Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

GMS-YOLO: An Efficient Lightweight Framework for Early-Stage Maize Tassel Detection in Complex Field Environments

Wei Wang
Anhui Science and Technology University
Xinwei Zhang
zhangxw@ahstu.edu.cn

Anhui Science and Technology University
Zeen Wang
Anhui Science and Technology University
Longfei Wang
Anhui Science and Technology University
Cheng Qian
Bengbu Triumph Display Technology Co., Ltd

Article

Keywords: maize tassel detection, YOLO11n, lightweight network, real-time recognition, edge deployment

Posted Date: May 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9113512/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Article

GMS-YOLO: An Efficient Lightweight Framework for Early-Stage Maize Tassel Detection in Complex Field Environments

Wei Wang [1], Xinwei Zhang [1,2,3,*], Zeen Wang [1], Longfei Wang [1], Cheng Qian [4]

[1] College of Mechanical Engineering, Anhui Science and Technology University, Chuzhou 233900, China; wwei0314@163.com (W.W.); W121707230509@163.com (Z.W.); yjs2024142@ahstu.edu.cn (L.W.); 3365813569@qq.com(C.Q.)
[2] Anhui Provincial Joint Key Laboratory for Common Key Technologies in Grain Storage Processes and Equipment, Chuzhou 233100, China
[3] Anhui Province Key Laboratory of Functional Agriculture and Functional Food, Anhui Science and Technology University, Chuzhou 233100, China
[4] Bengbu Triumph Display Technology Co., Ltd., Bengbu 233000, China
* Correspondence: zhangxw@ahstu.edu.cn

Abstract

Real-time and precise detection of maize tassels is crucial for improving the quality of mechanical detasseling in maize seed production. Existing research on tassel detection methods during the tasseling stage still has limitations. Therefore, this paper proposes a maize tassel recognition model, GMS-YOLO, based on an improved YOLO11n for real-time detection on embedded platforms. First, the original convolutional modules in the backbone network are replaced with GhostConv, and the existing C3k2 module is upgraded to C3GhostV3, reducing the model's parameter count and computational complexity. To compensate for potential performance loss due to lightweighting, an improved Multi-Scale Efficient Channel Attention (MSECA) module is introduced into the feature extraction network to enhance dynamic feature fusion capabilities for tassel detection. Additionally, the complex background noise suppression module C2GAM is integrated, and the original loss function is replaced with the SIoU loss function to accelerate model convergence. Experimental results show that the improved GMS-YOLO model achieves a maize tassel recognition accuracy of 94.2%, a recall rate of 87%, and an average precision of 93.4%. Compared to the original YOLO11n model, accuracy, recall, mean average precision, GFLOPs, and parameter count improved by 5.4, 3.5, 3.7, 10.77, and 8.61 percentage points, respectively. When deployed on a Raspberry Pi 5, the model achieves an inference speed of 13.92 FPS—a 23.19% increase—meeting real-time detection requirements. These results demonstrate that the model provides an efficient solution for real-time detection of tasseling maize tassels in resource-constrained environments.

Keywords: maize tassel detection; YOLO11n; lightweight network; real-time recognition; edge deployment

1. Introduction

As a core global staple crop, maize's yield stability and hybrid seed quality directly impact seed industry security and the sustainable development of the agricultural economy [1]. During hybrid seed production, removing tassels from female plants is a critical step to ensure seed purity and prevent varietal degradation [2]. Cross-pollination is the key to guaranteeing the expression of superior traits and enhancing both yield and quality [3]. The tasseling stage marks maize's critical transition from vegetative to reproductive growth, demanding highly precise timing for tasseling removal. Delayed or incomplete removal directly causes self-pollination in female plants, significantly reducing seed purity and final yield [4]. Against this backdrop, developing precise and efficient detection technologies for maize tassels during tasseling is vital for optimizing removal strategies and ensuring hybrid seed quality [5].

Under natural field conditions, ear recognition faces multiple challenges. Primarily, ears exhibit significant size variations during the heading stage, making it difficult to capture features of small targets. Additionally, ears share highly similar colors with leaves and stems (all within the green spectral range) and are highly susceptible to leaf occlusion, with occlusion rates often exceeding 30%. [6]. In recent years, deep learning methods have become the mainstream technology for tassel detection. For instance, Liu et al. [7] proposed adapting the Faster R-CNN model to drone-scale characteristics to enhance detection accuracy. Wang et al. [8] fine-tuned the YOLOX model using transfer learning, effectively improving maize tassel detection accuracy and generalization capabilities in drone imagery. To further meet edge-deployment demands for agricultural equipment, Li et al. [9] replaced the backbone network of RTMDet-tiny with the lightweight MobileNetv3. Combining the CBAM attention module with the EIoU loss function, they constructed the MLCE-RTMDet model with only 3.9 million parameters, achieving efficient detection in manual tassel removal scenarios. Although lightweight networks improve detection speed, they are prone to feature loss when handling small targets and complex backgrounds. To address this, Ma et al. [10] introduced SPD-Conv modules to replace traditional stride convolutions in YOLOv7-tiny and embedded the ECA-Net attention mechanism, significantly reducing false negative rates caused by excessively small or occluded tassels. Chen et al. [11] further focused on multi-varietal morphological variations and field background interference. Building upon YOLOv9m, they designed the Interactive Cross-layer Fusion (ICFFE) and Multi-scale Axial Awareness (MAA) modules, proposing the MAAFENet with enhanced interference resistance. Additionally, Wei et al. [12], Song et al. [13], and Chen et al. [14] respectively enhanced feature discrimination and robustness in complex environments by introducing Coordinate Attention (CA), Spatial Attention (SE), and Spatial Attention Residual Module (RESAM). While these optimization strategies improve maize tassel detection accuracy, existing models remain limited to mature tassels with distinct features. They face high false-negative risks when handling early-stage tasseling targets, which exhibit ambiguous features, extremely small scales, and high occlusion susceptibility.

Currently, the YOLO series models have become one of the most widely adopted mainstream solutions in the field of agricultural object detection, owing to their end-to-end detection capabilities, efficient inference speeds, and flexible room for improvement. Teng et al. [15] enhanced the robustness of wheat ear recognition in high-density scenes through adaptive feature fusion and channel attention mechanisms; Shen et al. [16] incorporated depthwise grouped convolutions to optimize YOLOv8, effectively adapting it for tomato disease detection under complex lighting conditions. With algorithmic iterations, recent research has further expanded YOLO's application in precision agricultural operations. Peng et al. [17] introduced Deep Separable Convolution (DWConv) and Adaptive Down-sampling (ADown) upon YOLO11n, achieving significant lightweighting while maintaining weed detection accuracy in maize fields, providing a reference for edge device deployment; Jia et al. [18] enhanced feature capture for minute objects in complex backgrounds through an augmented C3k2 module; Du et al. [19] proposed the YOLO-WAS model, which effectively addressed low recognition accuracy and redundant parameters in complex scenarios by reconstructing YOLO11's down-sampling process and incorporating lightweight structures. Li et al. [20] balanced model speed and hardware requirements through lightweight backbone networks, resolving fruit occlusion issues. Furthermore, concerning the precise detection of specific locations during detasselling operations, KAPŁAN et al. [21] validated the effectiveness of YOLO models in segmenting bud grafting sites in nurseries. Niu et al. [22] further achieved accurate identification of green targets and precise positioning of cutting points in complex environments, providing technical insights for spatial perception of maize detasselling points. These studies demonstrate that YOLO series models, through attention mechanism fusion, optimized feature extraction modules, and lightweight reconstruction, can flexibly adapt to diverse crop morphological characteristics and field detection scenarios. This provides a reliable technical foundation for the visual recognition research of the maize detasseling device presented herein.

To address challenges such as low accuracy in maize tassel recognition due to complex background interference, low operational efficiency of models on embedded platforms, and limited hardware deployment resources, this study proposes an improved GMS-YOLO model. The main contributions of this research are as follows:

1. Development of a lightweight detection framework: A novel GMS-YOLO model is proposed, integrating a lightweight backbone with multi-scale attention mechanisms. By reconstructing the feature extraction process, the model significantly reduces parameters and computational complexity while maintaining high detection precision.
2. Optimization for complex agricultural scenes: By incorporating the Multi-Scale Efficient Channel Attention (MSECA) and C2GAM modules, the model's ability to highlight effective features and suppress background noise is enhanced. This enables robust tassel detection even under conditions of leaf occlusion and varying illumination.
3. Edge-side deployment and validation: The proposed model is specifically optimized for and successfully deployed on the Raspberry Pi 5 embedded platform. Experimental results demonstrate a superior balance between accuracy and real-time performance, achieving an mAP of 93.4% with an inference speed of 13.92 FPS.

Practical application in smart farming: This research not only provides core technological support for the visual guidance of mechanized maize detasseling devices but also offers a crucial theoretical foundation and practical reference for constructing intelligent monitoring systems for other field crops in resource-constrained environments.

