conditions.

II. MATERIALS AND METHODS

In this study, a vision-based tomato harvesting decision system is developed, and its overall workflow is illustrated in Fig. 1. The system follows a unified "acquisition-detection-decision" system, allowing a complete closed-loop process from greenhouse tomato image acquisition to harvesting decision generation. The system consists of three main modules:

1. Dataset construction, which captures tomato images in the greenhouse environment and constructs the experimental dataset;
2. Ripeness detection, which leverages the proposed BIIE-YOLOv10 model to localize tomato fruits and classify their ripeness stages in each image, producing ripeness detection results at the individual fruit level;
3. Harvesting decision, which constructs fruit clusters based on single-fruit detection results and outputs cluster-level ripeness assessments to support subsequent harvesting strategy formulation.

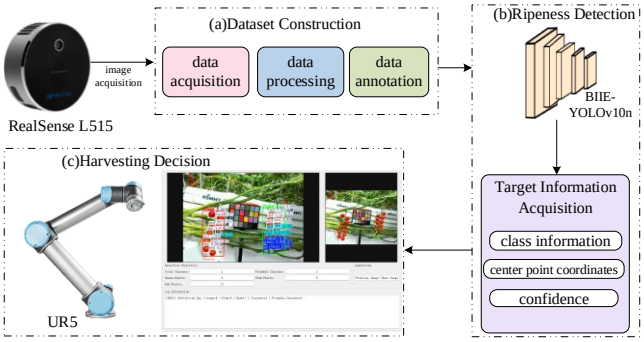


Fig. 1. Schematic illustration of the harvesting robot decision-making process: Module (a) represents the data acquisition process; Module (b) shows tomato ripeness detection based on the proposed BIIE-YOLOv10 model; Module (c) illustrates the generation of harvesting decisions based on tomato cluster ripeness information.

A. Dataset Construction

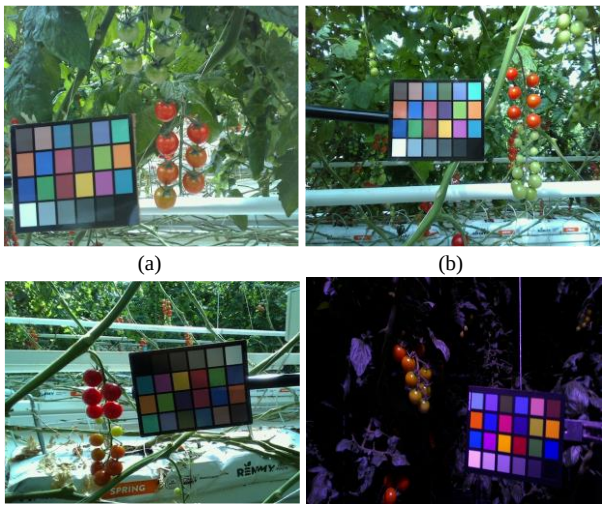
1. Data Acquisition

The tomato image dataset used in this study was collected at the Jingwa Agricultural Science and Technology Demonstration Park in Pinggu District, Beijing, under real greenhouse cultivation conditions. Image acquisition was performed with an Intel RealSense L515 camera, positioned approximately 0.4 m to 0.6 m from the tomato plants to ensure sufficient coverage of individual fruits while maintaining practical harvesting distances.

To comprehensively capture variations in fruit appearance, images were collected across different ripeness stages, spatial distributions, and illumination conditions. Daytime images were acquired on April 10 and April 24, 2024, from 9:00 a.m. to 7:00 p.m., covering morning, noon, and afternoon periods under both sunny and overcast weather conditions, resulting

in 447 images. Nighttime images were collected on April 20, from 7:00 p.m. to 11:00 p.m., under artificial supplementary lighting, yielding 131 images. As a result, the dataset encompasses various illumination scenarios, including normal illumination, front lighting, backlighting, and nighttime supplementary lighting, which are representative of real greenhouse operating environments. In total, 578 images of Sweet 100 red cherry tomatoes were collected, as illustrated in Fig. 2.

All images were uniformly resized to 640×640 pixels before training, ensuring input consistency for the detection model while preserving essential tomato features and reducing computational complexity, thus facilitating efficient and stable model training.



(a) (b) (c) (d)

Fig. 2. Images of the greenhouse tomato dataset: (a) Normal illumination; (b) Front lighting; (c) Backlighting; (d) Night supplementary lighting.

2. Dataset Annotation

Based on practical production requirements, tomatoes were classified into three ripeness stages, namely green, pink, and red, as shown in Table I, and the classification was determined based on changes in fruit skin color.

TABLE I. INSTANCES AND DESCRIPTIONS OF TOMATOES AT VARIOUS RIPENESS STAGES.

	Green	Pink	Red
Example pictures			
Quantities	2923	4106	3781
Description	Greenish peel	The pericarp shows a marked change from green to yellowish, yellowish, or a combination of both.	Dark red or reddish peel

Tomatoes at different ripeness stages are suitable for different processing, storage, and marketing strategies. The dataset was manually annotated using the open source LabelImg tool [21]. During the annotation, a minimum bounding rectangle was applied to each tomato, as illustrated in Fig. 3.

(a) (b)

Fig. 3. Examples of annotated tomato images collected in greenhouse environments, covering (a) day and (b) night. The white bounding boxes represent red tomatoes, the light-cyan boxes denote pink tomatoes, and the blue boxes indicate green tomatoes.

All annotations were repeatedly checked to ensure accuracy. The dataset was randomly divided into training, testing, and validation sets in a ratio of 8 to 1 to 1, and data augmentation techniques including translation, flipping, and scaling were applied to enhance the model's feature learning capacity and generalization performance. The augmented dataset was used for tomato ripeness detection.

B. Illumination Correction - CSEC

To mitigate the impact of illumination variations in greenhouse environments on tomato ripeness recognition, the CSEC model is introduced to perform illumination correction on acquired images; for detailed implementation, see [22]. Traditional color normalization and histogram-based correction methods are designed primarily based on pixel-level statistics or visual priors. Due to their reliance on fixed rules or global mapping relationships, these approaches struggle to simultaneously correct overexposed and underexposed regions. The CSEC method analyzes the opposing color shift patterns present in overexposed and underexposed regions and constructs pseudo-normal color features as references. Based on this representation, it estimates and corrects the color shifts in different exposure regions, enabling the collaborative optimization of brightness adjustment and color correction.

