ORIGINAL RESEARCH Open access

UAV-Based Maize Phenotyping Using Spatial-Channel Reconstruction Convolution for Tassel Detection

Lukas Weber[1*], Sophie Janssen[2], Jonas Richter[1], Elena Fischer[3], Maximilian Schulz[4]

Abstract

With the ongoing advancement of precision agriculture, detecting objects in UAV-acquired aerial imagery has become a critical technology for maize phenotyping. Intricate backgrounds weaken the model's capacity to extract maize tassel features, while sacrificing computational efficiency to strengthen feature representation hinders deployment on UAV platforms. To achieve a favorable trade-off between model compactness and deployability, an improved architecture incorporating spatial-channel convolution is proposed. First, a dedicated maize-breeding UAV was constructed, through which maize tassel imagery was successfully gathered. Second, Spatial and Channel Reconstruction Convolution (SCConv) was inserted into the YOLOv8 baseline's neck, reducing computational burden while preserving detection fidelity. Finally, the custom-built maize tassel dataset and the publicly accessible Maize Tasseling Stage (MTS) dataset were utilized to train and assess the improved model. The experiments demonstrated that the improved model yielded a precision of 92.2%, recall of 84.3%, and mAP@0.5 of 91.7%, operating at 7.3 G floating-point operations (FLOPs) with a model footprint of 5.16 MB. Compared with the baseline, the improved model delivered gains of 3.2%, 3.4%, and 3.4% in precision, recall, and mAP@0.5, while saving 0.8 G FLOPs in computation and 0.79 MB in size. When measured against YOLOv10n, the improved model's precision, recall, and mAP@0.5 are higher by 1.8%, 3.1%, and 2.9%, respectively, while computation drops by 0.3 G FLOPs and model size shrinks by 0.42 MB. The refined model achieves high accuracy, demonstrates superior performance on UAV aerial images captured in complex scenarios, and provides a methodological foundation for practical deployment. It further facilitates maize tassel detection and shows promise for application within maize breeding pipelines.

Keywords Precision agriculture, Object detection, Maize-breeding UAV, Phenotyping perception

*Correspondence:
Lukas Weber
lukas.weber@uni-stuttgart.de
[1] Department of AI for Sugar Beet and Rapeseed Yield Optimization, Faculty of Agricultural Sciences, University of Stuttgart, Stuttgart, Germany
[2] Department of Computational Agronomy and Phenomics, Faculty of Agriculture, Eindhoven University of Technology, Eindhoven, Netherlands
[3] Department of Smart Farm Automation and Robotics, Faculty of Agriculture, University of Basel, Basel, Switzerland
[4] Department of Machine Learning for Harvest Time Prediction, Faculty of Agriculture, University of Freiburg, Freiburg, Germany

Introduction

Amidst the global wave of agricultural modernization, precision agriculture has established itself as a fundamental paradigm for overcoming the constraints of traditional farming and enabling high-caliber agricultural development. Its essential mechanism relies on intelligent sensing and data-informed decision-making to meticulously manage crop growth cycles, thereby substantially amplifying both productivity and resource efficiency [1, 2]. As the primary manifestation of the low-altitude economy, UAVs are increasingly expanding and replicating their industrial model across agriculture and related sectors, injecting renewed energy into forestry and agriculture [3, 4]. Crop images captured from low-altitude

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

UAVs provide field-scale data that help drive the development of agricultural information systems [5]. Serving as vital instruments for on-site precision agriculture, maize-breeding Unmanned Aerial Vehicles (UAVs) tailored to specific crop settings have replaced human labor in repetitive, physically demanding operations (e.g., field surveillance and phenotypic recording), thereby shortening breeding timelines and facilitating the selection of superior varieties [6].

As a crop of immense importance, maize breeding bears direct consequences for food security and the long-term viability of agriculture [7]. The perception of maize phenotypes, a cornerstone of breeding programs, aims to accurately capture characteristics related to yield and stress tolerance, such as tassel branching patterns, plant height progression, and foliar disease extent [8]. Maize tassels are strongly linked to both yield and grain quality, acting as significant phenotypes and evaluative metrics throughout the breeding cycle [9]. Older methodologies present clear drawbacks: manual measurement suffers from low throughput and subjectivity, while ground-deployed sensors fall short in covering the phenotypic heterogeneity present at the field scale. By contrast, the UAV aerial image detection approach rooted in computer vision has risen as a mainstream solution in maize-breeding phenotyping, drawing strength from its non-destructive nature, wide-area coverage, and real-time responsiveness [10]. Breeding UAVs, as automated equipment for crop phenotype collection and detection, deliver an intelligent answer to modern breeding tasks and have become a crucial means and platform for uncovering and generating elite maize germplasm [11].

Aerial object detection via UAVs constitutes a vital link in enabling UAVs to carry out low-altitude missions [12, 13] effectively. The accurate target recognition capabilities of UAVs propel progress in the agricultural domain. Functioning as the principal environmental perception apparatus of UAVs [14], object detection directly determines phenotyping precision and serves as the bedrock of UAV-dependent phenotyping systems. Deep-learning-based techniques, with their innate ability to learn and extract features from complex settings, outperform classical detectors reliant on manually crafted features in both accuracy and generalizability [15, 16]. Within complex field environments featuring variable illumination, plant intermingling, and weed encroachment, single-stage architectures like the You Only Look Once (YOLO) series [17], prized for their speed and end-to-end design, demonstrate significant potential for UAV applications, with maize tassel recognition as a prime example [18].

Nonetheless, phenotyping models face significant obstacles within maize-breeding UAV contexts. UAV aerial imagery backgrounds are characterized by a high degree of complexity arising from lighting dynamics and occlusion. Existing models often seek performance gains by stacking additional layers and expanding feature map widths, thereby escalating both computational expense and model storage requirements [19]. Limited by the hardware capabilities of onboard UAV systems, these larger models are challenging to deploy, creating a disconnect between state-of-the-art algorithms and real-world field demands [20]. Consequently, the pursuit of lightweight YOLO-derived UAV models without compromising detection precision is of both theoretical consequence and practical value [21], particularly for maize phenotyping, which imposes strict demands for high accuracy and on-site operability.

Factors such as fluctuating lighting conditions and mutual occlusion within complex backgrounds complicate the extraction of maize tassel features, whereas inflating model scale to improve representational power is counterproductive for UAV deployment. Aiming to reconcile detection accuracy with model lightweighting and resolve the tension between computational efficiency and practical deployability, this study presents an enhanced YOLO-based maize phenotyping method that incorporates Spatial and Channel Reconstruction Convolution (SCConv). This technique effectively suppresses feature redundancy in both spatial and channel dimensions while maintaining the model's ability to extract maize tassel characteristics. The resulting optimized UAV aerial detection model exhibits reduced computational demands and a smaller footprint, delivering a lightweight yet precise tool for crop phenotyping detection and underpinning the advancement of precision agriculture in breeding-oriented scenarios. Furthermore, aerial maize tassel data were collected for model training, providing a methodological benchmark for UAV-based object detection efforts and easing the path toward deployment on UAV platforms. The specific research activities adopted in this work comprise the following:

· A quadrotor UAV platform tailored for maize breeding was constructed, proving capable of collecting aerial crop phenotypic information and achieving effective maize tassel image acquisition;

© 2023 The Author(s). | Published by Agro Intelligence Research Press

· The SCConv module was integrated into the YOLOv8 framework to create a crop phenotyping perception model, lessening spatial and channel-wise feature redundancy;

· The enhanced model was trained and benchmarked using the custom-built maize tassel dataset, furnishing a methodological reference for UAV-enabled maize breeding phenotype perception tasks.

Related Works

YOLO series architectures have risen to prominence as the standard methodology for UAV-based aerial object detection, thanks to their end-to-end processing and high-speed inference. Research efforts have pursued targeted refinements of YOLO models to accommodate the specific demands encountered in maize-breeding detection contexts. Lightweight convolutional approaches deliver an effective answer for UAV detection workloads. As the central mechanism for streamlined hardware deployment, lightweighting techniques aim to reduce both the computational footprint and the overall model size while safeguarding detection fidelity wherever feasible. Falahat and Karami [22] devised a modified YOLOv5n variant that leverages the Gaussian Error Linear Unit (GELU) activation, incorporates depth-wise convolutional operations, and integrates a spatial dropout component, thereby mitigating overfitting and enhancing the network's ability to learn intricate feature representations. Training and assessment are based on the Maize Tassel Detection Counting (MTDC) dataset, with outcomes underscoring the speed advantages conferred by the enhanced model for UAV aerial detection and enumeration workflows. Niu *et al.* [23] outlined a framework for UAV-captured image object recognition and tallying built atop the YOLOv8 architecture. Lightweight construction and swift inference are realized through the Partial Convolution (PConv) unit. At the same time, representational power is enhanced via a hybrid module, ACmix, that fuses self-attention mechanisms with convolutional operations. The Convolution with Transformer Attention Module (CTAM) is additionally woven into the pipeline to foster richer semantic interchange and yield accurate, resource-efficient aerial target identification. Pu *et al.* [24], grappling with shortcomings in both latency and precision plaguing contemporary real-time recognition systems, embedded the Global Attention Mechanism (GAM) within the YOLOv7 backbone, appended Group Shuffle Convolution (GSConv) along with Variety of View Group Shuffle Cross Stage Partial (VoV-GSCSP) blocks, substituted the native loss formulation with Scaled Intersection over Union (SIoU), and unveiled Tassel-YOLO, tailored for autonomous UAV aerial target spotting and quantification. Yu *et al.* [25], motivated by morphological variability rooted in individual divergence among specimens, proposed a UAV-oriented model, PConv-YOLOv8, that serves as a methodological touchstone for subsequent aerial detection efforts.