2. Materials and Methods

2.1. Dataset Construction

2.1.1. Target Definition and Characteristic Analysis of the Early Male Spike in Maize

This study focuses on maize tassels in the early stage. Unlike mature tassels that have open branches and clear shapes, early tassels look like small "spikelets" just coming out from the top leaves. Their side branches are not yet grown, so they look like thin, simple clusters. This makes it hard for the model to find a clear shape or enough detail (Figure 1a). In the field, these early tassels are light green or yellowish-green, which is almost the same color as the leaves and stems around them. This lack of contrast makes the tassels look hidden, making it difficult for the computer to tell them apart from the background [23]. On the other hand, mature tassels are much larger, have fully open branches, and a clear texture. They stay at the very top of the plant, and their color becomes lighter, making them easy to see against the green leaves (Figure 1b). Also, early tassels are very small and are often covered by nearby leaves, sometimes by more than 30%. Because only a few pixels of the tassel are visible and their shapes change as they grow, the model can easily get confused by the environment, leading to missed targets or wrong positions (Figure 1c).



(a) Early (b) Maturity (c) Obstruction

Figure 1. Typical morphological characteristics and visual disturbances of maize tassels

2.1.2. Dataset Acquisition

The maize tassel image data collection site is located at the experimental base in Wuyi Town, Nanqiao District, Chuzhou City, Anhui Province (118°24′38″E, 32°7′36″N). The cultivated variety is Fengnuo 168 (FN168), with a planting spacing of 40 cm between plants and 60 cm between rows. Data collection occurred from July 6 to July 15, 2025, using an iPhone 15 Plus with an image resolution of 5712×4284 pixels, saved in JPG format. To ensure image accuracy and minimize perspective distortion, the camera was mounted on a tripod at a fixed height of 1.5-2m above the tassel apex, maintaining a vertical downward shooting angle. To enhance sample diversity, the collection included scenes under natural conditions during both sunny and rainy weather. Numerous complex field images featuring leaf shading and other obstructions were retained during collection. After manual screening and cleanup, a total of 1,535 valid raw images were obtained.

2.1.3. Dataset Creation

To accelerate model training and facilitate analysis of target dimensions, the collected images were first compressed to a resolution of 640×640 pixels. Subsequently, LabelImg software was used to annotate the image data, generating TXT label files compliant with the YOLO format. To enrich the training samples, data augmentation was applied to the dataset, including brightness adjustment, random rotation, shadow addition, and contrast modification. The augmented images expanded the dataset to 7,675 images. The tassel dataset was randomly divided into training, validation, and test sets at an 8:1:1 ratio (Figure 2).



(a) Brightness Adjustment (b) Random Rotation (c) Add Shadows (d) Adjustment of Contrast

Figure 2. Image Feature Enhancement Results

2.2. Model Improvements

2.2.1. GMS-YOLO Module

The YOLO11 model studied in this paper undergoes an architectural upgrade based on YOLOv5[24] and YOLOv8[25], significantly enhancing feature extraction capabilities through the introduction of the C3k2 module and C2PSA layer[26].

To make the model run better on small agricultural devices, we improved YOLO11 and called it GMS-YOLO. First, in the main part of the network (the backbone), we used the C3GhostV3 module instead of the original C3k2. This module uses simple linear changes to create more features with less work. This reduces the number of parameters and makes the model faster, but still keeps the important information. For the maize field work, we used GhostConv in most parts of the backbone. We only kept standard convolutions in the first layer (the Stem) to make sure we don’t lose any important data at the start. At the end of the backbone, we added a C2GAM attention part. It helps the model focus on the important parts of the image and ignores the messy background, which provides better information for the next steps.

In the neck and detection head parts, we also use C3GhostV3 to make the whole structure smaller and faster. When we combine features from different scales, we add the MSECA module after each Concat step. This module helps the model better understand the mixed features by looking at the channels. It makes the model better at finding very small tassels and parts that are not clear. With these changes, GMS-YOLO uses fewer parameters while keeping the fast data flow of DWConv [27]. This achieves a dynamic balance between detection accuracy and inference speed, making it better suited for real-time applications like maize tassel detection.

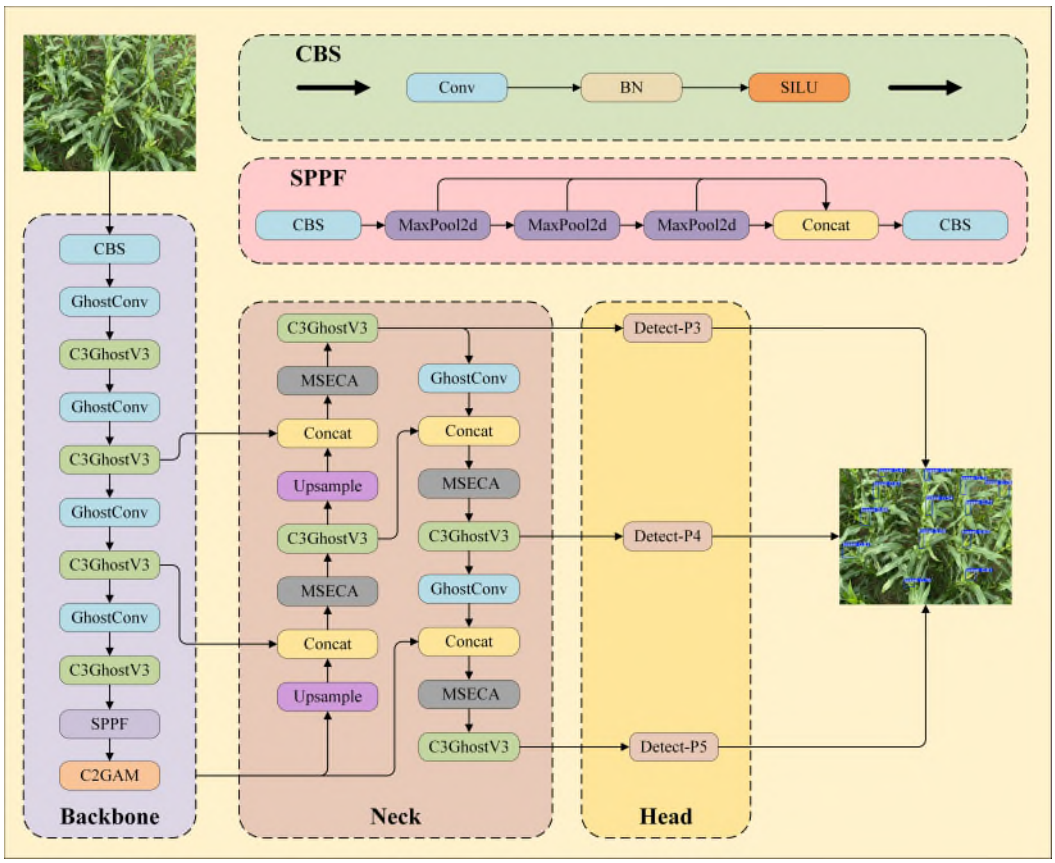


Figure 3. Network structure of GMS–YOLO

2.2.2. C3Ghostv3 Module

To address the practical constraints of limited computational power and storage space in embedded edge devices within agricultural operating environments, this paper designs a lightweight feature extraction module, C3GhostV3, based on GhostNetV3 [28] theory. This module replaces the C3k2 module in the original YOLO11 backbone network to enable real-time deployment of the maize tassel detection model at the edge. The GhostNetV3 architecture is shown in Figure 4.

Figure 4 shows the design of C3GhostV3, which uses GhostBottleneckV3 as its main part. Instead of using standard, heavy convolutions, this design combines simple convolutions with cheap linear changes. First, it uses a 1 × 1 GhostModule convolution to create basic feature maps. Then, it uses a fast 3 × 3 depthwise convolution (DWConv) to create more features. By putting these features together, the model uses fewer parameters and less power while keeping the same amount of information. To help the model see a larger area, we added a Decoupled Fully-Connected (DFC) attention part. It looks at the image in both horizontal and vertical directions to help the model better find the important parts of the maize tassels.

C3GhostV3 also keeps the two-path design from the original CSP structure. One part of the data goes through the GhostBottleneckV3 for deep processing, while the other part goes through a simple side connection. Looking at the data in Table 2, this new design reduces the number of parameters by 20.2% compared to the standard YOLO11n backbone. By using this structure and the DFC part together, the model becomes much smaller but still keeps the important information it needs to see.

