RESEARCH

Plant Methods

Open Access

UAS-based MT-YOLO model for detecting missed tassels in hybrid maize detasseling



Jiangtao Qi^{1,3}, Chenchen Ding^{1,2}, Ruirui Zhang^{2*}, Yuxin Xie⁴, Longlong Li², Weirong Zhang^{1,3} and Liping Chen^{4*}

Abstract

Accurate detection of missed tassels is crucial for maintaining the purity of hybrid maize seed production. This study introduces the MT-YOLO model, designed to replace or assist manual detection by leveraging deep learning and unmanned aerial systems (UASs). A comprehensive dataset was constructed, informed by an analysis of the agronomic characteristics of missed tassels during the detasseling period, including factors such as tassel visibility, plant height variability, and tassel development stages. The dataset captures diverse tassel images under varying lighting conditions, planting densities, and growth stages, with special attention to early tasseling stages when tassels are partially wrapped in leaves—a critical yet underexplored challenge for accurate detasseling. The MT-YOLO model demonstrates significant improvements in detection metrics, achieving an average precision (AP) of 93.1%, precision of 93.3%, recall of 91.6%, and an F1-score of 92.4%, outperforming Faster R-CNN, SSD, and various YOLO models. Compared to the baseline YOLO v5s, the MT-YOLO model increased recall by 1.1%, precision by 4.9%, and F1-score by 3.0%, while maintaining a detection speed of 124 fps. Field tests further validated its robustness, achieving a mean missed rate of 9.1%. These results highlight the potential of MT-YOLO as a reliable and efficient solution for enhancing detasseling efficiency in hybrid maize seed production.

Keywords Agricultural automation, Detasseling process, Hybrid maize seed production, Missed tassel detection, MT-YOLO

*Correspondence:

- Ruirui Zhang
- zhangrr@nercita.org.cn
- Liping Chen
- chenlp@nercita.org.cn
- ¹ College of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China
- ² Intelligent Equipment Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

³ Key Laboratory of Bionic Engineering, Ministry of Education, Jilin University, Changchun 130022, China

⁴ National Research Center of Intelligent Equipment for Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

Introduction

Maize is a crucial staple crop, accounting for approximately 35.7% of the world's total grain production [1], and serves as a primary resource for animal feed, fuel, and industrial raw materials. Hybrid maize, renowned for its increased yield and strong adaptability, dominates over two-thirds of global maize planting areas [2], with 95% of maize planting areas in China utilizing hybrid maize seeds [3]. To meet China's annual demand of 1.1 billion kilograms of hybrid maize seeds, it is essential to produce hybrid maize seeds every year, as their hybrid advantage diminishes from the second generation [4, 5]. Among the critical steps to ensure seed purity, detasseling-the removal of tassels from female maize parents-prevents self-pollination and ensures hybridization quality [6]. Despite the availability of advanced grounddetasseling machines, such as the Hagie 204SP from John Deere, the BD864 from Bourgoin, and China's 3CX-8A



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

[7], terrain fluctuations and high planting density often result in missed tassels after bulk detasseling. Traditionally, the detection of detasseling quality and the removal of missed tassels rely on labor-intensive manual methods, which are prone to subjectivity and inefficiencies. Considering the short 7–10 day window for detasseling during maize seed production, there is an urgent need for intelligent detection technologies and equipment to enhance the efficiency and precision of identifying and addressing missed tassels, thereby improving automation and reducing reliance on manual labor.

In recent years, UASs have gained prominence in agriculture and forestry, particularly for crop remote sensing and plant protection, due to their speed, maneuverability, terrain adaptability, and minimal soil compaction and plant damage [8–13]. In maize seed detasseling, Chen et al. [14] introduced a UAS capable of detecting and removing missed tassels, offering an intelligent, unmanned solution. Key technologies in detasseling UASs include missed tassel identification and positioning, detasseling path planning, control systems, and execution components, with the primary challenge being accurate missed tassel detection.