C. Ripeness Detection-BIIE-YOLOv10n

Based on the YOLOv10n framework [23], this study constructs a lightweight tomato ripeness detection model termed BIIE-YOLOv10n. To address challenges such as insufficient representation of ripeness-critical information across different feature hierarchies, limited efficiency in multi-scale feature fusion, and suboptimal bounding box regression, the model is systematically improved from three aspects: feature extraction, feature fusion, and bounding box regression.

First, an attention module named ICSAFF is designed and embedded into the C2f module of the backbone network. Inspired by SENet, ICSAFF extends the iAFF mechanism by integrating iterative channel and spatial attention, thereby enhancing the collaborative representation of multi-scale semantic information and fine-grained details. Second, an IBiFPN structure is introduced in the feature fusion stage, where only the first five layers of the original BiFPN bidirectional feature transmission architecture are retained. Learnable fusion weights are incorporated to adaptively balance cross-scale feature contributions, enabling effective multi-scale information interaction. Finally, the Inner-EIoU loss function is adopted for bounding box regression. By

jointly incorporating internal overlap constraints and scale-sensitive penalty terms, the proposed loss accelerates regression convergence and improves localization accuracy for tomato targets at different scales. The overall network architecture is illustrated in Fig. 4.

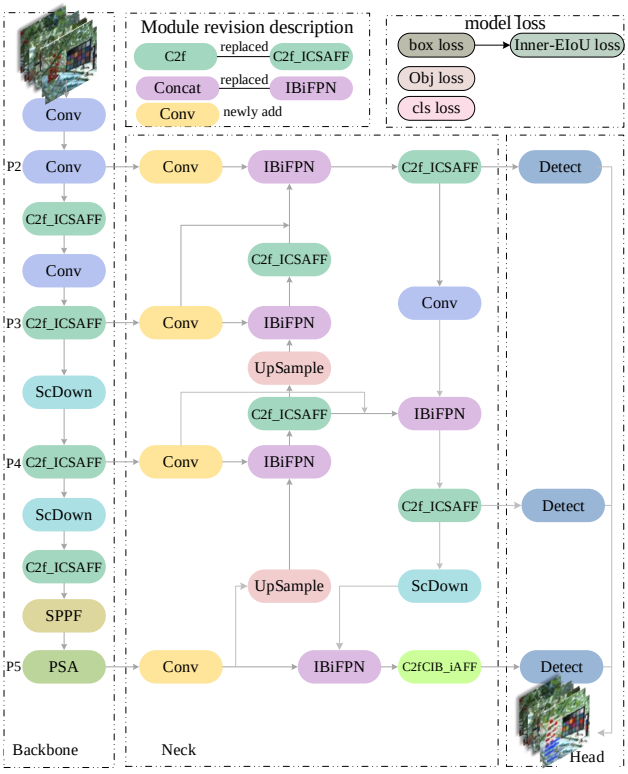


Fig. 4. The BIIE-YOLOv10n framework for tomato-related detection. The images extract features through the backbone network C2f_ICSAFF; the neck adopts IBiFPN for cross-scale feature fusion; and the head utilizes the Inner-EIOU loss function to predict classes and bounding boxes. The module revision description emphasizes the replaced and newly introduced components, and the model loss block summarizes the overall loss formulation.

1. Feature Extraction Model - ICSAFF

In this study, an attention mechanism termed ICSAFF is proposed and integrated into the C2f module, as illustrated in Fig. 5.

The input to the module is a tensor X with dimensions (B, C, H, W), where B represents the batch size, C the number of channels, and H and W the height and width of the feature map. The module adopts a two-stage iterative feature fusion strategy. In the first stage, local spatial characteristics are extracted using a sequence of 1×1 and 3×3 convolutions, while global contextual characteristics are obtained through adaptive average pooling followed by a 1×1 convolution. Considering the small-batch training setting, group normalization (GN) is consistently employed throughout the module to improve normalization stability.

On this basis, an SENet-based channel attention mechanism is introduced to compress global features and generate channel-wise attention weights, thereby adaptively emphasizing discriminative feature channels. Subsequently, the local and global features are fused via a weighted residual scheme, yielding an intermediate feature representation.

In the second stage, the module follows the same fusion framework as the first stage, but replaces the 3×3 convolution with a 5×5 convolution in the local feature extraction branch to enlarge the receptive field and facilitate multi-scale spatial feature modeling. The global attention and feature fusion operations remain identical to those in the first stage. After the two-stage fusion process, the final output of the ICSAFF module integrates multi-scale spatial information and adaptive channel-weighted features. The computational procedure of the proposed ICSAFF module is summarized in Algorithm 1.

Algorithm 1. ICSAFF Module.

Input: Feature tensor X

Output: Fused feature tensor Z

1. $Y_1, Y_2 \leftarrow chunk(Conv1(X), 2)$Apply Conv1 to X, then split into two branches
2. Initialize Y_{list} as empty
3. for i = 1 to 2 do
4. Apply the i^{th} Bottleneck to Y2
5. Append Y_{next} to Y_{list}
6. end for
7. Y_{csp} ←Conv2(concat(Y_{list})) Concatenate all features and apply Conv2
8. First-stage-MS-CSAM
9. $X_a \leftarrow X + Y_{csp}$ Initial element-wise addition
10. $X_l \leftarrow LocalAtt(X_a)$ Local attention (multi-scale 3×3)
11. $X_g \leftarrow GocalAtt(X_a)$ Global attention (via global pooling)
12. $X_a \leftarrow X_l + X_g$ Combine local and global attention
13. $W \leftarrow sigmoid(ChannelAtt(X_a) \times X_{lg})$ Channel weighting + Sigmoid
14. $X_i \leftarrow X \times W + Y_{csp} \times (1 - W)$ First-stage weighted fusion
15. Second-stage-MS-CSAM
16. $X_{l2} \leftarrow LocalAtt2(X_i)$ Local attention (5×5)
17. $X_{g2} \leftarrow GocalAtt2(X_i)$ Global attention
18. $X_{lg2} \leftarrow X_{l2} + X_{g2}$ Combine attention outputs
19. $W \leftarrow sigmoid(ChannelAtt(X_i) \times X_{lg2})$ Channel weighting + Sigmoid again
20. $Z \leftarrow X \times W_2 + Y_{csp} \times (1 - W_2)$ Second-stage weighted fusion
21. Return Z