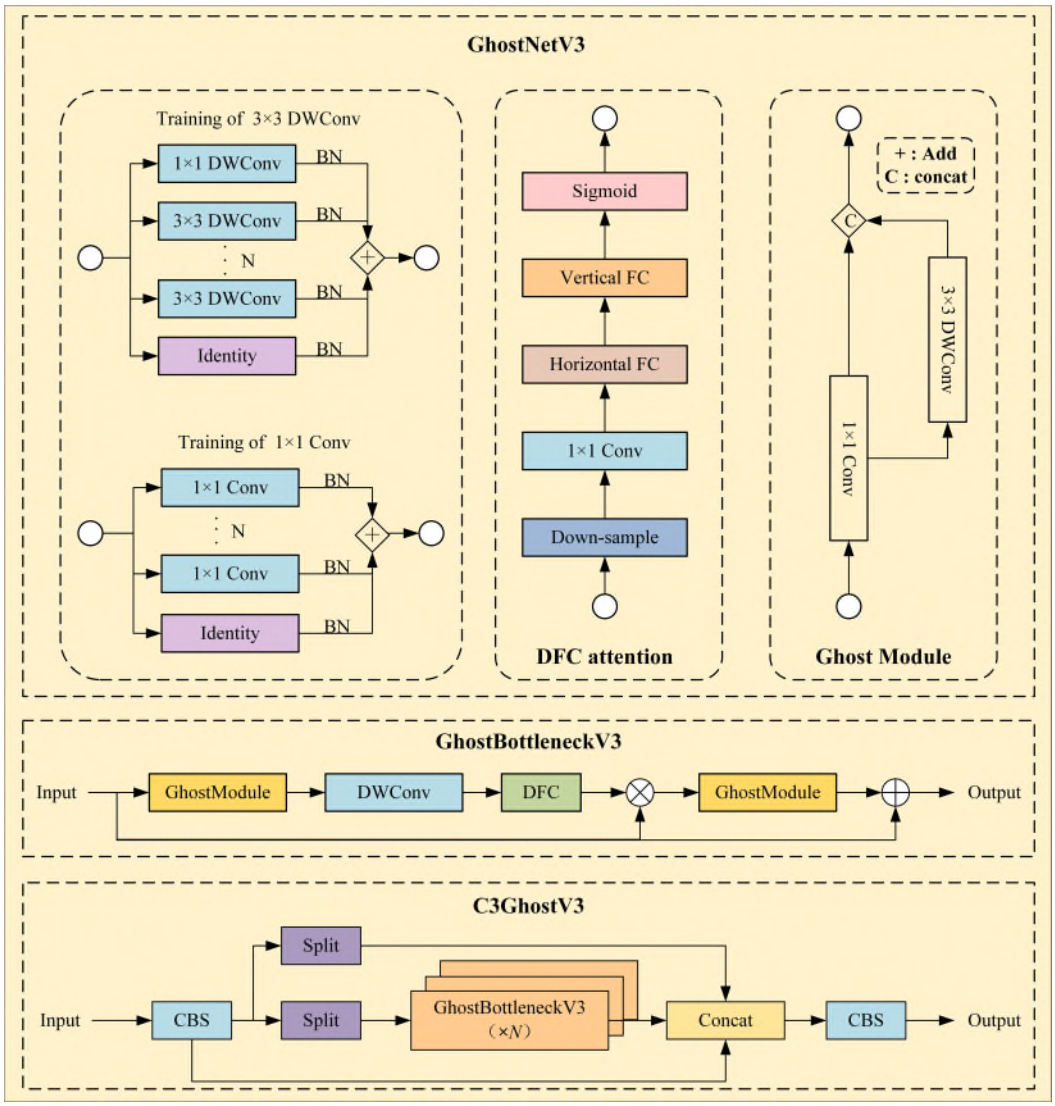


Figure 4. Network structure of GhostnetV3 and the proposed improved C3GhostV3 module.

2.2.3. C2GAM Module

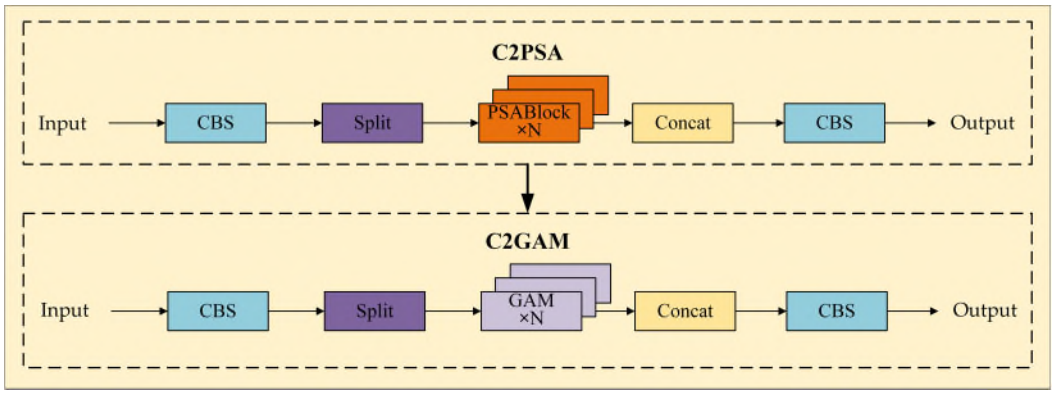


Figure 5. Structure of C2PSA and the improved C2GAM

To address the challenges of detecting maize tassels—which are small, exhibit significant size variation, and are prone to leaf occlusion—this study proposes a targeted enhancement to the C2PSA module within the YOLO11 backbone network. This enhancement introduces a Cross-Stage Global Attention Module (C2GAM). This module retains the CSP (Cross Stage Partial)

segmented architecture of C2PSA to preserve its lightweight nature and gradient flow advantages. However, it introduces a Global Attention Mechanism (GAM) within the feature processing branch to replace the original PSA module.

Traditional attention mechanisms (such as CBAM [29] and SE [30]) often suffer from feature loss during information compression due to the separation of channel and spatial features, thereby limiting the model's visual receptive field and global interaction capabilities. To address this, this study introduces the GAM module. As illustrated in Figure 6, GAM employs a 3D arrangement and multilayer perceptrons to amplify global-level interaction representations while reducing information diffusion. Comprising a channel attention submodule and a spatial attention submodule connected in series, it collectively enhances the capture of subtle structures within the tassel [31].

As shown in Figure 6, the channel part of the module keeps the features in three dimensions. Instead of just averaging everything together, it changes the order of the data (using Permute) to keep the information about channels, width, and height. Then, it uses a small network with two layers to help the model understand how these different parts relate to each other. We use a setting called "r" to help the model find better connections. After that, it changes the data back to its original shape and uses a Sigmoid function. Simultaneously, the spatial attention submodule focuses on accurately modeling spatial information. To prevent spatial information loss caused by max pooling operations, this module eliminates pooling layers and instead employs two consecutive 7×7 large-kernel convolutional layers. The first convolutional layer applies the same reduction ratio r as the channel module to fuse spatial contextual information, while the second layer restores the number of channels. This design effectively expands the receptive field, enabling the model to accurately locate the tassel boundary based on contextual information, even when facing leaf occlusion [32].

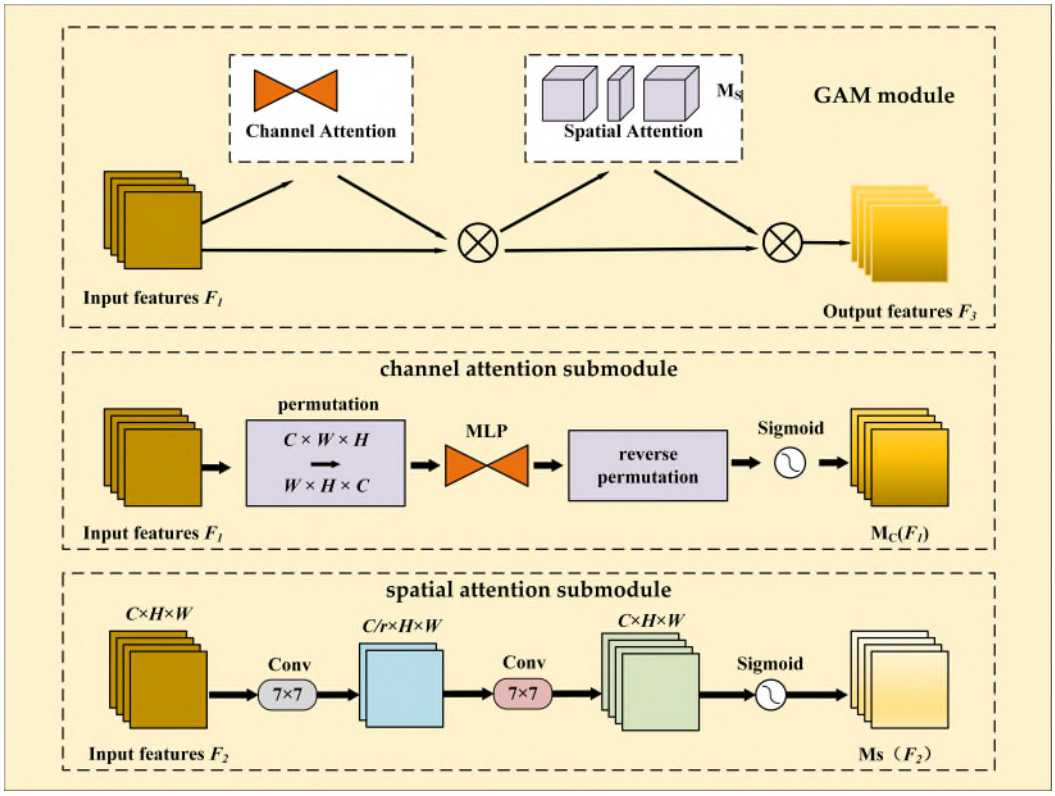


Figure 6. Structural diagram of the GAM and its internal components

2.2.4. MSECA Module

During the feature fusion stage of YOLO11, feature maps from different depths often exhibit redundant information after concatenation via Concat. To enable the network to adaptively discern

the importance of deep semantic features versus shallow detail features, the traditional Efficient Channel Attention (ECA) module achieves local interactions between channels through one-dimensional convolutions. However, its *k* value remains fixed according to the channel dimension C via formula (2), resulting in a limited receptive field. Finally, the Sigmoid activation function generates the weights M_e for each channel, and the original feature map undergoes weighted calibration[33]. The calculation process is as follows:

$$F' = M_e(F) \otimes F \tag{1}$$

In the equation, F represents the input feature map, F′ denotes the feature map processed by ECA, Me signifies the channel attention weights computed by ECA, and ⊗ indicates channel-wise multiplication.

$$k = \psi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{2}$$

In real farm work, the camera needs to see both small details like the texture of the tassel and its overall shape. Using only one scale makes it hard to catch all the changes from when the tassel first comes out to when it fully opens. This often leads to the model missing tassels when they are blurry or hidden behind leaves.

