Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.081818

ARTICLE

Freshness Detection of Plasma Treated Tomato Using CFL-YOLOv8n

Shaohuang Bian[1,#], Qinxiu Gao[1,#], Shan Su[1], Weifeng Wang[1] and Feng Huang[2,*]

[1]College of Information and Electronic Engineering, China Agricultural University, Beijing, China
[2]College of Science, China Agricultural University, Beijing, China

*Corresponding Author: Feng Huang. Email: huangfeng@cau.edu.cn

[#]These authors contributed equally to this work

Received: 09 March 2026; Accepted: 11 May 2026; Published: 23 July 2026

ABSTRACT: Tomato, as a globally important crop, its freshness directly affects postharvest quality, market value, and consumer acceptance. Traditional tomato freshness evaluation mainly relies on manual inspection and experience-based judgment, which is time-consuming, labor-intensive, and inefficient. Meanwhile, plasma technology has shown promising potential in agricultural preservation due to its safety and effectiveness, making the evaluation of tomato freshness after plasma treatment particularly important. In recent years, with the rapid development of deep learning technology, non-destructive detection methods based on image analysis have become important tools for agricultural product quality assessment. This study proposes an improved YOLOv8n-based model (named CFL-YOLOv8n) for tomato freshness detection. The method optimizes the C2f module by introducing a RetBlock residual structure, adopts an FDPN-DASI feature pyramid for cross-scale feature fusion, and incorporates the Focaler-IoU loss function to enhance bounding box localization accuracy. Additionally, a lightweight LSCD detection head is proposed to replace the original module to reduce parameters. It shows that the improved CFL-YOLOv8n model performs better performance for tomato freshness detection than comparison models, achieving mAP@.50 and mAP@.50:.95 of 90.0% and 88.5%, improved by 3.4% and 3.4% compared to the original YOLOv8n model, while significantly reducing parameters and computational complexity. The ablation experiments confirmed the effectiveness of each improved module. The proposed method provides an efficient solution for tomato freshness detection and offers technical support for evaluating the preservation effects of plasma treatment under controlled experimental conditions. Future work will focus on expanding the dataset scale and validating the model under more complex real-world environments.

KEYWORDS: Tomato freshness detection; YOLOv8; object detection; deep learning; plasma-assisted preservation

1 Introduction

Tomato is one of the most widely cultivated and consumed horticultural crops worldwide and is of considerable economic importance in agricultural production and supply chains. With the development of intelligent sensing technologies, machine learning, hyperspectral imaging, and computer vision have been increasingly used for tomato quality monitoring, defect discrimination, and grading, enabling automated and nondestructive evaluation [1,2]. However, tomato quality is highly dynamic during storage and transportation, and freshness deterioration often involves subtle and complex changes, making accurate assessment challenging. Therefore, the development of reliable and efficient freshness evaluation methods is of great significance for reducing postharvest losses and improving the commercial value of tomato products.

To delay postharvest deterioration, a variety of preservation strategies, such as low-temperature storage, edible coatings, modified atmosphere packaging, and plasma-based treatments, have been investigated.

Nevertheless, regardless of the preservation strategy used, practical application still depends on rapid, objective, and accurate methods for freshness assessment. It should be noted that plasma treatment does not produce a directly distinguishable visual label, but mainly influences the progression of freshness degradation during storage. Such effects are gradually reflected in surface features such as spots, wrinkles, and mold. Therefore, the proposed model serves as an objective vision-based tool to quantify these appearance changes and support the evaluation of plasma-assisted preservation effects. In this context, artificial intelligence (AI) has become an important driving force in modern agriculture, with broad applications in crop monitoring, disease diagnosis, yield estimation, and quality evaluation [3]. Recent studies have further shown that deep learning can simultaneously support disease detection, severity prediction, and crop loss estimation in agricultural systems, demonstrating its strong potential for comprehensive visual analysis and decision support in real production scenarios [4].

From the perspective of computer vision, tomato freshness detection is essentially a fine-grained image recognition task. The visual differences between different freshness states are often subtle and are mainly reflected in local and heterogeneous patterns, such as slight surface spots, wrinkles, texture degradation, and early decay symptoms. Similar challenges have been widely discussed in agricultural imaging studies, where fine-grained visual analysis and discriminative feature extraction are crucial for distinguishing subtle inter-class differences [5]. At the same time, the robustness of agricultural vision models remains a major concern, because real-world environments usually involve illumination variation, occlusion, cluttered backgrounds, and domain shifts, all of which may significantly degrade model performance [6]. Therefore, effective tomato freshness detection requires not only strong feature extraction ability, but also robust multi-scale representation and good adaptability to complex agricultural scenes.

With the rapid development of deep learning, convolutional neural network (CNN)-based detection methods have shown remarkable advantages in agricultural vision tasks, including fruit detection, disease recognition, ripeness identification, and quality grading [7,8]. In particular, the YOLO series has become one of the most widely used real-time detection frameworks because of its favorable balance between accuracy and inference efficiency [9,10]. Among these methods, YOLOv8 has further improved network architecture and detection performance, while its lightweight version, YOLOv8n, offers fast inference and convenient deployment, making it particularly suitable for agricultural applications with limited computational resources. However, when applied to tomato freshness detection, YOLOv8n still exhibits several limitations, including insufficient extraction of subtle visual cues, inadequate multi-scale feature fusion, redundancy in the detection head, and limited localization accuracy in complex scenes. These shortcomings restrict its effectiveness in fine-grained freshness detection tasks.

To address these challenges, this study proposes an improved YOLOv8n-based model, termed CFL-YOLOv8n, for tomato freshness detection. The proposed method enhances feature extraction capability by introducing a RetBlock residual structure into the C2f module. In addition, an FDPN-DASI feature pyramid is designed to strengthen cross-scale feature fusion, while the Focaler-IoU loss function is adopted to improve bounding box regression accuracy. Furthermore, a lightweight LSCD detection head is employed to reduce model parameters and computational complexity. It should be noted that the contribution of this study is not to introduce completely new generic operators, but to construct a task-oriented and lightweight detection framework for tomato freshness assessment under plasma-treatment conditions. Starting from the limitations of YOLOv8n in fine-grained agricultural detection, this work performs a coordinated redesign of the backbone, neck, detection head, and regression loss, so that the model can better balance feature sensitivity, localization accuracy, and deployment efficiency. The main contributions of this study are summarized as follows:

(1) An improved YOLOv8n-based tomato freshness detection framework (CFL-YOLOv8n) is proposed, which enhances detection accuracy while maintaining lightweight characteristics suitable for real-time agricultural applications.
(2) A RetBlock-enhanced C2f module is designed to improve feature extraction capability for fine-grained freshness characteristics.
(3) An FDPN-DASI feature fusion structure is introduced to strengthen multi-scale feature representation and improve detection robustness.
(4) A lightweight LSCD detection head and Focaler-IoU loss function are incorporated to reduce model complexity while improving localization accuracy.

