Journal of Artificial Intelligence and Technology, 2026, 6, 567-576
https://doi.org/10.37965/jait.2026.0971

RESEARCH ARTICLE

Intelligent Detection of Fruit Decay Severity Using YOLO with Hybrid Attention Mechanisms

Hao Peng[1,2] and Mideth Abisado[1]
[1]College of Computing and Information Technologies, National University, Manila, Philippines
[2]School of Information and Electromechanical Engineering, Hunan International Economics University, Changsha, China

(Received 16 October 2025; Revised 17 November 2025; Accepted 24 December 2025; Published online 24 June 2026)

***Abstract*:** Accurate detection of fruit decay severity is crucial in agricultural supply chains for minimizing economic losses, ensuring food safety, and optimizing storage–transportation workflows. Traditional detection methods relying on manual inspection or simple image processing suffer from low efficiency and poor robustness. This study proposes an improved YOLO (You Only Look Once)-based method for fruit decay severity detection. A tailored data augmentation strategy is employed to enhance model adaptability to complex scenarios. Furthermore, Squeeze-and-Excitation (SE) channel attention and Convolutional Block Attention Module (CBAM) hybrid attention modules are integrated into the YOLOv8 backbone to strengthen feature learning for decay regions. Experiments on public fruit datasets compare the performance of YOLOv5, YOLOv8, and their attention-enhanced variants (YOLOv8 + SE and YOLOv8 + CBAM). The results demonstrate that the YOLOv8 + CBAM model achieves 92.3% mAP@0.5, a 3.9% improvement over the baseline, with superior performance in small-target detection and scenarios with complex background interference. This study innovatively combines multi-scale feature fusion and attention mechanisms, providing an efficient and precise solution for agricultural automation detection.

***Keywords*:** attention mechanism; CBAM module; fruit decay detection; object detection; YOLO

I. INTRODUCTION

A. RESEARCH BACKGROUND

In the context of global agricultural modernization and intelligent transformation, the quality detection and grading technology of fruits, as critical components of the agricultural supply chain, have become focal points in agricultural management. Fruit decay, a key factor affecting agricultural product quality, not only directly leads to economic losses but also poses potential risks to food safety. FAO reports that approximately 14% of food ($400 billion annually) is lost after harvest, with fruits particularly vulnerable to post-harvest decay [1]. Furthermore, fruit decay can also affect consumer purchasing intentions and market confidence, further exacerbating economic losses. Traditional methods for fruit decay detection primarily rely on manual visual inspection or simple image processing techniques (threshold segmentation and texture feature analysis). However, these methods exhibit significant limitations: manual inspection is inefficient and prone to subjective variations, while simple image processing techniques struggle to cope with complex scenarios involving illumination variations, background interference, and minute decay regions. Therefore, the development of an efficient and precise automatic detection technology for fruit decay severity is of paramount importance for enhancing agricultural supply chain efficiency, ensuring food safety, and promoting sustainable agricultural development.

Agricultural Supply Chain Efficiency: Fruit decay stands as a pivotal factor contributing to significant disruptions within agricultural supply chains. According to a study conducted by Porat [2], storage losses attributable to fungal decay in "Gala" and "Fuji" apples, which were stored under commercial conditions in Brazil, escalated to a substantial range of 18.3% to 26.6%. These inefficiencies not only impose a direct financial burden by inflating logistics costs but also act as a barrier, restricting market access for producers and suppliers. The ramifications extend beyond immediate economic losses, as they impede the seamless flow of goods from farms to consumers, thereby affecting the overall competitiveness and resilience of the agricultural sector. Moreover, the accumulation of decayed fruits in storage facilities can lead to a deterioration of storage conditions, further exacerbating the problem and potentially causing cross-contamination among healthy produce.

Food Safety Risks: The presence of decayed fruits poses a severe threat to food safety, as they may harbor carcinogenic substances such as aflatoxins. A comprehensive survey undertaken in Turkey by Kayahan *et al.* [3] unveiled that approximately 57% of dried fruit samples, including figs and raisins, exhibited aflatoxin levels surpassing the stringent limits set by the European Union. Prolonged consumption of such contaminated fruits can markedly elevate the risk of developing liver cancer, underscoring the urgent need for improved detection and prevention strategies. This finding not only highlights the gravity of the food safety issue but also casts a spotlight on the inherent delays and limitations associated with traditional detection methods, which often rely on manual inspection or rudimentary chemical tests. These conventional approaches are not only time-consuming and labor-intensive but also prone to human error and variability, thereby compromising the accuracy and reliability of the results.

Sustainability Imperatives: Automated detection systems can reduce labor dependency and resource waste. After introducing machine vision-based automatic decay identification equipment

Corresponding author: Mideth Abisado (e-mail: mbabisado@national-u.edu.ph).

into the palm fruit sorting system in Kenya, the efficiency of manual sorting increased by 68%, the waste rate decreased by 22%, and annual labor costs were saved by approximately 25% [4].

By integrating the YOLO (You Only Look Once) model with attention mechanisms, this study aims to develop an efficient and precise automated fruit decay detection technology, providing critical technical support for optimizing agricultural supply chains, ensuring food safety, and advancing sustainable agricultural development.

B. DEVELOPMENT OF OBJECT DETECTION TECHNOLOGY

With the continuous advancement of machine learning technologies, it has increasingly become a vital research approach for template detection, owing to its powerful capabilities in automated feature learning, pattern recognition, and adaptation to complex data structures. Machine learning enables the extraction of intricate features from raw data, facilitating the identification of predefined templates even under noisy or variable conditions. For example, machine learning and deep learning models have been applied to fruit defect and decay detection in post-harvest processing, demonstrating their utility in extracting subtle visual features of rot or damage even under challenging imaging conditions [5]. Similarly, object detection frameworks such as YOLO have been used for multi-class fruit ripeness and defect classification under real-world harvest/field conditions, highlighting ML's role in assessing the quality and decaying status of fruits [6,7]. In addition, lightweight versions of YOLO architectures have been introduced for low-power edge computing in agricultural fruit monitoring, enabling robust decay/defect detection in fruit supply chain environments [8], while the fusion of UNet and Swin Transformer architectures has advanced lesion segmentation in medical imaging by accurately detecting biological templates [9]. These applications underscore machine learning's versatility and transformative potential in template detection across diverse domains.

In recent years, deep learning-based object detection technology has achieved revolutionary advancements, with the YOLO series algorithms standing out due to their exceptional balance of speed and accuracy, becoming a focal point in both industrial and academic research. Since the inception of YOLOv1 in 2015, the series has undergone multiple iterative optimizations:

YOLOv5 and YOLOv8, as widely adopted detectors in industrial applications, exhibit distinct trade-offs in terms of speed, accuracy, and feature fusion. In the context of fruit decay detection, YOLOv8 demonstrates significant advantages for detecting small-sized and ambiguous targets through its improved anchor-free prediction approach [10–12].

