

Article

A Phenotyping Perception Mechanism of Fusing Spatial and Channel Reconstruction Convolution Employing Maize-Breeding UAV Visual Images

Huanzhe Wang ¹, Jian Chen ^{1,*}, Xiqing Wang ² and Shuaisong Zhang ^{2,3}¹ Embodied AI and Agricultural Robotics Center, College of Engineering, China Agricultural University, Beijing 100083, China² College of Biological Sciences, China Agricultural University, Beijing 100193, China³ Shandong Key Laboratory of Intelligent Agricultural Equipment in Hilly and Mountainous Areas, Jinan 250100, China

* Correspondence: jchen@cau.edu.cn

Highlights

What are the main findings?

- A phenotyping perception approach fusing spatial and channel reconstruction convolutions was proposed. With a model size of 5.16 MB, the method achieved a maize tassel detection mAP@0.5 of 91.7%.
- A quadrotor UAV for breeding was built, and visual image data of maize tassel phenotypes in actual breeding fields was collected.

What is the implication of the main finding?

- It realizes a balance between the accuracy of the detection model and the edge deployment capability of UAVs, providing a technical reference for maize tassel detection algorithms.
- It offers a methodological basis for acquiring crop phenotypic data using UAV visual perception technology, and has application potential in maize-breeding work.



Academic Editors: Inkyu Sa and Fei Liu

Received: 24 September 2025

Revised: 18 November 2025

Accepted: 22 November 2025

Published: 30 November 2025

Citation: Wang, H.; Chen, J.; Wang, X.; Zhang, S. A Phenotyping Perception Mechanism of Fusing Spatial and Channel Reconstruction Convolution Employing Maize-Breeding UAV Visual Images. *Drones* **2025**, *9*, 830. <https://doi.org/10.3390/drones9120830>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract

As precision agriculture advances, UAV-based aerial image object detection has emerged as a pivotal technology for maize-phenotyping perception operations. Complex backgrounds reduce the model's performance in extracting features of maize tassels, while sacrificing model computation complexity to improve feature expression is detrimental to deployment on UAVs. To achieve a balance between the model size and deploy ability, an enhanced model incorporating spatial-channel convolution is proposed. First, a maize-breeding UAV was built, and the collection of maize tassel image data was realized. Second, Spatial and Channel Reconstruction Convolution (SCConv) was integrated into the neck network of the YOLOv8 baseline model, reducing the model computation complexity while maintaining the detection accuracy. Finally, the constructed maize tassel dataset and public Maize Tasseling Stage (MTS) dataset were used for the training and evaluation of the enhanced model. The results showed that the enhanced model achieved a precision of 92.2%, recall of 84.3%, and mAP@0.5 of 91.7%, with 7.3 G floating-point operations (FLOPs) and a model size of 5.16 MB. Compared with the original model, the enhanced model exhibited respective increases of 3.2%, 3.4%, and 3.4% in precision, recall, and mAP@0.5, along with respective reductions of 0.8 G FLOPs in computation complexity and 0.79 MB in model size. Compared with YOLOv10n, the precision, recall, and mAP@0.5 of the enhanced model are increased by 1.8%, 3.1%, and 2.9%, respectively, and the model computation is

reduced by 0.3 G FLOPs, and the model size is reduced by 0.42 MB. The improved model is accurate, performs better on UAV aerial images in complex scenarios, and provides a methodological basis for deployment. It also supports maize tassel detection and holds potential for application in maize breeding.

Keywords: precision agriculture; maize-breeding UAV; phenotyping perception; object detection; Spatial and Channel Reconstruction Convolution

1. Introduction

With the global advancement in agricultural modernization, precision agriculture has become a core method for breaking traditional farming bottlenecks and achieving high-quality agricultural development. Its core logic relies on intelligent perception and data-driven decision-making for refined crop growth cycle regulation, significantly boosting production and resource utilization efficiency [1,2]. As the main form of low-altitude economy, UAVs are expanding and replicating their industrial model to agriculture and other application fields, giving new kinetic energy to the development of the agriculture and forestry industries [3,4]. Crop images taken aerially by low-altitude UAVs provide farmland data, which is conducive to promoting the construction of agricultural informatization [5]. As key carriers for on-field precision agriculture, maize breeding Unmanned Aerial Vehicles (UAVs), specialized for specific crop scenarios, have replaced manual labor in repetitive, labor-intensive tasks (e.g., on-field monitoring, and phenotypic collection), shortening maize-breeding cycles and supporting elite variety selection [6].

As a significant crop, maize breeding is directly related to food safety and agricultural sustainability [7]. Maize-phenotyping perception, critical in breeding, aims to accurately acquire traits related to yield and stress resistance, e.g., tassel branches, plant height dynamics, and leaf disease severity [8]. Maize tassels are closely related to their yield and quality, which are important phenotypes and evaluation indicators in the breeding process [9]. Traditional methods have limitations: manual measurement is inefficient and subjective, while ground sensors lack coverage for field-scale phenotypic heterogeneity. In contrast, the UAV aerial image detection method based on computer vision has become a mainstream approach in maize-breeding phenotyping, leveraging the advantages of non-destructiveness, large-scale coverage, and real-time performance [10]. Breeding UAVs, as a kind of equipment for the automatic collection and detection of crop phenotypes, provide an intelligent solution for intelligent breeding tasks, and become a key vehicle and way to discover and create superior maize germplasm resources [11].

UAV aerial object detection is a key link for UAVs to effectively realize low-altitude operation [12,13]. The accurate target recognition technology of UAVs supports agricultural development progress. As the core environmental perception tool of UAVs [14], object detection determines phenotyping accuracy, forming the basis of UAV-based phenotyping systems. Deep-learning methods, with adaptive feature learning and complex scenario extraction capabilities, outperform traditional handcrafted feature-based detection in accuracy and generalization [15,16]. For complex field environments with dynamic light, plant overlap, and weed mixing, single-stage algorithms like the You Only Look Once (YOLO) series [17], which are fast and end-to-end, show broad prospects in breeding UAVs, such as maize tassel recognition [18].

However, phenotyping models face bottlenecks in maize-breeding UAVs. The background of UAV aerial images features a high complexity in terms of light and occlusion. And existing models boost performance via more layers and feature map dimensions,

increasing the computation complexity and model size [19]. Constrained by the UAV on-board hardware, these models are hard to deploy, disconnecting advanced algorithms from field needs [20]. Thus, optimizing YOLO-based UAV models for a lightweight design while ensuring the detection accuracy is theoretically and practically significant [21], especially for maize phenotyping requiring a high precision and on-site deployment.

Factors such as light and occlusion in complex backgrounds make it difficult to extract features of maize tassels, while the approach of increasing the model size to enhance feature expression is not conducive to deployment on UAVs. To address the balance between detection accuracy and model lightweighting, while resolving the contradiction between efficiency and deployment, this study proposes an improved YOLO-based maize-phenotyping method integrating Spatial and Channel Reconstruction Convolution (SCConv). This method effectively reduces the feature redundancy of the model in the spatial and channel dimensions while ensuring the model's ability to extract maize tassel features. The optimized UAV aerial object detection model has a reduced computational complexity and model size, providing a lightweight and high-precision tool for crop-phenotyping detection and supporting the development of precision agriculture in the breeding field. Additionally, aerial maize tassel data are acquired for model training, providing a methodological reference for UAV-based object detection tasks and facilitating the deployment and application of the model on UAV platforms. The specific research schemes adopted in this paper include the following:

- a. A quadrotor UAV platform for maize breeding was built, which could acquire aerial data of crop phenotypes and realize the effective collection of maize tassel images;
- b. The SCConv module was introduced into the YOLOv8 model to construct a crop-phenotyping perception model, reducing the feature redundancy of the model in the spatial and channel dimensions;
- c. The improved model was trained and evaluated using the constructed maize tassel dataset, providing a methodological reference for the task of perceiving UAV-based maize-breeding phenotypes.

2. Related Works

With the advantages of end-to-end processing and a fast detection speed, YOLO series models have become the mainstream framework for UAV aerial object detection. To meet the practical demands in maize-breeding detection scenarios, existing studies have carried out improvement work focusing on YOLO models. Lightweight convolution provides an efficient solution for UAV detection tasks. As the core means for achieving efficient deployment on hardware, lightweight technology aims to reduce the computation complexity and model size while maintaining the detection accuracy as much as possible. Falahat et al. [22] proposed an improved YOLOv5n by using the Gaussian Error Linear Unit (GELU) activation function, adding depth-wise convolution, and introducing a spatial-dropout layer, which effectively prevents overfitting and improves the learning ability for complex features. The Maize Tassel Detection Counting (MTDC) dataset serves as the basis for model training and evaluation. It indicates that the improved model has the advantage of rapidity in the task of UAV aerial object detection and counting. Niu et al. [23] proposed a UAV aerial images object recognition and counting method utilizing the YOLOv8 framework. The model achieves a lightweight design and fast detection by using the Partial Convolution (PConv) module and improves the feature extraction capability by adding the mixed module integrating self-attention and convolution, named ACmix. The model also integrates the Convolution with Transformer Attention Module (CTAM) to improve the semantic exchange and realize the efficient and accurate detection of aerial targets. Pu et al. [24], addressing the insufficiency of existing models in speed and

accuracy for real-time recognition, introduced the Global Attention Mechanism (GAM) to the YOLOv7 model, adding the Group Shuffle Convolution (GSCConv) and Variety of View Group Shuffle Cross Stage Partial (VoV-GSCSP) modules, replacing the original loss function with Scaled Intersection over Union (Siou), and proposed Tassel-YOLO for the automatic detection and counting of UAV aerial targets. Yu et al. [25], in response to the diversity of appearances resulting from individual differences, proposed a UAV-based model called PConv-YOLOv8, which provides a method reference for aerial object detection tasks.