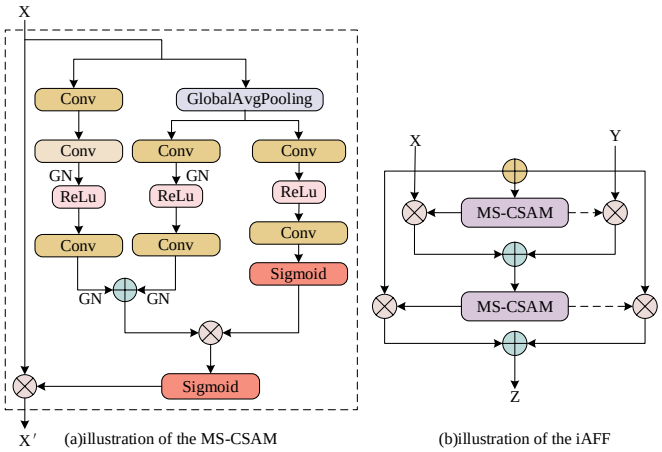


Fig. 5. Two core components of the ICSAFF module are presented: (a) The MS-CSAM sub-module integrates SENet channel attention, generating enhanced features through local/global feature extraction and channel weighting; (b) The iAFF sub-module fuses features in an iterative manner.

2. Multi-Scale Feature Fusion Model - IBiFPN

In this study, an IBiFPN structure was introduced to enhance multi-scale feature fusion. By jointly exploiting bottom-up feature aggregation and top-down semantic enhancement, IBiFPN enables effective interaction among features at different scales. In addition, learnable fusion weights were incorporated to adaptively balance the contributions of multi-scale features, thereby strengthening contextual information modeling, as illustrated in Fig. 6.

To further enhance the expressive capability of feature fusion, the original ReLU activation function was replaced by the Swish activation function. Due to its smooth and self-gated nonlinearity, Swish improves the model's feature selection ability. The Swish function is defined as

$$swish(x) = x \times \sigma(x) = x \times \frac{1}{1+e^{-x}}, \quad (1)$$

where $\sigma(x)$ is the Sigmoid function.

In the neck of the network, a multi-stage BiFPN-based feature fusion structure was constructed. Initially, three convolutional layers were introduced to align channel dimensions and generate multi-scale feature representations, which served as input for subsequent bidirectional fusion. The BiFPN modules then progressively fuse feature maps across different pyramid levels, enabling high-level semantic information to be effectively propagated to lower layers while fine-grained spatial details are fed back to higher layers. Through this bidirectional feature flow, consistent and robust feature representations are established across small-, medium-, and large-scale targets. Consequently, the proposed neck structure significantly enhances multi-scale feature representation capability, providing effective feature support for tomato detection under complex greenhouse conditions.

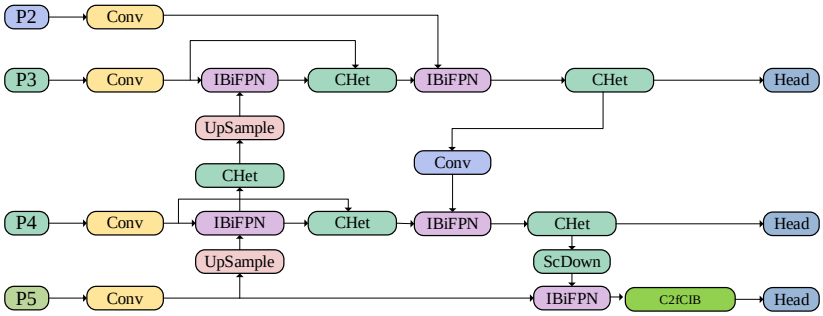


Fig. 6. Architecture of the IBiFPN module, illustrating multi-scale inputs from P2 to P5 and bidirectional top-down and bottom-up fusion paths with learnable fusion weights.

3. Loss Function-Inner-EIoU

In this study, the *EIoU loss* was introduced, and *Inner-IoU* was adopted to replace traditional IoU. By incorporating auxiliary bounding boxes and a scaling factor, this approach accelerates regression convergence and enhances the model's generalization capacity for tomatoes of different scales:

$$a_{l-gt} = x_{gt} - \frac{w_{gt} \times r}{2}, \quad (2)$$

$$a_{r-gt} = x_{gt} + \frac{w_{gt} \times r}{2}, \quad (3)$$

$$a_{t-gt} = y_{gt} - \frac{h_{gt} \times r}{2}, \quad (4)$$

$$a_{b-gt} = y_{gt} + \frac{h_{gt} \times r}{2}, \quad (5)$$

$$a_l = x - \frac{w \times r}{2}, \quad (6)$$

$$a_r = x + \frac{w \times r}{2}, \quad (7)$$

$$a_t = y - \frac{h \times r}{2}, \quad (8)$$

$$a_b = y + \frac{h \times r}{2}, \quad (9)$$

$$inter = \left(\min\left(a_{r-gt}, a_r\right) - \max\left(a_{l-gt}, a_l\right)\right) \times \times \left(\min\left(a_{b-gt}, a_b\right) - \max\left(a_{t-gt}, a_t\right)\right), \quad (10)$$

$$union = \left(w_{gt} \times h_{gt}\right) \times \left(r\right)^2 + \left(w \times h\right) \times \left(r\right)^2 - inter, \quad (11)$$

$$IoU_{inner} = \frac{inter}{union}, \quad (12)$$

where a_{l-gt}, a_{r-gt}, a_{t-gt}, a_{b-gt}, and a_l, a_r, a_t, a_b represent the left, right, top, and bottom boundaries of the ground-truth box and the predicted box, respectively. (x_{gt}, y_{gt}) and (x, y) denote the center coordinates of the ground truth and predicted boxes, and w_{gt}, h_{gt} and w, h represent their widths and heights. The r is a scaling factor used to adjust the size of the auxiliary bounding boxes.