Meanwhile, attention mechanisms, owing to their unique advantages in feature enhancement and noise suppression, have been widely applied in the field of object detection. Typical modules are given as follows:

Squeeze-and-Excitation (SE) Module: By explicitly modeling inter-channel dependencies, this module adaptively recalibrates channel-wise feature responses, thereby strengthening the learning of target region features.

Convolutional Block Attention Module (CBAM): Combining channel attention and spatial attention mechanisms, this module captures critical features in both channel and spatial dimensions through global average pooling and max pooling operations, enabling more refined feature optimization.

The integration of these attention modules has significantly improved model performance in complex scenarios, offering new technological pathways for specialized tasks such as fruit decay detection.

C. RESEARCH GAPS AND CONTRIBUTIONS

Despite the remarkable achievements of existing YOLO models in general object detection, they still face numerous challenges in the specific task of fruit decay detection:

1. Small-Target Miss Detection: Early-stage fruit decay often manifests as minute spots or regions, which traditional models struggle to effectively capture due to limitations in receptive field size or feature representation capabilities.
2. Complex Background Interference: Complex fruit surface textures, variable illumination conditions, and occlusions lead to increased false detection rates, as models may misidentify background regions as decayed areas.
3. Ambiguity in Decay Severity Grading: The subtle feature differences between mild and moderate decay necessitate models with enhanced feature discrimination capabilities to achieve precise grading.

To address these challenges, this study proposes the following innovations:

1. Tailored Data Augmentation Strategies: Customized techniques such as Mosaic, MixUp, and small-target copy-paste are employed to enhance model adaptability to multi-scale targets, complex backgrounds, and illumination variations, specifically tailored to the characteristics of fruit decay detection scenarios.
2. YOLOv8-Based Attention Mechanism Fusion: SE channel attention modules are embedded into the YOLOv8 backbone to strengthen the learning of decay region features. Simultaneously, CBAM hybrid attention modules are integrated post-feature pyramid to suppress background noise and enhance feature representation capabilities.

Multi-Model Comparative Experiments and Performance Analysis: Systematic evaluations of YOLOv5, YOLOv8, and their attention-enhanced variants (YOLOv8 + SE and YOLOv8 + CBAM) are conducted on public fruit decay datasets. The effectiveness of the proposed improvements is validated through a combination of quantitative metrics (mAP, precision, and recall) and qualitative analyses (visualization of detection results).

D. PAPER ORGANIZATION

The subsequent sections of this paper are organized as follows:

Section II presents the related work. Section III provides a detailed introduction to the dataset construction process, the specific implementation methods of data augmentation strategies, and the technical details of the model improvement schemes. Section IV elaborates on the principles of experimental design, the rationale for selecting evaluation metrics, and the experimental results of multi-model performance comparisons. Section V conducts an in-depth analysis of the experimental results, discussing the performance disparities of the models under different scenarios and their limitations. Section VI summarizes the entire research work, extracts the main innovations, and proposes future research directions and prospects.

II. RELATED STUDIES

A. AGRICULTURAL PRODUCT FRESHNESS EVALUATION TECHNIQUES

Conventional Approaches: Historical investigations predominantly employed colorimetric and textural analysis frameworks. For example, fruit ripeness assessment was conducted via epidermal pigmentation monitoring, while textural characterization utilized algorithms such as Gray-Level Co-occurrence Matrix (GLCM) for discriminating between intact and deteriorated tissue regions. Nevertheless, these methodologies demonstrate limited adaptability when confronted with heterogeneous backgrounds, nonuniform illumination patterns, or partial occlusion by adjacent objects. Their performance degrades significantly in detecting subtle morphological anomalies, including localized fungal colonies or epidermal micro-deformations. In recent years, hyperspectral imaging combined with deep learning has offered a robust alternative. Varga *et al.* [13] demonstrated high-accuracy ripeness classification outperforming conventional methods by exploiting spectral features beyond visible color.

Neural Network Advancements: The advent of convolutional neural networks (CNNs) has substantially elevated detection precision in recent epochs. Architectures like ResNet and EfficientNet have demonstrated multi-tiered maturity classification capabilities across diverse produce types, including apples and bananas [14]. Within the object detection domain, the YOLO family has garnered considerable traction due to its temporal efficiency: YOLOv3 pioneered defect localization through bounding box regression mechanisms; subsequent YOLOv5 iterations prioritized model parsimony to facilitate edge deployment. Modern research trajectories have augmented YOLOv8 through attention module integration, such as coordinate attention or SE/IRMB mechanisms, thereby enhancing feature extraction fidelity for low-contrast or occluded targets, as evidenced by enhanced detection of small-sized fruits like plums in cluttered orchard settings [15].

B. ARCHITECTURAL EVOLUTION OF YOLO-BASED DETECTORS

YOLOv5 establishes an equilibrium between computational velocity and diagnostic accuracy via its modular architecture incorporating CSP-DarkNet feature extractors and adaptive anchor dimension calibration, solidifying its industrial relevance [16]. YOLOv6 introduces bidirectional feature pyramid networks to bolster multi-scale target recognition, while YOLOv7 employs convolutional reparameterization techniques for model compression [17].

The contemporary YOLOv8 variant adopts anchor-free detection paradigms coupled with dynamic label assignment protocols, mitigating dependency on predefined anchor geometries and exhibiting superior generalization across complex scenarios such as densely packed fruit clusters [18]. YOLOv8 achieves enhanced detection performance through three core innovations: first, it employs a CSPDarknet-based backbone with C2f modules that replaced YOLOv5's C3 blocks for improved gradient flow and feature reuse. This is analogous to architecture advances seen in improved lightweight YOLO variants.

Second, the neck network utilizes Path Aggregation Network (PANet) for multi-scale feature fusion [19]. The integration of Mosaic and MixUp adaptive data synthesis strategies significantly improves model robustness against photometric variations and occlusion scenarios. Finally, by adopting Weighted Intersection over Union (WIoU) loss functions instead of traditional CIoU, it effectively mitigates missed detection issues in crowded fruit arrangements [20]. These three improvements collectively optimize the model's adaptability to complex agricultural detection environments.

III. DATASET CONSTRUCTION AND AUGMENTATION STRATEGIES

A. DATASET CONSTRUCTION

This study combines a public fruit decay dataset with a self-constructed dataset to create a diverse dataset encompassing various fruit types and decay severities. Details are as follows:

1). DATA SOURCES:. Public Dataset: The Fruits-360 extended dataset is selected as the core data source, containing high-resolution images of apples, tomatoes, and bananas, covering different decay stages and background conditions.