The attention mechanism, which helps models focus on key features and suppress background interference, is a critical technology for improving the detection accuracy [26]. For detection tasks in complex scenarios such as occlusion and varying lighting, the attention mechanism effectively enhances the model detection performance under complex conditions. Guang et al. [27], for the fast detection problem in the real natural environment, added enhanced deformable convolution and a bi-level routing attention mechanism to construct DBi-YOLOv8, which was used in the real complex operational environment. Jia et al. [28], in response to the limitations of the conventional manual detection approach, such as the considerable subjectivity, substantial workload, and poor timeliness, use UAVs to collect aerial images and build a dataset, and construct an aerial object detection model named CA-YOLO by adding a Coordinate Attention (CA) mechanism to YOLOv5. It indicated that the regression localization ability has been enhanced and the detection capability has been improved. Chen et al. [29], for the problems of UAV in accurately detecting overlapping aerial targets and the insufficient detection ability under bright light conditions, constructed the aerial object detection network RESAM-YOLOv8n by introducing the Residual Scale Attention Module (RESAM) and using large-size aerial image data. The enhanced model focuses more intently on target features, ignores irrelevant redundant information, and improves the detection performance.

With the strong global semantic modeling capability, the Transformer provides a new direction for improving the accuracy of UAV detection models. Currently, relevant studies mainly combine Transformer modules with traditional convolutional models to balance the model's ability to capture global semantics and extract local features [30,31], better meeting the application requirements of UAV aerial object detection. For the detection task of maize tassels and leaves, Yang et al. [32] optimized the Kernels Network (K-Net) framework using the Transformer backbone and dynamic loss reweighting strategy. They proposed the Diffusion-enhanced K-Net for Tassels and Leaves (DiffKNet-TL) model, which improves the model's ability to capture dynamic phenotypes across scales and time. Occlusion and interference in complex backgrounds disrupt the continuity of local features. The global dependency of the self-attention mechanism leads to the loss of shallow-layer detail features. These factors cause networks based on the Real-Time Detection Transformer (RT-DETR) to have detection biases. To overcome these limitations in UAV object detection, Zhu et al. [33] proposed the Multi-Scale Maize Tassel-RTDETR (MSMT-RTDETR) model by optimizing the RT-DETR network and feature fusion mechanism, which achieves a stable detection performance for maize tassels. Based on the YOLOv8 model, Tang et al. [34] introduced the Swin Transformer module to improve the ability of extracting semantic features. Meanwhile, they adopted the Bidirectional Feature Pyramid Network (BiFPN) to enhance the accuracy of multi-scale target detection, constructing the STB-YOLO network for object detection in aerial images of maize tassels.

Some studies have also investigated other types of models within the YOLO series, expanding and exploring the adaptability of different models in object detection scenarios [35,36]. This provides more flexible solutions for the application of YOLO models in breeding UAV detection tasks. Lupión et al. [37] built ThermalYOLO by fine-tuning

YOLOv3, which has a higher bounding box recognition rate and Intersection over Union (IoU) value to accurately achieve the 2D pose estimation from multi-view thermal images. Instead of the Path Aggregation Network (PANet), Doherty et al. [38] used BiFPN to propose a YOLOv5-based detection algorithm called BiFPN-YOLO. Song et al. [39], for the occlusion phenomenon and diversity problems in aerial images, used UAVs to obtain aerial data in different periods and construct the object detection model SEYOLOX-tiny to realize the accurate detection under real operating conditions. To improve the detection robustness, Tang et al. [40] proposed an anchor-free detection algorithm, Circle-YOLO, which accurately predicts and does not require the design of anchor parameters.

Beyond the mainstream technical directions, some research efforts have focused on non-YOLO models and anchor-free algorithms, specifically addressing maize tassel detection issues. These studies not only provide new research ideas for crop-phenotyping recognition but also offer diversified solutions for maize-breeding UAVs. Liu et al. [41] constructed an aerial target image dataset using UAVs and trained the object detection model using the Faster Region-based Convolutional Neural Network (Faster R-CNN), Residual Network (ResNet), and Visual Geometry Group Network (VGGNet) as the feature extraction networks, respectively, to construct an aerial object detection network model. Ye et al. [42] constructed the Maize Tassels Detection and Counting UAV (MTDC-UAV) dataset using the target images acquired by UAVs, and proposed an aerial object detection network fusing global and local information, called FGLNet, which effectively improves the model performance through weighted mechanisms. To address the balance between the model size and overall performance in existing detection tasks, Yu et al. [43] constructed a novel lightweight feature aggregation network for maize tassels called TasselLFANet using an Encoder, Decoder, and Dense Inference. The approach also introduced multiple receptive fields to capture diverse feature representations and integrated the Multi-Efficient Channel Attention (Mlt-ECA) module to enhance accurate feature detection.

3. Materials and Methods

3.1. Maize-Breeding UAV Construction

The UAV built in this study could effectively acquire image data of maize tassels and is applicable to crop phenotypic detection tasks. The experimental platform for maize-breeding UAVs is composed of five core units, mainly including the control system, navigation system, communication system, power system, and energy system. These units work in synergy to achieve precise flight control and task execution of the UAV. The system composition of the designed maize-breeding UAV is shown in Figure 1.

Figure 2 illustrates the overall structure of the maize-breeding UAV. The maize-breeding UAV achieves miniaturization while ensuring operational endurance. With an overall wheelbase of 350 mm and a weight of 1.4 kg, the breeding UAV maintains a flight endurance of 15–20 min. The selected on-board computer, measuring $89 \times 56 \times 1.6$ mm and weighing 62 g, uses the RK3588s multi-core processor as its main control chip. It could provide the breeding UAV with 6Tops of edge-computing power, offering computational support for the UAV's autonomous flight, real-time detection, and analysis. The on-board 1080P high-definition gimbal RGB camera, with dimensions of $49 \times 57.5 \times 47.1$ mm and a weight of approximately 55 g, is equipped with an adjustable dual-axis stabilized gimbal. It could stably capture high-definition phenotypic images of maize plants, facilitating subsequent phenotypic data analysis.

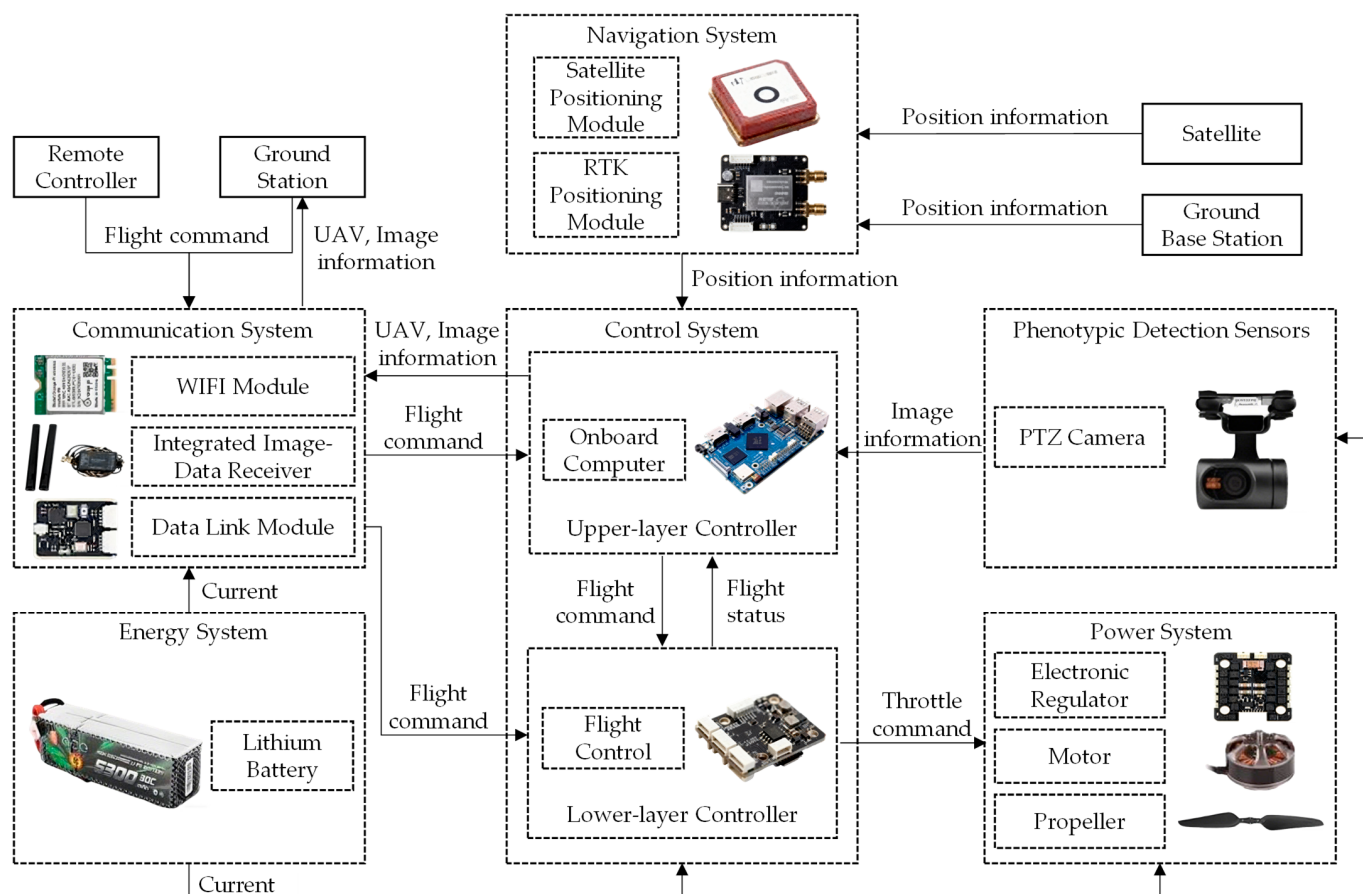


Figure 1. The hardware composition of the designed maize-breeding UAV for tassel detection.