EIoU inherits the advantages of *CIoU* while introducing key improvements. It retains CIoU's consideration of overlap area, center distance, and aspect ratio, and additionally incorporates a penalty term for the width and height differences between the predicted and ground-truth boxes. The formula is as follows

$$L_{inner} - EIoU = 1 - IoU_{inner} + \frac{\rho^2\left(a, a_{gt}\right)}{\left(w_c\right)^2 + \left(h_c\right)^2} + + \frac{\rho^2\left(w, w_{gt}\right)}{\left(w_c\right)^2} + \frac{\rho^2\left(h, h_{gt}\right)}{\left(h_c\right)^2}, \quad (13)$$

where $\rho^2(w, w_{gt})$ and $\rho^2(h, h_{gt})$ denote the squared differences in width and height between the predicted and ground-truth boxes, and w_c and h_c are the width and height of the minimum enclosing rectangle containing both boxes.

EIoU evaluates the similarity between predicted and ground-truth boxes from multiple dimensions. Its width-height penalty guides the model to adjust the predicted boxes more reasonably in terms of size, addressing CIoU's limitations in handling objects with large aspect ratio variations. *Inner-EIoU* further stabilizes gradient responses for high-quality samples by introducing auxiliary bounding boxes, enhancing the model's bounding box regression capability in complex and highly overlapping scenarios. This provides a more robust characteristic foundation for subsequent tomato ripeness estimation.

III. EXPERIMENTS

A. Experimental Environment and Parameter Settings

All experiments were conducted on a Windows 10 platform equipped with an Intel Core i7-12700 CPU, 16 GB of RAM, and a NVIDIA GeForce RTX 2060 GPU. The model was implemented using the PyTorch 2.0.1 deep learning framework with Python 3.9.19, and the training environment was managed via Anaconda. CUDA 11.8 and cuDNN 8.7.0 were employed to accelerate model training and inference.

The network was trained for 400 epochs with a batch size of 8, an initial learning rate of 0.01, and an IoU threshold of 0.7. The optimizer used a momentum factor of 0.937 and a weight decay of 0.0005. An early stopping mechanism was applied, terminating training if no improvement was observed over 50 consecutive epochs, to prevent overfitting and reduce unnecessary computation.

For fruit cluster construction, the DBSCAN clustering algorithm was employed to group the detected individual fruits based on their spatial proximity. The neighborhood radius parameter was set to 120 pixels to define the spatial threshold for clustering and the minimum number of samples was set to 1 to ensure that isolated fruits or sparsely distributed fruits could still be considered valid clusters.

B. Evaluation Metrics

To evaluate the proposed algorithm for detecting tomatoes at different ripeness stages and guiding harvesting decisions, four metrics were used: precision (P), recall (R), F1-score, mAP. These metrics are defined in detail in previous work [24].

P measures the proportion of correctly predicted positives among all predicted positives, while R measures the proportion of true positives correctly identified, reflecting the model's completeness. F1-score balances precision and recall, and the mAP averages the precision across all classes.

C. Harvesting Decision

In greenhouse tomato harvesting, the harvesting unit is typically based on fruit trusses. Within a truss, the ripeness of individual fruits exhibits spatial variation: fruits at the top receive sufficient light and ventilation, becoming ripe earlier, whereas fruits at the bottom are limited in light and nutrients and thus ripe later. If harvesting decisions are made solely based on the ripeness of the top or middle fruits, unripe fruits at the bottom may be prematurely harvested, leading to reduced fruit quality and overall yield.

To allow an effective transition from single-fruit detection to truss-level harvesting decisions, this study first employs the BIIE-YOLOv10 model to obtain single-fruit detection results, including the two-dimensional spatial coordinates (x, y) of each fruit center and its ripeness classification. Subsequently, the DBSCAN clustering algorithm is applied to spatially cluster individual fruits, grouping adjacent fruits into trusses, thereby forming practical harvesting units. The DBSCAN clustering uses the neighborhood radius and minimum number of points to control the cluster density, ensuring robust and reproducible clustering results.

For truss-level ripeness assessment, the ripeness of the lowest fruit within each truss is adopted as the criterion. As shown in Fig. 7, the ripeness of a truss is determined by the fruit located at the lowest spatial position within the cluster: if this fruit is classified as ripe red by BIIE-YOLOv10, the entire truss is considered pickable; otherwise, the truss is deemed unripe. This strategy ensures that all fruits within a truss meet the harvesting standard.

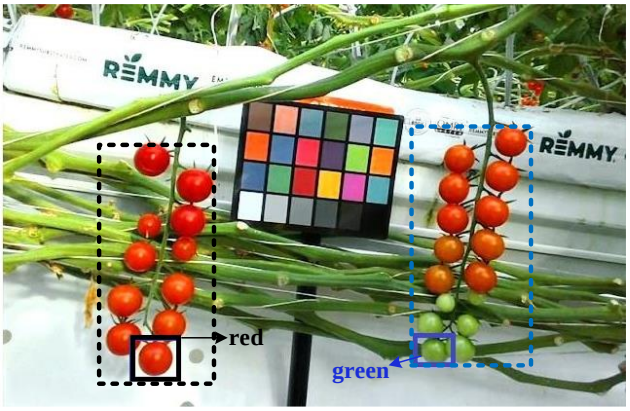


Fig. 7. Ripeness determination at the cluster level. The ripeness of each fruit cluster is based on the fruit located at the lowest position within the cluster: if the lowest-positioned fruit is ripe, the entire cluster is considered ready for harvest; otherwise, the cluster is deemed unripe and not harvested. For example, in the black box, the fruit is red, so the entire cluster is judged ready for harvest. In the blue box, the fruit is green, so the cluster is considered unripe.

IV. RESULTS

This section provides a systematic evaluation of the proposed BIIE-YOLOv10n model for tomato ripeness detection. The experimental design consists of the following parts. First, comparison experiments were performed to evaluate the proposed model against mainstream object detection algorithms, using the evaluation metrics to analyze classification accuracy. Second, ablation experiments were conducted to systematically verify the individual and combined effects of the IBiFPN, ICSAFF, and Inner-EIoU modules, analyzing their contributions to precision, recall, and overall detection performance. Finally, a harvesting robot decision experiment was carried out to identify the ripeness stages of tomato clusters and generate picking recommendations, demonstrating the practical applicability of the model in agricultural automation scenarios.

A. Comparative Experiments

To further validate the performance of the proposed method for tomato ripeness detection, a series of comparative experiments were conducted. Several mainstream object detection algorithms were selected for comparison, including YOLOv10s, YOLOv8n, YOLOv8s, and SSD. After thorough testing, the detection performance of each model is summarized in Table II.