Self-Constructed Dataset: To supplement the diversity of the public dataset, additional fruit images are collected through actual acquisition and web crawling techniques, encompassing various lighting conditions, shooting angles, and decay morphologies.

2). DATA ANNOTATION STANDARDS. Decay Region Bounding Box Annotation: The LabelImg tool is employed to precisely annotate decay regions in each image, generating PASCAL VOC-formatted XML label files that record bounding box coordinates.

3). DATA COLLECTION RESULTS. Decayed Fruit Recognition Dataset: This comprises 3,304 images, each accompanied by an XML-formatted bounding box annotation file and a TXT-formatted decay severity label file.

Category Division: The dataset covers three fruit types—apples, tomatoes, and bananas—each further categorized into "good fruit" (good_apple/good_tomato/good_banana) and "decayed fruit" (bad_apple/bad_tomato/bad_banana) classes, with the latter including samples of varying decay severities, as shown in Fig. 1.

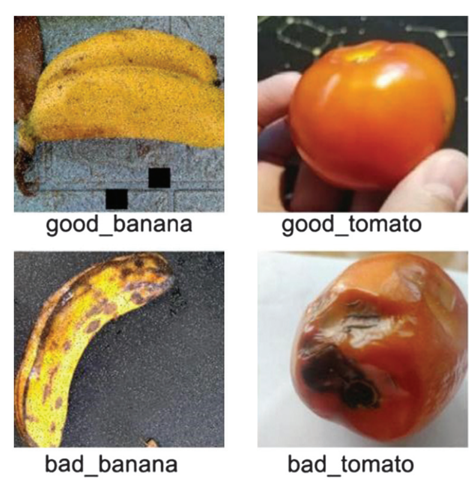

Fig. 1. Dataset of different rotten fruits.

Fig. 2. Data augmentation.

B. DATA AUGMENTATION STRATEGIES

To enhance the model's adaptability to complex scenarios and generalization performance, this study employs data augmentation strategies encompassing two categories of methods: spatial transformations and brightness adjustments.

1). BASELINE AUGMENTATIONS. Spatial Transformations: Images undergo horizontal/vertical flipping or random rotation within the range of -15° to 15°, simulating varied camera angles during acquisition.

Brightness Adjustment: Image brightness is randomly modulated by ±20%–50%, improving the model's robustness against illumination variations.

2). EMPIRICAL VALIDATION. By comparing original images with their augmented counterparts (as illustrated in Fig. 2), the impact of data augmentation on image content preservation, object scale consistency, and background complexity diversification can be visually assessed. This systematic augmentation framework effectively expands sample diversity while maintaining semantic integrity, thereby mitigating overfitting risks in training.

IV. MODEL METHODS AND IMPROVEMENT DESIGN

A. BASELINE MODEL ARCHITECTURES

This study designs and optimizes fruit decay severity detection models based on the YOLO series algorithms, with a focus on YOLOv5 and YOLOv8 as baseline models, and introduces structural improvements tailored to the characteristics of fruit decay detection tasks. Although YOLOv6 and YOLOv7 introduce notable improvements, preliminary tests revealed that YOLOv8 consistently outperformed them on fruit decay datasets, particularly in small-target detection and real-time inference speed. Additionally, YOLOv8's anchor-free mechanism and support for attention integration made it more adaptable to the hybrid attention framework in this study.

1). YOLOv5 MODEL ARCHITECTURE:. YOLOv5 adopts a hierarchical design with the following core components:

Backbone: CSPDarknet53
An improved version of Darknet53 incorporates Cross-Stage Partial (CSP) connections, which reduce computational complexity and enhance feature extraction capabilities through gradient flow segmentation.

Neck: PANet Feature Pyramid
This employs the PANet, enhancing multi-scale feature representation through bidirectional (bottom-up and top-down) path augmentation.

Head: Anchor-Based Detection Head
This utilizes preset anchor boxes for bounding box regression and classification, enabling target localization and recognition.

2). YOLOv8 MODEL ARCHITECTURE:. The network architecture of YOLOv8 is distinctive. The neck network utilizes an enhanced PANet feature fusion mechanism, enabling adaptive cross-scale feature fusion through learnable weights. In the detection head, YOLOv8 adopts a decoupled head design, separating classification and regression tasks to enhance detection accuracy and efficiency. The overall optimized structure allows YOLOv8 to excel in object detection tasks.

In decay detection tasks, YOLOv8's anchor-free design allows more flexible localization of irregular decay shapes, while the decoupled head improves classification accuracy in differentiating mild and moderate decay. Its robustness against variable lighting and background textures also makes it ideal for orchard and storage scenarios.

B. ATTENTION MECHANISM IMPROVEMENT SCHEMES

To further enhance the model's sensitivity to decay region features, this study integrates two attention mechanism modules into YOLOv8 and analyzes their functional mechanisms.

The SE module dynamically weights feature channels by explicitly modeling inter-channel dependencies, amplifying decay-related channels while suppressing irrelevant background channels [16], as shown in Fig. 3.

SE modules enhance feature representability through channel attention but have limited utilization of spatial information [21,22].

The CBAM module first generates a channel attention map, followed by a spatial attention map, which are element-wise multiplied and applied to the input features. This joint optimization in channel and spatial dimensions effectively suppresses background noise and focuses on decay regions, as shown in Fig. 4.

CBAM modules, combining channel and spatial attention, more accurately localize decay regions, particularly in complex backgrounds [23,24].

C. IMPROVING THE YOLO MODEL USING ATTENTION MECHANISMS

In this experiment, in addition to adopting YOLOV5 and YOLOV8 as the baseline experimental models, respectively, we also attempted to incorporate two different attention mechanisms—SE and CBAM—into the YOLOV8 model to optimize its training performance. The SE attention mechanism primarily enhances feature representation by modeling the dependencies between channels [25], while the CBAM attention mechanism combines channel attention and spatial attention, enabling it to simultaneously focus on important channels and spatial regions within the feature map [26].

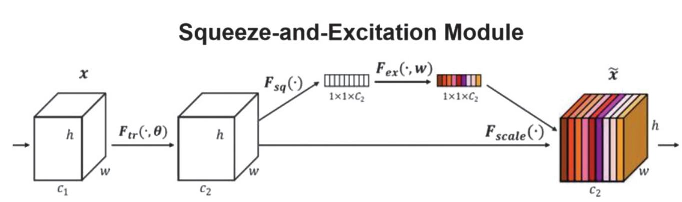

Fig. 3. SE model.