Figure 2. The schematic diagram of the overall external structure of the maize-breeding UAV.

3.2. Maize Tassel Dataset Creation

To ascertain the efficacy in various application contexts including maize-breeding operations and phenotype sensing, this paper selected maize tassels as the detection target for training and evaluation. The maize data images were collected in Quzhou County, Handan City, Hebei Province, China ($114^{\circ}98' \text{ E}$, $36^{\circ}66' \text{ N}$). The experimental area has a total area of approximately 3000 square meters. This region has a temperate monsoon climate, characterized by hot and rainy summers, and cold and dry winters. The summer maize was sown on 15 June 2025, with the variety of MY73. The planting density follows a spacing of 18 cm between plants and 60 cm between rows.

The maize data were collected on 9 August 2025, with maize images captured at different time periods throughout the day. The maize data collection equipment is a maize-

breeding UAV equipped with a camera, and the camera resolution is 1920×1080 pixels. The maize-breeding UAV flew at a constant speed in manual mode, with the flight altitude set to 4 m, and the camera angle was adjusted to face straight forward and tilt downward at a certain angle. The UAV data acquisition process is shown in Figure 3. The UAV flew over each maize row along the field rows and covered the entire maize-breeding field in an “S”-shaped flight path. When the UAV flew above the maize plants, the camera was used to photograph the tops of the maize plants in order to obtain image data of maize tassels. The UAV flies along an “S”-shaped trajectory to achieve row-by-row coverage along the maize field without collection dead ends, while using a conical field of view to capture images of maize tassels.

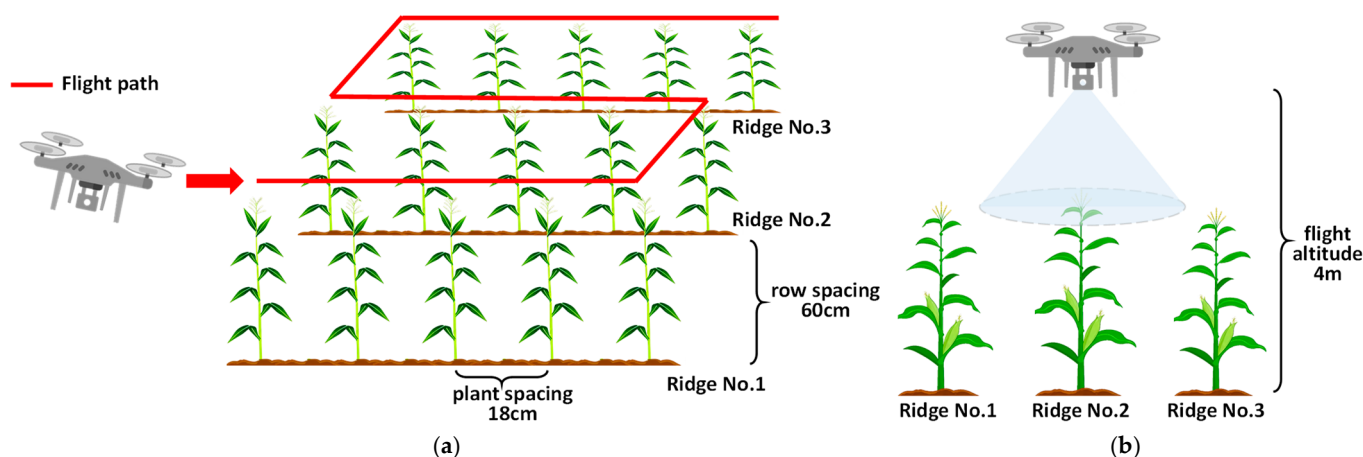


Figure 3. The collection process schematic of aerial tassels data by the maize-breeding UAV. Note: (a) represents the UAV flight path and the parameter of maize planting, and (b) represents the UAV flight altitude and the camera’s field of view.

To minimize the impact of maize plant oscillation, the image acquisition process was conducted in windless weather. The acquired maize tassels images exhibited a variety of real and complex operation scenarios, such as brightness, dimness, and partial occlusion. Since the original size of the images is not suitable for direct input into the model, this study uniformly scales the data to a size of 640×640 to ensure normal training.

For dataset acquisition, 500 images were randomly selected from the collected raw data. We employed the LabelImg annotation tool for labeling all the maize tassels in the acquired images. Meanwhile, we divided the dataset into training set, validation set, and test set according to the ratio of 8:1:1, which contained 400, 50, and 50 UAV aerial images, respectively. In addition to collecting field maize images independently, this study also adopted the Maize Tasseling Stage (MTS) dataset [44], which is a publicly available dataset (<https://github.com/islls/MTS-master-JX.git>, accessed on 3 November 2025). The MTS dataset has a wide range of data sources, covering different regions and years, and holds significant reference value in the field of maize tassel detection. This dataset consists of 7 image sequences containing maize tassels, selected from maize field scenarios in three provinces, which are Hebei, Henan, and Shandong. The sequences are Hebei2011, Hebei2014, Henan2011, Henan2012, Shandong2011_1, Shandong2011_2, and Shandong2012_2013, with a time span of 2011 to 2014. The MTS dataset contains a total of 160 maize images, including 90 for the training set, 30 for the validation set, and 40 for the test set, along with 5240 annotated instances of maize tassels. Since instances are what affect the loss value during model training, and for small-sized targets such as maize tassels, multiple instances exist in a single image. Therefore, in the maize tassel detection task, we should consider the instance-count level rather than only focusing on the image-count level.

The instance-count in the maize tassel dataset is shown in Table 1. Both our maize tassel dataset and MTS dataset contain only one type of instance. Figure 4 displays the dataset samples of maize tassel images. We used the constructed maize tassel dataset and the MTS dataset to train and evaluate the object detection model.

Table 1. The overview of our maize tassel dataset and MTS dataset.

Dataset	Our Maize Tassel Dataset		MTS Dataset	
	Image-Count	Instance-Count	Image-Count	Instance-Count
Training set	400	6882	90	3396 #
Validation set	50	871	30	1095 #
Test set	50	867	40	749
Total dataset	500	8620	160	5240

Note: “#” represents the instance-count in the training set and validation set divided by us, which means the value of the MTS dataset will vary with different division methods.

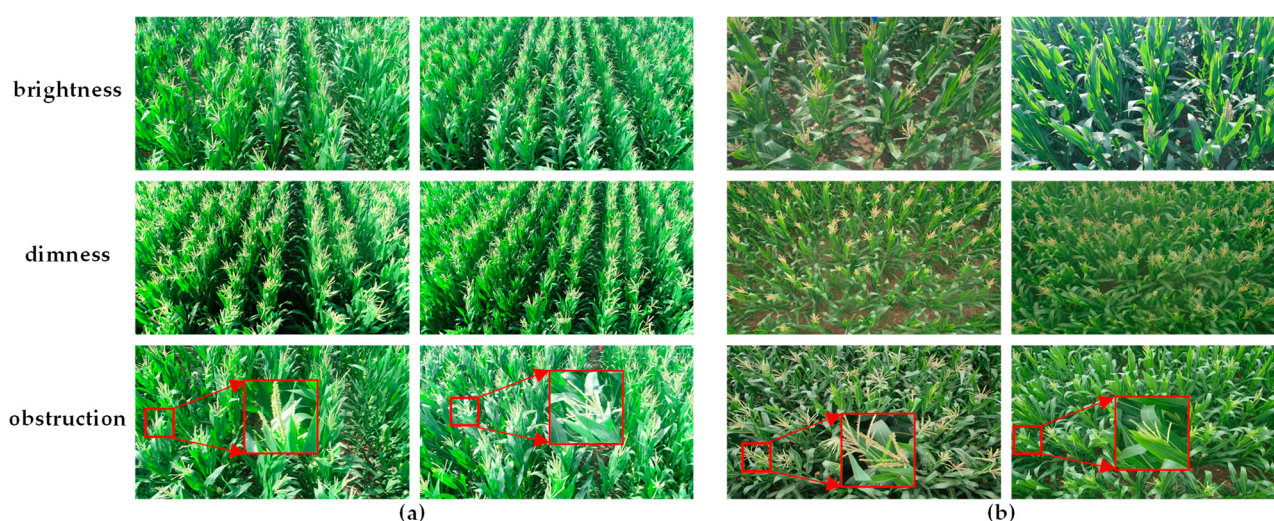


Figure 4. The dataset samples of the maize tassel images. Note: (a) represents the tassels obtained by the maize-breeding UAV in this paper, and (b) represents the tassels from the public MTS dataset.

3.3. Fusion Approach of SCConv for Improvement of the Neck Network

The YOLOv8 model [45] has been chosen as the fundamental model for object detection. Since the real detection background is usually complex, the existing model enhances performance at the expense of computation complexity and model size, which makes model large and not favorable to be deployed in UAVs. SCConv [46], which serves as an effective convolutional module, successfully minimizes the network redundancy of spatial and channel dimensions under the premise of model performance. Therefore, SCConv is applied to the baseline model.