In this study, we propose a multi-scale version of ECA, called the MSECA (Multi-Scale Efficient Channel Attention) module. As shown in Figure 10, we use different convolutional kernels with k=3, 5, and 7. This creates a simple mechanism to pick up features of maize tassels at different growth stages. The small kernel (k=3) focuses on details like the points where the tassel branches out, which helps the model find small or distant targets. The large kernel (k=7) looks at a wider area to see the relationship between the tassel and the top leaves, helping the model find the target even when it is covered by leaves. The middle kernel (k=5) helps connect these two scales. By using these parallel kernels instead of just one, the model can better handle targets of different sizes and focus on the tassels more accurately.

Through MSECA's multi-scale processing, the model can simultaneously perceive different details of the tassel. It can discern fine local features using small-scale kernels while observing the relationship between the tassel and surrounding leaves via large-scale kernels. This approach enables the model to avoid misclassification due to incomplete feature detection when encountering leaf occlusion or poor lighting conditions.

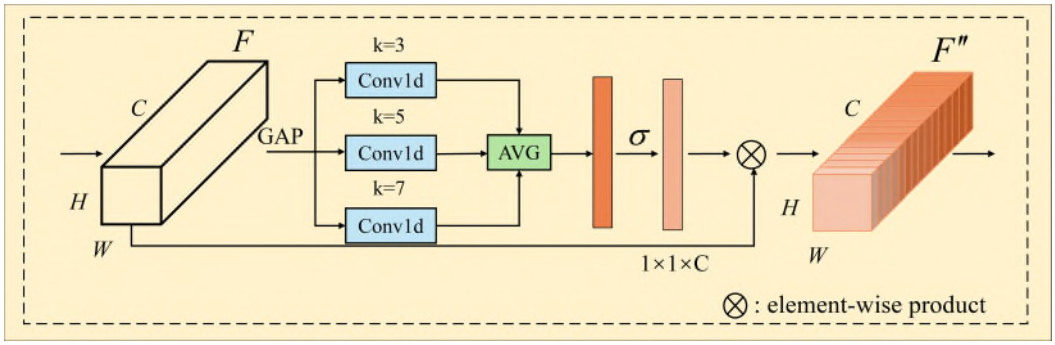


Figure 7. Structural diagram of the improved MSECA module

2.2.5. Optimization of Bounding Box Regression Based on SIoU

YOLO11 employs CIoU [34] as its bounding box regression loss function. This function comprehensively evaluates overlap area, center distance, and aspect ratio. However, it exhibits limitations when applied to maize tassels as detection targets. This is primarily due to the tassels' irregular morphology, high aspect ratio, and diverse growth angles, while CIoU lacks constraints on the directionality between predicted and ground truth bounding boxes.

To address the aforementioned issues, the SIoU loss function is introduced to replace CIoU in the model. SIoU [35] redefines the bounding box regression strategy by incorporating an "angular cost," comprising four core components: angular loss (L_a), distance loss (L_d), shape loss (L_s), and bounding box intersection-over-union loss (L_{IoU}). The innovation of SIoU lies in introducing a directional alignment mechanism that leverages a synergistic loss between angle and distance. The angle loss aims to reduce the geometric drift between the center of the predicted bounding box (B) and the center of the ground truth bounding box (B^{GT}). As shown in Equation (3), this mechanism prioritizes the regression of the predicted box onto the nearest horizontal or vertical axis of the ground truth box by minimizing α. As illustrated in Figure 11, the loss is minimal when angle α is 0, and maximal when α is $\pi/4$. The distance loss is defined as in Equation (4). As α approaches 0, the contribution of distance loss significantly diminishes. Conversely, as α approaches $\pi/4$, the distance loss contribution increases substantially. The shape loss is given by Equation (5), accounting for the aspect ratio dimension between the predicted and ground truth boxes. The bounding box intersection-over-union loss is given by Equation (6), with Figure 12 illustrating the contribution relationship of each component. The overall loss function L_{SIoU} is defined by Equation (7). Considering the morphological variations in maize tassels, the aspect ratio dimension factor between the predicted and ground truth boxes simplifies computational complexity while comprehensively capturing the positional discrepancy between them.

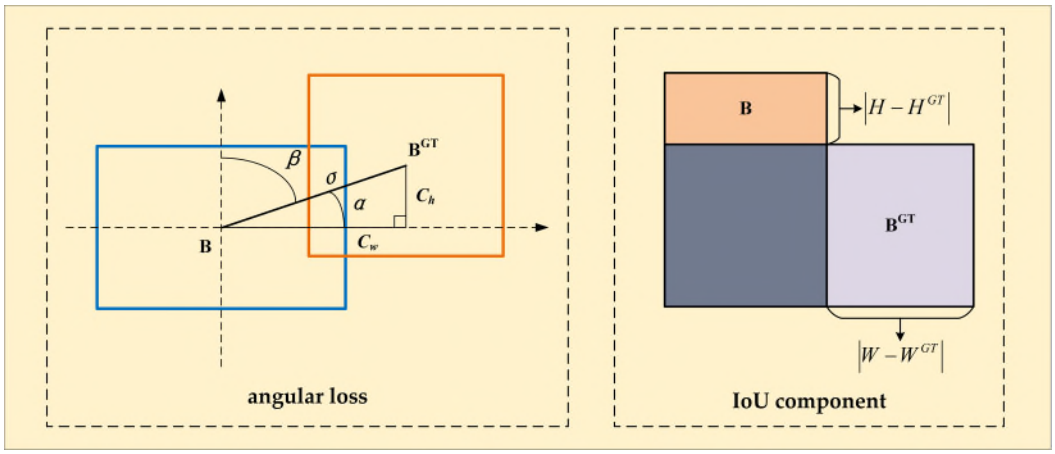


Figure 8. Example diagram of angular loss calculation and Schematic of the relation of IoU component contribution.

$$L_a = 1 - 2 \times \sin^2\left(\arcsin x - \frac{\pi}{4}\right) \tag{3}$$

$$\begin{cases} L_d = \sum_{t=x,y} \left(1 - e^{(2-L_a)\rho_t}\right) \\ \rho_x = \left(\frac{b_{cx}^{gt} - b_{cx}}{c_w}\right)^2, \rho_y = \left(\frac{b_{cy}^{gt} - b_{cy}}{c_h}\right)^2 \end{cases} \tag{4}$$

$$\begin{cases} L_s = \sum_{t=w,h} \left(1 - e^{-w_t}\right)^{\theta} \\ w_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}, w_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases} \tag{5}$$

$$L_{IoU} = 1 - IoU \tag{6}$$

$$L_{SIoU} = 1 - \mathrm{IoU} + \frac{L_d + L_s}{2} \tag{7}$$

In the equation, x denotes the ratio of the height of the rectangle formed by connecting the centre points of the predicted bounding box and the ground truth bounding box to the straight-line distance between the centre points. C_w and C_h represent the width and height of the minimum bounding rectangles for the ground truth and predicted bounding boxes, respectively. w, h, w^{gt}, and h^{gt} denote the width and height of the ground truth and predicted bounding boxes, respectively.θ is a hyperparameter, typically set within the range of 2 to 6. IoU signifies the intersection-over-union ratio between the predicted and ground truth bounding boxes.

2.3. Test Environment and Evaluation Criteria

2.3.1. Test Platform and Parameter Configuration

The hardware configuration for the model training platform is as follows: i5-14600 KF CPU, NVIDIA GeForce RTX 3060 GPU, 12GB graphics memory. The software environment consists of the Windows 11 operating system (64-bit), Python 3.11 as the programming language, the PyTorch 2.7.0 framework, and CUDA 11.8. Training parameters are detailed in Table 1.

Table 1. Training parameters

Parameters	Numerical/Type
Optimizer	SGD
Optimizer Momentum	0.937
Optimizer Weight Decay Coefficient	0.000 5
Initial learning rate	0.01
Number of training batches	8
Number of Iteration Cycles	300

2.3.2. Evaluation indicators

To comprehensively evaluate the performance of models in detecting maize tassels, this paper employs precision (P), recall (R), and mean average precision (mAP) to assess detection capabilities. Model size is measured using parameters and floating-point operations (FLOPs), while detection accuracy is quantified by mean average precision (mAP). mean average precision (mAP) to demonstrate the model's detection capability. Model size is measured using parameters and floating-point operations (FLOPs), while inference speed is evaluated using frames per second (FPS).

3. Test Results and Analysis

3.1. Analysis of YOLO11 Model Ablation Studies

The ablation test results are shown in Table 2. As indicated in the table, the improved C3GhostV3 module reduces computational complexity. The C2GAM mechanism enhances global attention, enabling the model to capture richer feature contexts. Subsequently, the MSECA attention module further strengthens inter-channel feature integration capabilities. The SIoU loss function optimizes angular costs, thereby improving regression prediction accuracy and model fitting precision.