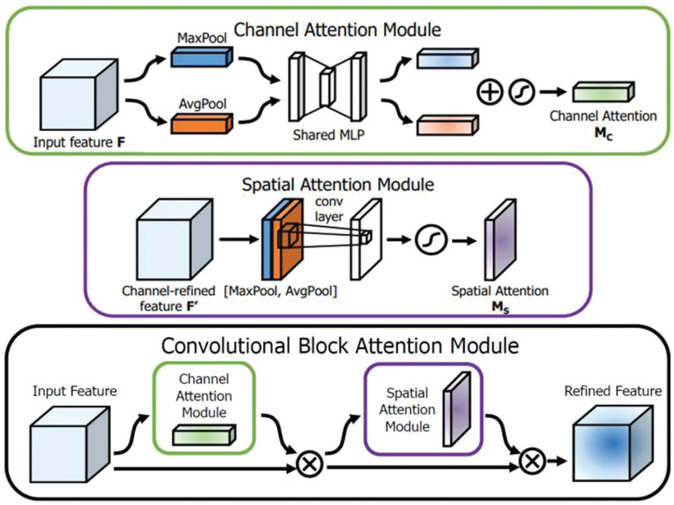

Fig. 4. CBAM model.

To enhance the detection accuracy of the YOLOv8 model for fruit rot characteristics, this study developed two independent improved models based on the SE attention mechanism and the CBAM, respectively. Specifically, the SE-based improved model addresses the challenge of multi-scale target detection in fruit rot analysis (ranging from tiny mildew spots to large-scale rotten areas) by integrating an SE attention module into the Neck section of YOLOv8 (after the C2f modules in Upsample, P3, P4, and P5 layers). This module compresses spatial information through global average pooling and learns nonlinear inter-channel relationships via fully connected layers, enabling adaptive adjustment of channel weights across feature maps of different scales. For instance, it enhances channels associated with small rot spots in the P3 layer while suppressing background noise channels in the P5 layer, thereby optimizing the modeling of cross-scale channel correlations. The improved model structures with the addition of the SE attention mechanisms are shown in Fig. 5.

In contrast, the CBAM-based improved model targets fine-grained texture features of fruit rot areas (e.g., mildew patches and

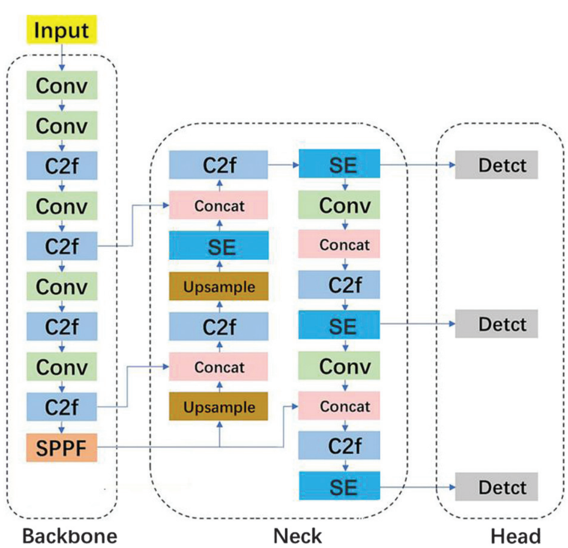

Fig. 5. YOLOv8 + SE model.

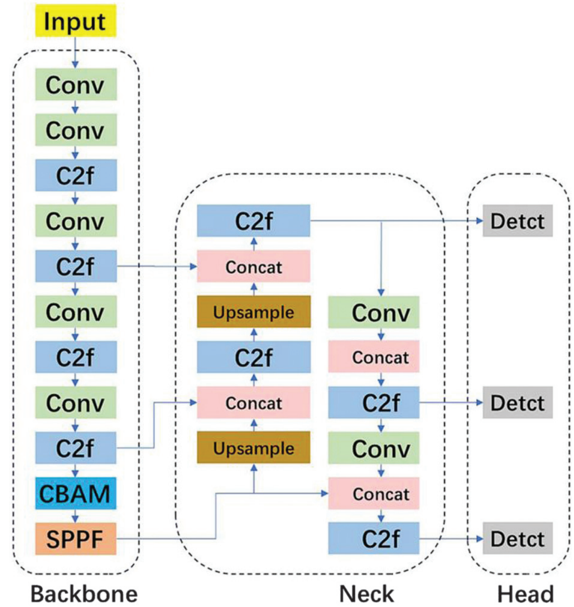

Fig. 6. YOLOv8 + CBAM model.

soft rot spots) by embedding a CBAM attention module at the end of YOLOv8's backbone network (after the final C2f module). This module utilizes a spatial attention branch to focus on local boundaries of rotten regions (e.g., spot edges) and a channel attention branch to amplify weights of channels highly correlated with rot characteristics (e.g., color and texture channels), thereby improving the discriminative capability of high-dimensional feature maps. The improved model structures with the addition of the CBAM attention mechanisms are shown in Fig. 6.

V. EXPERIMENTS

A. EXPERIMENTAL SETUP

Hardware Environment: Experiments were conducted on an NVIDIA RTX 3090 GPU (24GB VRAM), paired with an Intel i9-10900K CPU and 64GB DDR4 RAM.

Training Strategies: Learning Rate Adjustment: Cosine annealing was employed, with an initial learning rate of 0.001 dynamically adjusted during training.

Random Seed Limitation: This experiment employed a fixed random seed for model initialization. While this approach has certain limitations—such as the potential for different initializations to affect model convergence—it adheres to conventional practices in rapid prototyping development.

Multi-Scale Training: Input image sizes were randomly adjusted (640×640 to 1280×1280) to enhance robustness against scale variations.

Mixed Precision Training: FP16 mixed precision was enabled to accelerate convergence and reduce VRAM usage.

B. EXPERIMENTAL PARAMETERS

Dataset division: 70% is allocated for training data, and 30% for validation data.

Batch Size: 16 (balanced for GPU memory efficiency and gradient stability).

Training Epochs: 300 (convergence verified via mAP@0.5 plateau).

Optimizer: AdamW—preferred for its decoupled weight decay and robustness in vision tasks.

Weight Decay: 0.01 (L2 regularization to prevent overfitting).

Momentum: N/A (AdamW does not use momentum, replaced by adaptive learning rate).

Initial Learning Rate: 0.001 (cosine annealing adjusted dynamically during training).

Learning Rate Scheduler: Cosine annealing (warm restarts at epochs 75, 150, 225).

C. EVALUATION METRICS

Core Metric:
mAP@0.5 (mean average precision with IoU threshold 0.5). Calculates the average precision (AP) for all classes at an IoU threshold of 0.5, reflecting the model's comprehensive detection capability.
Auxiliary Metrics:

1) Precision: It is to measure the model's ability to avoid false positives (e.g., misidentifying shadows as decay). The formula is given as:

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (1)$$

where TP (true positive) is the number of correctly detected decayed regions and FP (false positive) includes misidentifying healthy regions as decayed or duplicate detections of the same decayed region.