3.3.1. Fusion Approach of Spatial Reconstruction Unit of SCConv

The components of SCConv mainly include the Spatial Reconstruction Unit (SRU) and the Channel Reconstruction Unit (CRU). The input feature is first transformed into spatial-refined feature by SRU, and then transformed into channel-refined features by CRU. The SRU module reduces input feature redundancy in spatial dimension by S Separation–Reconstruction strategy. The flowchart of SRU is shown in Figure 5.

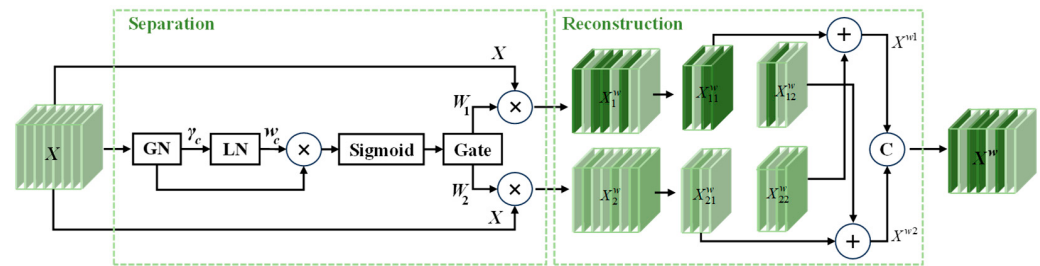


Figure 5. The flowchart of SRU for spatial feature reconstruction in maize tassel detection model.

Firstly, separation operation isolates feature maps containing more information from those containing less information to correspond to each other with the spatial content.

The information content in feature maps is evaluated by Group Normalization (GN) operation, and its calculation formula is as follows:

$$X_{\text{out}} = \text{GN}(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (1)$$

where X indicates input feature, and X_{out} indicates output feature of GN. μ and σ denote the mean and variance, respectively. γ and β both denote the hyperparameters. ε denotes the constant which is very small.

The trainable parameter γ_i measures the difference between pixels, i.e., the richness of spatial information, and then operated Layer Normalization (LN). It is calculated as follows:

$$W_\gamma = \{w_i\} = \frac{\gamma_i}{\sum_{j=1}^C \gamma_j}, i, j = 1, 2, \dots, C \quad (2)$$

where C denotes channel count.

The information and non-information weight of features, i.e., W_1 and W_2 , are obtained by setting the Gate threshold. The weight is calculated as follows:

$$W = \text{Gate}(\text{Sigmoid}(W_\gamma(\text{GN}(X)))) \quad (3)$$

where Sigmoid denotes the activation function.

Then, the reconstruction operation sums the features with more information and those with less information, which generates features and achieves space saving.

The obtained weights W_1 and W_2 multiply elements with X , respectively, to generate weighted features, which are merged by cross-reconstruction to obtain X_1^w and X_2^w . After connecting, the spatial-refined feature maps X^w are obtained. It is calculated as follows:

$$\begin{cases} X_1^w = W_1 \otimes X \\ X_2^w = W_2 \otimes X \\ X_{11}^w \oplus X_{22}^w = X^{w1} \\ X_{21}^w \oplus X_{12}^w = X^{w2} \\ X^{w1} \cup X^{w2} = X^w \end{cases} \quad (4)$$

where $\otimes$, $\oplus$, and $\cup$ denote multiplication, addition, and union operation, respectively.

3.3.2. Fusion Approach of Channel Reconstruction Unit of SCConv

The CRU reduces input feature redundancy in channel dimension by adopting the Split–Transformation–Fusion strategy. The flowchart of CRU is shown in Figure 6.

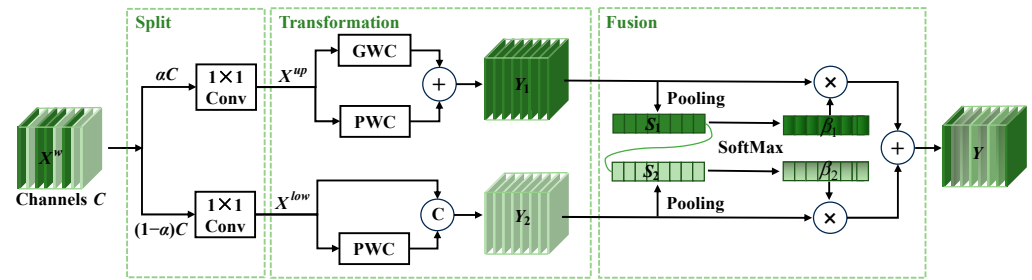


Figure 6. The flowchart of CRU for channel feature reconstruction in maize tassel detection model.

Firstly, the split operation takes the spatial-refined feature X^w output from the SRU as the input to the CRU, and splits this feature into two components. The channels of two features are compressed using 1×1 convolution to produce X^{up} and X^{low} , respectively.

Secondly, the transformation uses X^{up} for the rich feature extraction, which performs Group-Wise Convolution (GWC) and Point-Wise Convolution (PWC), respectively, summing the results to obtain Y_1 . The supplementary process takes X^{low} as input, conducts pointwise convolution, and combines the result with X^{low} to obtain Y_2 . The output is calculated as follows:

$$Y_1 = M^G X^{up} + M^{P1} X^{up} \quad (5)$$

$$Y_2 = M^{P2} X^{low} \cup X^{low} \quad (6)$$

where M^G and M^P denote the weight matrix of GWC and PWC, respectively.

Finally, the fusion operation is performed by combining Y_1 and Y_2 to generate Y . After Y_i undergoes the Pooling operation, it fuses global information to obtain S_i . Then, through the SoftMax operation, the feature weight vector β_i is generated, which is used to obtain the final refined-feature from CRU, i.e., Y . The Y and β_i is calculated as follows:

$$Y = \beta_1 Y_1 + \beta_2 Y_2 \quad (7)$$

$$\beta_i = \text{SoftMax}(S_i) = \text{SoftMax}(\text{Pooling}(Y_i)) \quad (8)$$

The YOLOv8 model mainly consists of three parts: the backbone, the neck, and the head network. The baseline YOLOv8 model incorporates SCConv into the C2f module of its neck network to construct the phenotyping perception detection model. Specifically, the original standard convolution in the baseline model is replaced with the SCConv to construct the SCConv Block module, and the C2f module in the neck network is modified into the C2f-SCConv module. Figure 7 shows the improved model based on SCConv. Under the premise of accuracy, SCConv module effectively reduces model parameters, conserves computational resources, and improves hardware requirements for equipment, which is conducive to deployment and application aboard UAVs.

The object detection usually realizes the position prediction by the rectangular bounding box regression. The regression process takes the degree of regional overlap degree between boxes as loss. The IoU loss is calculated as follows:

$$L_{IoU} = 1 - \text{IoU} = 1 - \frac{W_0 H_0}{S_u} \quad (9)$$

where S_u denotes the area of concatenated region of boxes, and W_0 and H_0 denote the width and height of the intersection region of boxes, respectively.

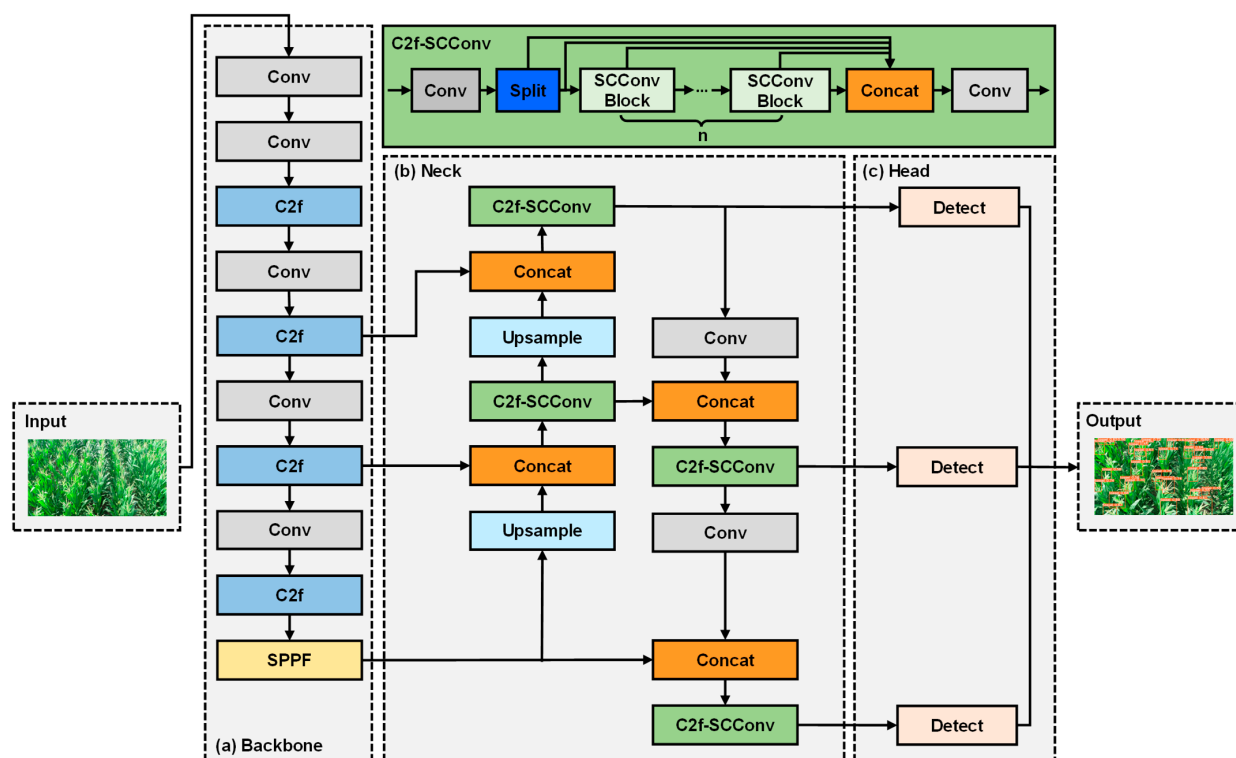


Figure 7. The flowchart of the enhanced YOLOv8 model using SCConv for maize tassel detection.