The results indicate that the proposed method achieves competitive performance while maintaining a relatively small model size. Specifically, the proposed model attains a P of 82.6 %, which is higher than that of YOLOv8s at 80.5 %, YOLOv8n at 75.9 %, YOLOv10s at 81.4 %, and SSD at 81.5 %, with improvements of 2.1 percentage points, 6.7 percentage points, 1.2 percentage points, and 1.1 percentage points, respectively. These results demonstrate that the proposed approach can more accurately determine tomato ripeness and effectively reduce false detections. In terms of R, the proposed model achieves a value of 80.5 %, which is

higher than that of YOLOv10s and YOLOv8s at 80.1 %, YOLOv8n at 77.7 %, and SSD at 67.6 %, indicating that the proposed method is capable of detecting most of tomato ripeness samples. Regarding mAP50, the proposed model reaches a high accuracy of 85.4 %, representing an improvement of 3.1 percentage points compared to YOLOv10s, 3.8 percentage points over YOLOv8s, and an improvement of 6.7 percentage points compared to YOLOv8n, YOLOv8s, and SSD. This indicates that the proposed model exhibits superior overall performance in multi class detection tasks and can more effectively distinguish different tomato ripeness stages.

Moreover, the proposed model achieves the highest F1-score of 82.0 %, demonstrating a better balance between precision and recall. This indicates that the model can maintain high detection accuracy while identifying a larger number of samples. Finally, the model size of the proposed approach is 12.1 MB, which is larger than that of YOLOv8n at 6.3 MB, but significantly smaller than those of YOLOv10s, YOLOv8s, and SSD. This suggests that the proposed model requires fewer hardware resources.

Overall, the experimental results demonstrate that the proposed model achieves a favorable balance between detection performance and computational complexity, delivering high accuracy with relatively low model complexity.

TABLE II. COMPARISON RESULTS OF DIFFERENT NETWORKS.

Network	P/%	R/%	mAP50/%	F1/%	Model Size(M)
BIIE-YOLOv10n	**82.6**	**80.5**	**85.4**	**82.0**	12.1
YOLOv10s	81.4	80.1	82.3	80.0	16.6
YOLOv8n	75.9	77.7	78.7	77.0	**6.3**
YOLOv8s	80.5	80.1	81.6	80.0	22.6
SSD	81.5	67.6	78.7	73.6	23.8

B. Ablation Experiments

To verify the effectiveness of the BIIE-YOLOv10n model in detecting tomatoes at different ripeness stages in complex growth environments, YOLOv10n was adopted as the baseline model, and the IBiFPN, ICSAFF, and Inner-EIoU modules were gradually integrated to conduct ablation experiments. The experimental results are presented in Table III.

TABLE III. ABLATION EXPERIMENT OF EACH MODULE IN THE PROPOSED METHOD.

Model	IBiFPN	ICSFAFF	Inner EIoU	F1/%	P/%	R/%	mAP50/%
YOLOv10n	x	x	x	78.0	81.3	75.3-	80.4
	√	x	x	80.0	82.3	78.7	84.5
	√	√	x	80.0	84.2	75.3	83.1
	√	√	√	82.0	82.6	80.5	85.4

The baseline YOLOv10n model achieved an initial F1-score of 78.0 %, a precision of 81.3 %, a recall of 75.3 %, and an mAP50 of 80.4 %. After introducing the IBiFPN module, the model performance improved significantly, with the F1-score increasing to 80.0 %, precision increasing to 82.3 %, recall improving to 78.7 %, and mAP50 increasing to 84.5 %. These results indicate that the IBiFPN module effectively alleviates the limitations of traditional feature concatenation structures, such as unidirectional information flow and imbalanced feature distribution. By employing adaptive weight allocation, IBiFPN enhances multi scale feature fusion and improves both detection accuracy and target recall for tomato ripeness recognition.

Building upon this configuration, further integration of the ICSAFF module maintained the F1-score at 80.0 %, while the precision increased to 84.2 %. However, the recall decreased to 75.3 %, and the mAP50 reached 83.1 %. This suggests that the ICSAFF module enables the model to focus more effectively on key spatial regions and channel characteristics of tomato fruits, suppressing redundant background information. While this attention mechanism further improves detection precision, it introduces a certain trade-off that affects recall performance.

Finally, after incorporating the Inner-EIoU loss function, the complete BIIE-YOLOv10n model achieved an F1-score of 82.0 %, a precision of 82.6 %, a recall of 80.5 %, and an mAP50 of 85.4 %. The slight decrease in precision can be attributed to the Inner-EIoU loss function that increases the number of positive predictions in scenarios with imbalanced positive and negative samples, particularly for tomatoes with blurred boundaries or severe occlusion, to enhance recall. Although this introduces a small number of false detections, the simultaneous improvement in both the F1-score and the mAP50 demonstrates that the proposed loss function effectively enhances the model's adaptability to complex growth environments.

Overall, through the sequential integration of the three modules, the proposed BIIE-YOLOv10n model achieves substantial improvements in key performance metrics, significantly enhancing the accuracy and robustness of tomato ripeness detection under complex greenhouse conditions.

C. Visualization Analysis

Figure 8 presents a visual comparison between the baseline model and the proposed BIIE-YOLOv10n model in greenhouse scenarios, including cases of background interference, fruit occlusion, and transitional ripeness stages.

In background interference and transitional ripeness stage scenarios, the baseline model struggles to effectively distinguish target objects from the background or to discriminate fine-grained ripeness features under these conditions, resulting in missed detections and ripeness misclassification. The affected targets are highlighted with red and blue bounding boxes, respectively. In contrast, the improved model leverages the ICSAFF attention mechanism to effectively emphasize target-related features while suppressing background interference, enabling stable detection of tomato targets and demonstrating superior robustness under complex background conditions.

In fruit occlusion scenarios, where tomatoes grow densely and occlude each other, the baseline model exhibits limited

ability to separate adjacent targets, leading to missed detections, as indicated by the yellow bounding boxes. By comparison, the improved model employs the IBiFPN multi-scale feature fusion module to more accurately separate overlapping fruits, significantly reducing missed detections and highlighting its advantage in densely populated target scenes.