Attention-centric mechanisms, which steer model focus toward salient cues while attenuating distractors, constitute a pivotal tool for boosting detection fidelity [26]. In demanding perception scenarios characterized by occlusions and fluctuating illumination, attention modules materially enhance model resilience under taxing conditions. Guang *et al.* [27], addressing the challenge of expeditious recognition in unconstrained natural settings, embedded augmented deformable convolutions alongside a bi-level routing attention scheme to assemble DBi-YOLOv8, which was fielded in genuine, operationally complex environments. Jia *et al.* [28], reacting to the pitfalls of traditional human-in-the-loop inspection—namely pronounced subjectivity, onerous labor demands, and sluggish responsiveness—commissioned UAV platforms to acquire overhead imagery and curate a dedicated corpus, subsequently erecting an aerial detection architecture termed CA-YOLO through the fusion of a Coordinate Attention (CA) module into the YOLOv5 pipeline. Findings confirmed that regressive localization aptitude was reinforced and overall detection capability sharpened. Chen *et al.* [29], targeting the twin challenges of accurately resolving mutually occluded aerial entities and the degraded discriminatory power under intense luminous flux, designed the RESAM-YOLOv8n detection network by integrating the Residual Scale Attention Module (RESAM) and harnessing expansive high-resolution aerial imagery. The fortified architecture focuses on object-level signatures, discards extraneous background clutter, and ultimately elevates detection benchmarks.

Equipped with powerful global semantic encoding capabilities, the Transformer paradigm opens a new trajectory for raising the precision ceiling of UAV detection frameworks. The predominant strand of contemporary scholarship marries Transformer-style modules with classical convolutional skeletons, seeking a balance between holistic contextual

© 2023 The Author(s). | Published by Agro Intelligence Research Press

comprehension and fine-grained local feature distillation [30, 31], thereby aligning more closely with the operational prerequisites of UAV aerial object perception. In the specific context of maize inflorescence and foliage detection, Yang *et al.* [32] reconfigured the Kernels Network (K-Net) scaffold by adopting a Transformer backbone and a dynamic loss reweighting regimen. The resulting Diffusion-enhanced K-Net for Tassels and Leaves (DiffKNet-TL) framework enhances the architecture's ability to chart dynamic phenotypic trajectories across both spatial granularities and temporal windows. Mutual obstruction and environmental noise in cluttered scenes fracture the continuity of localized feature streams. The holistic dependency encoded in self-attention operators leads to the erosion of low-level textural cues. Taken together, these phenomena introduce systematic detection biases in pipelines anchored on the Real-Time Detection Transformer (RT-DETR). To mitigate these impediments in UAV-borne object detection, Zhu *et al.* [33] introduced the Multi-Scale Maize Tassel-RTDETR (MSMT-RTDETR) formulation by streamlining both the RT-DETR backbone and the associated feature fusion schema, thereby achieving consistent and reliable detection metrics for maize tassel instances. Working from a YOLOv8 base, Tang *et al.* [34] imported the Swin Transformer module to magnify semantic feature extraction fidelity. Simultaneously, they embraced the Bidirectional Feature Pyramid Network (BiFPN) paradigm to enhance multi-scale target discrimination, culminating in the STB-YOLO network, purpose-built for object detection in overhead maize tassel captures.

A contingent of investigations has also explored alternative instantiations within the broader YOLO lineage, charting the versatility and scope of distinct model variants across diverse detection settings [35, 36]. Such explorations yield a richer palette of deployment options for YOLO-based solutions in breeding-oriented UAV surveillance scenarios. Lupión *et al.* [37] fabricated ThermalYOLO by targeted fine-tuning of the YOLOv3 architecture, achieving elevated bounding-box recall rates and Intersection over Union (IoU) scores for faithfully reconstructing 2D poses from multi-perspective thermographic streams. Eschewing the Path Aggregation Network (PANet) design, Doherty *et al.* [38] instead operationalized the BiFPN topology to advance a YOLOv5-derived detector termed BiFPN-YOLO. Song *et al.* [39] addressed occlusion phenomena and object diversity challenges inherent in aerial captures by deploying UAVs to collect multi-temporal overhead imagery and subsequently developed the SEYOLOX-tiny detection framework to deliver trustworthy identification under authentic field conditions. Seeking to improve detection robustness, Tang *et al.* [40] proposed an anchor-free detection paradigm, Circle-YOLO, capable of precise localization without manual anchor tuning.

Venturing beyond established methodological channels, select research threads have pivoted toward non-YOLO formulations and anchor-free algorithmic families, with a dedicated emphasis on maize tassel recognition challenges. These inquiries not only inject fresh conceptual vantage points into crop phenotyping workflows but also supply an eclectic toolkit of solutions for UAV-assisted maize breeding programs. Liu *et al.* [41] amassed a corpus of aerial target captures through UAV sorties and proceeded to train detection architectures using Faster Region-based Convolutional Neural Network (Faster R-CNN), Residual Network (ResNet), and Visual Geometry Group Network (VGGNet) as alternate feature extraction substrates, thereby instantiating an aerial object detection modeling pipeline. Ye *et al.* [42] compiled the Maize Tassels Detection and Counting UAV (MTDC-UAV) repository from UAV-sourced target snapshots. They introduced an aerial detection framework, dubbed FGLNet, that harmonizes global contextual signals with local details and materially boosts model efficacy through learned weighting policies. Grappling with the inherent tension between model compactness and holistic performance in prevailing detection regimes, Yu *et al.* [43] developed an innovative lightweight feature aggregation backbone for maize tassels, termed TasselLFANet, organized into an Encoder, a Decoder, and a Dense Inference stage. The design further exploits diverse receptive field configurations to capture heterogeneous feature encodings and embeds the Multi-Efficient Channel Attention (Mlt-ECA) unit to support high-fidelity feature detection.

Materials and Methods

Maize-breeding UAV construction

The aerial platform engineered for this research demonstrated strong competency in gathering maize tassel imagery and proved transferable to broader crop phenotyping workflows. Five foundational subsystems form the backbone of the

© 2023 The Author(s). | Published by Agro Intelligence Research Press

experimental maize-breeding UAV, specifically the flight controller, positioning module, data link, drive train, and onboard power source. Harmonized interplay among these elements permits accurate trajectory governance and operational fulfillment. The architectural breakdown of the custom-built maize-breeding UAV is visualized in **Figure 1**.