This paper is organized as follows. Section 2 provides a comprehensive review and analysis of related studies on tomato freshness detection. Section 3 describes the proposed methodology in detail. Section 4 presents the experimental results and corresponding discussions. Section 5 summarizes the main findings and conclusions of this study.

2 Related Work

In recent years, CNN-based methods have been widely applied to various agricultural vision tasks, including fruit recognition, disease detection, and maturity assessment. Liu et al. [11] employed an improved DenseNet architecture to achieve tomato maturity detection under complex environments. Mukhiddinov et al. [12] developed a YOLOv4-based approach to classify fresh and rotten states across multiple fruit and vegetable categories. Valentino et al. [13] proposed a lightweight CNN model for fruit freshness evaluation. However, most existing studies rely on generic network architectures, primarily focus on binary classification between “fresh” and “rotten” states, and rarely involve the freshness state with spots and wrinkles during gradual quality degradation.

Although object detection models have achieved notable progress in agricultural vision tasks, they still face a series of structural challenges in fine-grained detection. Existing detection frameworks can generally be categorized into two-stage and one-stage approaches. Two-stage methods, represented by Faster Region-based Convolutional Neural Network (Faster R-CNN) [14], rely on a Region Proposal Network (RPN) to generate candidate regions and achieve high detection accuracy; however, their substantial computational overhead limits their applicability in real-time sorting scenarios. Mask R-CNN, proposed by He et al. [15], further extends Faster R-CNN by incorporating instance segmentation capabilities, but at the cost of even higher computational complexity. In contrast, one-stage detectors such as the YOLO series [16] and Single Shot MultiBox Detector (SSD) [17] offer significantly faster inference speeds. Nevertheless, their backbone networks and feature fusion strategies were originally designed for generic objects in natural scenes. As a result, they often struggle with tomato surface defects characterized by small pixel proportions and weak visual features, leading to insufficient feature extraction and loss of fine-grained semantic information. Although YOLOv3, proposed by Redmon and Farhadi [18], and YOLOv4, developed by Bochkovskiy et al. [19], introduced improvements in multi-scale prediction, their capability to detect subtle and localized defects remains limited.

In tomato images, fruit scales vary significantly due to differences in imaging angles and individual growth characteristics, while surface defects exhibit diverse scales as well. This results in a dual-scale challenge involving both fruit-level and defect-level representations. The Feature Pyramid Network (FPN) [20] and its extended variant Path Aggregation Network (PANet) [21] adopt fixed feature fusion strategies that lack adaptive capability across scales. Although the Bidirectional Feature Pyramid Network (BiFPN) [22] introduces learnable fusion weights, its feature sources remain constrained to limited fusion paths, hindering effective integration of cross-scale contextual information. This limitation negatively affects comprehensive perception ranging from whole-fruit structures to localized surface defects. Subsequent studies, such as the Augmented Feature Pyramid Network (AugFPN) [23] introduced consistency loss and residual feature enhancement to alleviate semantic gaps and reduce detail loss. Nevertheless, the adaptive perception capability of these approaches for dual-scale targets in complex agricultural environments remains insufficient and warrants further improvement.

To meet the deployment requirements of agricultural edge devices, model lightweighting has become an inevitable choice. Models such as YOLO typically employ structurally repetitive and independent detection heads for outputs at different scales, leading to architectural redundancy and a large number of parameters [24]. Lightweight networks such as MobileNet [25] and ShuffleNet [26] reduce computational cost through depthwise separable convolutions; however, this often comes at the expense of feature representation capability. In complex detection tasks, such compromises tend to cause a significant decline in detection accuracy, making it difficult to achieve an optimal balance between computational efficiency and model performance.

Additionally, insufficient localization accuracy directly impacts the reliability of detection results. In practical scenarios, tomatoes are often densely arranged, with object bounding boxes exhibiting high overlap. Furthermore, the boundaries of states such as "wrinkled" and "moldy" are inherently blurred, imposing higher demands on localization precision. Mainstream IoU series loss functions, such as DIoU [27] and its extension CIoU, apply identical optimization strategies to all samples, lacking targeted treatment for hard examples, such as those with low overlap or ambiguous boundaries. While GIoU proposed by Rezatofighi et al. [28] and DIoU and CIoU by Zheng et al. [29] have introduced improvements in optimization direction, they still exhibit significant limitations in highly overlapping and boundary-ambiguous scenarios.

In summary, although deep learning-based object detection methods have demonstrated promising potential, they still face numerous challenges in the fine-grained detection of tomato freshness states. First, the backbone network exhibits insufficient modeling capability for local defect details and relative positional relationships among features, limiting its perceptual capacity for fine-grained features. Second, most feature fusion mechanisms rely on simple concatenation or weighting, making it difficult to balance global semantics and local details, thereby restricting the model's perceptual accuracy for targets of varying sizes. Third, traditional detection head structures are redundant, with high parameter counts and computational costs, which hinders their deployment on agricultural edge devices. Fourth, in scenarios where tomatoes are densely arranged or defect boundaries are blurred, conventional IoU series loss functions such as CIoU are insensitive to gradient variations in low-quality prediction boxes, making precise bounding box regression challenging.

3 Materials and Methods

3.1 Dataset

As physicochemical properties of PAW depend on plasma strategy, plasma-activated water was produced using arc plasma (450 W) with treatment times of 1, 3, and 5 min, generating PW1, PW3, and PW5 groups, respectively. Untreated samples (CK) and tap water treatment (TAP) were used as controls. Uniform cherry tomatoes were manually selected, randomly divided into five groups, and sprayed with 10 mL of the corresponding solution per tray after PAW was cooled to room temperature. All samples were stored at room temperature and monitored for 25 days, with images recorded every three days.

A total of 1044 tomato images from five groups were collected over 25 days and annotated into four categories, fresh, spot, wrinkle, and mouldy, using X-AnyLabeling, with examples of the labeling results shown in Fig. 1. "Fresh" refers to tomatoes with smooth skin and no visible defects; "spot," "wrinkle," and "mouldy" represent tomatoes with surface spots, skin wrinkling, and decay or mold growth, respectively, with the visually dominant feature used for labeling when multiple symptoms appeared. Bounding boxes were used to mark targets and annotation files were converted from XML to TXT format. After annotation, data augmentation including brightness enhancement, random flipping, and horizontal mirroring with synchronized annotation updates expanded the dataset from 1044 to 2088 images, which were then divided into training and testing sets at an 8:2 ratio, with 1672 images used for training and 416 images used for testing. It should be noted that the dataset was collected under controlled conditions to reduce environmental interference and to facilitate the observation of freshness changes after plasma treatment. Therefore, it does not yet cover broader variations such as multiple tomato varieties, complex backgrounds, occlusions, or diverse lighting conditions. These limitations will be addressed in future work.