The YOLO (You Only Look Once) model, introduced by Redmon et al. [15], revolutionized object detection by offering a single-stage framework that combines high speed with competitive accuracy. Unlike traditional twostage detectors like Faster R-CNN, YOLO processes the entire image in a single forward pass, making it ideal for real-time applications. Over the years, YOLO has evolved through multiple iterations, with each version improving accuracy, speed, and computational efficiency. The latest versions, such as YOLOv7 [16], YOLOv8 [17], and the recently released YOLOv11 [18], have set new benchmarks in real-time object detection, particularly in complex environments. YOLO's modular design, featuring a backbone for feature extraction, a neck for feature aggregation, and a head for detection, allows for easy customization, making it a popular choice in agricultural applications. Its balance between speed and accuracy has been effectively demonstrated in tasks such as crop monitoring, pest detection, and yield estimation [19, 20]. Furthermore, the integration of attention mechanisms [21] and lightweight architectures has enhanced YOLO's capabilities, enabling it to handle challenges such as dense foliage and variable lighting conditions.

In maize tassel detection, Liu et al. [22] optimized the Faster R-CNN model for UAV imagery, while Alzadjali et al. [23] developed the TD-CNN model, which provided faster training and simpler frameworks. Onestage DL algorithms, such as YOLO and SSD, have further enhanced real-time detection capabilities. Liu et al. [24] improved YOLO v5 to achieve a mean average precision (mAP) of 44.7%, while Song et al. [25] integrated the SENet attention mechanism into YOLO_X, increasing mAP to 95%. Similarly, Liang et al. [26] proposed the SSD_mobileNet model, combining faster detection speeds and compact size with high accuracy. Pu et al. [27] introduced Tassel-YOLO, a UAV-based model for maize tassel detection and counting, demonstrating the extensive potential of YOLO models in agricultural applications.

Despite these advancements in maize tassel detection, previous studies have predominantly focused on tassels in common maize or those that have already shed pollen, which are more readily identified due to their distinct coloration and open morphology. However, detecting missed tassels during the critical detasseling stage in hybrid maize seed production remains a significant challenge. These tassels are often partially or entirely enclosed by green leaves, resulting in indistinct coloration that blends with the surrounding foliage. Moreover, their variable morphology and the dense planting configurations frequently lead to occlusion by neighboring leaves. Additionally, inconsistent lighting conditions and shadowing within the maize canopy further complicate detection by obscuring tassels or introducing visual artifacts.

Given these challenges, ensuring high accuracy in identifying missed tassels is critical for maintaining the purity of hybrid maize seeds. This study addresses these issues by integrating drones and artificial intelligence technologies to develop and optimize detection algorithms. The proposed approach aims to enhance detection precision across large-scale fields, enabling rapid identification of missed tassels to support manual or mechanical detasseling operations, ultimately improving the efficiency and quality of hybrid seed production. To achieve this, the study focuses on the following objectives:

- Develop an MT-YOLO Model: Design and optimize an MT-YOLO model using UAS and deep learning technologies to detect missed tassels during the detasseling stage of hybrid maize production.
- (2) Construct a Dataset of Missed Tassels: Develop a specialized dataset featuring missed tassels under varied lighting conditions, planting densities, and tasseling stages to ensure robust model training and evaluation.
- (3) Enhance Detection Accuracy and Speed: Improve the YOLOv5 model for higher precision and speed in detecting missed tassels in complex field environments, addressing challenges such as occlusion, morphological variability, and environmental interference.



Fig. 1 Workflow diagram

(4) Validate Model Performance: Conduct large-scale field tests to assess the accuracy, efficiency, and adaptability of the MT-YOLO model in real-world detasseling scenarios.

Materials and methods

Overview and workflow of the UAS-based missed tassel detection system

In hybrid maize seed production, detasseling is essential for maintaining seed purity. However, mechanical detasseling systems often leave behind missed tassels, requiring additional inspection and removal. Traditionally, this process has relied on manual labor, which is both time-consuming and labor-intensive. To address these limitations, this study introduces an efficient solution by integrating UASs with deep learning technology for the automated detection of missed tassels. As illustrated in the workflow diagram (Fig. 1), after mechanical detasseling, UASs equipped with advanced computer vision systems survey the fields to detect missed tassels. The MT-YOLO model processes the captured images in real-time, accurately identifying missed tassels and providing georeferenced location data to guide subsequent manual or drone-based detasseling operations. This system demonstrates significant potential to enhance operational efficiency and detection accuracy while reducing reliance on labor-intensive methods.