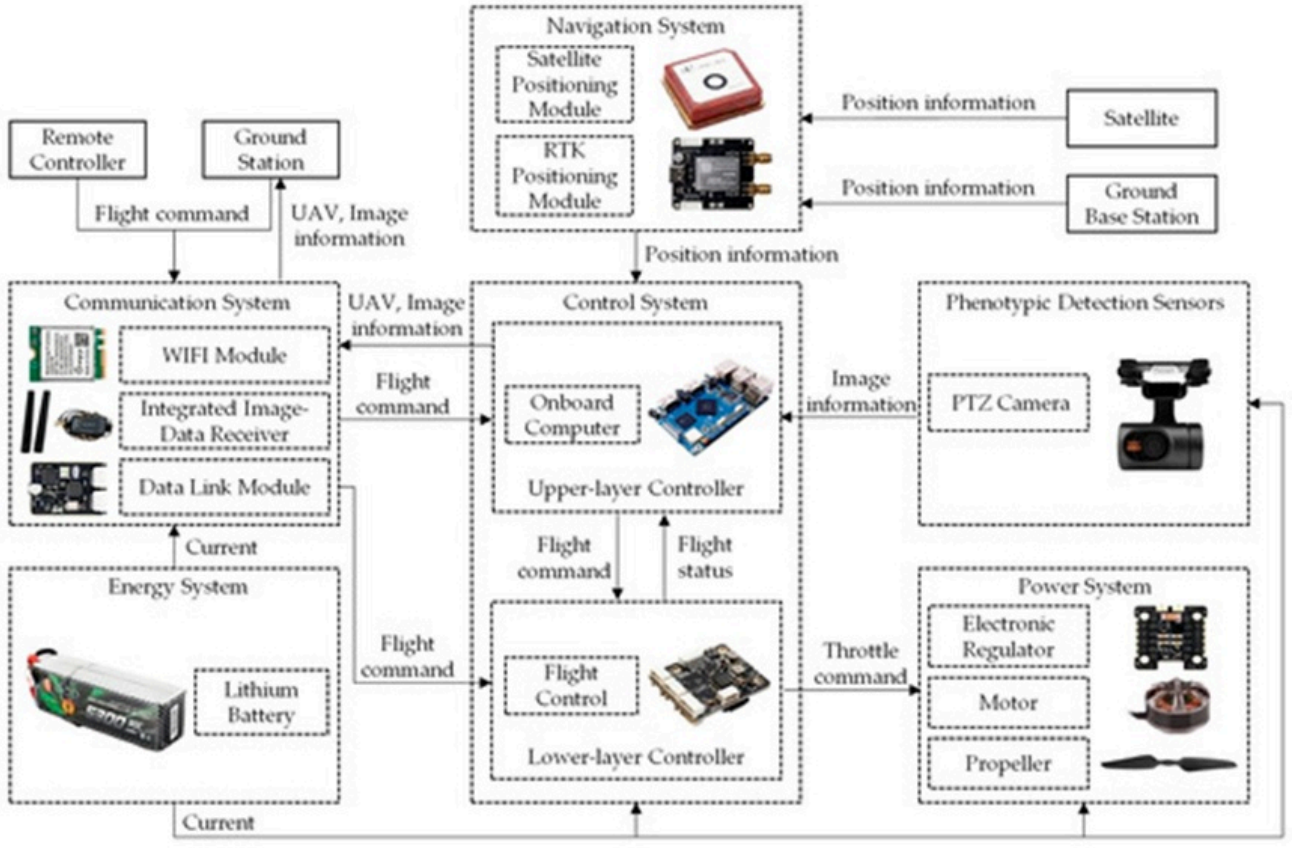

Figure 1. Hardware components of the designed maize-breeding UAV for tassel detection.

Figure 2 illustrates the full-scale structural layout of the maize-breeding UAV. Shrinking the physical envelope was accomplished without compromising airborne persistence. Sporting a diagonal motor span of 350 mm and tipping the scales at 1.4 kg, the platform delivers between 15 and 20 minutes of continuous flight. The companion computer selected for integration, with a footprint of 89 × 56 × 1.6 mm and a mass of merely 62 g, runs on the RK3588s multi-core silicon as its primary logic engine. This unit endows the UAV with 6 Tbps of onboard edge-computing throughput, shouldering the processing load for autonomous guidance, live object recognition, and downstream analytics. A 1080P high-definition gimbal-mounted RGB imaging payload, measuring 49 × 57.5 × 47.1 mm and weighing roughly 55 g, features a tunable two-axis stabilized cradle. It consistently captures crisp phenotypic snapshots of maize specimens, supporting subsequent phenotypic evaluation.

© 2023 The Author(s). | Published by Agro Intelligence Research Press



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

Maize tassel dataset creation

To gauge practical utility across varied deployment circumstances, including maize breeding and phenotypic monitoring, this investigation designated maize tassels as the object class for algorithm training and benchmarking. Source imagery was gathered in Quzhou County, within Handan City, Hebei Province, China (114°98′ E, 36°66′ N). The test plot extends across roughly 3000 square meters. The zone falls under a temperate monsoon climate regime, typified by sultry, precipitation-heavy summers and frigid, moisture-poor winters. Summer maize was sown on 15 June 2025 with the MY73 hybrid. Spacing followed an intra-row distance of 18 cm and an inter-row width of 60 cm.

Field data were collected on 9 August 2025 across multiple daytime intervals. The imaging tool was the maize-breeding UAV fitted with its onboard camera, running at a pixel array of 1920 × 1080. The drone was piloted at a constant velocity in manual operation mode, held at a ceiling of 4 m, with the lens pitched to gaze directly ahead while angled downward by a set degree. The workflow for airborne data collection is mapped out in **Figure 3**. The craft coursed over every planted row, blanketing the entire breeding parcel via an “S”-pattern sweep. While hovering above the plant canopy, the camera was triggered to snap the uppermost plant structures, procuring raw maize tassel imagery. The serpentine routing ensures contiguous, row-by-row coverage, devoid of unsampled gaps, while exploiting a conical field of regard to capture tassel visuals.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

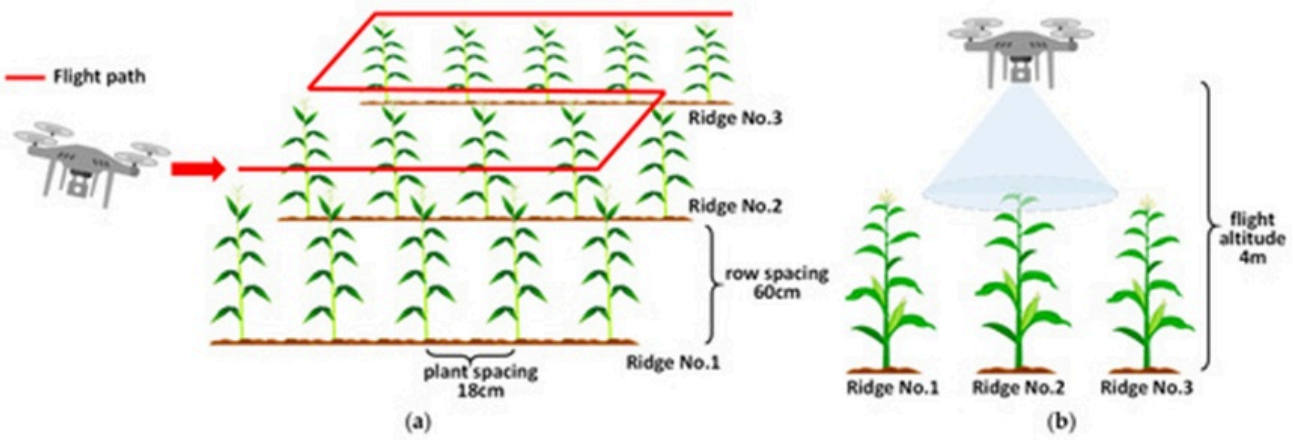

Figure 3. Schematic of the collection process for aerial tassel data by the maize-breeding UAV. Note: (a) represents the UAV flight path and the parameter of maize planting, and (b) represents the UAV flight altitude and the camera's field of view.

To mitigate plant sway, all recording missions were conducted during still-air intervals. The harvested tassel imagery spanned a range of authentic and demanding field scenarios, from intense sunlight and deep shadow to partial line-of-sight blockage. Recognizing that native image resolutions are impractical as direct network inputs, all frames in this study were uniformly rescaled to a 640 × 640-pixel canvas.

To assemble the working corpus, 500 stills were randomly selected from the pool of collected raw captures. The LabelImg tagging application was used to delineate every visible tassel in the curated set. Alongside this, the dataset was split into training, validation, and test sets at 8:1:1, yielding 400, 50, and 50 UAV-sourced frames, respectively. Beyond in-house field captures, this work also drew upon the Maize Tasseling Stage (MTS) dataset [44], an open-access resource (https://github.com/islls/MTS-master-JX.git, accessed on 3 November 2025). Characterized by richly varied origins encompassing diverse geographies and growing seasons, the MTS benchmark holds substantial referential weight in maize tassel recognition research. The repository aggregates 7 image sequences bearing maize tassels, curated from field trials situated in three provinces: Hebei, Henan, and Shandong. These sequences carry identifiers Hebei2011, Hebei2014, Henan2011, Henan2012, Shandong2011_1, Shandong2011_2, and Shandong2012_2013, collectively spanning the years 2011 through 2014. In aggregate, the MTS collection comprises 160 frames, partitioned into 90 training, 30 validation, and 40 test examples, paired with 5240 annotated tassel occurrences. Because it is the object instances themselves that drive loss computations during optimization, and because minute objects like tassels tend to cluster multiple times within a single frame, an instance-centric viewpoint supersedes an image-centric tally in this task domain. The distribution of labeled instances across the maize tassel corpus is catalogued in **Table 1**. Both the self-constructed dataset and the MTS dataset are single-class. **Figure 4** exhibits representative visual excerpts from the assembled collections. Both the in-house dataset and the MTS dataset served jointly as the basis for model optimization and performance auditing.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