The Complete Intersection over Union (CIoU) [47] loss integrally takes into account geometric aspects including centroid distance, aspect ratio, and intersecting region of boxes. The object detection model enhanced with SCConv proposed in this paper adopts the CIoU loss as the regression function, which is calculated as follows:

$$L_{CIoU} = L_{IoU} + R_{CIoU} \quad (10)$$

where R_{CIoU} denotes penalty term of loss function.

3.4. Experiments Settings and Evaluation Indicators

The experimental platform used in the paper is a single computer. The CPU is Intel(R) Core(TM) i7-14700HX@2.10 GHz. The GPU is NVIDIA GeForce RTX 4060. The operating system of software environment is Windows 11. The programming language is Python 3.8.19. The deep-learning framework is PyTorch 2.0.0. The version of CUDA and Torchvision are 11.8 and 0.15.0, respectively.

YOLOv8 model was chosen as the baseline model in this paper. Improvement and training operations were performed on the original model to obtain the enhanced model. Table 2 presents the configurations of parameters for the model training.

Table 2. The configurations of the experimental parameters for the maize tassel detection model.

Model Parameters	Setting Values
Epochs	100
Batch size	16
Image size	640 × 640
Workers	4
Optimizer	SGD
Initial learning rate	0.01
Momentum	0.937
Weight decay	0.0005

The overall performance of the enhanced model was evaluated using precision (P), recall (R), and mean average precision (mAP). The model size and floating-point operations (FLOPs) were chosen for further measurement.

Precision and recall are calculated as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (12)$$

where TP, FP, and FN represent true positive, false positive, and false negative, respectively. They represent the object count accurately predicted as positive samples, mistakenly predicted as positive samples, and incorrectly predicted as negative samples.

The mAP is calculated as follows:

$$\text{mAP} = \sum_{i=1}^n AP_i / n, \quad AP = \int_0^1 P(R) dR \quad (13)$$

where AP denotes the average precision, serving as a measure of the region encompassed by the precision–recall curve and the axes. n denotes the object categories in dataset. mAP@0.5 denotes the AP for all categories when IoU threshold set to 0.5.

4. Results

4.1. Fusion-Based Neck Network Improvement Experiments

In this paper, we optimized the YOLOv8n baseline model by introducing the SCConv module. Integrating SCConv into the neck network significantly decreased the model parameters while preserving the feature extraction capability. To confirm the efficacy of the SCConv module in improving performance, Switchable Atrous Convolution (SAConv) [48], Omni-Dimensional Dynamic Convolution (ODConv) [49], and SCConv were selected for comparison experiments. Table 3 defines the outcomes aimed at enhancing the neck network of the baseline model.

Table 3. The results of the neck network improvement experiments for maize tassel detection.

Convolution Types	mAP@0.5/%	FLOPs/G	Model Size/MB
Standard Convolution	88.3	8.1	5.95
SAConv	90.3	7.5	6.50
ODConv	89.8	7.6	7.31
SCConv	91.7	7.3	5.16

As shown in Table 3, compared with other convolutions, the mAP@0.5, model operation, and model size of the enhanced model using SCConv were 91.7%, 7.3 G FLOPs, and 5.16 MB, respectively. In comparison to the initial model, the enhanced model was reduced by 0.8 G FLOPs in the model computation and decreased by 0.79 MB in the model size. It indicated that incorporating SCConv decreased the model parameters while maintaining the feature extraction capability, and facilitated deployment and utilization.

4.2. Ablation Experiments of Enhanced Detection Model

To confirm the effect of the improvement points, the YOLOv8n was selected as the baseline model and controlled each improvement point through ablation experiments. Table 4 exhibits the ablation experiments outcomes. Figure 8 displays the training outcomes of the enhanced model.

Table 4. The results of the improved point ablation experiments for maize tassel detection.

No.	SCConv	P/%	R/%	mAP@0.5/%	FLOPs/G	Model Size/MB
1	×	89.0	80.9	88.3	8.1	5.95
2	✓	92.2	84.3	91.7	7.3	5.16

Note: “✓” and “×”, respectively, indicate whether the improvement point has been adopted.

The outcomes of the baseline model, i.e., YOLOv8n, comprised the first group, which was used as a comparison benchmark for other experiments. The precision, recall, and mAP@0.5 of the baseline model were 89.0%, 80.9%, and 88.3%, respectively. Meanwhile, the model operation was 8.1 G FLOPs and the model size was 5.95 MB.

The second group represented the outcome when adding SCConv. The precision, recall, and mAP@0.5 of the improved object detection model were 92.2%, 84.3%, and 91.7%, respectively. Meanwhile, the operation and model size of the improved model were 7.3 G FLOPs and 5.16 MB, respectively. In comparison to the initial model, the precision, recall, and mAP@0.5 of the improved model had increase of 3.2%, 3.4%, and 3.4%, respectively. Meanwhile, the operation and model size decreased by 0.8 G FLOPs and 0.79 MB, respectively. These results demonstrated that the computation complexity and model size of the improved model had applicability, which was favorable for the model to be deployed in UAVs and had more potential for maize tassel detection in practical scenarios.

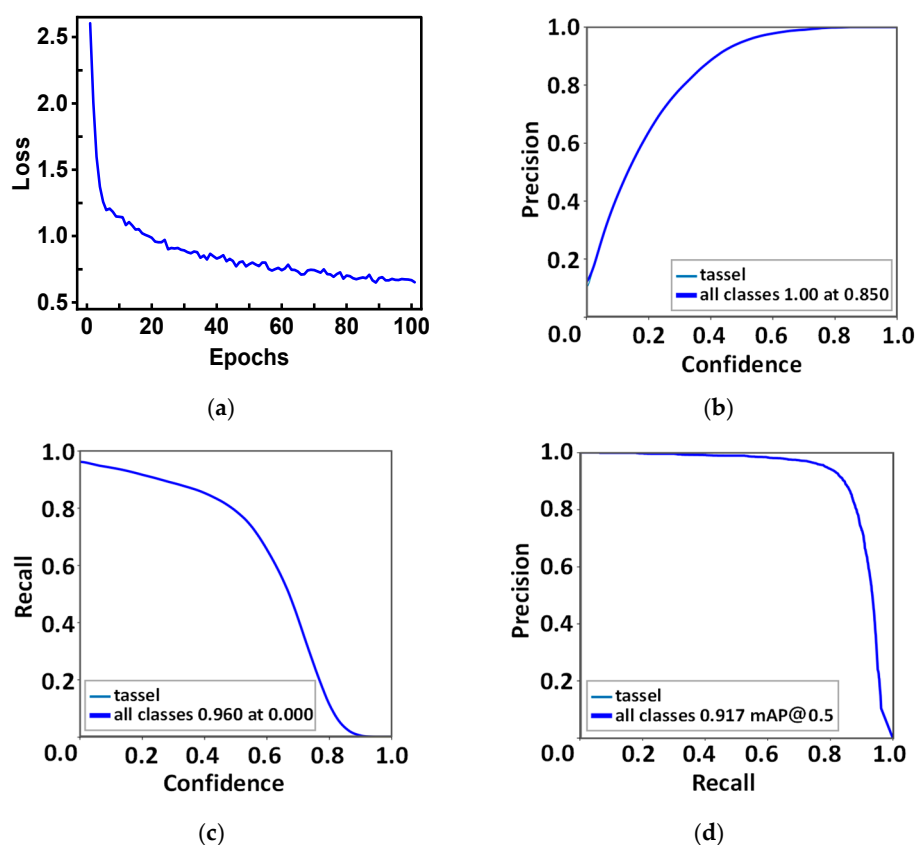


Figure 8. The training results of the maize tassel detection model using SCConv. Note: (a) represents the loss values during model training, (b) represents precision–confidence curve, (c) represents recall–confidence curve, and (d) represents precision–recall curve.

In Figure 8a, the curve shows a trend of rapidly decreasing and then stabilizing, indicating that the enhanced model gradually converges during training, the feature learning effect for maize tassels continuously improves, and the final loss stabilizes at a low level, reflecting the effectiveness of model training. In Figure 8b, as the confidence

threshold increases, the precision gradually rises, indicating that the enhanced model has a low recognition error rate for maize tassels under high confidence, and performs well in precision. Figure 8c shows that, as the confidence threshold increases, the recall rate gradually decreases, indicating that the enhanced model could cover most real tassel targets under low confidence and has a strong recall capability. The area enclosed by the curve and the coordinate axes in Figure 8d is relatively large, indicating that the enhanced model achieves an effective trade-off between precision and recall, and has a good comprehensive detection performance.

4.3. Comparison Experiments for Enhanced Detection Model

To further confirm the effectiveness of the model in maize tassel detection, the models were trained using the constructed aerial maize tassel image dataset and the public maize tassel dataset MTS, and the different model is contrasted with the enhanced YOLOv8n model with identical experimental conditions and parameter configurations. We selected algorithms such as Faster-RCNN, SSD, RT-DETR, YOLOv5n, YOLOv8n, YOLOv9t, YOLOv10n, and YOLOv11n for training the maize tassel detection model. Table 5 elaborates the outcomes of the comparative experiments for different models.

Table 5. The results of the comparative experiments for different maize tassel detection models.