In fruit instance separation scenarios, the baseline model tends to merge multiple adjacent fruits into a single detection due to insufficient bounding box regression accuracy and severe instance adhesion. By introducing the Inner-EIoU loss function, the improved model jointly constrains the center distance, the scale discrepancy, and the internal overlap between the predicted and ground-truth bounding boxes during regression. This design enhances the model's sensitivity to the geometric relationships among overlapping instances during training, thus improving the accuracy and stability of bounding box regression and enabling effective separation of highly overlapping fruit instances.

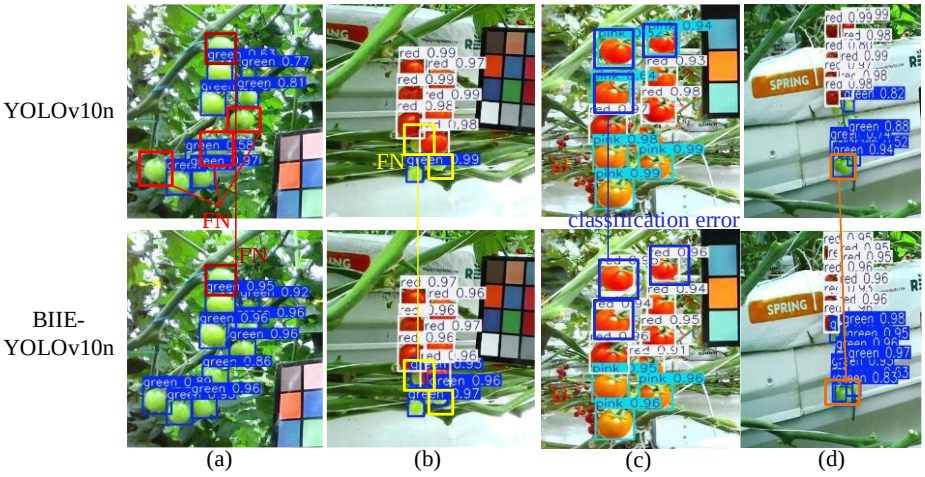


Fig. 8. Detection visualization results under different conditions, including (a) background interference, (b) fruit overlap, (c) transitional ripeness stage, and (d) fruit instance separation. For each scenario, the upper panel shows the results of the baseline model, while the lower panel corresponds to the proposed BIIE-YOLOv10n model. The red boxes indicate missed detections caused by background interference, the yellow boxes denote missed detections due to occlusion, the blue boxes represent misclassification errors, and the orange boxes additionally highlight the separation of highly overlapping fruit instances.

Overall, the visual comparison results indicate that the proposed BIIE-YOLOv10n model exhibits superior detection stability and ripeness recognition accuracy compared to the baseline model in complex greenhouse scenarios, further validating the effectiveness of the improved module for fine-grained fruit detection and practical agricultural applications.

D. Harvesting Decision Results

Table IV compares the harvesting decision performance of the proposed model under six illumination and fruit growth scenarios. Across these scenarios, harvesting decision accuracy ranged from 40.0 % to 65.0 %. In contrast, recall remained consistently high, ranging from 87.5 % to 92.9 %, indicating that the model was able to identify and cover the majority of actual harvestable trusses, with minimal missed harvesting. Specifically, under common greenhouse conditions, such as backlighting, front lighting, and leaf interference, the model maintained a relatively stable harvesting decision accuracy while achieving high recall, demonstrating robust overall performance. Under nighttime artificial lighting, although harvesting decision accuracy decreased to 40.0 %, recall remained at 88.9 %, indicating that performance reduction was primarily due to increased false positives rather than missed harvestable trusses. In scenarios involving fruit occlusion and overlap, recall exceeded 92 %, further confirming the model's ability to reliably identify harvestable trusses in complex spatial arrangements.

TABLE IV. PICKING DECISION EXPERIMENT.

Truss Index	Scenario	Ground-Truth Harvestable Trusses	Correct Harvesting Decisions	Predicted Harvestable Trusses	TP	FP	FN	P/%	R/%
1	Normal lighting	11	10	17	10	7	1	58.8	90.9
2	Backlighting	14	13	20	13	7	1	65.0	92.9
3	Front lighting	13	12	19	12	7	1	63.2	92.3
4	Leaf interference	8	7	11	7	4	1	63.6	87.5
5	Nighttime artificial lighting	27	24	60	24	36	3	40.0	88.9
6	Fruit occlusion and overlap	13	12	22	12	10	1	54.5	92.3

Figure 9 presents the experimental results of the tomato ripeness and harvesting decision system. In the experiments, the system was able to automatically detect each tomato truss in an image and classify the ripeness of individual fruits within each truss. Based on the ripeness information of the fruits, the system determined the harvesting feasibility of each truss. Quantitative analysis shows that the system accurately records the number of each fruit type, the total number of trusses, and the number of pickable trusses. The experimental results indicate that most of trusses were detected correctly and their ripeness was accurately assessed, with high consistency in both the fruit classification and

harvesting decisions. Furthermore, the visualized results allow for intuitive observation of fruit ripeness distribution within each truss, providing reliable experimental support and a decision-making reference for precise harvesting of greenhouse tomatoes.

Overall, the results indicate that the model maintains high harvesting decision accuracy across varying illumination and complex fruit arrangements.

Fig. 9. Visualization of the tomato harvesting decision system. It provides the number of each fruit type, the total truss count, and the pickable truss count, intuitively showing the distribution of fruits and the status of each truss.

V. DISCUSSION

For tomato ripeness detection under the same three-class classification setting, existing studies have achieved varying degrees of mAP improvement through network structure optimization: NVW-YOLOv8s [25] achieved an increase of approximately 4.3 %, Lightweight YOLO [26] improved by 0.5 %, and ToRLNet [27] achieved an enhancement of approximately 0.7 %. Although these methods perform reasonably well in limited-category tasks, the magnitude of improvement remains relatively modest. In contrast, the proposed BIIE-YOLOv10n model achieves a more substantial improvement in mAP of approximately 5 %, indicating that it can perform tomato ripeness recognition more stably and accurately under complex environmental conditions.

The ICSAFF module introduced in this study effectively enhances the discriminative representation of target regions through a multi-dimensional feature interaction mechanism. Compared with the CBAM mechanism employed by Yu *et al.* [28], ICSAFF can more effectively suppress responses from non-target regions and strengthen fruit edges and local fine-grained features under complex backgrounds, thus achieving a more balanced improvement between precision and recall. This indicates that relying solely on channel attention and spatial attention to enhance features along a single dimension may be insufficient to capture relevant fine-grained differences, whereas multi-dimensional feature co-modeling is more conducive to improving the stability of recognition during transitional ripeness stages.