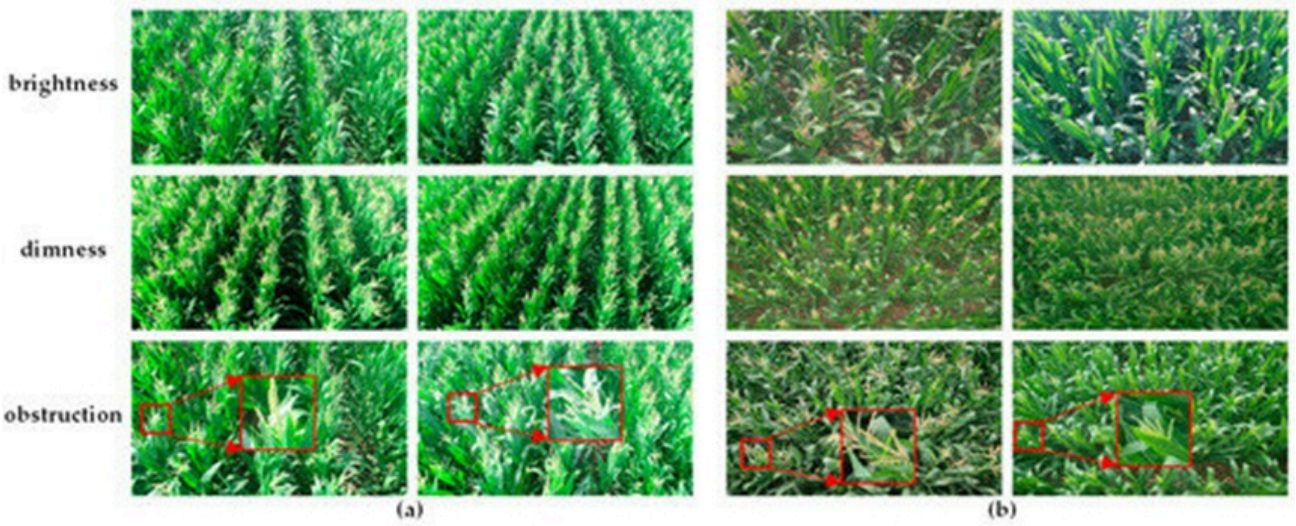

Figure 4. Dataset samples of maize tassel images. Note: (a) represents the tassels obtained by the maize-breeding UAV in this paper, and (b) represents the tassels from the public MTS dataset.

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

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

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.

Fusion approach of SCConv for improvement of the neck network

The YOLOv8 detector [45] was adopted as the base recognition engine. Because genuine operating backdrops are seldom pristine, contemporary architectures routinely increase their computational demands and parameter counts to achieve higher fidelity, yielding bulky configurations that resist streamlined UAV integration. SCConv [46], functioning as a resource-efficient convolutional construct, demonstrably suppresses both channel-direction and spatial-direction redundancies without degrading representational quality. Consequently, SCConv was embedded into the reference detector.

Fusion approach of spatial reconstruction unit of SCConv

Internally, SCConv decomposes into a Spatial Reconstruction Unit (SRU) followed by a Channel Reconstruction Unit (CRU). Input activations are initially passed through SRU to produce spatially refined maps, and thereafter channeled through CRU to yield channel-refined maps—the SRU sub-block trims spatial-domain excess through a dedicated Separate-then-Reconstruct protocol. The logic flow governing SRU is sketched in **Figure 5**.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

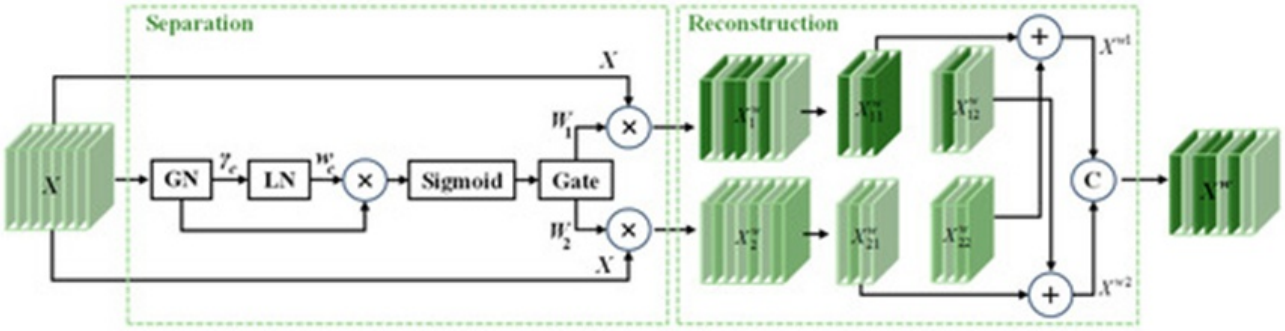

Figure 5. The flowchart of the SRU for spatial feature reconstruction in the maize tassel detection model.

In the first phase, the separation step partitions activation maps into those relatively abundant in informational signal and those comparatively devoid of it, establishing a dichotomy aligned with spatial statistics.

The informational magnitude carried within each activation map is appraised via Group Normalization (GN), governed by the following relation:

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

Where X represents the incoming feature tensor, and X_out its normalized counterpart. The terms μ and σ represent the computed mean and standard deviation, respectively. γ and β denote affine transform coefficients, while ε is a small positive constant that ensures numerical stability.

The learned scaling factor γi captures pixel-wise variation as a proxy for spatial information density and is subsequently normalized via Layer Normalization (LN). The computation is given by:

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

Where C indexes the channel dimensionality.

Informative versus uninformative salience weights, expressed as W_1 and W_2, emerge once a gating threshold is imposed. The weighting rule follows:

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

Where Sigmoid denotes the logistic activation nonlinearity.

Following separation, the reconstruction phase merges information-rich and information-sparse activation subsets, synthesizing refined output tensors and realizing spatial economy.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

The derived weight matrices W_1 and W_2 each undergo element-wise product with the original input X, forming two weighted activation branches. These branches are then interlaced via a cross-reconstruction fusion, generating and . Their concatenation produces the ultimate spatially reconstructed feature bundle . This is encapsulated as:

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

Where ⊗, ⊕, and ∪ respectively denote element-wise multiplication, element-wise summation, and channel-wise concatenation.

Fusion approach of channel reconstruction unit of SCConv

Channel-wise redundancy within the incoming feature maps is tackled by the CRU through a procedure organized around three sequential stages: splitting, transforming, and fusing. A visual breakdown of the CRU workflow is provided in **Figure 6**.

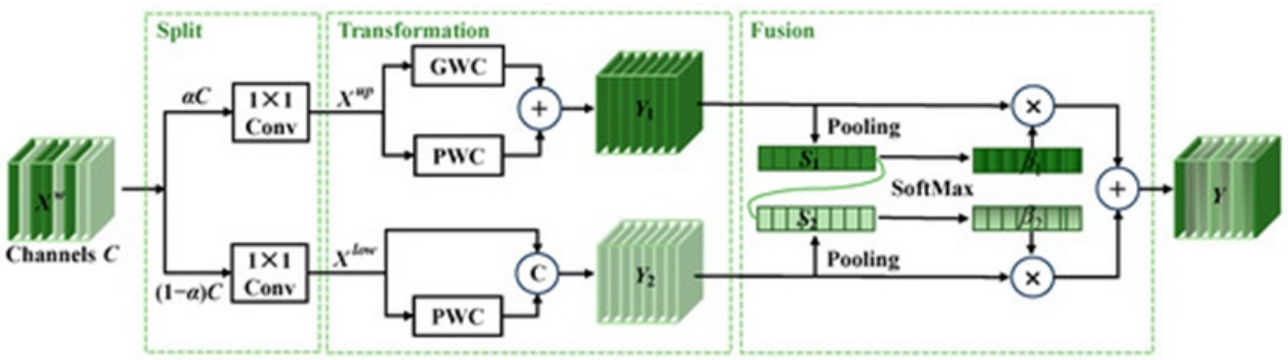

Figure 6. The CRU flowchart for channel-feature reconstruction in the maize tassel detection model.

The procedure begins with a split operation, where the spatially purified feature map X^w arriving from the SRU is accepted as the CRU's input and divided into two distinct portions. A 1 × 1 convolutional squeeze is applied to the channel dimensions of both resultant tensors, yielding the compressed representations X^{up} and X^{low}.

Moving to the transformation phase, X^{up} serves as the substrate for in-depth feature mining. It is routed through two parallel branches—Group-Wise Convolution (GWC) and Point-Wise Convolution (PWC)—whose outputs are then summed to form Y_1. In a secondary processing thread, X^{low} is forwarded through a point-wise convolution, and the resulting activation is combined with the unmodified X^{low} to construct Y_2. The formal specification is:

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

© 2023 The Author(s). | Published by Agro Intelligence Research Press

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

Within which M^G and M^P correspond to the learnable kernel tensors of the GWC and PWC operations, respectively.