2) Recall: It is to evaluate the model's probability of missing detections (e.g., ignoring small mold spots). The formula is given as:

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (2)$$

where FN (false negative) refers to undetected true decayed regions.

As the harmonic average of precision and recall, it is more suitable for evaluating class-imbalanced scenarios (in this dataset, the "Severe Decay" class accounts for a smaller proportion).

VI. EXPERIMENTAL RESULTS

A. QUANTITATIVE EVALUATION RESULTS

To better detect the freshness and decay of fruits, this study employs four models: YOLOv5, YOLOv8, YOLO + SE, and YOLO + CBAM. All models demonstrated satisfactory performance during the experimental training process, and the comparison of their evaluation metrics is presented in Table I:

Table I. Comparison of optimal experimental results

Model	mAP@0.5	Precision	Recall
YOLOv5	85.2	88.7	82.3
YOLOv8	88.4	91.2	86.1
YOLOv8 + SE	89.7	92.1	87.3
YOLOv8 + CBAM	92.3	93.4	89.2

Based on the experimental results table, we can summarize the comparison of evaluation metric data for each model as follows:

- YOLOv5 vs. YOLOv8: YOLOv8 achieves a 3.2% improvement in mAP@0.5, attributed to its anchor-free mechanism and hybrid backbone design.
- YOLOv8 + SE vs. YOLOv8 + CBAM: CBAM further improves mAP@0.5 by 2.6%, with a more significant advantage in mAP@0.5:0.95 (+3.5%), demonstrating the superior adaptability of hybrid attention mechanisms to complex scenarios.

As shown in the comparison, the YOLO + CBAM model achieved the best experimental results.

B. ABLATION STUDY ON DATA AUGMENTATION AND ATTENTION MODULES

To validate the contribution of each component, we conducted a systematic ablation study covering data augmentation and attention mechanisms. As shown in Table II, removing augmentation from YOLOv8 results in a noticeable performance drop (mAP@0.5 = 84.1%). Traditional baseline augmentation improves mAP@0.5 to 86.4%, while applying Mosaic and MixUp individually yields gradual improvements. The full augmentation strategy increases mAP@0.5 to 88.4% and recall to 86.1%, demonstrating enhanced generalization to small and occluded decay spots. Introducing SE and CBAM further improves mAP@0.5 by 1.3% and 3.9%, respectively, with CBAM achieving the best overall performance, especially in mAP@0.5:0.95 (+5.2%), highlighting its advantage in fine-grained decay severity recognition (Table II).

C. COMPARISON OF CURVE

To better compare the differences and changes among various models, this paper will present the experimental results of each model through the following line graphs.

Firstly, it shows the comparison of the mAP50 curve changes for each model, as depicted in Figs. 7–10:

The mAP50 (mean average precision at 50% IoU) is a commonly used evaluation metric in object detection tasks, representing the AP at an IoU threshold of 0.5. From the mAP50

Table II. Ablation study

Model	mAP@0.5	mAP@0.5:0.95	Precision	Recall
YOLOv8 (no augmentation)	84.1	61.2	88.4	79.3
YOLOv8 + baseline augmentation	86.4	63.5	89.6	82.7
YOLOv8 + Mosaic only	87.2	64.1	90.1	84.9
YOLOv8 + MixUp only	87.6	64.4	90.7	85.3
YOLOv8 + full augmentation	88.4	65.8	91.2	86.1
YOLOv8 + full aug + SE	89.7	66.9	92.1	87.3
YOLOv8 + full aug + CBAM	92.3	71.0	93.4	89.2

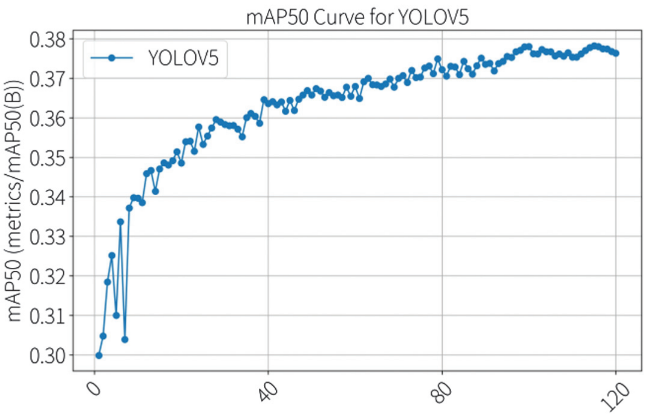

Fig. 7. mAP50 curve for YOLOv5.

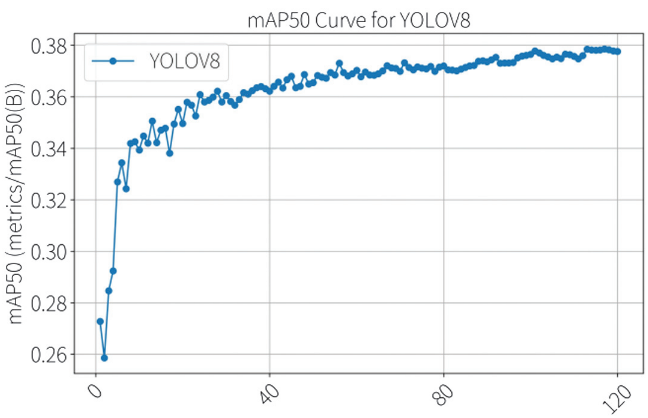

Fig. 8. mAP50 curve for YOLOv8.

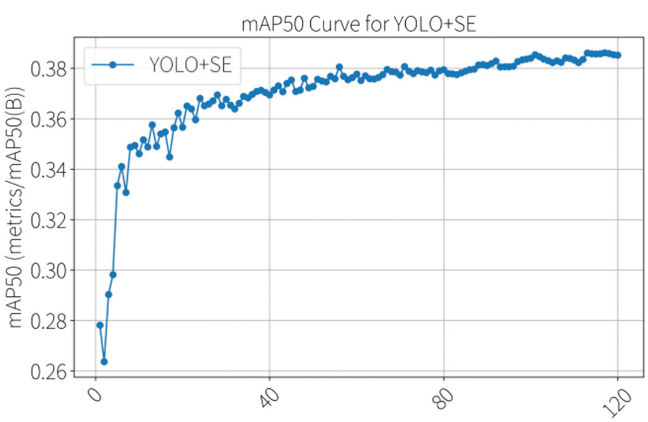

Fig. 9. mAP50 curve for YOLOv8 + SE.