As shown in Table 2, in Experiment 2, replacing the original C3k2 module with C3GhostV3 in the YOLO11n network reduced the model parameters to 10% of the original. The model's precision, recall, and average precision all decreased, indicating that while lightening the model, some redundant features were removed, leading to a decline in detection performance. In Experiment 3, after adding C2GAM, the model's precision, recall, and average precision all improved, demonstrating that the C2GAM module can effectively compensate for feature loss and enhance model recognition accuracy. Experiment 4 introduced the MSECA module, maintaining the model's parameter count while achieving significant improvements across all metrics compared to Experiment

2. This demonstrates enhanced feature extraction without incurring additional computational overhead. Experiment 5 combined C3GhostV3, C2GAM, and MSECA modules, achieving an average precision of 0.629 while preserving a compact model size. Experiment 6 incorporated the SIoU loss function for regression optimization, achieving optimal performance across all metrics. Compared to the original YOLO11n baseline, precision, recall, and average precision improved by 6.1, 4.2, and 10 percentage points, respectively, while model parameters decreased by 8.61%. This fully validates the effectiveness and complementary nature of each improved module in enhancing maize tassel detection performance.

Table 2. Ablation Experiment

Experiment	C3GhostV3	C2GAM	MSECA	SIOU	P/%	R/%	mAP50-95/%	Parameter/M
1	×	×	×	×	88.8	83.5	58.2	2.67
2	√	×	×	×	86.0	80.5	46.6	2.13
3	√	√	×	×	92.1	87.5	62.6	2.42
4	√	×	√	×	91.7	85.5	60.8	2.13
5	√	√	√	×	93.1	86.7	62.9	2.37
6	√	√	√	√	94.2	87.0	64.0	2.44

3.2. Comparison of Improvements Across Different Network Backbones

This experiment uses an enhanced YOLO11 object detection network as the base architecture, replacing mainstream lightweight feature extraction backbone networks such as MobileNetV3, MobileNeXt, EfficientNet, ShuffleNetV2, and MobileVit. While keeping all parameters constant except for the backbone network, the experiment compares the training performance of different backbone networks.

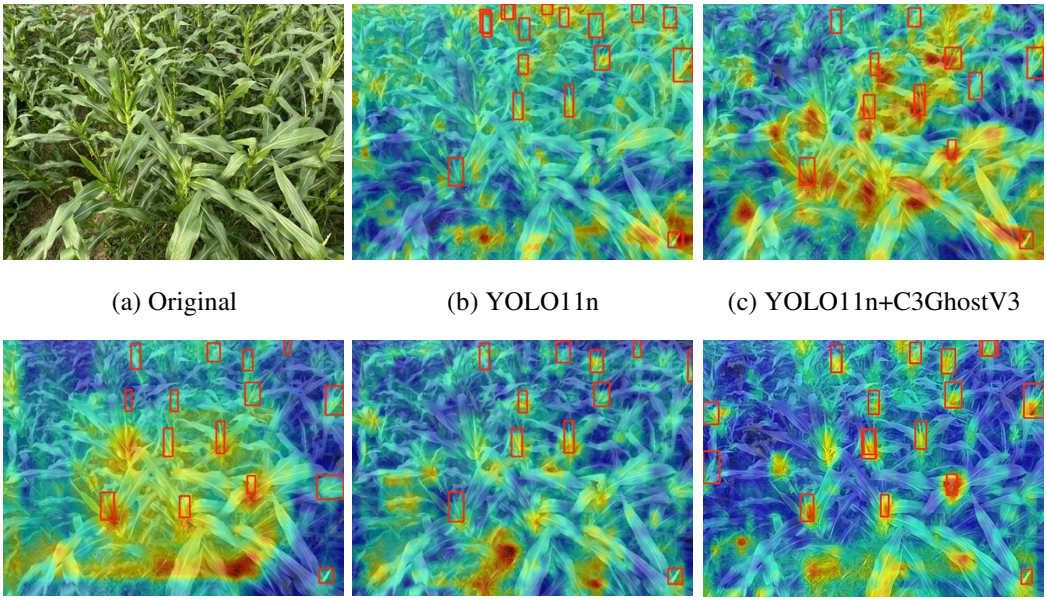
Table 3. Comparison of Results with Different Backbone Networks

Model	mAP50/%	mAP50-95/%	Parameter/M	GFLOPs
YOLO11	89.7	58.2	2.67	6.5
+MobileNetV3	75.3	38.7	1.41	3.4
+MobileNeXt	70.2	36.1	1.72	4.7
+EfficientNet	80.7	42.8	0.72	2.4
+ShuffleNetV2	55.9	23.4	0.58	1.8
+MobileVit	67.7	30.7	1.45	9.1
+C3GhostV3	88.3	46.6	2.13	5.4

As shown in Table 3, the improved C3GhostV3 approach demonstrates significant advantages in both model training accuracy and average precision compared to MobileNetV3, MobileNeXt, EfficientNet, ShuffleNetV2, and MobileVit. The average precision improved by 13, 18.1, 7.5, 32.4, and 20.6 percentage points, respectively, compared to other networks. Although some lightweight networks have substantially fewer parameters than our method, their feature extraction capabilities are inadequate and fail to meet deployment requirements. This approach reduces model parameters by 20.2% (from 2.67M to 2.13M) and computational complexity by 16.9% (from 6.5G to 5.4G) while preserving core feature extraction capabilities.

3.3. Heatmap Comparison

To further observe the feature extraction logic and anti-interference capabilities of the GMS-YOLO model in real field environments, this paper employs Grad-CAM (Gradient-weighted Class Activation Mapping) technology to conduct heatmap visualisation analysis on models at different refinement stages. Figure 9 illustrates the evolution of attention distribution from the original image to the final refined model within the same scene. In the figure, redder hues indicate higher model attention, while blue hues denote lower attention.



(a) Original (b) YOLO11n (c) YOLO11n+C3GhostV3

(d) YOLO11n+C3GhostV3+ MSECA (e) YOLO11n+C3GhostV3+ MSECA+C2GAM (f) YOLO11n+C3GhostV3+ MSECA+C2GAM+SIOU

Figure 9. Heatmap of Ablation Experiments for the Tassel Detection Model

Looking at Figure 9, the focus of the standard YOLO11n model is quite spread out. Although it finds most tassels, many yellow-green spots also show up on the surrounding leaves. This means the basic model struggles to tell the difference between tassel textures and leaf textures, making it easy for the background to interfere. This is why it often makes mistakes (Figure 9b). After adding the C3GhostV3 module, the heatmap shows that the model still keeps the important features even though it is much smaller. By removing extra, unneeded data, the focus actually gets a bit better than the base model, and some background noise starts to disappear (Figure 9c). When we add the MSECA part, there is a clear change. The red areas, which show where the model is looking most, start to shrink and move toward the center of the tassel. There are also fewer messy spots on the leaves. This shows the multi-scale design helps the model focus on tassels of different sizes while ignoring the background (Figure 9d). After adding the C2GAM module, the model becomes much more sure of itself. The red areas in the heatmap become darker with sharper edges. This means the model is not just looking at small details but also using the whole picture to tell the difference between hidden tassels and leaves in front of them (Figure 9e). Compared to the others, our final model shows the cleanest results. The background is mostly dark blue, which means the model is ignoring the messy field environment. The focus is exactly on each tassel, showing a bright red color even for tassels that are far away, very small, or mostly covered by leaves (Figure 9f).

3.4. Comparison with Different Models

To check if the GMS-YOLO model is better for finding maize tassels, we compared it with other popular models: YOLOv5n, YOLOv6n, YOLOv8n, YOLOv10n, and the standard YOLO11n. The results are in Table 4. The data shows that our model has the highest precision, recall, and mAP. Compared to the basic YOLO11n, our model's average precision went up by 3.7%, and its mAP increased by 5.8%. Our mAP is also 0.7% higher than YOLOv8n and 2.0% higher than YOLOv5n. In terms of size, our model is very small. Its computing load is only 5.8 GFLOPs, which is the lowest in the group and 10.2% less than YOLOv8n. It has 2.44 million parameters, which is 22.3% fewer than YOLOv8n and 12.4% fewer than YOLO11n.

Figure 10 shows how different models perform as training goes on. In all metrics, including Precision, Recall, mAP50, and mAP50-95, the GMS-YOLO model (yellow line) stays at the top from the beginning. Its curves are smooth and stable, and it performs better than other models from

YOLOv5n to YOLO11n. This shows that changing the model structure helps it work better in messy farm fields. GMS-YOLO uses the C3GhostV3 module to reduce unneeded work and the MSECA module to combine features better. Also, the C2GAM part helps it ignore the background. This design makes the model both accurate and small enough to run on devices like the Raspberry Pi.