Figure 1: Examples of tomato freshness status label.

3.2 The Proposed Improved CFL-YOLOv8n

As shown in Fig. 2, this study develops the CFL-YOLOv8n model based on YOLOv8n. The overall framework consists of improvements in the backbone, neck, detection head, and loss function to enhance feature representation and detection efficiency for tomato freshness assessment. The proposed CFL-YOLOv8n is designed as a coordinated optimization pipeline rather than a simple combination of independent modules. Specifically, CRBP in the backbone enhances contextual and spatial feature representation for subtle surface variations. Based on these enriched features, FDPN-DASI in the neck improves multi-scale feature interaction and preserves fine-grained details such as spots and wrinkles. The LSCD head further reduces redundancy while maintaining detection capability in dense scenarios. Finally, Focaler-IoU refines the localization process by adaptively emphasizing samples with different overlap qualities. Therefore, the four components operate at different stages of the detection pipeline, forming a progressive optimization process from feature extraction to prediction and regression, which jointly improves performance for fine-grained tomato freshness detection. This coordinated design distinguishes the proposed method from simple module concatenation.

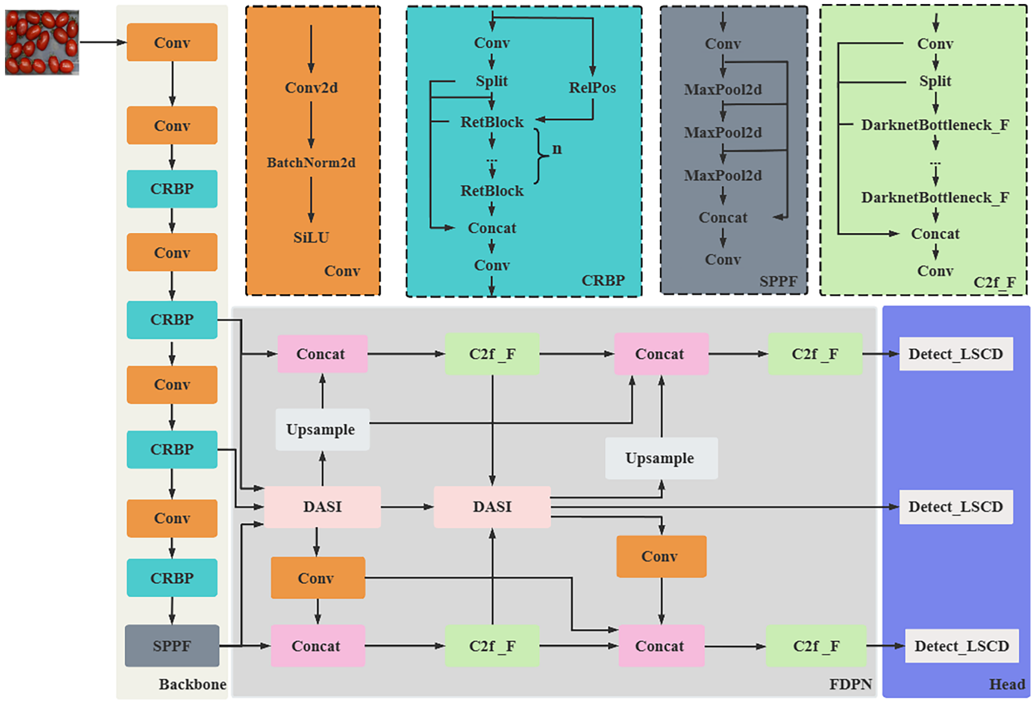

Figure 2: CFL-YOLOv8n model structure.

3.2.1 Context-Aware Residual Block with Relative Position (CRBP) Module in Backbone

In this experiment, the YOLOv8n backbone primarily uses convolution and concatenation operations for feature extraction, but it neglects the positional information of tomatoes in images. Factors like their placement and spatial relationships are vital for judging freshness. The original C2f module struggles to capture these relative-position cues, limiting detection accuracy in complex layouts or occlusions. Additionally, the Bottleneck within C2f is ineffective at modeling long-range dependencies, hindering information integration across different tomato parts and scales.

To resolve these issues, we replace the C2f modules with the Context-aware Residual Block with Relative Position (CRBP) module. CRBP maintains the C2f structure while incorporating relative-position information through RelPos [30] and substituting the Darknet-Bottleneck with RetBlock, as shown in Fig. 3.

In object detection, RelPos enhances the feature map by providing pairwise relative positions, improving the network's ability to perceive spatial relationships and boosting representational power. Its incorporation into CRBP is essential for tomato detection, as it delivers vital location information. The RetBlock employs Manhattan Self-Attention [31] to measure relative positions, explicitly integrating this positional information into attention weights to enhance both localization and recognition.

CRBP follows the C2f pipeline: the input feature map first passes through a Conv layer that doubles the channel count, then splits into two parts. One part is reserved for concatenation, while the other flows through a sequence of RetBlocks, where RelPos infuses positional cues for deeper feature extraction. Finally, processed features and the reserved branch are concatenated and compressed by another Conv layer to produce the output. Each RetBlock receives n RelPos maps of varying spatial ranges. Consequently, the feature processed by a CRBP module can be expressed as follows:

$$CRBP(X) = \frac{\sum_{i=1}^{n}[RetBlock_i(X) + RelPos_i(X)]}{n} \tag{1}$$

here, n is the number of RetBlocks and X denotes the input feature.

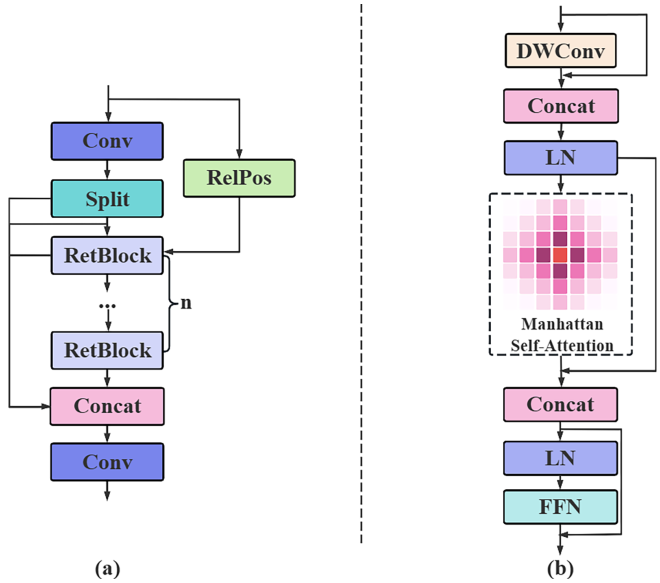

Figure 3: (**a**) Context-aware residual block with relative position (CRBP) module, (**b**) RetBlock structure.