Image acquisition and dataset construction

The maize tassel images used in this study were captured in July 2023 at the National Precision Agriculture Research and Demonstration Base in Changping District, Beijing, China (116°26′ 18.84″ E, 40°10′ 55.17″ N) (Fig. 2). Photography sessions were conducted twice daily: in the morning (09:00-11:00) and in the afternoon (15:00-18:00) to ensure a variety of lighting conditions, such as brighter illumination in the morning and softer lighting in the afternoon. A DJI Phantom 4 RTK unmanned aerial vehicle was used, equipped with a gimbal camera positioned vertically to the ground. Flights were conducted at altitudes of 3 m, 5 m, and 7 m to capture maize tassel images at different heights above the canopy. A total of 9000 tassel images were collected, and after manual screening for quality and relevance, 886 high-resolution (5,472×3,648 pixels) visible-light images were retained for further analysis. This combination of varying lighting conditions and multiple altitudes enhanced the dataset's utility for training models designed to detect missed tassels effectively.

To enhance model training efficiency and accuracy, the distorted edges of each image were cropped, retaining a central region of 2560×2560 pixels. To further facilitate training, the cropped images were divided into smaller patches of 640×640 pixels. The cropped dataset was annotated using LabelImg software, where each maize tassel was enclosed in a bounding box defined by the minimum enclosing rectangle. XML-format label files were generated, containing image dimensions (width, height, and channels) and bounding box details, such as



Fig. 2 Location of the image acquisition area and data acquisition process

the target category and coordinates of the top-left and bottom-right vertices. To improve model generalizability [28], data augmentation techniques—brightness adjustment, vertical flipping, horizontal flipping, random noise injection, and translation transformations—expanded the annotated dataset to 7,300 images. The XML-format annotations were subsequently converted to YOLO-compatible TXT format. Finally, the dataset was randomly split into training, validation, and test sets in an 8:1:1 ratio, ensuring balanced representation for robust model evaluation.

From an agronomic perspective, maize tassels during the detasseling period can be classified into three stages: early tasseling, middle tasseling, and late tasseling [29]. Early tasseling refers to the period before tassel emergence, while middle tasseling is characterized by the male ear being drawn but not yet dispersing pollen. The late tasseling stage occurs when there is full tassel emergence and pollen shedding (Fig. 3). Detasseling, which involves removing the tassels from female maize plants, typically begins in the early tasseling stage with the use of ground-based machinery for large-scale removal. However, this method can result in missed tassels. Later, UAVs are deployed to detect and remove any missed tassels, regardless of their developmental stage. This study focuses on the early and middle tasseling stages, capturing images during these two periods for tassel detection. The constructed dataset includes images from both stages, all labeled under the class "detassel". The model's detection performance is evaluated based on its ability to accurately identify and remove tassels from both the early and middle stages, ensuring comprehensive detasseling.

During the early and middle tasseling stages, maize tassels are in critical developmental phases that present distinct challenges for detection. In the early tasseling stage, the tassels are still enclosed by green leaves and have not fully emerged from the maize canopy. In the middle tasseling stage, although the male ear begins to form, the tassels are still developing and have not yet shed pollen. These stages are crucial for effective detasseling, but missed tassels during these periods are difficult to detect due to several factors. The inconspicuous color features of these tassels, which are concealed within the green foliage, make them blend seamlessly with the surrounding plant material. Their indistinct morphology further complicates detection, as the tassels are not fully visible or defined. Finally, the position and shadowing of the tassels, often hidden beneath the maize leaves and obscured by overlapping foliage, add to the difficulty of identifying them, especially under varying light conditions during field operations.

The maize tassel dataset used in this study was analyzed to accurately detect individual tassels and understand their distribution. It consisted of 7300 images of maize seed production tassels, totaling 97,041 labeled tassels (Fig. 4). Due to varying height settings during image capture, the number of tassels per image ranged from 1 to 74, with an average of 13 tassels per image. The most common image contained four tassels, appearing in 576 images, or approximately 7.88% of the dataset.