The concluding fusion stage merges Y_1 and Y_2 into a unified tensor Y. Global contextual statistics S_i are distilled from each Y_i via a pooling operator. A SoftMax transformation then converts these statistics into a weighting vector β_i, which selectively emphasizes the final channel-refined output tensor Y. The defining expressions for Y and β_i are:

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

$$\beta_t = \mathrm{SoftMax}(S_t) = \mathrm{SoftMax}(\mathrm{Pooling}(Y_t)) \quad (8)$$

The YOLOv8 design can be conceptually divided into three functional blocks: the backbone for feature extraction, the neck for feature aggregation, and the detection head. The proposed phenotyping perception detector is realized by embedding SCConv into the C2f modules residing in the neck of the baseline YOLOv8. More precisely, the vanilla convolution operator in the reference model is replaced with SCConv to form the SCConv Block, and the original neck C2f component is remodeled into C2f-SCConv. **Figure 7** offers an overview of the SCConv-enhanced network. By maintaining accuracy standards, the SCConv unit meaningfully compresses the total parameter count, reduces the computational budget, and lowers the hardware threshold for edge devices, thereby promoting seamless migration onto UAV-borne systems.

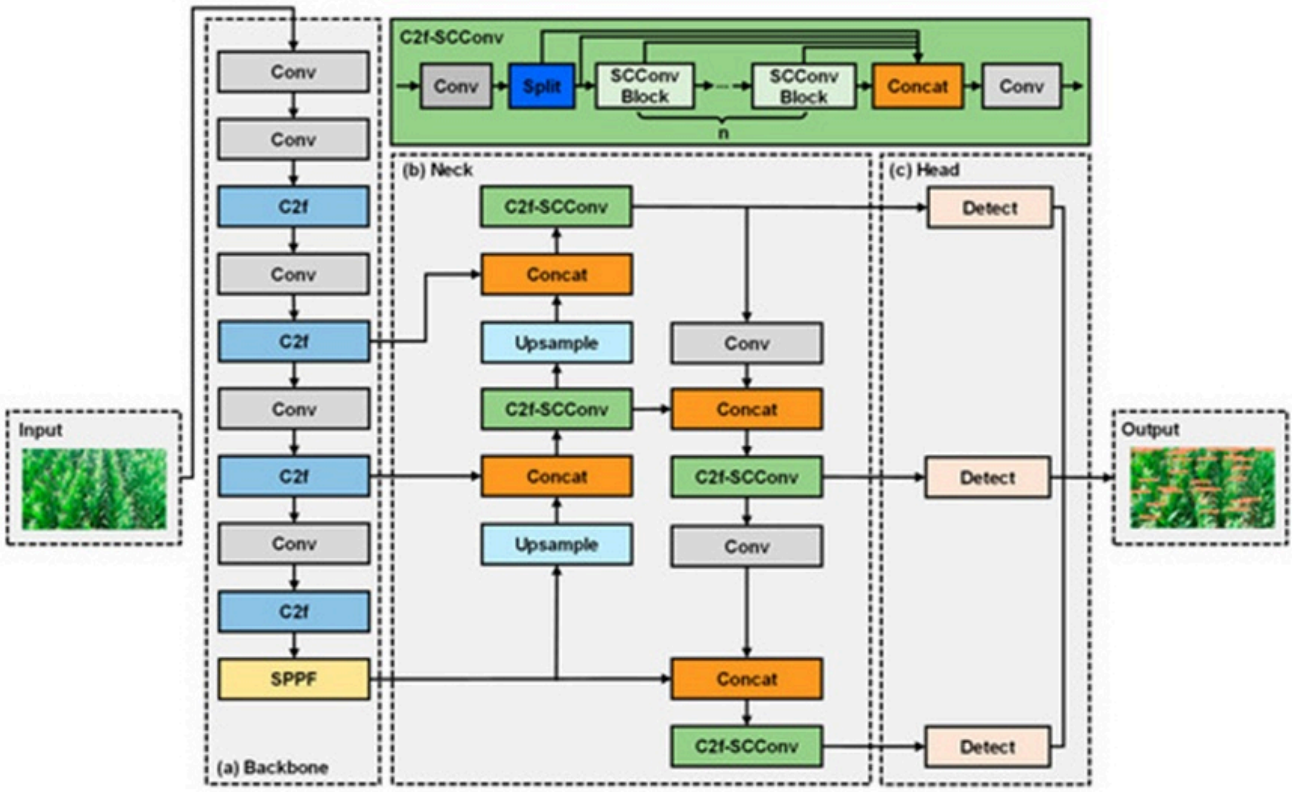

Figure 7. The flowchart of the enhanced YOLOv8 model using SCConv for maize tassel detection.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Conventional object detectors typically estimate spatial locations via axis-aligned bounding box regression. The regression objective measures the spatial agreement between predicted and reference rectangles. The Intersection over Union (IoU) metric follows the formulation:

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

Where S_u captures the total area of the smallest enclosing rectangle spanning both boxes, while W_0 and H_0 describe the width and height of the shared region, respectively.

Moving beyond simple overlap, the Complete Intersection over Union (CIoU) [47] criterion integrates supplementary geometric relationships—namely, the separation between box centers, the consistency of aspect ratios, and the extent of intersection. The SCConv-enhanced detection model introduced in this manuscript adopts CIoU as its bounding box regression loss, expressed by:

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

Where R_{CIoU} stands for the additive regularization term augmenting the base IoU calculation.

Experiment settings and evaluation indicators

All experiments reported herein were executed on a single desktop machine. The central processing unit is an Intel(R) Core(TM) i7-14700HX operating at 2.10 GHz, paired with an NVIDIA GeForce RTX 4060 graphics processor. The software environment runs on Windows 11. Development was conducted in Python 3.8.19, leveraging the PyTorch 2.0.0 deep learning framework. The supporting CUDA and Torchvision toolkit versions are 11.8 and 0.15.0, respectively.

This study anchors its investigation on the YOLOv8 architecture as the reference point. The original detector was iteratively refined and retrained to produce the final enhanced model. The hyperparameter set governing the training run is tabulated in **Table 2**.

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

Training parameter	Configured value
Number of epochs	100
Batch size	16
Input image resolution	640 × 640
Data loader workers	4
Optimization algorithm	Stochastic Gradient Descent (SGD)
Starting learning rate	0.01
Momentum coefficient	0.937
Regularization (Weight decay)	0.0005

© 2023 The Author(s). | Published by Agro Intelligence Research Press

The global assessment of the improved detector used the metrics precision (P), recall (R), and mean average precision (mAP). Additionally, model size on disk and floating-point operations (FLOPs) were tracked as secondary descriptors of efficiency.

The definitions for precision and recall are:

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

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

Where the quantities TP, FP, and FN refer to the counts of true positive, false positive, and false negative predictions, respectively. In sequence, these tally the objects correctly assigned to the positive category, mistakenly assigned to the positive category, and wrongly assigned to the negative category.

The mAP score is computed through:

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

In this expression, AP denotes average precision, which quantifies the area under the precision–recall curve. The symbol n indicates the total number of object classes within the dataset. The notation mAP@0.5 denotes the mean AP across all categories evaluated at an IoU matching threshold of 0.5.

Results and Discussion

Fusion-based neck network improvement experiments

The present work improved the YOLOv8n baseline by integrating the SCConv module. Incorporating SCConv into the model's neck section produced a substantial contraction in parameter volume while fully retaining the model's capacity for discriminative feature learning. To benchmark the effectiveness of SCConv against competing approaches, comparative trials were orchestrated pitting Switchable Atrous Convolution (SAConv) [48], Omni-Dimensional Dynamic Convolution (ODConv) [49], and SCConv against one another. The comparative findings from the neck network augmentation study are collated in **Table 3**.

Table 3. Results of neck network improvement experiments for maize tassel detection.

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

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Reviewing the data in **Table 3**, the SCConv-augmented variant achieved an mAP@0.5 of 91.7% and a computational load of 7.3 G FLOPs. It occupied 5.16 MB of storage, outperforming the alternative convolution schemes in the efficiency-accuracy trade-off. Compared with the unmodified starting model, the SCConv-enhanced version trimmed 0.8 G FLOPs from the processing budget and shaved 0.79 MB from the stored model weight file. These observations confirm that embedding SCConv reduces parameterization while maintaining strong feature extraction capabilities and streamlining the path toward field deployment and practical use.

Ablation experiments of enhanced detection model

To disentangle and verify the specific contribution of each design modification, the YOLOv8n baseline was taken as the control, and individual improvement components were methodically activated through a series of ablation runs. The resulting metrics are compiled in **Table 4**. **Figure 8** plots the training performance curves of the enhanced model.