Table 4. Comparison of Maize Tassel Recognition Results of YOLO Series Models

Model	P/%	R/%	mAP50/%	mAP50-95/%	GFLOPs	Parameters/M
YOLO11n	88.8	83.5	89.7	58.2	6.5	2.67
YOLOv10n	88.2	83.2	90.4	58.1	6.8	2.27
YOLOv8n	90.4	86.6	92.7	58.3	8.1	3.01
YOLOv6n	87.7	81.1	86.3	51.0	11.8	4.23
YOLOv5n	91.2	83.9	91.4	55.4	7.1	2.5
GMS-YOLO	94.2	87.0	93.4	64.0	5.8	2.44

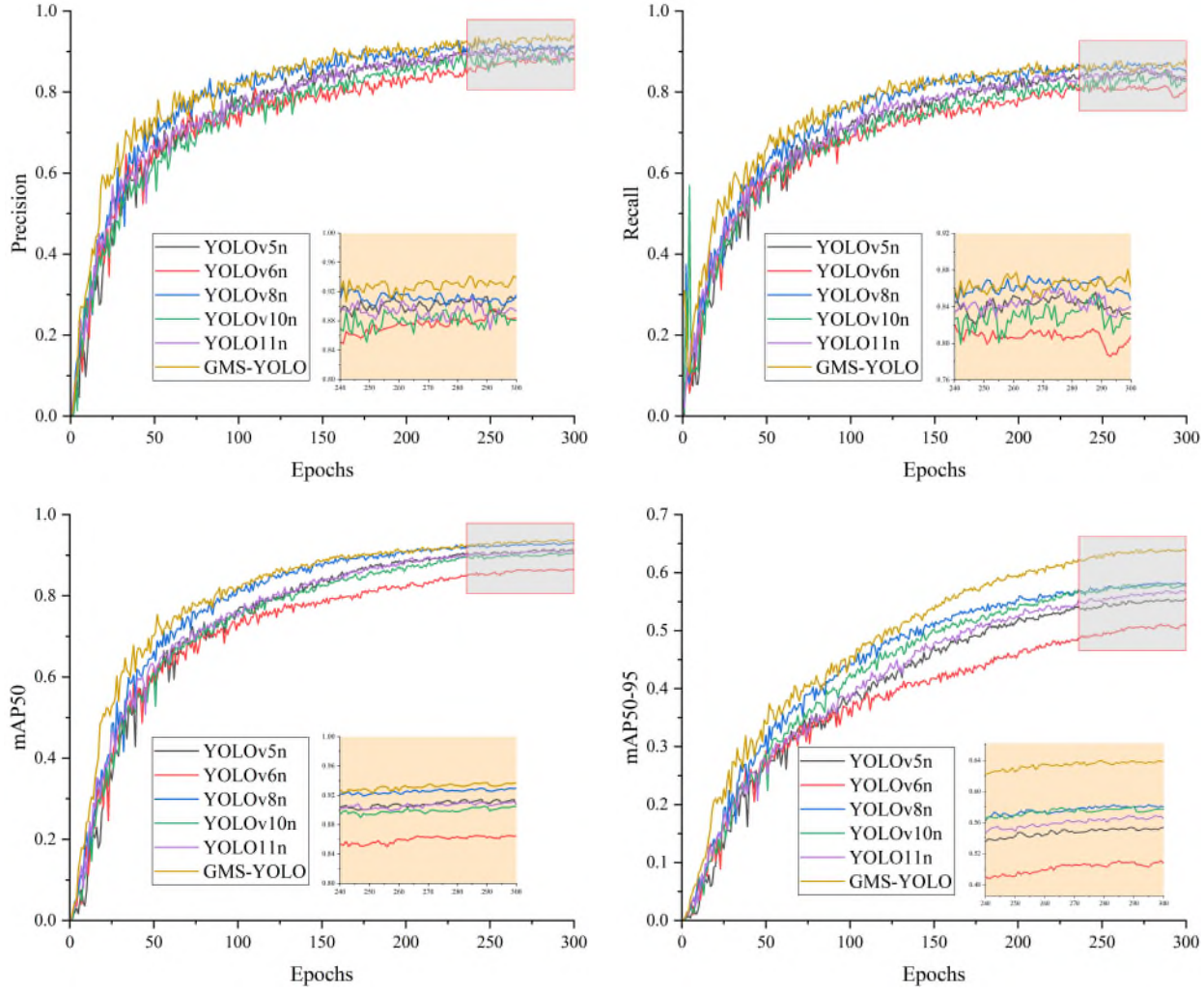


Figure 10. Variation trend of the accuracy of different models

3.5. Feature Visualisation Analysis in Natural Environments

Figure 11 illustrates the comparative performance of different detection models in spike detection (missed detections marked with red circles). As evident from Figure 11, under leaf-obscured conditions, the YOLOv5n, YOLOv6n, YOLOv8n, YOLOv10n, and YOLO11n models missed 5, 6, 5, 7, and 9 tassels, respectively, while the GMS-YOLO model missed 2 tassels. Under low-light and high-light conditions, the other models missed 2, 1, 1, 3, and 6, and 4, 4, 5, 2, and 5 tassels,

respectively, whereas the GMS-YOLO model detected all tassels without omission. Combining Table 6 and Figure 13 reveals that the enhanced GMS-YOLO model significantly improves tassel detection coverage. It not only captures more marginal and small-scale tassels but also positions detection boxes more accurately against targets, markedly reducing missed detections and localisation errors.

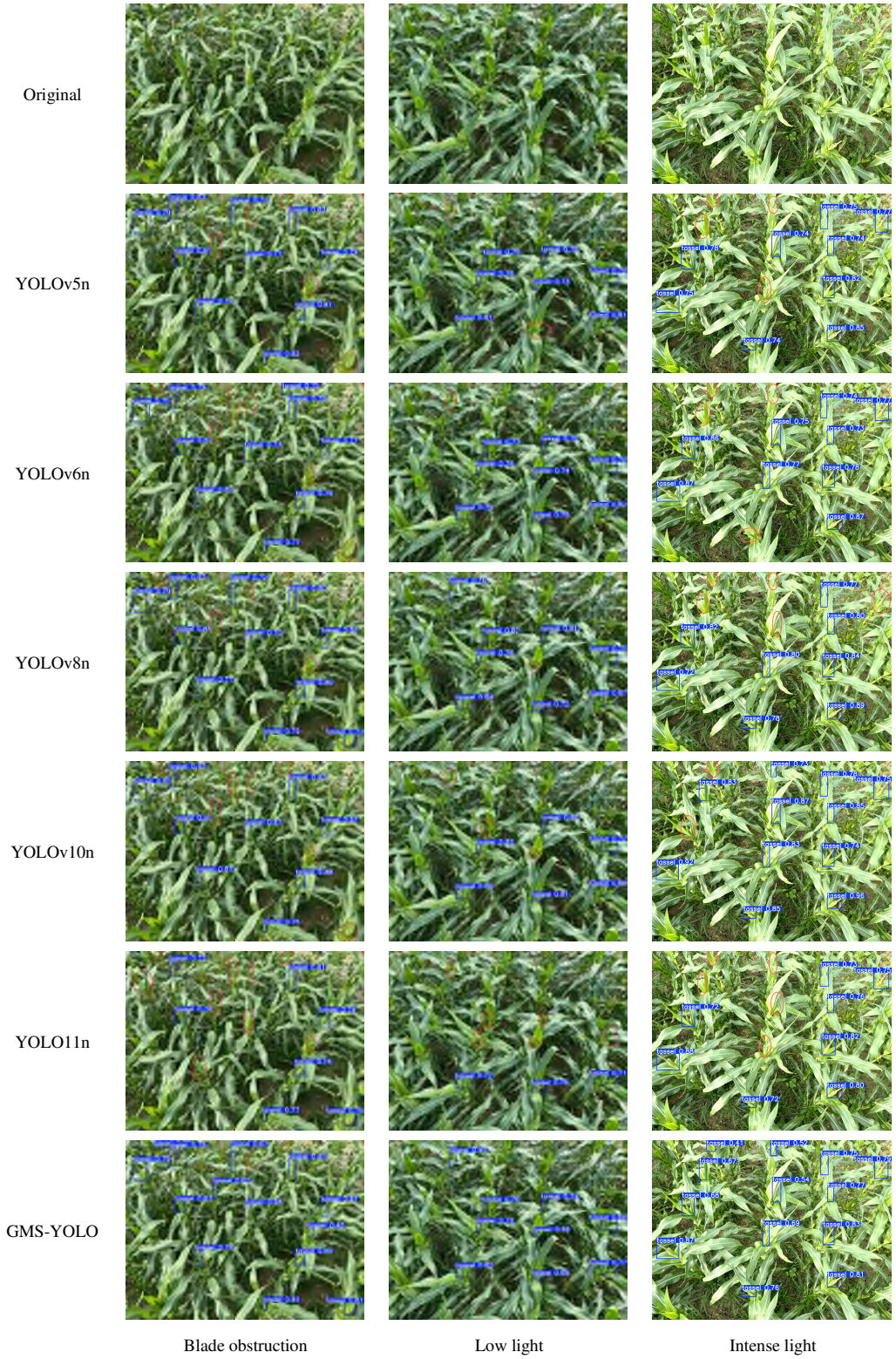


Figure 11. Comparison of Detection Effects of the Tassel Detection Model on the Test Set

4. Design of a Real-time Maize Tassel Detection System

4.1. System Construction

The Raspberry Pi 5 serves as the control core of the castration device, running the YOLO algorithm to detect and identify maize tassels in real-time from video streams captured by the camera. The system connects to a high-resolution camera module via USB, leveraging the bus's versatility to rapidly integrate imaging devices and enable real-time data transmission. Regarding the algorithm development process, the GMS-YOLO model was first constructed and trained on a local deep learning workstation equipped with an NVIDIA GPU. Subsequently, the trained model was deployed onto the Raspberry Pi embedded platform.

4.2. System Overview

The embedded main control board of this system utilizes a Raspberry Pi 5 as the edge computing chip for the upper-level control unit. This device features dual-band Wi-Fi, Bluetooth 5.0, and Gigabit Ethernet capabilities, enabling high-speed data transmission via its USB 3.0 interface. Its 40-pin GPIO interface provides exceptional control capabilities and expandability for system actuators.

Visual data acquisition utilizes a high-frame-rate USB industrial camera. This camera connects to the Raspberry Pi via USB 3.0, transmitting real-time video streams of maize plants in the field to the main control unit. The system can be powered through a Type-C interface connected to an independent power supply module. The overall hardware integration solution is illustrated in Figure 12.