No.	Model	P/%	R/%	mAP@0.5/%	FLOPs/G	Model Size/MB
1	Faster-RCNN	68.2	59.5	66.2	134.8	106.40
2	SSD	74.5	63.2	72.8	57.4	94.20
3	RT-DETR	91.5	83.7	90.6	108.6	63.45
4	YOLOv5n	83.4	76.8	82.4	8.7	5.01
5	YOLOv8n	89.0	80.9	88.3	8.1	5.95
6	YOLOv9t	87.8	79.4	87.2	7.8	4.73
7	YOLOv10n	90.4	81.2	88.8	7.6	5.58
8	YOLOv11n	88.7	80.6	86.5	6.7	5.35
9	Ours	92.2	84.3	91.7	7.3	5.16

As shown in Table 5, Faster-RCNN achieves a precision of 68.2%, a recall of 59.5%, and an mAP@0.5 of 66.2%, with a model size of 106.40 MB, showing poor performance in both detection and lightweighting. SSD has a precision of 74.5%, a recall of 63.2%, and an mAP@0.5 of 72.8%, with a model size of 94.20 MB; its performance is improved compared to Faster-RCNN but still not suitable for lightweight deployment. RT-DETR reaches a precision of 91.5%, a recall of 83.7%, and an mAP@0.5 of 90.6%, with a computation of 108.6 G and a model size of 63.45 MB, demonstrating good detection performance but insufficient lightweighting. YOLOv5n has a precision of 83.4%, a recall of 76.8%, and an mAP@0.5 of 82.4%, with a computation of 8.7 G and a model size of 5.01 MB, highlighting lightweighting but having a gap in performance. YOLOv8n achieves a precision of 89.0%, a recall of 80.9%, and an mAP@0.5 of 88.3%, with a computation of 8.1 G and a model size of 5.95 MB; its performance is better than that of YOLOv5n but still has room for improvement. YOLOv9t has a precision of 87.8%, a recall of 79.4%, and an mAP@0.5 of 87.2%, with a computation of 7.8 G and a model size of 4.73 MB, excelling in lightweighting but lacking in detection performance. YOLOv10n reaches a precision of 90.4%, a recall of 81.2%, and an mAP@0.5 of 88.8%, with a computation of 7.6 G and a model size of 5.58 MB, showing a balanced performance in detection and lightweighting but not the optimal. YOLOv11n has a precision of 88.7%, a recall of 80.6%, and an mAP@0.5 of 86.5%, with a computation of 6.7 G and a model size of 5.35 MB, having the lowest computation but moderate detection performance. Ours model achieves a precision of 92.2%, a recall of 84.3%, and an mAP@0.5 of 91.7%, with a computation of 7.3 G and a model size of 5.16 MB. Compared with the baseline model, the precision, recall, and mAP@0.5 of the enhanced model are increased by

3.2%, 3.4%, and 3.4%, respectively, and the model computation is reduced by 0.8 G FLOPs, and the model size is reduced by 0.79 MB. Compared with YOLOv10n, the precision, recall, and mAP@0.5 of the enhanced model are increased by 1.8%, 3.1%, and 2.9%, respectively, and the model computation is reduced by 0.3 G FLOPs, and the model size is reduced by 0.42 MB. It achieves the best detection performance among all models while maintaining lightweight advantages, realizing the optimal balance between detection performance and lightweighting, and is more suitable for deployment and practical application on platforms such as maize-breeding UAV.

4.4. Visualization and Analysis of Enhanced Detection Model

To demonstrate its effectiveness, a visualization analysis is conducted and the outcomes are depicted in Figure 9. After a visual comparison, we proved the applicability and effectiveness of the enhanced YOLOv8n using SCConv in different operational scenarios.

As shown in Figure 9, Faster-RCNN has frequent missed detections in bright scenes, with maize tassels going unrecognized; in dim scenes, the missed detections are severe, and it could barely detect tassels in low-light conditions; in occluded scenes, tassels are generally missed, showing an extremely weak detection capability, with frequent missed detections and poor performance in complex scenarios. SSD has obvious missed detections in bright scenes, with the incomplete recognition of some tassels; in dim scenes, the number of missed detections is relatively large, and its detection accuracy drops significantly under dim light; in occluded scenes, severely missed detections occur for occluded tassels. RT-DETR has few missed detections in bright scenes, with an acceptable detection performance; in dim scenes, the missed detections are obvious and its ability to recognize tassels is insufficient; in occluded scenes, there are certain missed detections. Overall, it performs well in bright scenes, but missed detections remain prominent in dim and occluded scenes. YOLOv5n has few missed detections in bright scenes, with good detection results; in dim scenes, the number of missed detections increases, and its performance drops significantly under dim light; in occluded scenes, its ability to recognize targets is limited, leaving room for improvement in model performance. YOLOv8n has few missed detections in bright scenes, with a good detection performance; in dim scenes, there are certain missed detections and its performance decreases; in occluded scenes, there are partial missed detections and its target recognition is unstable. It has a relatively good overall performance but still has missed detections in complex scenarios. YOLOv9t has few missed detections in bright scenes; in dim scenes, the number of missed detections is relatively large and its performance is insufficient under dim light; in occluded scenes, there are obvious missed detections. YOLOv10n has very few missed detections in bright scenes, with an excellent detection performance; in dim scenes, there are certain missed detections; in occluded scenes, there are partial missed detections. It has a relatively good overall performance but still has room for improvement in complex scenarios. YOLOv11n has few missed detections in bright scenes, with a high detection accuracy; in dim scenes, the number of missed detections is relatively large and its detection performance is insufficient; in occluded scenes, there are obvious missed detections. Comparing the visualization performances of model detection, the results further emphasized that the enhanced phenotyping detection model could detect maize tassels that were partially obscured by maize leaves or branches. Meanwhile, the improved model still achieved the accurate detection of maize tassels under different scenarios, such as brightness, dimness, and occlusion. This demonstrated that the enhanced model has applicability, effectiveness, and better detection performance, which is accurate and reliable for maize tassel detection and has certain advantages in real complex operation scenarios. In summary, the performance of the enhanced model is accurate and reliable. While maintaining detection accuracy, the model operation and model size significantly

decreased, which is conducive to being deployed and applied aboard maize-breeding UAVs, and provides a methodological reference for object detection technology.

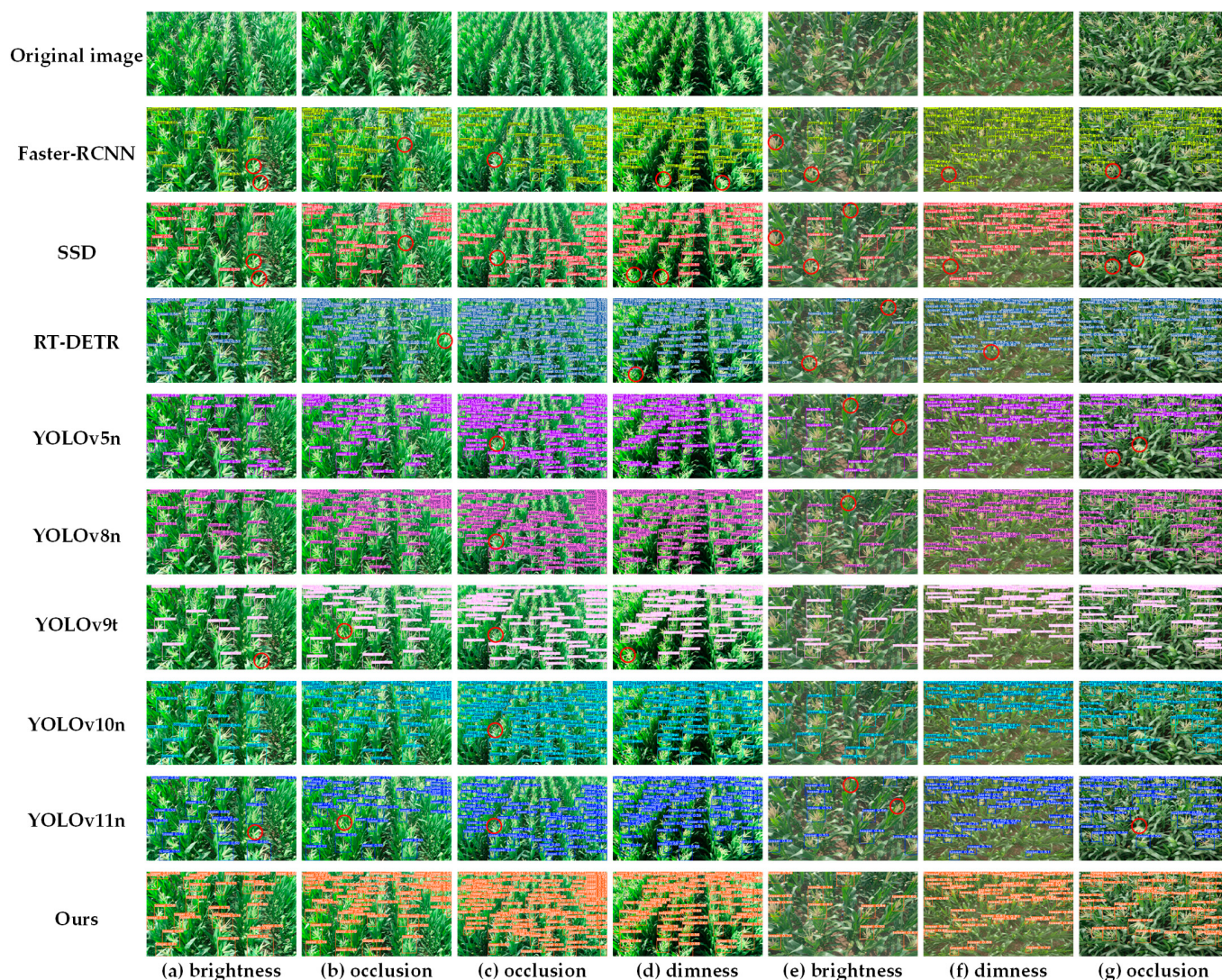


Figure 9. The visualization results of different models in maize tassel detection task. Note: (a–d) represent the tassel images from the dataset of this paper, and (e–g) represent the tassel images from the MTS dataset, which include scenarios such as brightness, dimness, and occlusion. The red circle indicates the missed detection cases of the model.