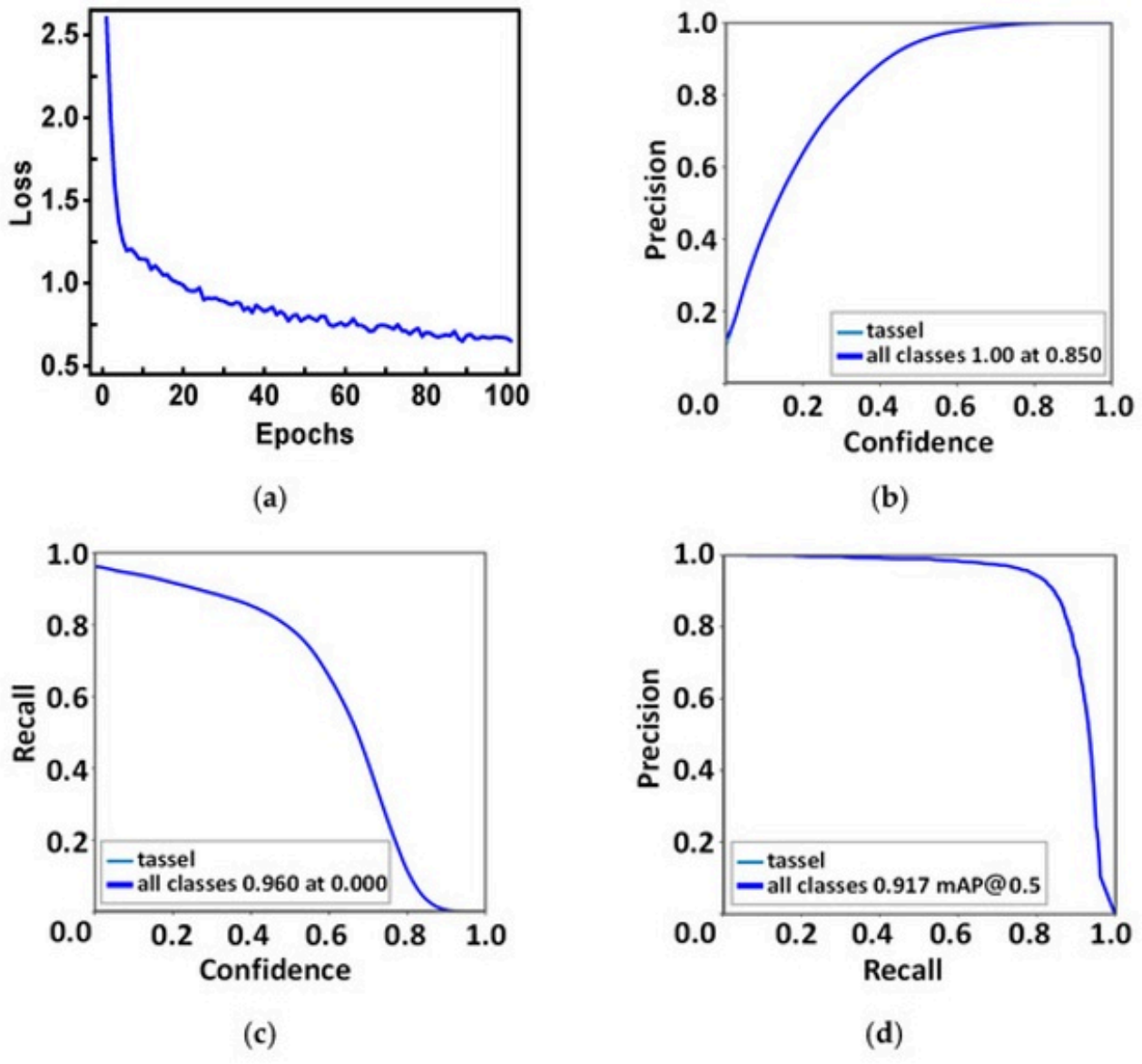

Figure 8. Training results for the maize tassel detection model using SCConv. Note: (a) represents the loss values during model training, (b) represents the precision–confidence curve, (c) represents the recall–confidence curve, and (d)

© 2023 The Author(s). | Published by Agro Intelligence Research Press

represents the precision–recall curve.

Table 4. Results from improved point ablation experiments for maize tassel detection.

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

Note: "√" and "×", respectively, indicate whether the improvement point has been adopted.

The first experimental row corresponds to the untouched YOLOv8n baseline, serving as the reference benchmark for all subsequent comparisons. This reference model reached precision, recall, and mAP@0.5 values of 89.0%, 80.9%, and 88.3%, respectively. Its inference complexity registered at 8.1 G FLOPs, with the weight footprint measuring 5.95 MB.

The second row shows the performance metrics after integrating the SCConv module. The augmented detector achieved a precision of 92.2%, a recall of 84.3%, and an mAP@0.5 of 91.7%. Its computational demand and storage requirements dropped to 7.3 GFLOPs and 5.16 MB, respectively. Relative to the initial configuration, these figures represent absolute percentage-point gains of 3.2% in precision, 3.4% in recall, and 3.4% in mAP@0.5. Simultaneously, the processing workload contracted by 0.8 G FLOPs, and the model's disk image compressed by 0.79 MB. This body of evidence establishes that the optimized model couples reduced computational intensity and a more compact footprint with superior detection scores, characteristics that strongly favor embedding on resource-constrained UAV hardware and demonstrate heightened suitability for maize tassel perception in real operational environments.

Examining the subplots in **Figure 8**: panel (a) traces a loss descent that rapidly decays before leveling out, a pattern consistent with stable convergence where the model progressively internalizes tassel-specific features and eventually settles at a minimal error floor, confirming training soundness. Panel (b) reveals that precision climbs as the classifier confidence threshold is raised, denoting a low false-alarm rate under strict detection criteria and indicating reliable precision behavior. Panel (c) illustrates a declining recall trend as the confidence requirement tightens, suggesting that under lenient thresholds, the detector successfully retrieves the vast majority of genuine tassel occurrences, highlighting its strong recall response. The substantial area beneath the curve traced in panel (d) signals that the enhanced detector strikes a productive compromise between precision and recall, delivering well-rounded, holistic detection quality.

Comparison experiments for enhanced detection model

A broader empirical validation was undertaken to substantiate the model's merits in the maize tassel detection domain. Both the custom-assembled aerial maize tassel corpus and the publicly accessible MTS dataset served as training substrates. Every candidate detector was evaluated head-to-head against the SCConv-augmented YOLOv8n variant, with all runs governed by uniform experimental conditions and identical parameterization. The pool of benchmark algorithms spanned Faster-RCNN, SSD, RT-DETR, YOLOv5n, YOLOv8n, YOLOv9t, YOLOv10n, and YOLOv11n, each trained to convergence for the maize tassel detection objective. A comprehensive summary of the resulting performance figures is presented in **Table 5**.

Table 5. Results of comparative experiments for different maize tassel detection models.

No.	Model	Model size/MB	FLOPs/G	mAP@0.5/%	R/%	P/%
1	Faster-RCNN	106.40	134.8	66.2	59.5	68.2
2	SSD	94.20	57.4	72.8	63.2	74.5
3	RT-DETR	63.45	108.6	90.6	83.7	91.5

© 2023 The Author(s). | Published by Agro Intelligence Research Press

4	YOLOv5n	5.01	8.7	82.4	76.8	83.4
5	YOLOv8n	5.95	8.1	88.3	80.9	89.0
6	YOLOv9t	4.73	7.8	87.2	79.4	87.8
7	YOLOv10n	5.58	7.6	88.8	81.2	90.4
8	YOLOv11n	5.35	6.7	86.5	80.6	88.7
9	Ours	5.16	7.3	91.7	84.3	92.2

Turning to the data compiled in **Table 5**, Faster-RCNN mustered only a 68.2% precision score, a 59.5% recall, and a 66.2% mAP@0.5, while demanding 106.40 MB of storage—evidence of lackluster results on both the detection quality and model compactness fronts. SSD improved marginally to 74.5% precision, 63.2% recall, and 72.8% mAP@0.5, occupying 94.20 MB; though representing a step forward from Faster-RCNN, its bulk precludes practical deployment on edge hardware. RT-DETR achieved strong 91.5% precision, 83.7% recall, and 90.6% mAP@0.5, yet these figures came at a steep price: 108.6 G FLOPs and a 63.45 MB footprint—a clear case of detection prowess undercut by poor efficiency. YOLOv5n yielded 83.4% precision, 76.8% recall, and 82.4% mAP@0.5, with 8.7 G FLOPs and 5.01 MB, positioning it as a lightweight contender but one that sacrifices considerable detection capability. The baseline YOLOv8n produced 89.0% precision, 80.9% recall, and 88.3% mAP@0.5, consuming 8.1 G FLOPs and 5.95 MB—a measurable advance over YOLOv5n, though not yet at the performance ceiling. YOLOv9t contributed 87.8% precision, 79.4% recall, and 87.2% mAP@0.5, with a lean 7.8 G FLOPs and 4.73 MB footprint; its strength lies in miniaturization, but detection metrics fall behind. YOLOv10n achieved 90.4% precision, 81.2% recall, and 88.8% mAP@0.5, drawing 7.6 G FLOPs and weighing 5.58 MB—a well-rounded performer, yet not the most favorable outcome achievable. YOLOv11n achieved 88.7% precision, 80.6% recall, and 86.5% mAP@0.5 at a computational cost of 6.7 G FLOPs and a model size of 5.35 MB; its processing efficiency is unmatched, but detection quality remains merely adequate. The detector proposed in this work delivers 92.2% precision, 84.3% recall, and 91.7% mAP@0.5, alongside a 7.3 G FLOPs processing load and a 5.16 MB storage requirement. In a direct comparison with the YOLOv8n baseline, improvements of 3.2, 3.4, and 3.4 percentage points were observed in precision, recall, and mAP@0.5, respectively. At the same time, 0.8 G FLOPs were trimmed from the computational budget, and 0.79 MB were carved off the model's disk image. Set against YOLOv10n, precision, recall, and mAP@0.5 rose by 1.8, 3.1, and 2.9 percentage points, with corresponding reductions of 0.3 G FLOPs and 0.42 MB. In summary, the proposed detector outperforms all competing architectures in detection accuracy while maintaining a lightweight, efficient profile—representing the most favorable trade-off between perceptual quality and resource economy and thus being distinctly more amenable to integration and field operation aboard maize-breeding UAV platforms.