3.2.2 Focus Diffusion Pyramid Network with DASI (FDPN-DASI) Structure in Neck

During tomato freshness assessment, subtle cues such as color and texture are crucial. The original neck of YOLOv8 has significant limitations, as uneven multi-scale feature fusion allows dominant scales to overshadow others, impairing the accuracy of local details. To address these issues, we propose the FDPN-DASI structure. This architecture introduces a Dimension-Aware Selective Integration (DASI) module [32] based on the Focused Diffusion Pyramid Network (FDPN), replacing the original Neck component. As shown in Fig. 4, this redesign transforms fixed, scale-wise summation into a learnable, dimension-aware process, preventing the loss of small defects while retaining the deep semantics needed to identify decayed regions. This enhances the recall and localization of tiny targets and fine-grained features in tomato freshness detection.

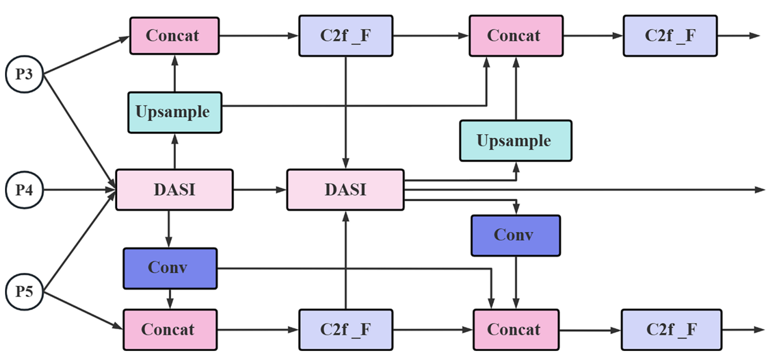

Figure 4: Focus diffusion pyramid network with DASI (FDPN-DASI) structure diagram.

The DASI module is a key component of the Focus Diffusion Pyramid Network with DASI (FDPN-DASI), designed to adaptively select and fuse multi-scale features, enhancing the detection of small targets like spots, decay, wrinkles, and mold. As shown in Fig. 5, this adaptive mechanism allows the model to modify its fusion strategy based on target size and characteristics, improving the localization of small defects in complex backgrounds. The DASI module takes three sets of aligned feature maps as inputs: high-level features F_h, low-level features F_l, and the features from the current layer F_u. Within the module, the process begins by splitting each input feature map along the channel dimension into four equal segments. This operation produces the segmented feature groups $(h_i)_{i=1}^{4}$, $(l_i)_{i=1}^{4}$, and $(u_i)_{i=1}^{4}$, where each segment has dimensions of $R^{H\times W\times \frac{C}{4}}$. An adaptive feature selection step is then performed. For each segment i, a weight α_i is computed by applying a $Sigmoid$ function to u_i, mapping its values to the range $[0, 1]$. Based on the learned weight α_i, the module adaptively selects whether to emphasize the corresponding high-level segment h_i or low-level segment l_i during the feature fusion process. The formulas of the α_i and the u_i are expressed as follows:

$$\alpha_i = Sigmoid\,(u_i) \tag{2}$$

$$u_i = \alpha l_i + (1-\alpha)h_i \tag{3}$$

here, the weight α_i is obtained by applying an activation function to α_i. If $\alpha > 0.5$, the model tends to prioritize low-dimensional features (detailed information); conversely, if $\alpha < 0.5$, it leans toward selecting high-dimensional features (contextual information).

Subsequently, the fused feature segments $u_i^{'}$ are concatenated along the channel dimension to obtain the integrated feature map $F_u^{'}$. The formulas of the $F_u^{'}$ is formulated as follows:

$$F_u^{'} = [u_1^{'}, u_2^{'}, u_3^{'}, u_4^{'}] \tag{4}$$

here, $u_i^{'}$ denotes the selectively aggregated result for each partitioned segment, and $F_u^{'}$ is generated by concatenating $(u_i)_{i=1}^{4}$ along the channel dimension.

Finally, feature enhancement is performed by passing the fused feature map $F_u^{'}$ through a 1×1 convolutional layer to further adjust the channel dimension and strengthen its representational capacity. Subsequently, batch normalization ($BatchNorm$) and a Rectified Linear Unit ($ReLU$) activation function are applied, resulting in the final output feature map $\widehat{F_u}$. The formula is as follows:

$$\widehat{F_u} = \delta(B(Conv(F_u^{'}))) \tag{5}$$

here, the operations $Conv()$, $B()$, and $\delta()$ denote convolution, batch normalization ($BatchNorm$), and the ReLU activation function, respectively, ultimately yielding the output result $\widehat{F_u}$.

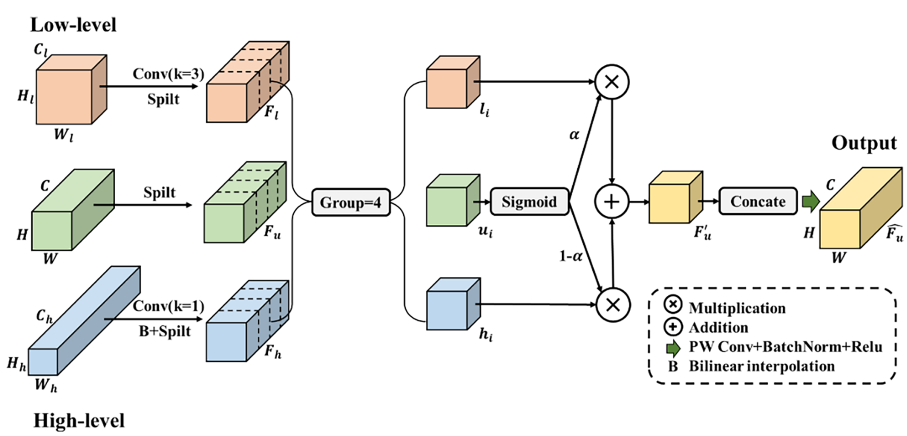

Figure 5: Dimension-aware selective integration (DASI) module.

3.2.3 Lightweight Detection Head LSCD

The YOLOv8 model's original detection head features three independent branches, leading to high parameter counts and limited multi-scale feature sharing. This is problematic for tomato freshness detection on handheld devices, as it increases complexity and can misclassify common features like wrinkles or mold. To improve efficiency and accuracy, this paper proposes a Lightweight Shared Convolutional Detection head (LSCD) to replace the original detection head [33]. The LSCD structure is shown in Fig. 6.