5. Discussion

Against the backdrop of precision agriculture development, UAV-based aerial object detection has become a pivotal technology for maize-phenotyping perception operations. To address the core limitations of the existing models, which are the large model size and poor deploy ability, this study proposes an enhanced YOLOv8 model integrated with SCConv, and achieves certain technical contributions.

First, a dedicated quadrotor UAV platform for maize breeding was constructed, enabling the effective acquisition of aerial maize tassel images. Second, the SCConv was integrated into the neck network of the YOLOv8 model, which effectively reduces feature redundancy, which realized the optimization of accuracy improvement and lightweight enhancement. The enhanced model achieves a precision of 92.2%, a recall of 84.3%, and an mAP@0.5 of 91.7%, with a computation of 7.3 G FLOPs and a size of 5.16 MB. Compared with the baseline model, its precision, recall, and mAP@0.5 are increased by 3.2%, 3.4%,

and 3.4%, respectively, while the model computation is reduced by 0.8 G FLOPs and the model size by 0.79 MB. Compared with the YOLOv10n, the model computation and model size are reduced by 0.3 G FLOPs and 0.42 MB, respectively. Third, the model exhibits a strong adaptability to complex scenarios. The constructed maize tassel dataset and the MTS dataset were used to train and evaluate the enhanced model, and it maintained a high detection accuracy in UAV aerial images under the brightness, dimness, and occlusion condition, which verified its practical value in field maize-phenotyping operations.

The contradiction between the detection accuracy, computational efficiency, and deploy ability has long been a bottleneck in the field of UAV-based crop-phenotyping detection. Most existing models either pursue a high accuracy at the cost of excessive computation, e.g., RT-DETR, and Faster-RCNN, or prioritize lightweighting but sacrifice the detection stability in complex field scenarios, e.g., YOLOv5n. In contrast, the enhanced model improves this issue by leveraging the feature compression capability of SCConv, which not only outperforms mainstream lightweight models in key metrics but also maintains a compact model size and low model computation, making it compatible with UAV platforms.

From a practical perspective, this study provides two key supports for maize breeding and precision agriculture. On one hand, it offers a lightweight and high precision tool for maize tassel phenotyping, improving the efficiency of UAV-based phenotyping data acquisition. On the other hand, the methodological framework constructed in this study could be extended to other UAV-based crop object detection tasks, providing a reference for resolving similar performance and deploy ability contradictions in the phenotyping detection of food crops such as wheat, rice, and soybeans.

6. Conclusions

Addressing the limitations of the existing object detection models, which had a high computation complexity and large model size that hindered deployment and application, we proposed a object detection model that integrated Spatial and Channel Reconstruction Convolution. The SCConv module was introduced into the neck network of the YOLOv8-based model. The constructed maize tassels dataset and MTS dataset were used for the training of the improved model. The results demonstrated that the precision, recall, and mAP@0.5 of the enhanced model were 92.2%, 84.3%, and 91.7%, respectively. Meanwhile, the model operation was 7.3 G FLOPs and the model size is 5.16 MB. In comparison to the baseline model, the precision, recall, and mAP@0.5 of the improved model achieved increase of 3.2%, 3.4%, and 3.4%, respectively. Meanwhile, the model operation decreased by 0.8 G FLOPs and the model size decreased by 0.79 MB. Compared with YOLOv10n, the precision, recall, and mAP@0.5 of the enhanced model are increased by 1.8%, 3.1%, and 2.9%, respectively, and the model computation is reduced by 0.3 G FLOPs, and the model size is reduced by 0.42 MB.

Although the study has achieved certain results, there are still some aspects that need to be improved. At present, this study faces challenges such as a single type of crop application, limited data diversity, and insufficient multi-source data fusion. In the future, the proposed approach would be extended to food crop breeding such as wheat, rice, and soybeans. UAVs would be used to collect crop data across different geographical regions, phenotypic types, and growth cycles. Meanwhile, sensor data of types such as LiDAR, multispectral, and thermal infrared would be fused simultaneously.

Author Contributions: Conceptualization, J.C.; methodology, H.W. and J.C.; software, H.W. and S.Z.; validation, H.W.; formal analysis, H.W., J.C., X.W. and S.Z.; investigation, H.W.; resources, H.W., X.W. and S.Z.; data curation, H.W. and S.Z.; writing—original draft preparation, H.W., J.C., X.W. and S.Z.; writing—review and editing, H.W., J.C., X.W. and S.Z.; visualization, H.W. and S.Z.; supervision, J.C.

and X.W.; project administration, J.C. and X.W.; funding acquisition, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 12426518 and No. 52472463) and the National Key Research and Development Plan Project of China (No. 2022YFD2001405).

Data Availability Statement: The data will be made available upon request.

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

References

1. Li, D.; Zhang, W.; Liu, Y. STRM-KD: Semantic Topological Relation Matching Knowledge Distillation Model for Smart Agriculture Apple Leaf Disease Recognition. *Expert Syst. Appl.* **2025**, *263*, 125824. [\[CrossRef\]](#)
2. Yan, Y.; Song, F.; Sun, J. The Application of UAV Technology in Maize Crop Protection Strategies: A Review. *Comput. Electron. Agric.* **2025**, *237*, 110679. [\[CrossRef\]](#)
3. Muhmad Kamarulzaman, A.M.; Wan Mohd Jaafar, W.S.; Mohd Said, M.N.; Saad, S.N.M.; Mohan, M. UAV Implementations in Urban Planning and Related Sectors of Rapidly Developing Nations: A Review and Future Perspectives for Malaysia. *Remote Sens.* **2023**, *15*, 2845. [\[CrossRef\]](#)
4. Peng, Y.; Yan, H.; Rao, K.; Yang, P.; Lv, Y. Distributed Model Predictive Control for Unmanned Aerial Vehicles and Vehicle Platoon Systems: A Review. *Intell. Robot.* **2024**, *4*, 293–317. [\[CrossRef\]](#)
5. Zhu, B.; Zhang, Y.; Sun, Y.; Shi, Y.; Ma, Y.; Guo, Y. Quantitative Estimation of Organ-Scale Phenotypic Parameters of Field Crops through 3D Modeling Using Extremely Low Altitude UAV Images. *Comput. Electron. Agric.* **2023**, *210*, 107910. [\[CrossRef\]](#)
6. Hu, G.; Ren, Z.; Chen, J.; Ren, N.; Mao, X. Using the MSFNet Model to Explore the Temporal and Spatial Evolution of Crop Planting Area and Increase Its Contribution to the Application of UAV Remote Sensing. *Drones* **2024**, *8*, 432. [\[CrossRef\]](#)
7. Erenstein, O.; Jaleta, M.; Sonder, K.; Mottaleb, K.; Prasanna, B.M. Global Maize Production, Consumption and Trade: Trends and R&D Implications. *Food Secur.* **2022**, *14*, 1295–1319. [\[CrossRef\]](#)
8. Jiang, C.; Ren, H.; Ye, X.; Zhu, J.; Zeng, H.; Nan, Y.; Sun, M.; Ren, X.; Huo, H. Object Detection from UAV Thermal Infrared Images and Videos Using YOLO Models. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102912. [\[CrossRef\]](#)
9. Zhang, Y.; Dong, X.; Wang, H.; Lin, Y.; Jin, L.; Lv, X.; Yao, Q.; Li, B.; Gao, J.; Wang, P.; et al. Maize Breeding for Smaller Tassels Threatens Yield Under a Warming Climate. *Nat. Clim. Change* **2024**, *14*, 1306–1313. [\[CrossRef\]](#)
10. Zhu, H.; Liang, S.; Lin, C.; He, Y.; Xu, J. Using Multi-Sensor Data Fusion Techniques and Machine Learning Algorithms for Improving UAV-Based Yield Prediction of Oilseed Rape. *Drones* **2024**, *8*, 642. [\[CrossRef\]](#)
11. Shu, M.; Fei, S.; Zhang, B.; Yang, X.; Guo, Y.; Li, B.; Ma, Y. Application of UAV Multisensor Data and Ensemble Approach for High-Throughput Estimation of Maize Phenotyping Traits. *Plant Phenom.* **2022**, *2022*, 9802585. [\[CrossRef\]](#)
12. Ni, J.; Chen, Y.; Tang, G.; Shi, J.; Cao, W.; Shi, P. Deep Learning-Based Scene Understanding for Autonomous Robots: A Survey. *Intell. Robot.* **2023**, *3*, 374–401. [\[CrossRef\]](#)
13. Zhang, Z.; Chen, J.; Xu, X.; Liu, C.; Hang, Y. Hawk-Eye-Inspired Perception Algorithm of Stereo Vision for Obtaining Orchard 3D Point Cloud Navigation Map. *CAAI Trans. Intell. Technol.* **2023**, *8*, 987–1001. [\[CrossRef\]](#)
14. Li, H.; Wang, P.; Huang, C. Comparison of Deep Learning Methods for Detecting and Counting Sorghum Heads in UAV Imagery. *Remote Sens.* **2022**, *14*, 3143. [\[CrossRef\]](#)
15. Badgujar, C.M.; Poulouse, A.; Gan, H. Agricultural Object Detection with You Only Look Once (YOLO) Algorithm: A Bibliometric and Systematic Literature Review. *Comput. Electron. Agric.* **2024**, *223*, 109090. [\[CrossRef\]](#)
16. Wu, D.; Liang, Z.; Chen, G. Deep Learning for LiDAR-Only and LiDAR-Fusion 3D Perception: A Survey. *Intell. Robot.* **2022**, *2*, 105–129. [\[CrossRef\]](#)
17. Wu, T.; Wang, L.; Xu, X.; Su, L.; He, W.; Wang, X. An Intelligent Fault Detection Algorithm for Power Transmission Lines Based on Multi-Scale Fusion. *Intell. Robot.* **2025**, *5*, 474–487. [\[CrossRef\]](#)
18. Seidaliyeva, U.; Akhmetov, D.; Ilipbayeva, L.; Matson, E.T. Real-Time and Accurate Drone Detection in a Video with a Static Background. *Sensors* **2020**, *20*, 3856. [\[CrossRef\]](#)
19. Alhafnawi, M.; Salameh, H.A.B.; Masadeh, A.E.; Al-Obiedollah, H.; Ayyash, M.; El-Khazali, R.; Elgala, H. A Survey of Indoor and Outdoor UAV-Based Target Tracking Systems: Current Status, Challenges, Technologies, and Future Directions. *IEEE Access* **2023**, *11*, 68324–68339. [\[CrossRef\]](#)
20. Li, X.; Yue, H.; Liu, J.; Cheng, A. AMS-YOLO: Asymmetric Multi-Scale Fusion Network for Cannabis Detection in UAV Imagery. *Drones* **2025**, *9*, 629. [\[CrossRef\]](#)
21. Duan, H.; Shi, F.; Gao, B.; Zhou, Y.; Cui, Q. A Novel Real-Time Intelligent Detector for Monitoring UAVs in Live-Line Operation on 10 KV Distribution Networks. *Intell. Robot.* **2025**, *5*, 70–87. [\[CrossRef\]](#)