Visualization and analysis of enhanced detection model

Qualitative evidence for the model's practical value was gathered through a visual comparison study, with representative outputs displayed in **Figure 9.** A frame-by-frame inspection across varying field conditions confirmed the robustness and practical effectiveness of the SCConv-enhanced YOLOv8n.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

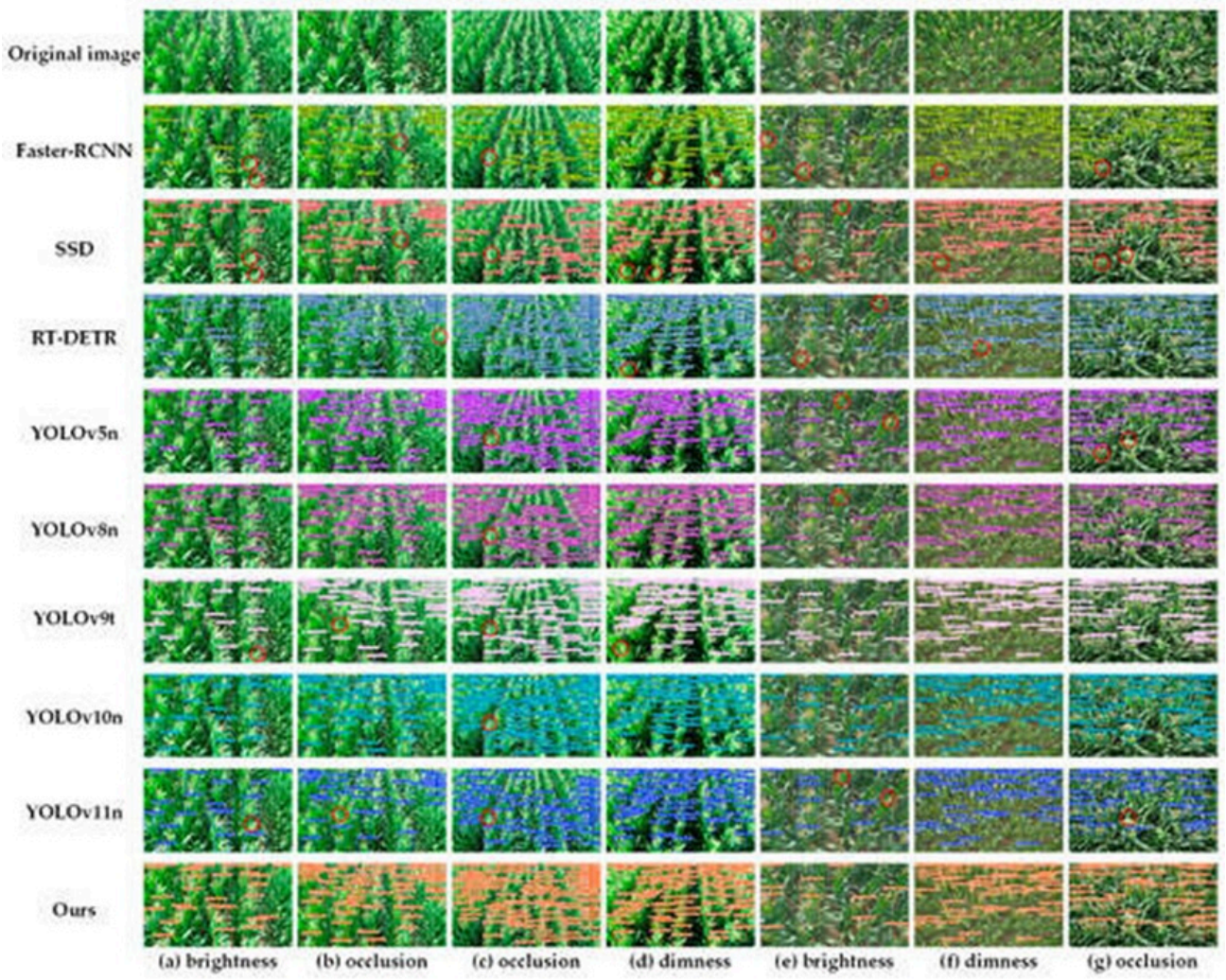

Figure 9. Visualization of results from different models in the maize tassel detection task. Note: (a–d) represent the tassel images from the dataset of this paper, and (e–g) represent the tassel images from the MTS dataset, which include scenarios such as brightness, dimness, and occlusion. The red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

The evidence in **Figure 9** reveals the following patterns. Under strong illumination, Faster-RCNN is plagued by recurring omissions where entire tassels escape detection; as lighting grows dim, the failure rate escalates dramatically, with the model proving nearly blind in low-light regimes; whenever occlusion is present, tassels are overwhelmingly ignored—a profile of extremely fragile perception, persistent false negatives, and negligible resilience to real-world complexity. SSD continues to miss targets noticeably in well-lit frames, producing incomplete tassel coverage; under dim conditions, the tally of undetected instances swells and accuracy crumbles as illumination wanes; occluded tassels suffer from severe, systematic non-detection. RT-DETR largely avoids omissions under favorable lighting, producing passable results; however, as lighting deteriorates, missed detections become conspicuous, and its capacity to discriminate tassels proves wanting; occluded scenarios still induce a non-trivial number of overlooked instances. Its strength is confined primarily to bright environments, with dim and occluded conditions exposing persistent vulnerabilities. YOLOv5n achieves relatively sparse missed detections in bright settings, but introducing dim light causes the omission count to multiply and performance to deteriorate sharply; under occlusion, its target discrimination ability is limited, indicating residual room for model refinement. YOLOv8n handles bright scenes with few missed detections and solid overall detection; in dim conditions, its performance erodes somewhat, leading to scattered omissions; occluded frames trigger partial non-detection and unstable target recognition. Although its aggregate performance is reasonably strong, gaps remain in complex scenarios. YOLOv9t remains reliable in bright light but falters notably as illumination fades, with a marked

© 2023 The Author(s). | Published by Agro Intelligence Research Press

increase in missed tassels; occlusion also induces clear detection failures. YOLOv10n exhibits near-flawless detection in bright environments; dim light provokes a modest number of missed instances, and occlusion leads to partial omissions—overall a respectable showing, albeit with room for further gains under challenging conditions. YOLOv11n delivers high precision in well-illuminated frames; however, dim lighting leads to a pronounced increase in missed detections, and occluded views result in clear failures.

Synthesizing the visual evidence, a key finding emerges: the enhanced phenotyping detection model consistently localized maize tassels that were partially concealed behind leaves or stalks. Equally important, the refined detector maintained accurate recognition across the full spectrum of tested conditions—bright, dim, and occluded alike. These observations collectively affirm the model's practical suitability, robust effectiveness, and enhanced detection quality. The detector proves both accurate and dependable for maize tassel perception, offering distinct advantages for real-world, operationally complex field environments. In conclusion, the enhanced model delivers a combination of precision and reliability, and critically, the improvements in detection fidelity were achieved alongside significant compression of both computational footprint and model storage. This dual advancement renders the detector well-suited for embedding and operational deployment on maize-breeding UAVs. It establishes a valuable methodological reference point for the broader field of object detection.

Within the broader movement toward precision agriculture, UAV-based aerial object detection has ascended to a central role in maize phenotyping and trait assessment pipelines. This investigation directly confronts the primary bottlenecks afflicting current detection architectures—namely, excessive model volume and poor deployability on resource-constrained platforms—by proposing an SCConv-integrated enhancement of the YOLOv8 framework, yielding several noteworthy technical advances.