At the feature fusion level, the proposed IBiFPN structure optimizes the information transmission paths of features at different scales by introducing a dynamic weighting mechanism. Compared with BiFPN or conventional path aggregation networks adopted in similar tasks [29], [30], IBiFPN demonstrates higher feature utilization efficiency in scenarios involving branch and leaf occlusion or fruit overlap. This adaptive multi-scale fusion approach helps mitigate the progressive attenuation of critical information during deep feature propagation, thereby enhancing the model's ability to perceive fine-grained differences in ripeness.

Furthermore, the introduced Inner-EIoU loss function improves the regression process from the perspective of bounding box geometric constraints. In greenhouse scenarios with densely distributed and mutually occluded fruits, Inner-EIoU effectively enhances the stability and consistency of target localization through fine-grained constraints on center distances and width-height differences, which is consistent with the findings of Wang *et al.* [31]. More accurate spatial localization provides a more reliable feature basis for subsequent ripeness classification, thus alleviating, to some extent, fluctuations in prediction caused by unstable localization.

In the harvesting decision experiments, the proposed method was able to stably identify harvestable tomato clusters in different scenarios, demonstrating a strong overall robustness. However, as shown in Fig. 10, cluster-level decision-making still exhibits certain limitations. On the one hand, due to the diverse spatial distribution of tomato clusters, the DBSCAN clustering algorithm relies on fixed thresholds and cannot automatically learn cluster-related features, which can cause multiple closely positioned mature clusters to be erroneously merged into a single cluster. On the other hand, the harvesting decision depends on identifying the least ripe fruit within each cluster, and under nighttime or locally low-light conditions, misclassification of this critical fruit can be amplified and propagated, thus affecting the final decision outcome.

(a) (b)

Fig. 10. Examples of harvesting decision results: (a) Multiple clusters mistakenly recognized as a single cluster; (b) Decision affected by an error in determining the lowest ripeness position of a cluster.

Despite these limitations, the experimental results indicate that the method provides stable fruit ripeness assessment and harvesting decision-making capabilities in complex greenhouse environments, offering practical technical support for automated tomato harvesting systems. Future research will further explore its generalization performance in different tomato varieties and cultivation regions.

VI. CONCLUSIONS

This study proposes a tomato ripeness detection system named BIIE-YOLOv10n, which addresses the challenges of tomato ripeness recognition under complex greenhouse conditions using a self-constructed tomato ripeness dataset. An improved BiFPN structure is used to replace the original neck network, allowing dynamic adjustment of feature

contributions on different scales and enhancing the network's ability to capture fine-grained target details. In addition, an improved ICSAFF module is embedded into the original C2f structure, which effectively suppresses background interference and fully exploits multi-dimensional feature information, thereby improving the accurate extraction and discrimination of tomato features at different ripeness stages. Furthermore, the introduction of the Inner-EIoU loss function alleviates the imbalance between positive and negative samples, leading to further improvement in overall model performance. The experimental results demonstrate that the proposed model achieves better performance in tomato ripeness detection, with a precision of 82.6 %, a recall of 80.5 %, an F1-score of 82.0 %, and an mAP50 of 85.4 %. Compared to the baseline model, precision, recall, F1-score, and mAP50 are improved by 1.3 percentage points, 5.2 percentage points, 4.0 percentage points, and 5.0 percentage points, respectively, indicating strong robustness under complex environmental conditions. In addition, the single-image inference time is 18.7 ms, satisfying real-time detection requirements. The proposed model enables accurate ripeness detection of tomatoes in complex backgrounds and supports corresponding harvesting decisions, thus facilitating the formulation of appropriate marketing strategies for tomatoes at different ripeness stages.

VII. FUTURE WORK

Although the proposed method demonstrates good performance and robustness in tomato ripeness detection and harvesting decision tasks, several limitations remain. On the one hand, the current model is trained on a single tomato variety with three ripeness categories, and its adaptability to different tomato varieties and finer-grained ripeness classifications can be further improved. On the other hand, cluster-level harvesting decisions rely on post-processing clustering algorithms, which may affect the stability and accuracy of harvesting decisions under complex spatial distributions.

To address these limitations, future research will focus on further advancements. In terms of dataset construction, we plan to increase the number of ripeness categories and include additional tomato varieties to enhance the model's adaptability and applicability in real-world production scenarios. Moreover, tomato cluster detection algorithms will be integrated into the model training process, guiding the network to actively learn the grouping features of tomato clusters. This integration is expected to enable the coordinated optimization of cluster-level harvesting decisions and ripeness detection, thus further improving the accuracy and reliability of harvesting decisions and providing stronger technical support for intelligent greenhouse tomato harvesting.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Z. Meng, X. Du, J. Xia, Z. Ma, and T. Zhang, "Real-time statistical algorithm for cherry tomatoes with different ripeness based on depth information mapping", *Computers and Electronics in Agriculture*, vol. 220, art. 108900, 2024. DOI: 10.1016/j.compag.2024.108900.

[2] J. Sun, X. He, M. Wu, X. Wu, J. Shen, and B. Lu, "Detection of tomato organs based on convolutional neural network under the overlap and occlusion backgrounds", *Machine Vision and Applications*, vol. 31, art. no. 31, pp. 1–13, 2020. DOI: 10.1007/s00138-020-01081-6.

[3] W. Chen, M. Liu, C. Zhao, X. Li, and Y. Wang, "MTD-YOLO: Multi-task deep convolutional neural network for cherry tomato fruit bunch maturity detection", *Computers and Electronics in Agriculture*, vol. 216, art. 108533, 2024. DOI: 10.1016/j.compag.2023.108533.

[4] W. Han, W. Hao, J. Sun, Y. Xue, and W. Li, "Tomatoes maturity detection approach based on YOLOv5 and attention mechanisms", in *Proc. of 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, 2022, pp. 1363–1371. DOI: 10.1109/ICCASIT55263.2022.9986640.

[5] S. A. Magalhães *et al.*, "Evaluating the Single-Shot Multibox Detector and YOLO deep learning models for the detection of tomatoes in a greenhouse", *Sensors*, vol. 21, no. 10, p. 3569, 2021. DOI: 10.3390/s21103569.