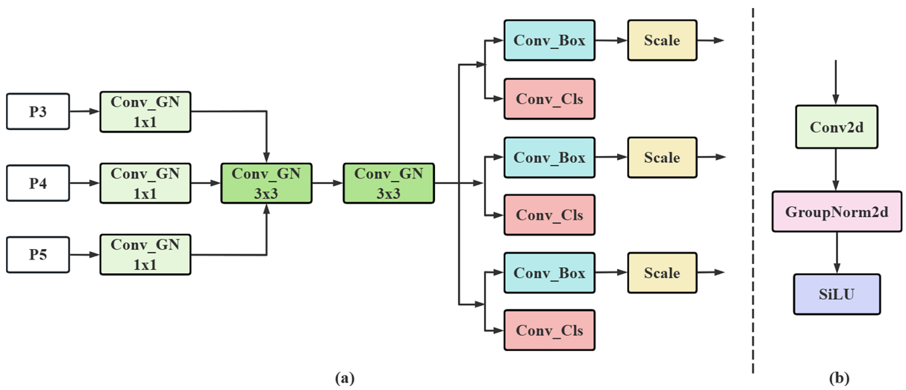

Figure 6: (**a**) Lightweight shared convolutional detection head (LSCD) structure, (**b**) Conv_GN structure.

The LSCD uses a shared convolutional layer design, where some convolutional operations are shared across detection layers. Group Normalization (GN) replaces Batch Normalization (BN) to reduce parameters and computational load, maintaining performance. Initially, input feature maps undergo a Conv_GN layer for transformation, followed by shared convolutions for feature extraction. Independent layers are used for bounding box prediction and category classification. A learnable scale factor fine-tunes feature magnitudes across different detection layers, enhancing adaptability without adding parameters.

The workflow involves processing multi-scale feature maps layer by layer. Each map is aligned to a uniform dimension using a 1×1 Conv_GN convolution. The aligned features are then processed by a shared lightweight convolutional module. After this, features split into two branches: one for regression (bounding box parameters) and one for classification (class probabilities). Finally, the results from both branches are concatenated to produce the final output.

3.2.4 Focaler-IoU Loss Function

The YOLOv8 model uses CIoU Loss and DFL Loss for bounding box regression, and binary cross-entropy (BCE) Loss for classification. CIoU Loss assesses overlapping areas, center-point distances, and aspect ratios between predicted and ground-truth boxes. However, in low-overlap scenarios, CIoU may inadequately represent regression loss, potentially decreasing model accuracy. Additionally, CIoU lacks adaptability to adjust bounding box regression based on the difficulty of individual samples, which might affect the model's generalization capability.

In object detection, Intersection over Union (IoU) measures the localization accuracy of detection boxes by calculating the ratio of the intersection area to the union area between the predicted bounding box and the ground-truth bounding box. The formula is:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{6}$$

here, A and B represent the predicted bounding box and the ground-truth bounding box, respectively. $|A \cap B|$ is the area of intersection, and $|A \cup B|$ is the union of A and B.

The calculation process of CIoU can be expressed by the following formula:

$$CIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v \tag{7}$$

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{8}$$

$$v = \frac{4}{\pi^2}(\arctan \frac{w_{gt}}{h_{gt}} - \arctan(\frac{w}{h}))^2 \tag{9}$$

here, b and b^{gt} represent the center points of the predicted bounding box A and the ground-truth bounding box B, respectively. ρ denotes the Euclidean distance between the two center points. c is the diagonal length of the smallest enclosing rectangle that contains both bounding boxes A and B. α is a balancing factor. v measures the consistency of aspect ratios, as illustrated in Fig. 7.

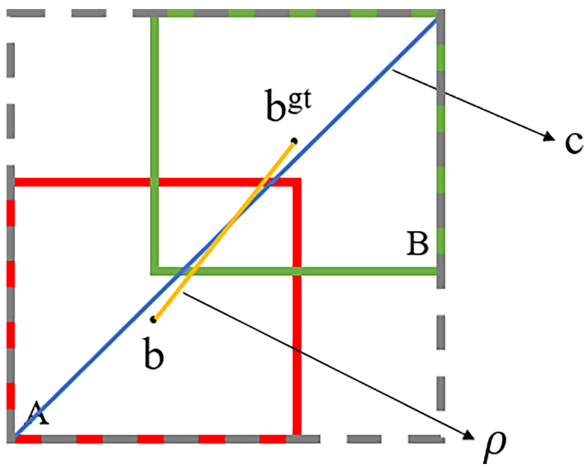

Figure 7: Schematic diagram of CIoU.

In this experiment, Focaler_IoU [34] is employed to replace CIoU. By incorporating a Focal mechanism, Focaler_IoU can adaptively handle different samples and sensitively capture loss variations for bounding boxes with low overlap, thereby enhancing the model's detection accuracy and generalization capability. The calculation formula for Focaler_IoU is as follows:

$$IoU^{focaler} = \begin{cases} 0, & IoU < d \\ \dfrac{IoU - d}{u - d}, & d \ll IoU \ll u \\ 1, & IoU > u \end{cases} \tag{10}$$

here, $IoU^{focaler}$ represents the reconstructed Focaler_IoU, while IoU refers to the original IoU value. The parameters d and u are the predefined lower and upper threshold values, respectively, with $[d, u] \in [0, 1]$. These thresholds partition the IoU value range: when $IoU < d$, $IoU^{focaler}$ is set to 0; when $IoU > u$, $IoU^{focaler}$ is set to 1. By adjusting the values of d and u, $IoU^{focaler}$ can be tailored to emphasize different types of regression samples, allowing for a more refined differentiation of their overlap degrees.

3.3 Experimental Environment and Model Hyperparameters

This study was conducted on a computer running the Windows 11 (64-bit) operating system, equipped with an Intel(R) Core(TM) i9-10900K CPU and an NVIDIA GeForce RTX 3090 GPU. The open-source deep learning framework PyTorch was used, with Python 3.9 as the programming language. The development environment was built using PyCharm and Anaconda, with CUDA 12.1 and cuDNN 8.9 for GPU acceleration. The CFL-YOLOv8n model was adopted for tomato freshness state detection. A total of 1672 images were used for training, and the input images were resized to 640 × 640 pixels. The initial learning rate was set to 0.01, with a weight decay coefficient of 0.0005 and a momentum of 0.937. The SGD optimizer was used during training with a batch size of 16 and a maximum of 200 epochs.