curves of the four models (YOLOV5, YOLOV8, YOLOV8 + SE, and YOLOV8 + CBAM), it is evident that YOLOV8 + CBAM performs the best throughout the training process, with its mAP50 value consistently higher than the other models. YOLOV8 + SE follows closely, while YOLOV8 and YOLOV5 show relatively poorer performance. This is because CBAM can dynamically adjust feature responses in both channel and spatial dimensions,

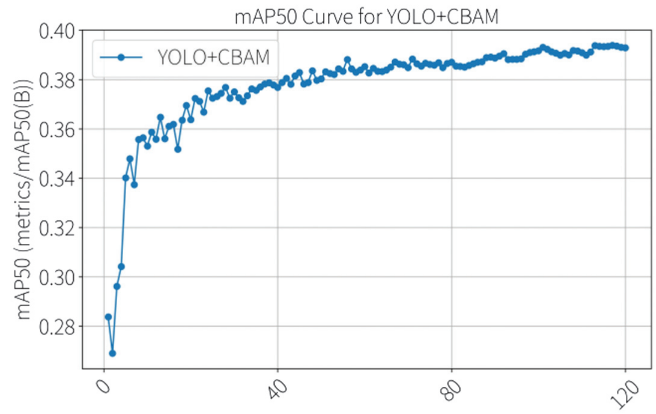

Fig. 10. mAP50 curve for YOLOv8 + CBAM.

enhancing the model's ability to capture critical features and thus improving detection accuracy. Although the SE module also enhances performance, its effect is not as significant as CBAM.

Second, it presents the comparison of the precision curve changes for each model, as shown in Figs. 11 to 14:

Precision is an important metric for evaluating the performance of object detection models, representing the proportion of true positives among the samples predicted as positive. From the precision curves of the four models, it can be observed that YOLOV8 + CBAM maintains a high precision level throughout the training

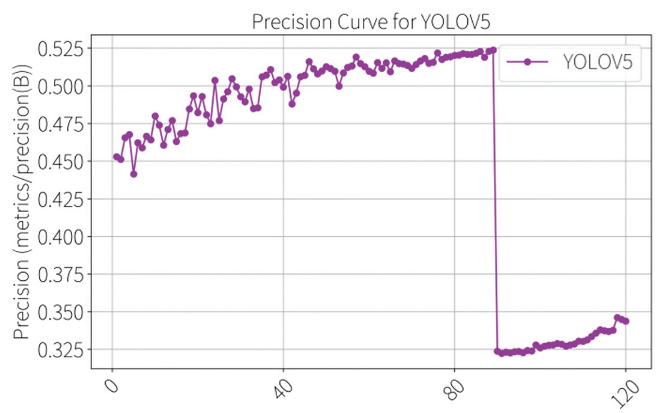

Fig. 11. Precision curve for YOLOv5.

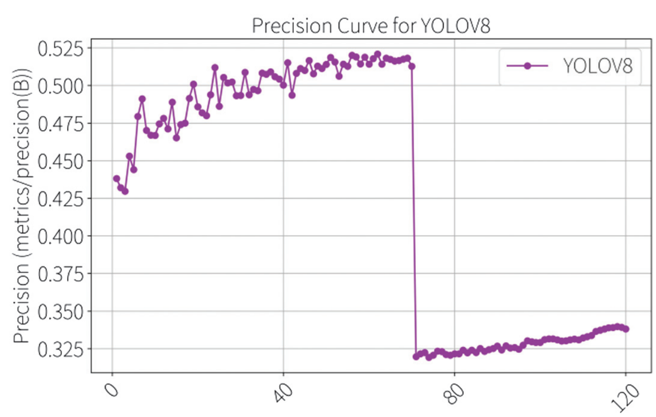

Fig. 12. Precision curve for YOLOv8.

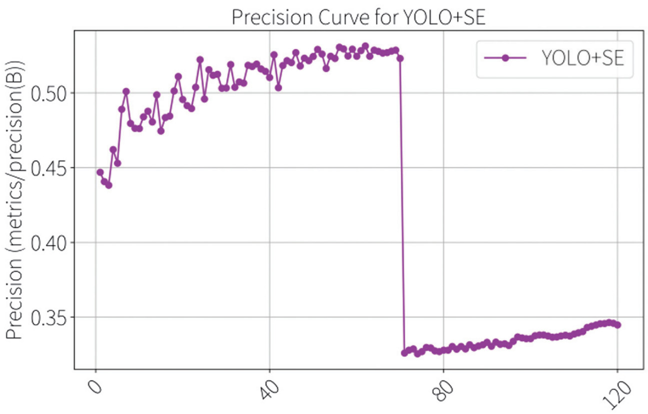

Fig. 13. Precision curve for YOLOv8 + SE.

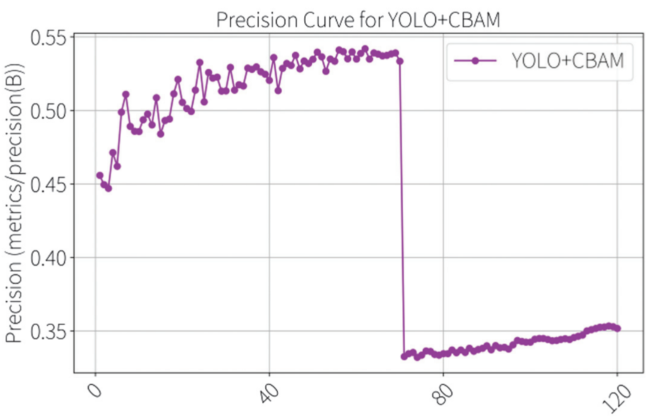

Fig. 14. Precision curve for YOLOv8 + CBAM.

process with minimal fluctuations. YOLOV8 + SE also shows high precision but with slightly more fluctuations. In comparison, YOLOV5 and YOLOV8 exhibit lower precision with greater fluctuations. This is because the CBAM module effectively reduces false detections through its attention mechanism, improving the model's prediction accuracy. Although the SE module also enhances precision, its stability is not as good as CBAM.

Finally, it displays the comparison of the recall rate curve changes for each model, as illustrated in Figs. 15 to 18:

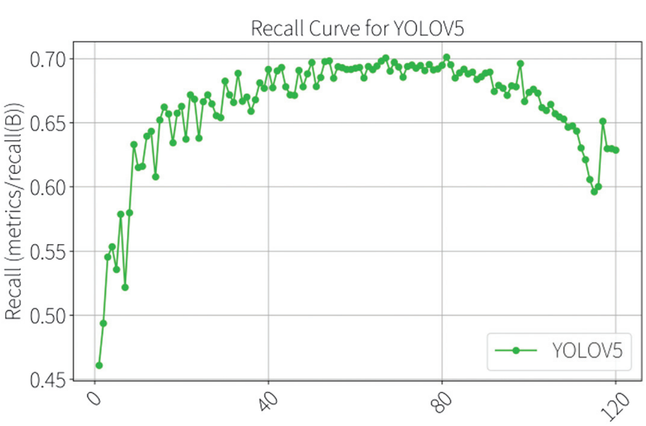

Fig. 15. Recall curve for YOLOv5.