[6] Y. Huang *et al.*, "A review of visual perception technology for intelligent fruit harvesting robots", *Frontiers in Plant Science*, vol. 16, art. no. 1646871, 2025. DOI: 10.3389/fpls.2025.1646871.

[7] N. T. Anderson *et al.*, "Estimation of fruit load in Australian mango orchards using machine vision", *Agronomy*, vol. 11, no. 9, p. 1711, 2021. DOI: 10.3390/agronomy11091711.

[8] S. Tu, Y. Xue, C. Zheng, Y. Qi, H. Wan, and L. Mao, "Detection of passion fruits and maturity classification using Red-Green-Blue Depth images", *Biosystems Engineering*, vol. 175, pp. 156–167, 2018. DOI: 10.1016/j.biosystemseng.2018.09.004.

[9] C. Yan and H. Li, "CAPNet: Tomato leaf disease detection network based on adaptive feature fusion and convolutional enhancement", *Multimedia Systems*, vol. 31, art. no. 178, pp. 1–25, 2025. DOI: 10.1007/s00530-025-01756-y.

[10] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation networks", in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.

[11] X. Jin, Y. Xie, X.-S. Wei, B.-R. Zhao, Z.-M. Chen, and X. Tan, "Delving deep into spatial pooling for squeeze-and-excitation networks", *Pattern Recognition*, vol. 121, art. 108159, 2022. DOI: 10.1016/j.patcog.2021.108159.

[12] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional feature fusion", in *Proc. of IEEE/CVF Winter Conf. on Applications of Computer Vision*, 2021, pp. 3560–3569. DOI: 10.1109/WACV48630.2021.00360.

[13] Q. Wang, Y. Hua, Q. Lou, and X. Kan, "SWMD-YOLO: A lightweight model for tomato detection in greenhouse environments", *Agronomy*, vol. 15, no. 7, p. 1593, 2025. DOI: 10.3390/agronomy15071593.

[14] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection", in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125. DOI: 10.1109/CVPR.2017.106.

[15] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection", in *Proc. of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10778–10787. DOI: 10.1109/CVPR42600.2020.01079.

[16] C. Guo, C. Tang, Y. Liu, X. Wang, and S. Wang, "Relative position detection of clustered tomatoes based on BlendMask-BiFPN", *Elektronika ir Elektrotechnika*, vol. 30, no. 4, pp. 52–60, 2024. DOI: 10.5755/j02.eie.38247.

[17] D. Cao *et al.*, "Research on apple detection and tracking count in complex scenes based on the improved YOLOv7-Tiny-PDE", *Agriculture*, vol. 15, no. 5, p. 483, 2025. DOI: 10.3390/agriculture15050483.

[18] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, "Focal and efficient IOU loss for accurate bounding box regression", *Neurocomputing*, vol. 506, pp. 146–157, 2022. DOI: 10.1016/j.neucom.2022.07.042.

[19] R. Li, Z. Ji, S. Hu, X. Huang, J. Yang, and W. Li, "Tomato maturity recognition model based on improved YOLOv5 in greenhouse", *Agronomy*, vol. 13, no. 2, p. 603, 2023. DOI: 10.3390/agronomy13020603.

[20] H. Zhang, C. Xu, and S. Zhang, "Inner-IoU: More effective intersection over union loss with auxiliary bounding box", *ArXiv*, 2023. DOI: 10.48550/arXiv.2311.02877.

[21] Tzutalin, "LabelImg: Image annotation tool", GitHub repository, 2015. [Online]. Available: https://github.com/tzutalin/labelImg

[22] M. Lu, R. da Silva Torres, F. Meng, and X. Wang, "CICE-YOLO: An improved YOLO-based network for tomato ripeness detection in Greenhouse", *Smart Agricultural Technology*, vol. 14, art. 101973, 2026. DOI: 10.1016/j.atech.2026.101973.

[23] A. Wang *et al.*, "YOLOv10: Real-time end-to-end object detection", in *Proc. of the 38th International Conference on Neural Information Processing Systems*, art. no. 3429, pp. 107984–108011, 2024. DOI:

10.52202/079017-3429.

[24] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms", in *Proc. of 2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, pp. 237–242. DOI: 10.1109/iwssip48289.2020.9145130.

[25] A. Wang *et al.*, "NVW-YOLOv8s: An improved YOLOv8s network for real-time detection and segmentation of tomato fruits at different ripeness stages", *Computers and Electronics in Agriculture*, vol. 219, art. 108833, 2024. DOI: 10.1016/j.compag.2024.108833.

[26] T. Zeng, S. Li, Q. Song, F. Zhong, and X. Wei, "Lightweight tomato real-time detection method based on improved YOLO and mobile deployment", *Computers and Electronics in Agriculture*, vol. 205, art. 107625, 2023. DOI: 10.1016/j.compag.2023.107625.

[27] H. Sun, X. Xi, A.-Q. Wu, and R.-F. Wang, "ToRLNet: A lightweight deep learning model for tomato detection and quality assessment across ripeness stages", *Horticulturae*, vol. 11, no. 11, p. 1334, 2025. DOI: 10.3390/horticulturae11111334.

[28] Y. Xing, H. Hu, J. Zhang, Z. Han, and J. Han, "CB-YOLO-DeepSORT: Real-time yield estimation for tomatoes in greenhouses", in *Proc. of the 15th IEEE International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, 2025, pp. 66–71. DOI: 10.1109/CYBER67662.2025.11168194.

[29] F. Hao, Z. Zhang, D. Ma, and H. Kong, "GSBF-YOLO: A lightweight model for tomato ripeness detection in natural environments", *Journal of Real-Time Image Processing*, vol. 22, art. no. 47, 2025. DOI: 10.1007/s11554-025-01624-y.

[30] M. Zhao *et al.*, "Intelligent detection of tomato ripening in natural environments using YOLO-DGS", *Sensors*, vol. 25, no. 9, p. 2664, 2025. DOI: 10.3390/s25092664.

[31] S. Wang *et al.*, "Lightweight tomato ripeness detection algorithm based on the improved RT-DETR", *Frontiers in Plant Science*, vol. 15, art. 1415297, 2024. DOI: 10.3389/fpls.2024.1415297.