3.4 Model Evaluation Indicators

In this study, the model's detection capability was evaluated using the metrics of mAP@.50, mAP@.50:.95, parameter count (Parameters), and computational complexity (FLOPs). mAP@.50 refers to the mean Average Precision (AP) across all categories when the Intersection over Union (IoU) threshold is set to 0.5. mAP@.50:.95 denotes the mean AP averaged over IoU thresholds ranging from 0.5 to 0.95. Among these, the mAP@.50:.95 criterion is stricter and provides a more comprehensive assessment of model performance. The parameter count represents the total number of trainable parameters in the model, reflecting its complexity and scale. The computational complexity metric measures the computational resources required for model operation, expressed in GFLOPs (Giga Floating-point Operations Per Second), i.e., billions of floating-point operations per second. The formula for calculating mAP is as follows:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP\left(i\right), \tag{11}$$

here N is the total number of categories, and AP_i is the AP value for the i-th category.

4 Experiments and Results

4.1 Comparison Experiment

To further validate the performance of the proposed CFL-YOLOv8n model, this experiment compared it with several mainstream detection models, including YOLOv7-Tiny, YOLOX-Tiny, Faster R-CNN, RT-DETR-r18, YOLOv10n, YOLOv11n, YOLOv12n, YOLOv13n, YOLO26n and the baseline YOLOv8n. To ensure fair comparison, all models were trained and evaluated using the same dataset, input resolution, optimizer, and training settings, without task-specific tuning for individual models. The comparative results are summarized in Table 1.

The experimental results indicate that the CFL-YOLOv8n model achieves superior performance in distinguishing different tomato freshness conditions compared with other models. The improved mAP@.50 and mAP@.50:.95 values indicate that the proposed model can more accurately recognize subtle freshness degradation symptoms, such as spots, wrinkles, and mould regions on tomato surfaces. Compared with the baseline YOLOv8n, CFL-YOLOv8n improves mAP@.50 and mAP@.50:.95 by 3.4% and 3.4%, respectively, demonstrating enhanced capability for evaluating tomato freshness changes under plasma treatment conditions.

In contrast to the more complex and computationally heavy Faster R-CNN, CFL-YOLOv8n reduces parameters to 2.056M and the computational cost to 6.5 GFLOPs. Additionally, when compared to lightweight models like YOLOv10n, CFL-YOLOv8n offers a better balance between parameter size and computational complexity, providing advantages in model efficiency and computational cost under the current experimental setting.

Table 1: Results of the comparison experiments.

Model	mAP@.50	mAP@.50:.95	Parameters (M)	GFLOPs
YOLOv7-Tiny	0.622	0.556	6.015	13.0
YOLOX-Tiny	0.809	0.754	5.035	7.5
Faster-RCNN	0.869	0.850	41.364	193.1
RT-DETR-r18	0.609	0.542	20.093	58.3
YOLOv10n	0.656	0.568	2.696	8.2
YOLOv11n	0.852	0.730	2.583	6.3
YOLOv12n	0.812	0.799	2.520	6.0
YOLOv13n	0.637	0.608	2.460	6.4
YOLO26n	0.826	0.811	2.505	5.8
YOLOv8n	0.866	0.851	3.006	8.1
CFL-YOLOv8n	0.900	0.885	2.056	6.5

4.2 Ablation Experiments

In this experiment, the proposed architectural innovations include the CRBP module for enhanced feature extraction in the backbone, the FDPN-DASI architecture for optimized multi-scale feature fusion in the neck, the Focaler-IoU loss function for improved bounding box regression, and the lightweight LSCD head for efficient detection. To investigate the impact of these improvements on model performance, ablation studies were conducted under consistent experimental conditions and hyperparameter settings. The results of the ablation experiments are presented in Table 2.

Table 2: Results of the ablation experiments.

Model	mAP@.50	mAP@.50:.95	Parameters (M)	GFLOPs
YOLOv8n	0.866	0.851	3.006	8.1
YOLOv8n+①	0.872	0.861	2.698	7.2
YOLOv8n+②	0.877	0.867	3.042	9.4
YOLOv8n+③	0.874	0.861	2.366	6.6
YOLOv8n+①+②	0.884	0.874	2.735	8.5
YOLOv8n+①+②+④	0.892	0.881	2.735	8.5
YOLOv8n+①+②+③+④	0.900	0.885	2.056	6.5

Note: ① is CRBP, ② is FDPN-DASI, ③ is LSCD, ④ is Focaler-IoU, and YOLOv8n+①+②+③+④ is CFL-YOLOv8n.

The experimental results demonstrate that each proposed module improves the model's ability to recognize tomato freshness conditions, with synergistic combinations yielding superior outcomes. As shown in Table 2, individual integration of CRBP, FDPN-DASI, and LSCD improves mAP@.50 by 0.6%, 1.1%, and 0.8%, respectively, indicating enhanced sensitivity to subtle surface features such as spots, wrinkles, and mould regions. Notably, the combined application of CRBP and FDPN-DASI achieves a 1.8% mAP@.50 improvement, suggesting effective complementarity between fine-grained feature extraction and multi-scale feature fusion. Further incorporation of Focaler-IoU and the complete LSCD head yields progressive gains, with the final CFL-YOLOv8n model achieving mAP@.50 of 0.900 and mAP@.50:.95 of 0.885—representing improvements of 3.4% and 3.4% over the baseline, respectively. These results indicate that the proposed model can more accurately identify and localize freshness degradation symptoms in tomatoes under plasma treatment conditions.

Regarding computational efficiency, the proposed model reduces parameters and GFLOPs to 2.056M and 6.5, compared to 3.006M and 8.1 for the baseline YOLOv8n. This 31.6% parameter reduction and 19.8% computational saving, coupled with significant accuracy gains, validate that the architectural innovations successfully balance detection precision with deployment efficiency for tomato freshness detection tasks.

These results further indicate that the performance improvement does not come from a single dominant modification, but from the complementary interaction among feature extraction, multi-scale fusion, lightweight prediction, and localization optimization. In other words, the four components contribute to different stages of the detection process and jointly improve the model's suitability for fine-grained freshness recognition.

To illustrate the impact of different modules on the model, we conducted ablation experiments under the same training environment and hyperparameter settings. The comparison of validation loss between the improved model CFL-YOLOv8n and the baseline model YOLOv8n is shown in Fig. 8a. The loss stabilizes after 200 epochs, with CFL-YOLOv8n exhibiting significantly lower validation loss than YOLOv8n while consistently maintaining a lower level. Therefore, the improved model achieves not only precise classification but also more accurate localization of targets in tomato freshness state detection.