The first contribution is the bespoke construction of a quadrotor UAV platform, purpose-built for maize breeding applications, that reliably captures high-quality overhead imagery of maize tassels. Building on this hardware foundation, the second contribution involves strategically embedding SCConv within the neck subnetwork of the YOLOv8 architecture. This design choice effectively suppresses redundant feature representations and simultaneously enables a coupled optimization that boosts detection accuracy while reducing model complexity. Quantitatively, the resulting detector reaches 92.2% precision, 84.3% recall, and 91.7% mAP@0.5, operating within a 7.3 G FLOPs compute envelope and occupying just 5.16 MB. Relative to the unmodified YOLOv8n baseline, these figures translate to absolute gains of 3.2% in precision, 3.4% in recall, and 3.4% in mAP@0.5, while 0.8 G FLOPs of computation and 0.79 MB of model weight are eliminated. Benchmarked against YOLOv10n, the proposal saves an additional 0.3 G FLOPs in processing and 0.42 MB in storage. The third contribution is the model's demonstrated adaptability to challenging and variable field conditions. The joint utilization of the self-constructed maize tassel dataset and the MTS repository for training and evaluation confirmed that the enhanced detector sustains high accuracy on UAV-captured aerial imagery across bright, dim, and occluded scenarios—robust empirical evidence of its practical value in authentic, in-field maize phenotyping operations.

The persistent tension among detection fidelity, processing economy, and ease of embedding has long constituted a critical impasse within UAV-driven crop phenotyping research. The preponderance of existing architectures gravitates toward one of two extremes: they either pursue elevated accuracy at a steep computational cost—RT-DETR and Faster-RCNN being representative cases—or champion compactness while accepting degraded detection reliability under the variable and demanding conditions of real-world fields, as exemplified by YOLOv5n. By contrast, the refined model introduced here mitigates this trade-off by leveraging SCConv's inherent feature-compression properties. The outcome is a detector that not only surpasses prevailing lightweight alternatives across key performance indicators but also maintains a restrained storage profile and low processing overhead, making it compatible with UAV-based deployment constraints.

When viewed through an applied lens, the present work furnishes two pivotal supports for maize breeding endeavors and the wider precision agriculture paradigm. First, it delivers a compact yet high-accuracy instrument for maize tassel phenotyping, substantially streamlining the throughput of UAV-enabled phenotypic data gathering. Second, the

© 2023 The Author(s). | Published by Agro Intelligence Research Press

methodological scaffold erected in this investigation lends itself to adaptation across other UAV-based crop object detection contexts, offering a blueprint for reconciling analogous tensions between performance and deployability in the phenotyping surveillance of staple crops, including wheat, rice, and soybean.

Conclusion

To address the shortcomings of current object detection frameworks—specifically, their burdensome computational demands and unwieldy model sizes that hinder practical integration and field deployment—we advanced a detection architecture that incorporates Spatial and Channel Reconstruction Convolution. The SCConv unit was woven into the neck subnetwork of a YOLOv8-derived backbone. Both the self-assembled maize tassel image corpus and the MTS benchmark were employed to optimize the enhanced detector. Experimental results showed that the improved model achieved precision, recall, and mAP@0.5 of 92.2%, 84.3%, and 91.7%, respectively. Concurrently, its computational throughput was 7.3 G FLOPs, and its storage size was 5.16 MB. Stacked against the baseline YOLOv8n, precision, recall, and mAP@0.5 rose by margins of 3.2%, 3.4%, and 3.4%, while processing load contracted by 0.8 G FLOPs and model bulk shrank by 0.79 MB. Benchmarked against YOLOv10n, the enhanced version delivered gains of 1.8% in precision, 3.1% in recall, and 2.9% in mAP@0.5, alongside savings of 0.3 G FLOPs in computation and 0.42 MB in model weight.

Notwithstanding the progress attained, several dimensions warrant continued refinement. Presently, the study faces constraints, including a focus on a single crop type, limited diversity in the underlying data, and a shortfall in integrating multi-source measurements. Looking ahead, the proposed methodology would be broadened to encompass food crop breeding programs for wheat, rice, and soybean. UAV platforms would be tasked with amassing crop observations spanning disparate geographic zones, phenotypic expressions, and developmental stages. In parallel, complementary sensor modalities—specifically LiDAR, multispectral, and thermal infrared—would be jointly fused into the analytical pipeline.

Acknowledgements

None

Conflict of interest

None

Financial support

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).

Ethics statement

None

Received: 06 Feb 2022 Revised: 12 Apr 2022 Accepted: 29 Apr 2022
Published online: 25 May 2023

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

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.

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.

3. Muhmad Kamarulzaman AM, Wan Mohd Jaafar WS, Mohd Said MN, Saad SNM, 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.

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.

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.

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.

7. Erenstein O, Jaleta M, Sonder K, Mottaleb K, Prasanna BM. Global maize production, consumption and trade: Trends and R&D implications. Food Secur. 2022;14:1295–319.

8. Jiang C, Ren H, Ye X, Zhu J, Zeng H, Nan Y, et al. Object detection from uav thermal infrared images and videos using yolo models. Int J Appl Earth Obs Geoinf. 2022;112:102912.

9. Zhang Y, Dong X, Wang H, Lin Y, Jin L, Lv X, et al. Maize breeding for smaller tassels threatens yield under a warming climate. Nat Clim Change. 2024;14:1306–13.

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.

11. Shu M, Fei S, Zhang B, Yang X, Guo Y, Li B, et al. Application of UAV multisensor data and ensemble approach for high-throughput estimation of maize phenotyping traits. Plant Phenom. 2022;2022:9802585.

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.

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.

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.

15. Badgujar CM, Poulose 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.

16. Wu D, Liang Z, Chen G. Deep learning for LiDAR-only and LiDAR-fusion 3D perception: A survey. Intell Robot. 2022;2:105-29.

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;5:474-87.

18. Seidaliyeva U, Akhmetov D, Ilipbayeva L, Matson ET. Real-time and accurate drone detection in a video with a static background. Sensors. 2020;20:3856.

19. Alhafnawi M, Salameh HAB, Masadeh AE, Al-Obiedollah H, Ayyash M, El-Khazali R, et al. A survey of indoor and outdoor UAV-based target tracking systems: Current status, challenges, technologies, and future directions. IEEE Access. 2023;11:68324-39.

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.

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.

22. Falahat S, Karami A. Maize tassel detection and counting using a YOLOv5-based model. Multimed Tools Appl. 2023;82:19521-38.

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.

24. Pu H, Chen X, Yang Y, Tang R, Luo J, Wang Y, et al. 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.

25. Yu X, Yin D, Xu H, Pinto Espinosa F, Schmidhalter U, Nie C, et al. Maize tassel number and tasseling stage monitoring based on near-ground and UAV RGB images by improved YOLOv8. Precis Agric. 2024;25:1800-38.

26. Liu D, Zhang Y, Ren Y, Pan D, He X, Cong M, et al. Multi-directional attention: A lightweight attention module for slender structures. Intell Robot. 2025;5:827-43.

27. Guan H, Deng H, Ma X, Zhang T, Zhang Y, Zhu T, et al. A corn canopy organs detection method based on improved DBi-YOLOv8 network. Eur J Agron. 2024;154:127076.

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.

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.

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;5:562-78.

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.

32. Yang G, Liu C, Li G, Chen H, Chen K, Wang Y, et al. DiffKNet-TL: Maize phenology monitoring with confidence-aware constrained diffusion model based on UAV platform. Comput Electron Agric. 2025;239:110936.

33. Zhu Z, Gao Z, Zhuang J, Huang D, Huang G, Wang H, et al. MSMT-RTDETR: A multi-scale model for detecting maize tassels in UAV images with complex field backgrounds. Agriculture. 2025;15:1653.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

34. Tang B, Zhou J, Li X, Pan Y, Lu Y, Liu C, et al. Detecting tasseling rate of breeding maize using UAV-based RGB images and STB-YOLO model. Smart Agric Technol. 2025;11:100893.

35. Lei T, Li G, Luo C, Zhang L, Liu L, Gates RS. An informative planning-based multi-layer robot navigation system as applied in a poultry barn. Intell Robot. 2022;2:313-32.

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;4:276-92.

37. Lupión M, Polo-Rodríguez A, Medina-Quero J, Sanjuan JF, Ortigosa PM. 3D human pose estimation from multi-view thermal vision sensors. Inf Fusion. 2024;104:102154.

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.

39. Song C, Zhang F, Li J, Xie J, Yang C, Zhou H, et al. Detection of maize tassels for UAV remote sensing image with an improved YOLOX model. J Integr Agric. 2023;22:1671-83.

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.

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.

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;17:4100-8.

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.

44. Ye J, Yu Z, Lin J, Lin L, Li H, Zhu X, et al. Maize tasseling stage automated observation method via semantic enhancement and context occlusion learning. IEEE Sens J. 2024;24:37259-74.

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); 2024; Chennai, India. New York (NY): IEEE; 2024. p. 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); 2023; Vancouver, BC, Canada. New York (NY): IEEE; 2023. p. 6153-62.

47. Zheng Z, Wang P, Ren D, Liu W, Ye R, Hu Q, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. IEEE Trans Cybern. 2021;52:8574-86.

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); 2021; Nashville, TN, USA. New York (NY): IEEE; 2021. p. 10213-24.

49. Li C, Zhou A, Yao A. Omni-dimensional dynamic convolution. arXiv. 2022;arXiv:2209.07947.

© 2023 The Author(s). | Published by Agro Intelligence Research Press