4.3. System Settings

The improved GMS-YOLO model described in this paper is deployed on a Raspberry Pi 5 device. After completing the hardware configuration, the operating system must be flashed onto the storage medium. For this purpose, the official 64-bit Raspberry Pi operating system based on Debian 12 (codenamed bookworm) was selected and flashed onto a high-speed MicroSD card using the Raspberry Pi burning tool, as shown in Figure 13.

4.4. Model Deployment

To evaluate the performance of the deployed models, both the original YOLOv11n and the improved GMS-YOLO were tested on Raspberry Pi hardware. Four video clips recorded with a mobile phone in a maize test field were selected for testing. The total video duration was 20 seconds, with a resolution of 1920×1440 pixels. The videos were downsampled to 640×640 pixels before inputting them into the models. The detection results comparison is shown in Table 5. Compared to the original model, the improved model demonstrated a significant advantage in detection speed. The frame rate on the Raspberry Pi 5 increased from 11.3 FPS to 13.92 FPS, representing a 23.19% improvement. Figure 14 shows four frames extracted from the detection video. Deploying this model on Raspberry Pi hardware enables real-time detection of maize tassels. The system effectively identifies tassel models with low computational cost and minimal memory usage, making it suitable for resource-constrained embedded devices.

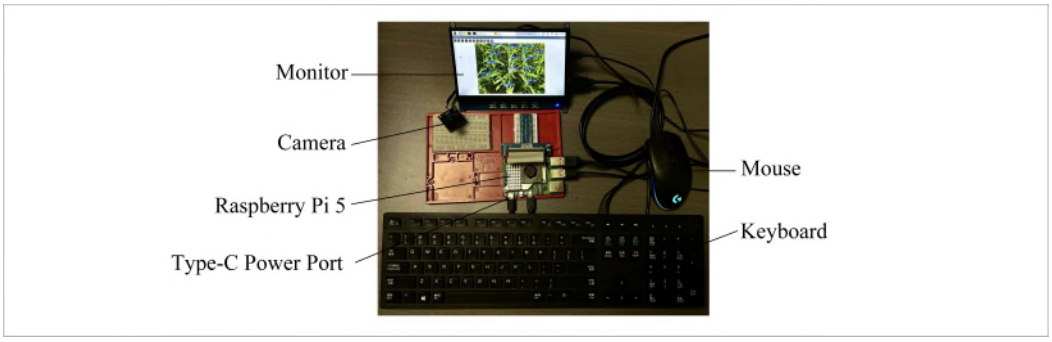


Figure 12. System Hardware

Table 5. Model Deployment Detection Results

Model	Parameters/M	FLOPs/G	Model size/MB	Speed(FPS)
YOLO11n	2.67	6.5	5.6	11.30
GMS-YOLO	2.44	5.8	4.9	13.92

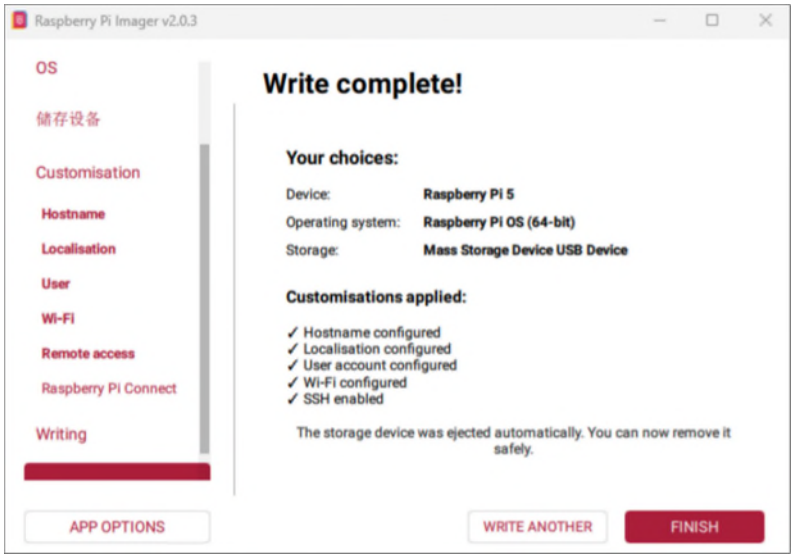
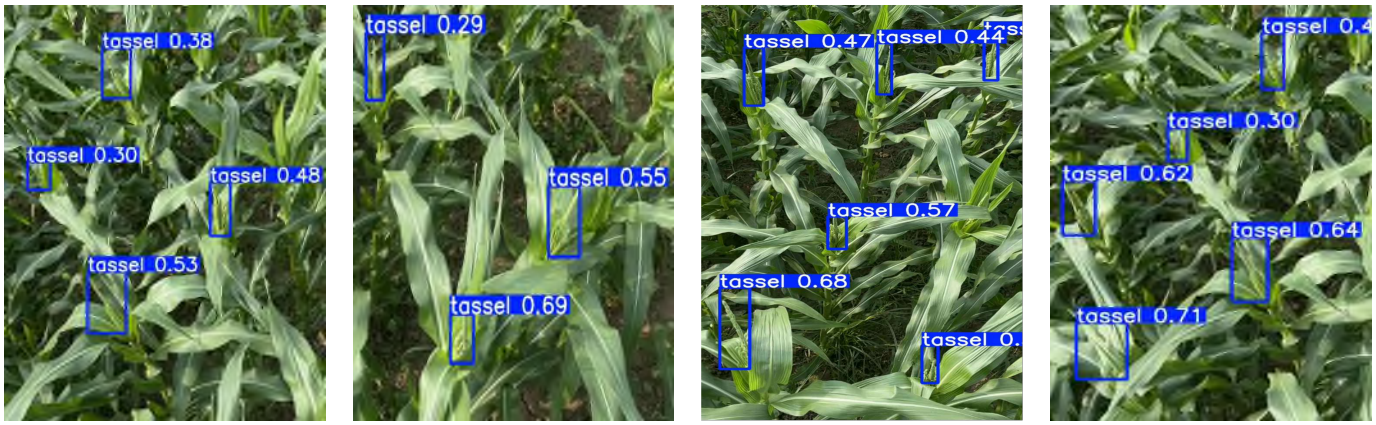


Figure 13. Raspberry Pi System Installation

Figure 14. Video Image Detection



5. Conclusion

This paper proposes a lightweight real-time object detection model, GMS-YOLO, for early-stage maize tassel detection. The model's key improvements encompass three aspects: The C3GhostV3 module replaces the original convolutional module in the backbone network. By generating redundant feature maps through low-cost linear transformations, it significantly reduces the model's parameter count and computational complexity while preserving feature representation integrity. The Multi-Scale Efficient Channel Attention (MSECA) module is introduced. Parallel multi-scale convolutional kernels enhance the model's dynamic fusion capability for high-level semantics and low-level details across different tassel scales, effectively reducing missed detections caused by complex leaf occlusions. Integrating the C2GAM global attention mechanism and adopting the SIoU loss function further amplifies the global interaction of effective features, suppresses irrelevant background noise interference, and simultaneously optimizes the accuracy of bounding box regression. Extensive experiments on a maize tassel dataset constructed in natural field environments demonstrate that the improved GMS-YOLO model achieves a mean average precision (mAP) of 93.4% and a precision rate of 94.2%. Compared to the original YOLOv1n model, this represents a 3.7 percentage point increase in mean average precision and an 8.61% reduction in parameter count. Inference speed on a Raspberry Pi 5 edge device reached 13.92 FPS. These results

fully demonstrate the effectiveness of the proposed method in balancing detection accuracy with lightweight computational efficiency.

Although our model achieves significant improvements in early-stage maize tassel detection, certain limitations remain. For instance, the training data in this study is restricted to the "Fengnuo 168" variety collected in Chuzhou City, potentially limiting the model's generalization capability for other maize varieties with distinct morphological features. Furthermore, the current model may still face challenges of misclassification or missed detection under extremely adverse weather conditions or when leaves exhibit severe overlapping occlusion. Addressing these limitations will be a key direction for future research. Future work will focus on enriching the dataset diversity, planning to include tassel image data from different geographical regions, more maize varieties, and various growth stages to further enhance the model's robustness and generalization capability in complex environments. Furthermore, integrating with the design and research of maize detasseling equipment for seed production, we plan to deploy the optimized GMS-YOLO model onto actual automated detasseling agricultural machinery and high-end chassis. This aims to validate and enhance the real-time performance and precise collaborative operation capabilities of this visual perception system in authentic field working environments.

6. Discussion

A lightweight identification method for early maize tassels, GMS-YOLO, based on improved YOLO11n, is proposed in this study and achieves excellent detection results on a field-collected dataset. The model optimizes the original YOLO11n architecture by integrating the C3GhostV3 module and the C2GAM global attention mechanism, while replacing the standard convolutions with GhostConv. These optimizations significantly improve the model’s Precision (P) and mAP, effectively reducing the number of parameters and model weight, thereby achieving an ideal balance between lightweight design and accuracy.