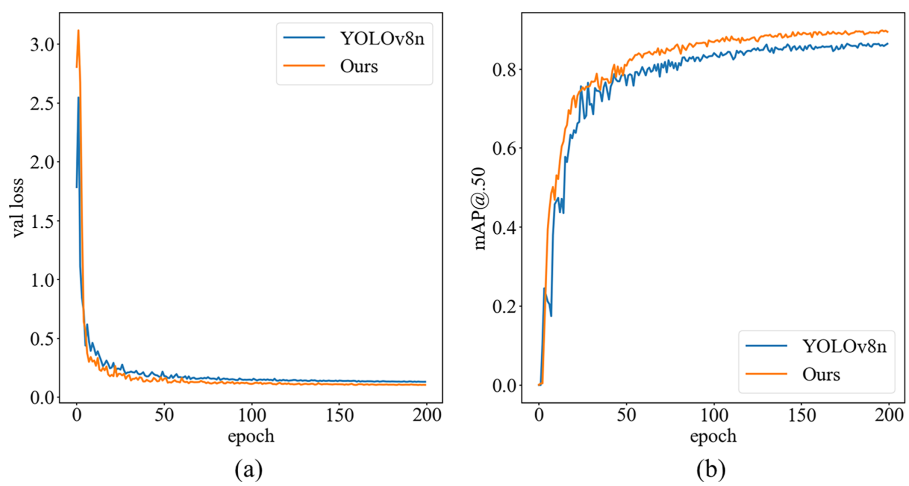

Figure 8: The comparison of the val loss (**a**) and mAP@.50; (**b**) curves of YOLOv8n and our module.

The comparison of the mAP@.50 metric before and after model improvement is presented in Fig. 8b. During the 0–50 epoch phase, YOLOv8n shows considerable fluctuation in loss, whereas the new model incorporating the four modules, although still experiencing some fluctuation, demonstrates markedly smaller variations compared to YOLOv8n. This indicates that the new modules enhance model stability. From 50 epochs, the loss gradually stabilizes, suggesting good model convergence. As shown in the figure, after 200 epochs of iteration, the CFL-YOLOv8n model outperforms the YOLOv8n model in both of the aforementioned metrics.

4.3 *Visualization of Detection Results from the CFL-YOLOv8n Model*

The trained CFL-YOLOv8n model was employed to detect the freshness state categories of tomatoes, with the detection results presented in Fig. 9. The model demonstrates the ability to clearly distinguish between different states, such as fresh, spot, wrinkle, and mouldy, and accurately identifies the freshness condition of each fruit. This indicates that the model exhibits excellent performance in detecting tomato freshness states.

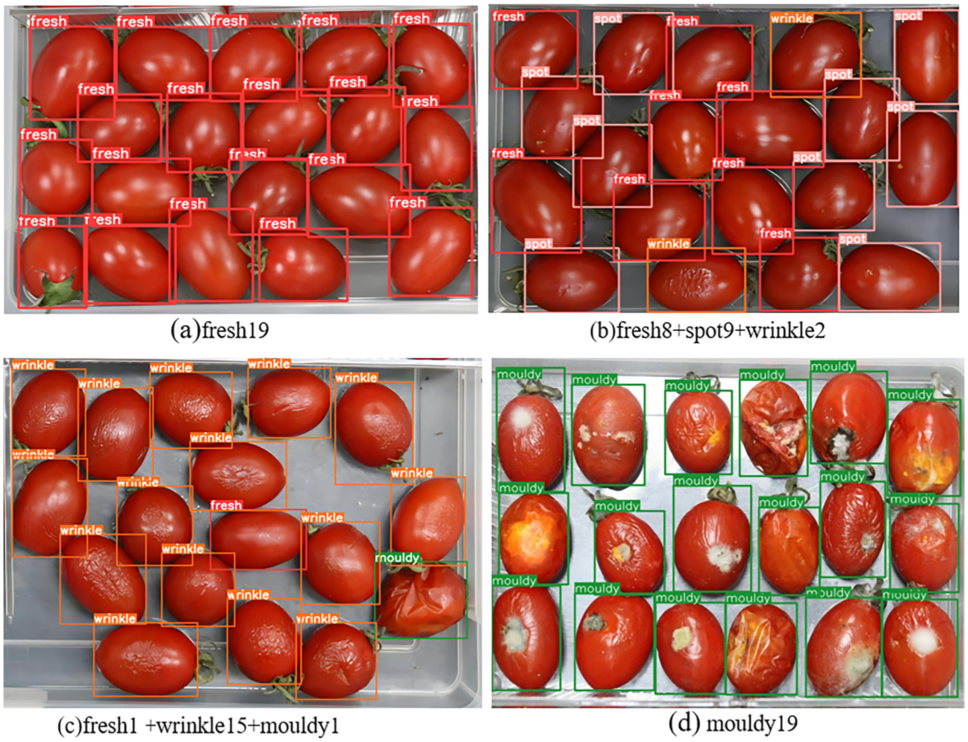

Figure 9: Visualization of CFL-YOLOv8n model test results. (**a**) Fresh 10; (**b**) fresh8+spot9+winkle2; (**c**) fresh 1+winkle15+mouldy1; (**d**) mouldy19.

5 Conclusion

Based on the YOLOv8n baseline network, this study proposes a tomato freshness detection model named CFL-YOLOv8n for assessing the freshness states of plasma-treated tomatoes, which significantly improves both detection accuracy and computational efficiency. The proposed model aims to address the challenge of accurately identifying subtle freshness degradation symptoms during plasma-assisted tomato preservation. Specifically, the CRBP module replaces the original C2f module in the backbone network, enhancing the perception of local defect features on tomato surfaces by integrating relative positional information. The FDPN-DASI structure is introduced during the feature fusion stage to construct an adaptive multi-scale feature fusion architecture, effectively improving the model's ability to recognize tomatoes and surface defects at varying scales. A lightweight detection head, LSCD, is adopted to reduce the parameter count through shared convolution and group normalization operations. Additionally, the Focaler-IoU loss function is employed to optimize the bounding box regression process, enhancing localization accuracy in densely arranged tomato scenarios.

Experimental results demonstrate that the proposed model can effectively distinguish different tomato freshness conditions, including fresh, spot, wrinkle, and mouldy states, thereby providing a reliable visual evaluation for plasma-treated tomato preservation. The parameter count and computational cost are significantly reduced to 2.056M and 6.5 GFLOPs, respectively, indicating good potential for efficient freshness monitoring applications. Furthermore, the tomato plasma freshness preservation evaluation system developed in this study validates the significant freshness preservation-enhancing effect of plasma treatment, providing an effective solution for tomato freshness detection under controlled experimental conditions and offering technical support for evaluating the preservation effects of plasma treatment. The results should be interpreted within the current experimental setting. For further improvement, expanding the dataset scale and validating the model under more complex and real-world conditions are important directions for future work, which will help further evaluate its robustness and practical applicability. In addition, incorporating explainability techniques such as Grad-CAM will be considered in future work to provide further insights into the model's predictions.

Acknowledgement: None.

Funding Statement: This work was partly supported by China Agricultural University Horizontal Project (No. 202405411010197 and 202405510810535).