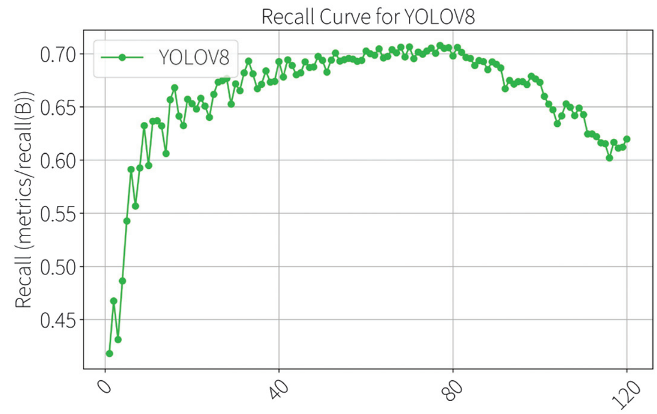

Fig. 16. Recall curve for YOLOv8.

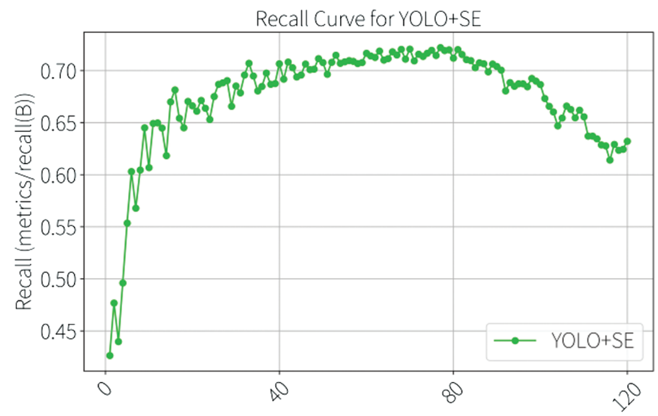

Fig. 17. Recall curve for YOLOv8 + SE.

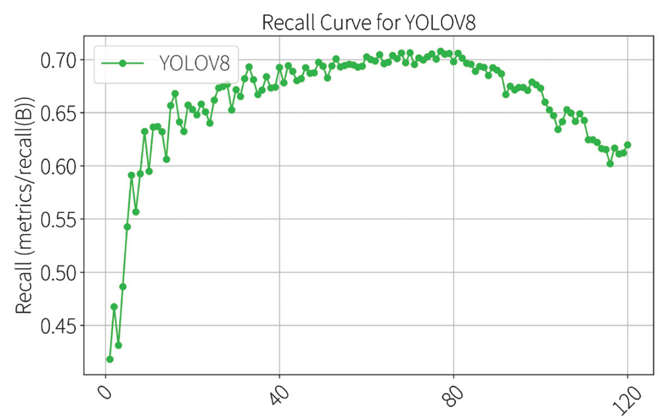

Fig. 18. Recall curve for YOLOv8 + CBAM.

Recall is another important metric for evaluating the performance of object detection models, representing the proportion of true positives that are correctly predicted as positive among all actual positive samples. From the recall curves of the four models, it can be seen that YOLOV8 + CBAM's recall rate continues to rise in the later stages of training and eventually stabilizes at the highest level. YOLOV8 + SE also shows a high recall rate but slightly lower than YOLOV8 + CBAM. YOLOV5 and YOLOV8 exhibit relatively lower recall rates. This is because the CBAM module enhances the model's ability to capture critical features, allowing more actual positive samples to be correctly detected, thereby improving the recall rate. Although the SE module also enhances recall, its effect is not as significant as CBAM.

D. ANALYSIS OF EXPERIMENTAL RESULTS

1). ROBUSTNESS ADVANTAGE OF YOLOv8 + CBAM. YOLOv8 + CBAM demonstrates superior performance in complex backgrounds (mAP@0.5 = 92.3%). The CBAM module effectively suppresses background noise (e.g., leaves and branches) through combined channel and spatial attention, enabling the model to focus on decay regions. Additionally, the model exhibits enhanced feature discrimination for mild and moderate decay grading (F1-score = 0.912), validating the sensitivity of hybrid attention mechanisms to subtle feature differences.

2). ANALYSIS OF THE APPLICATION EFFECTS OF DIFFERENT DATA AUGMENTATION METHODS. Quantitative analysis shows that the inclusion of MixUp and Mosaic results in a 6.8% increase in recall and a 4.3% increase in mAP@0.5 compared to baseline augmentation alone, demonstrating improved generalization in detecting small and obscured decay spots.

E. METHODOLOGICAL LIMITATION

The current study's primary constraint lies in the experimental design's susceptibility to initialization variance. While our core contributions regarding hybrid attention mechanisms remain valid, the quantitative results should be interpreted with caution due to the single-seed protocol.

VII. CONCLUSIONS

This study systematically evaluated the performance of YOLOv5, YOLOv8, and their attention-enhanced variants (YOLOv8 + SE, YOLOv8 + CBAM) for fruit decay detection. The results confirmed that the YOLOv8 + CBAM model, integrating a hybrid CBAM attention mechanism, achieves 92.3% mAP@0.5 detection accuracy on public datasets—a 3.9% improvement over the baseline model. Notably, in complex backgrounds and small-target detection scenarios, CBAM significantly suppresses background noise and focuses on decay regions through combined channel-spatial attention, demonstrating its advantages in feature enhancement and noise suppression. Ablation experiments showed that SE and CBAM modules improved model performance by 1.3% and 3.9%, respectively, with CBAM achieving a remarkable 5.2% gain in mAP@0.5:0.95, underscoring its critical role in capturing subtle feature differences for decay severity classification. Additionally, customized data augmentation strategies, including Mosaic, MixUp, and small-target copy-paste techniques, enhanced recall by 8.3% for small-target detection.

Future work will focus on three directions: first, for cross-modal data fusion, we plan to construct a paired dataset containing near-infrared (NIR) and visible light images by simultaneously capturing both modalities of the same fruit sample using multi-spectral imaging equipment. A cross-modal feature fusion network based on attention mechanisms will be employed to achieve multi-modal feature complementarity, enhancing the model's perception of internal decay structures—for instance, NIR images can penetrate fruit surfaces to reveal deeper decay information. Second, we will design dynamic decay severity grading algorithms that integrate temporal information (e.g., sequential detection results) or environmental parameters (e.g., temperature and humidity) to build a dynamic grading model capable of accurately predicting decay progression, thereby providing real-time support for agricultural decision-making. Third, in subsequent experiments, we will introduce five additional random seeds by conducting repeated training–validation experiments to statistically analyze the mean and standard deviation of model performance across seeds, quantifying the impact of initialization variance on result stability.