Firstly, the C3GhostV3 module replaces the C3k2 structure in the original backbone. By utilizing low-cost linear transformations to generate redundant feature maps based on GhostNetV3 theory, this module addresses the parameter redundancy typical of standard convolutional layers. Experimental data indicate that this improvement achieves a 20.2% reduction in parameter count compared to the original YOLO11n backbone, preserving feature representation integrity while enabling edge deployment. To address the challenges of "visual camouflage" and scale variation in early tassels, the MSECA (Multi-Scale Efficient Channel Attention) module is embedded in the Neck section. By employing parallel multi-scale kernels, MSECA enables the model to capture high-frequency texture details (e.g., tassel axis bifurcation) through small receptive fields while aggregating broader contextual information through large receptive fields to infer boundaries under leaf occlusion. Furthermore, the introduction of the C2GAM module at the backbone's terminal layer employs a 3D feature retention architecture to amplify global interaction and suppress background noise.

In the model comparison, GMS-YOLO works well for finding early maize tassels. It reached a Precision of 94.2% and an mAP of 93.4%. These results are 6.1% and 4.1% higher than the standard YOLO11n. The model size is only 4.9MB, so it can be used on farm robots or drones. However, there are some limits. Our data comes from the Fengnuo 168 variety in Chuzhou City. The model might not work as well for other maize types with different shapes. Also, very bad weather or leaves covering each other can still cause some mistakes in the field.

In future work, we will add more maize types and growth stages from different areas to our dataset. We also plan to put GMS-YOLO into real detasseling machines. This will help us test how well the system works in real farm environments.

References

1. You, N.; Dong, J.; Li, J.; Huang, J.; Jin, Z. Rapid early-season maize mapping without crop labels. Remote Sensing of Environment. **290**, 113496. (2023)

2. Gage, J.L.; Miller, N.D.; Spalding, E.P.; Kaeppler, S.M.; De Leon, N. TIPS: a system for automated image-based phenotyping of maize tassels. Plant Methods. **13**. (2024)
3. Gambin, B.L.; Trifunovic, S.; Archontoulis, S.V. Increased yield stability of U.S. Midwest maize hybrids released during the last 40 years of breeding. Field Crops Research. **333**, 110092. (2025)
4. Lu, H.; Cao, Z.; Xiao, Y.; Fang, Z.; Zhu, Y.; Xian, K. Fine-grained maize tassel trait characterization with multi-view representations. Computers and Electronics in Agriculture. **118**, 143–158. (2015)
5. Zhang, R.; Yang, J.; Chen, L.; Ding, C.; Li, L.; Zhang, L. Review of the detasseling techniques for maize (Zea mays L.) hybrid seed production. International Journal of Agricultural and Biological Engineering. **17**, 1–11. (2024)
6. Che, Y.; Wang, Q.; Li, S.; Li, B.; Ma, Y. Monitoring of maize phenotypic traits using super-resolution reconstruction and multimodal data fusion. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE). **37**, 169–178. (2021)
7. Liu, Y.; Cen, C.; Che, Y.; Ke, R.; Ma, Y.; Ma, Y. Detection of Maize Tassels from UAV RGB Imagery with Faster R-CNN. Remote Sensing. **12**, 338. (2024)
8. Wang, B.; Yang, G.; Yang, H.; Gu, J.; Zhao, D.; Xu, S.; Xu, B. UAV images for detecting maize tassel based on YOLO_X and transfer learning. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE). **38**, 53–62. (2022)
9. Li, J.; Du, J.; Zhang, H.; Guo, X.; Zhao, C. Maize Tassel Detection Algorithm after Artificial Emasculation Based on Lightweight MLCE - RTMDet. Transactions of the Chinese Society for Agricultural Machinery. **55**, 184–192, 503. (2024)
10. Ma, Z.; Luo, C.; Luo, W.; Wang, L.; Fang, X.; Li, H. Tassel Detection Method of Maize Germplasm Resources Based on Improved YOLO v7-tiny. Transactions of the Chinese Society for Agricultural Machinery. **55**, 290–297. (2024)
11. Chen, P.; Wang, X.; Zhang, J.; Wang, B.; Liang, D. Multi-varieties maize tassels detection method based on improved YOLOv9m. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE). **41**, 122–131. (2025)
12. Wei, J.; Wang, R.; Wei, S.; Wang, X.; Xu, S. Recognition of Maize Tassels Based on Improved YOLOv8 and Unmanned Aerial Vehicles RGB Images. Drones. **8**, 691. (2024)
13. Song, C.; Zhang, F.; Li, J.; Xie, J; Yang, C.; Zhou, H.; Zhang, J. Detection of maize tassels for UAV remote sensing image with an improved YOLOX Model. Journal of Integrative Agriculture. **22**, 1671–1683. (2022)
14. Chen, J.; Fu, Y.; Guo, Y.; Xu, Y.; Zhang, X.; Hao, F. An improved deep learning approach for detection of maize tassels using UAV-based RGB images. International Journal of Applied Earth Observation and Geoinformation. **130**, 103922. (2024)
15. Teng, G.; Yan, M.; Yao, J.; Li, J. Field wheat ear recognition and counting based on the improved YOLOv7 model. Journal of Chinese Agricultural Mechanization. **46**, 77-85. (2025)
16. Shen, Y.; Yang, Z.; Khan, Z.; Liu, H.; Chen, W.; Duan, S. Optimization of Improved YOLOv8 for Precision Tomato Leaf Disease Detection in Sustainable Agriculture. Sensors. **25**, 1398. (2025)
17. Peng, G.; Wang, K.; Ma, J.; Cui, B.; Wang, D. AGRI-YOLO: A Lightweight Model for Maize Weed Detection with Enhanced YOLO v11n. Agriculture. **15**. (2025)
18. Jia, L.; Yuan, X.; Chen, Z.; Wang, T.; Gao, L.; Gu, G.; Wang, X.; Wang, Y. TSE-YOLO: A Model for Tomato Ripeness Segmentation. Agriculture. **16**. (2025)
19. Du, X.; Zhang, X.; Li, T.; Chen, X.; Yu, X.; Wang, H. YOLO-WAS: A Lightweight Apple Target Detection Method Based on Improved YOLO11. Agriculture. **15**. (2025)
20. Li, M.; Zhang, J.; Liu, H.; Yuan, Y.; Li, J.; Zhao, L. A lightweight method for apple-on-tree detection based on improved YOLOv5. SIViP. **18**, 6713–6727. (2024)
21. Kapłan, M.; Wójcik, D.I.; Buczyński, K. Detection and Segmentation of Chip Budding Graft Sites in Apple Nursery Using YOLO Models. Agriculture. **15**. (2025)
22. Niu, Q.; Ma, W.; Diao, R.; Yu, W.; Wang, C.; Li, H.; Wang, L.; Li, C.; Wang, P. Research on Recognition of Green Sichuan Pepper Clusters and Cutting-Point Localization in Complex Environments. Agriculture. **15**. (2025)
23. Zhang, X.; Fan, J.; Guo, X.; Zhao, C. Research on male ear detection and tasseling stage identification of field maize based on faster R-CNN. Journal of Anhui Agricultural University. **48**, 849–856. (2021)
24. Li, S.; Zheng, C.; Wen, C.; Li, K. Method for Locating Missing Ratoon Sugarcane Seedlings Based on RGB Images from Unmanned Aerial Vehicles and Improving YOLO v5s. Transactions of the Chinese Society for Agricultural Machinery. **55**, 57–70. (2024)
25. Liu, S.; Zhang, W.; Hu, X.; Wang, L.; Song, Z.; Wang, J. Rice Planting Machinery Operation Quality Detection Based on Improved YOLO v8s. Transactions of the Chinese Society for Agricultural Machinery. **55**, 61–70. (2024)
26. Li, M.; Xiao, Y.; Zong, W. Multi-scale Chestnut Detection Method Based on Improved YOLO 11 Model. Transactions of the Chinese Society for Agricultural Machinery. **56**, 443–454. (2025)
27. Chollet, F. Xception: Deep Learning With Depthwise Separable Convolutions. 1251–1258.(2017)

28. Liu, Z.; Hao, Z.; Han, K.; Tang, Y.; Wang, Y. GhostNetV3: Exploring the Training Strategies for Compact Models. (2024)
29. Ye, B.; Feng, R.; Tang, T.; Hu, S.; Xue, H.; Yu, G. Ripeness detection of lotus seedpod in natural environment based on improved YOLOv10n. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE). **41**, 145–153. (2024)
30. Qiao, C.; Han, Z.; Zhang, Y.; Han, M.; Zhang, L. Method for Identifying Infection Structures of Cucumber Downy Mildew Based on Improved YOLO v7 Model. Transactions of the Chinese Society for Agricultural Machinery. **56**, 528–537. (2025)
31. Zhang, H.; Dai, C.; Ren, J.; Wang, G.; Teng, F.; Wang, D. Research on Grape Disease Identification Method Based on GhostNetV2 Improved YOLO v8 Model. Transactions of the Chinese Society for Agricultural Machinery, 1–11. (2024)
32. Liu, Y.; Shao, Z.; Hoffmann, N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. (2021)
33. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 11534–11542. (2024)
34. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. Proceedings of the AAAI Conference on Artificial Intelligence. **34**, 12993–13000. (2020)
35. Gevorgyan, Z. SIoU Loss: More Powerful Learning for Bounding Box Regression. (2022)

Author Contributions: Conceptualization, W.W.; methodology, W.W.; software, Z.W.; validation, C.J.; formal analysis, C.Q.; investigation, C.Q.; resources, W.W. and X.Z.; data curation, Z.W. and L.W.; writing—original draft preparation, W.W.; writing review and editing, X.Z. and C.J. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Anhui Provincial Key R&D Program (202104a06020001), the Key Project of Natural Science Research of Anhui Province (No. 2024AH050321).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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