Author Contributions: Conceptualization, Shaohuang Bian; Methodology, Shaohuang Bian and Feng Huang; Validation, Qinxiu Gao and Shan Su; Formal analysis, Shaohuang Bian and Weifeng Wang; Data curation, Qinxiu Gao; Writing—original draft, Qinxiu Gao, Shaohuang Bian and Shan Su; Writing—review and editing, Feng Huang; Resources, Feng Huang; Supervision, Feng Huang. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the authors.

Ethics Approval: Not applicable.

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

References

1. Ireri D, Belal E, Okinda C, Makange N, Ji C. A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing. Artif Intell Agric. 2019;2(1):28–37. doi:10.1016/j.aiia.2019.06.001.
2. Fass E, Shlomi E, Ziv C, Glickman O, Helman D. Machine learning models based on hyperspectral imaging for pre-harvest tomato fruit quality monitoring. Comput Electron Agric. 2025;229:109788. doi:10.1016/j.compag.2024.109788.
3. Vocaturo E, Rani G, Dhaka VS, Zumpano E. AI-driven agriculture: opportunities and challenges. In: 2023 IEEE International Conference on Big Data (BigData). 2023 [cited 2026 Apr 13]. Available from: https://ieeexplore.ieee.org/abstract/document/10386314.
4. Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: a survey. Comput Electron Agric. 2018;147(2):70–90. doi:10.1016/j.compag.2018.02.016.
5. Kundu N, Rani G, Dhaka VS, Gupta K, Nayaka SC, Vocaturo E, et al. Disease detection, severity prediction, and crop loss estimation in MaizeCrop using deep learning. Artif Intell Agric. 2022;6(11):276–91. doi:10.1016/j.aiia.2022.11.002.
6. Yang G, He Y, Yang Y, Xu B. Fine-grained image classification for crop disease based on attention mechanism. Front Plant Sci. 2020;11:600854. doi:10.3389/fpls.2020.600854.
7. Tian Q, Zhao G, Yan C, Yao L, Qu J, Yin L, et al. Enhancing practicality of deep learning for crop disease identification under field conditions: insights from model evaluation and crop-specific approaches. Pest Manag Sci. 2024;80(11):5864–75. doi:10.1002/ps.8317.

8. Wan P, Toudeshki A, Tan H, Ehsani R. A methodology for fresh tomato maturity detection using computer vision. Comput Electron Agric. 2018;146:43–50. doi:10.1016/j.compag.2018.01.011.
9. da Costa AZ, Figueroa HEH, Fracarolli JA. Computer vision based detection of external defects on tomatoes using deep learning. Biosyst Eng. 2020;190(3):131–44. doi:10.1016/j.biosystemseng.2019.12.003.
10. Zheng T, Jiang M, Li Y, Feng M. Research on tomato detection in natural environment based on RC-YOLOv4. Comput Electron Agric. 2022;198(10):107029. doi:10.1016/j.compag.2022.107029.
11. Liu J, Pi J, Xia L. A novel and high precision tomato maturity recognition algorithm based on multi-level deep residual network. Multimed Tools Appl. 2020;79(13):9403–17. doi:10.1007/s11042-019-7648-7.
12. Mukhiddinov M, Muminov A, Cho J. Improved classification approach for fruits and vegetables freshness based on deep learning. Sensors. 2022;22(21):8192. doi:10.3390/s22218192.
13. Valentino F, Wawan Cenggoro T, Natanael Elwirehardja G, Pardamean B. Energy-efficient deep learning model for fruit freshness detection. IAES Int J Artif Intell. 2023;12(3):1386. doi:10.11591/ijai.v12.i3.pp1386-1395.
14. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell. 2017;39(6):1137–49. doi:10.1109/TPAMI.2016.2577031.
15. He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. arXiv:1703.06870. 2017.
16. Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21–26; Honolulu, HI, USA. p. 6517–25. doi:10.1109/CVPR.2017.690.
17. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al. SSD: single shot MultiBox detector. In: Computer vision—ECCV 2016. Cham, Switzerland: Springer International Publishing; 2016. p. 21–37. doi:10.1007/978-3-319-46448-0_2.
18. Redmon J, Farhadi A. YOLOv3: an incremental improvement. arXiv:1804.02767. 2018.
19. Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: optimal speed and accuracy of object detection. arXiv:2004.10934. 2020.
20. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21–26; Honolulu, HI, USA. p. 936–44. doi:10.1109/CVPR.2017.106.
21. Liu S, Qi L, Qin H, Shi J, Jia J. Path aggregation network for instance segmentation. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 8759–68. doi:10.1109/cvpr.2018.00913.
22. Tan M, Pang R, Le QV. EfficientDet: scalable and efficient object detection. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. p. 10778–87. doi:10.1109/cvpr42600.2020.01079.
23. Guo C, Fan B, Zhang Q, Xiang S, Pan C. AugFPN: improving multi-scale feature learning for object detection. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. p. 12592–601. doi:10.1109/cvpr42600.2020.01261.
24. Ge Z, Liu S, Wang F, Li Z, Sun J. YOLOX: exceeding YOLO series in 2021. arXiv:2107.08430. 2021.
25. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861. 2017.
26. Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: an extremely efficient convolutional neural network for mobile devices. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 6848–56. doi:10.1109/CVPR.2018.00716.
27. Zheng Z, Wang P, Liu W, Li J, Ye R, Ren D. Distance-IoU loss: faster and better learning for bounding box regression. Proc AAAI Conf Artif Intell. 2020;34(7):12993–3000. doi:10.1609/aaai.v34i07.6999.
28. Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I, Savarese S. Generalized intersection over union: a metric and a loss for bounding box regression. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 658–66. doi:10.1109/CVPR.2019.00075.
29. Zheng Z, Wang P, Ren D, Liu W, Ye R, Hu Q, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. arXiv:2005.03572. 2021.

30. Shaw P, Uszkoreit J, Vaswani A. Self-attention with relative position representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human LanguageTechnologies, Volume 2 (Short Papers). Stroudsburg, PA, USA: ACL; 2018. p. 464–8. doi:10.18653/v1/n18-2074.
31. Katharopoulos A, Vyas A, Pappas N, Fleuret F. Transformers are RNNs: fast autoregressive transformers with linear attention. In: Proceedings of the 37th International Conference on Machine Learning; 2020 Jul 13–18; Virtual. p. 5156–65.
32. Xu S, Zheng S, Xu W, Xu R, Wang C, Zhang J, et al. HCF-net: hierarchical context fusion network for infrared small object detection. arXiv:2403.10778. 2024.
33. Li C, Li L, Jiang H, Weng K, Geng Y, Li L, et al. YOLOv6: a single-stage object detection framework for industrial applications. arXiv:2209.02976. 2022.
34. Zhang H, Zhang S. Focaler-IoU: more focused intersection over union loss. arXiv:2401.10525. 2024.