CONFLICTS OF INTEREST STATEMENT

The author(s) declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

[1] Food and Agriculture Organization of the United Nations, *The State of Food and Agriculture 2019: Moving Forward on Food Loss and Waste Reduction*. Rome: FAO, 2019. https://www.fao.org/4/ca6122en/ca6122en.pdf

[2] R. Porat *et al.*, "Characterization and quantification of postharvest losses of apple fruit stored under commercial conditions," *HortScience*, vol. 56, no. 5, pp. 608–616, 2021, DOI: 10.21273/HORTSCI15611-21.

[3] O. Kayahan, M. T. Yilmaz, and O. Ozcan-Tok, "Total aflatoxin and ochratoxin A levels, dietary exposure and cancer risk assessment in dried fruits in Türkiye," *Food Control*, vol. 141, p. 109183, 2022.

[4] J. N. Mwangi, S. W. Kariuki, and J. M. Gachohi, "Machine vision-based sorting of palm fruit: Efficiency and sustainability outcomes," *Comput. Electron. Agric.*, vol. 205, p. 107614, 2023.

[5] P. Afsharpour, T. Zoughi, and M. Deypir, "Robust deep learning method for fruit decay detection and plant identification: Enhancing food security and quality control," *Front Plant Sci.*, vol. 15, p. 1366395, 2024, DOI: 10.3389/fpls.2024.1366395.

[6] P. Kamat *et al.*, "Multi-class fruit ripeness detection using YOLO and SSD," *SN Appl. Sci.*, vol. 7, pp. 1–21, 2025, DOI: 10.1007/s42452-025-07617-7.

[7] Y. Lu *et al.*, "Strawberry defect identification using deep learning infrared–visible image fusion," *Agronomy*, vol. 13, no. 9, p. 2217, 2023.

[8] O. M. Lawal *et al.*, "Lightweight fruit detection algorithms for low-power computing," *IET Image Process.*, vol. 18, no. 9, pp. 2318–2328, 2024, DOI: 10.1049/ipr2.13098.

[9] X. Shi *et al.*, "Unet and swin transformer fusion network for lesion segmentation in biological kidney imaging," *Int. J. Pattern Recogn. Artif. Intell.*, vol. 36, no. 12, p. 2550021, 2025.

[10] R. Gai, Q. Liu, and Y. Zhang, "A detection algorithm for cherry fruits based on the improved YOLO-v4 model," *Neural Comput. Appl.*, vol. 35, pp. 24123–24136, 2023, DOI: 10.1007/s00521-023-08367-9.

[11] Y. Liu, K. Zhang, and F. Li, "FruitDet: Attentive feature aggregation for real-time fruit detection in orchards," *Sensors*, vol. 21, no. 12, p. 4137, 2021, DOI: 10.3390/s21124137.

[12] Z. Wang *et al.*, "Sc-yolov8 network with soft-pooling and attention for elevator passenger detection," *Appl. Sci.*, vol. 14, no. 8, pp. 2076–3417, 2024.

[13] L. A. Varga, J. Makowski, and A. Zell, "Measuring the ripeness of fruit with hyperspectral imaging and deep learning," In *2021*

International Joint Conference on Neural Networks (IJCNN), pp. 1–8, 2021. https://arxiv.org/pdf/2104.09808.

[14] A. M. Roy, R. Bose, and J. Bhaduri, "A fast accurate fine-grain object detection model based on YOLOv4 deep neural network," *Neural Comput. Appl.*, vol. 34, no. 5, pp. 3895–3921, 2022. https://arxiv.org/pdf/2111.00298.

[15] Z. Wei, M. Chang, and Y. Zhong, "Fruit freshness detection based on YOLOv8 and SE attention mechanism," *Acad. J. Sci. Technol.*, vol. 6, no. 1, pp. 195–197, 2024, DOI: 10.54097/ajst.v6i1.9125.

[16] D. Li *et al.*, "Cow-yolo: automatic cow mounting detection based on non-local cspdarknet53 and multiscale neck," *Int. J. Agric. Biol. Eng.*, vol. 17, no. 3, pp. 193–202, 2024.

[17] M. Hussain, "YOLO-v1 to YOLO-v8: The rise of YOLO in industrial defect detection," *Machines*, vol. 11, no. 7, p. 677, 2023, DOI: 10.3390/machines11070677.

[18] M. Elavarasu and K. Govindaraju, "Unveiling the advancements: YOLOv7 vs YOLOv8 in pulmonary carcinoma detection," *J. Rob. Control (JRC)*, vol. 5, no. 2, pp. 459–470, 2024, DOI: 10.18196/jrc.v5i2.20900.

[19] L. Yang, Y. Gu, and H. Feng, "Multi-scale feature fusion and feature calibration with edge information enhancement for remote sensing object detection," *Sci. Rep.*, vol. 15, no. 1, p. 15371, 2025.

[20] F. Liu *et al.*, "Rethinking the multi-scale feature hierarchy in object detection transformer (DETR)," *Appl. Soft Comput.*, vol. 175, p. 113081, 2025.

[21] D. Li, X. Zhang, and R. Ma, "Implementation and evaluation of spatial attention mechanism in apricot disease detection," *Appl. Sci.*, vol. 14, no. 5, p. 2034, 2024, DOI: 10.3390/app14052034.

[22] X. Song and V. Y. Mariano, "Fruit tree disease recognition based on residual neural network and attention mechanism," *J. Artif. Intell. Technol.*, vol. 4, no. 2, pp. 145–152, 2024.

[23] Z. Xu *et al.*, "Real-time identification of strawberry pests and diseases using an improved YOLOv8 algorithm," *Symmetry*, vol. 16, no. 10, p. 1280, 2024.

[24] X. Song, V. Y. Mariano, and X. Shi, "Apple disease recognition based on comprehensive application of residual neural network," *Mach. Learn. Artif. Intell.*, vol. 4, no. 2, pp. 33–42, 2023.

[25] Y. Tian *et al.*, "Detecting blueberry maturity under complex field conditions using improved yolov8," *Trans. Chin. Soc. Agric. Eng.*, vol. 40, no. 16, pp. 153–162, 2024.

[26] H. Yu *et al.*, "Research on an intelligent identification method for wind turbine blade damage based on cbam-bifpn-yolov8," *Processes*, vol. 12, no. 1, pp. 15–20, 2024.