22. Falahat, S.; Karami, A. Maize Tassel Detection and Counting Using a YOLOv5-Based Model. *Multimed. Tools Appl.* **2023**, *82*, 19521–19538. [\[CrossRef\]](#)
23. Niu, S.; Nie, Z.; Li, G.; Zhu, W. Multi-Altitude Corn Tassel Detection and Counting Based on UAV RGB Imagery and Deep Learning. *Drones* **2024**, *8*, 198. [\[CrossRef\]](#)
24. Pu, H.; Chen, X.; Yang, Y.; Tang, R.; Luo, J.; Wang, Y.; Mu, J. Tassel-YOLO: A New High-Precision and Real-Time Method for Maize Tassel Detection and Counting Based on UAV Aerial Images. *Drones* **2023**, *7*, 492. [\[CrossRef\]](#)
25. Yu, X.; Yin, D.; Xu, H.; Pinto Espinosa, F.; Schmidhalter, U.; Nie, C.; Bai, Y.; Sankaran, S.; Ming, B.; Cui, N.; et al. Maize Tassel Number and Tasseling Stage Monitoring Based on Near-Ground and UAV RGB Images by Improved YOLOv8. *Precis. Agric.* **2024**, *25*, 1800–1838. [\[CrossRef\]](#)
26. Liu, D.; Zhang, Y.; Ren, Y.; Pan, D.; He, X.; Cong, M.; Yu, G. Multi-Directional Attention: A Lightweight Attention Module for Slender Structures. *Intell. Robot.* **2025**, *5*, 827–843. [\[CrossRef\]](#)
27. Guan, H.; Deng, H.; Ma, X.; Zhang, T.; Zhang, Y.; Zhu, T.; Zhou, H.; Gu, Z.; Lu, Y. A Corn Canopy Organs Detection Method Based on Improved DBi-YOLOv8 Network. *Eur. J. Agron.* **2024**, *154*, 127076. [\[CrossRef\]](#)
28. Jia, Y.; Fu, K.; Lan, H.; Wang, X.; Su, Z. Maize Tassel Detection with CA-YOLO for UAV Images in Complex Field Environments. *Comput. Electron. Agric.* **2024**, *217*, 108562. [\[CrossRef\]](#)
29. 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. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *130*, 103922. [\[CrossRef\]](#)
30. Chen, X.; Wu, P.; Wu, Y.; Aboud, L.; Postolache, O.; Wang, Z. Ship Trajectory Prediction Via a Transformer-Based Model by Considering Spatial-Temporal Dependency. *Intell. Robot.* **2025**, *5*, 562–578. [\[CrossRef\]](#)
31. Bai, Q.; Gao, R.; Li, Q.; Wang, R.; Zhang, H. Recognition of the Behaviors of Dairy Cows by an Improved YOLO. *Intell. Robot.* **2024**, *4*, 1–19. [\[CrossRef\]](#)
32. Yang, G.; Liu, C.; Li, G.; Chen, H.; Chen, K.; Wang, Y.; Hu, X. DiffKNet-TL: Maize Phenology Monitoring with Confidence-Aware Constrained Diffusion Model Based on UAV Platform. *Comput. Electron. Agric.* **2025**, *239*, 110936. [\[CrossRef\]](#)
33. Zhu, Z.; Gao, Z.; Zhuang, J.; Huang, D.; Huang, G.; Wang, H.; Pei, J.; Zheng, J.; Liu, C. MSMT-RTDETR: A Multi-Scale Model for Detecting Maize Tassels in UAV Images with Complex Field Backgrounds. *Agriculture* **2025**, *15*, 1653. [\[CrossRef\]](#)
34. Tang, B.; Zhou, J.; Li, X.; Pan, Y.; Lu, Y.; Liu, C.; Ma, K.; Sun, X.; Chen, D.; Gu, X. Detecting Tasseling Rate of Breeding Maize Using UAV-Based RGB Images and STB-YOLO Model. *Smart Agric. Technol.* **2025**, *11*, 100893. [\[CrossRef\]](#)
35. Lei, T.; Li, G.; Luo, C.; Zhang, L.; Liu, L.; Gates, R.S. An Informative Planning-Based Multi-Layer Robot Navigation System as Applied in a Poultry Barn. *Intell. Robot.* **2022**, *2*, 313–332. [\[CrossRef\]](#)
36. Zhuang, T.; Liang, X.; Xue, B.; Tang, X. An In-Vehicle Real-Time Infrared Object Detection System Based on Deep Learning with Resource-Constrained Hardware. *Intell. Robot.* **2024**, *4*, 276–292. [\[CrossRef\]](#)
37. Lupión, M.; Polo-Rodríguez, A.; Medina-Quero, J.; Sanjuan, J.F.; Ortigosa, P.M. 3D Human Pose Estimation from Multi-View Thermal Vision Sensors. *Inf. Fusion* **2024**, *104*, 102154. [\[CrossRef\]](#)
38. Doherty, J.; Gardiner, B.; Kerr, E.; Siddique, N. BiFPN-YOLO: One-Stage Object Detection Integrating Bi-Directional Feature Pyramid Networks. *Pattern Recognit.* **2025**, *160*, 111209. [\[CrossRef\]](#)
39. 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. *J. Integr. Agric.* **2023**, *22*, 1671–1683. [\[CrossRef\]](#)
40. Tang, C.; Zhou, F.; Sun, J.; Zhang, Y. Circle-YOLO: An Anchor-Free Lung Nodule Detection Algorithm Using Bounding Circle Representation. *Pattern Recognit.* **2025**, *161*, 111294. [\[CrossRef\]](#)
41. 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 Sens.* **2020**, *12*, 338. [\[CrossRef\]](#)
42. Ye, J.; Yu, Z. Fusing Global and Local Information Network for Tassel Detection in UAV Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 4100–4108. [\[CrossRef\]](#)
43. Yu, Z.; Ye, J.; Li, C.; Zhou, H.; Li, X. TasselLFANet: A Novel Lightweight Multi-Branch Feature Aggregation Neural Network for High-Throughput Image-Based Maize Tassels Detection and Counting. *Front. Plant Sci.* **2023**, *14*, 1158940. [\[CrossRef\]](#)
44. Ye, J.; Yu, Z.; Lin, J.; Lin, L.; Li, H.; Zhu, X.; Feng, H.; Yin, H.; Zhou, H. Maize Tasseling Stage Automated Observation Method via Semantic Enhancement and Context Occlusion Learning. *IEEE Sens. J.* **2024**, *24*, 37259–37274. [\[CrossRef\]](#)
45. Varghese, R.; Sambath, M. YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. In Proceedings of the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, 18–19 April 2024; IEEE: New York, NY, USA, 2024; pp. 1–6.
46. Li, J.; Wen, Y.; He, L. SCConv: Spatial and Channel Reconstruction Convolution for Feature Redundancy. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023; IEEE: New York, NY, USA, 2023; pp. 6153–6162.
47. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation. *IEEE Trans. Cybern.* **2021**, *52*, 8574–8586. [\[CrossRef\]](#)

-
48. Qiao, S.; Chen, L.; Yuille, A. Detectors: Detecting Objects with Recursive Feature Pyramid and Switchable Atrous Convolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; IEEE: New York, NY, USA, 2021; pp. 10213–10224.
 49. Li, C.; Zhou, A.; Yao, A. Omni-Dimensional Dynamic Convolution. *arXiv* **2022**, arXiv:2209.07947. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.