Tassel image dataset



Fig. 3 Three Stages of tassels and dataset construction





Fig. 4 Statistical analysis of maize tassels in the dataset

Images with 10 or more tassels accounted for over 50% of the total. Statistical analysis revealed that tassels were typically densely clustered in most images, with occasional omissions or obstructions caused by overlapping leaves.

Construction of the missed tassel detection model YOLO v5 object detection algorithm

YOLOv5 is a single-stage object detection algorithm with a modular architecture comprising input, backbone, neck, and head networks. The backbone employs the CSPDarknet53 network structure enhanced with

cross-stage partial (CSP) modules and the C3 module for improved feature extraction and gradient flow. Additionally, it integrates a spatial pyramid pooling (SPP) layer to capture multi-scale features with diverse topologies, enabling robust detection across various object scales [30]. The neck utilizes a path aggregation network (PANet) to aggregate features from different hierarchical levels, facilitating improved object localization and classification. The detection head incorporates YOLO-specific layers for bounding box prediction and class label assignment, with the introduction of the SPPF (Spatial Pyramid Pooling-Fast) module in YOLOv5v7.0, enhancing multi-scale feature extraction while reducing computational cost [31]. Additionally, YOLOv5 leverages the CBS (Conv, Batch Normalization, SiLU) module, which optimizes learning efficiency through its streamlined design.

To address diverse detection requirements, YOLOv5 provides multiple model versions: YOLOv5n (nano), YOLOv5s (small), YOLOv5m (medium), YOLOv5l (large), and YOLOv5x (extra-large) [30]. Among these, YOLOv5n is the lightest and fastest, making it suitable for tasks with strict computational constraints, while YOLOv5s strikes a balance between speed, accuracy, and resource utilization. This versatility makes YOLOv5n and YOLOv5s particularly suitable for UAS-based applications, such as missed tassel detection, where real-time performance is critical. Specifically, YOLOv5s is often preferred for agricultural detection tasks due to its slightly higher detection accuracy, ensuring reliable performance even in complex field environments.

The preference for YOLOv5 over newer versions like YOLOv8 and YOLOv11 is attributed to its optimal balance of speed, accuracy, and computational efficiency, which are crucial for UAS-based agricultural applications. Although newer versions offer advanced features such as anchor-free detection and dynamic feature pyramids, their higher computational demands and increased model complexity make them less practical for resourceconstrained environments [32]. In contrast, YOLOv5 combines a lightweight architecture with a modular design and multiple model variants (e.g., YOLOv5s), enabling real-time performance without sacrificing accuracy. This capability has been validated in studies such as Pu et al. [27] and Song et al. [25], which highlight its effectiveness in agricultural tasks like crop monitoring and pest detection. Additionally, YOLOv5's proven reliability, ease of deployment, and robust community support further establish it as the ideal choice for dynamic agricultural applications, including missed tassel detection in hybrid maize seed production.

Improving tassel detection in dense plantings through enhanced feature extraction

In the main hybrid maize seed production regions of China, such as the Northwest (e.g., Xinjiang), a high planting density of 100,000 to 130,000 plants per hectare is commonly employed. This approach is designed to ensure high germination rates and maximize yield potential. However, this dense planting pattern increases leaf and tassel shading during the tasseling stage, complicating accurate tassel detection and detasseling operations. The resulting overlap between adjacent plants elevates the risk of missed tassels and false detections. To mitigate this challenge, integrating an attention mechanism that dynamically enhances the weight of relevant tassel features within complex environments can significantly improve feature extraction and detection accuracy [33].

Song et al. [25] proposed a maize tassel detection model that enhances feature extraction using SENet, a channel attention mechanism. SENet applies global average pooling (GAP) to each channel and uses two fully connected (FC) layers to generate channel weights. While this improves feature extraction, the dimensionality reduction introduced by SENet reduces the effectiveness of capturing dependencies across channels. To overcome this limitation, Wang et al. [21] introduced ECANet, which avoids dimension reduction and focuses on local cross-channel interactions to improve efficiency and model performance.

Figure 5 shows the structure of the ECANet module. The module first performs GAP on the input feature map F' (dimensions: $H \times W \times C$) to obtain the average feature value for each channel, thereby forming a feature map (Conv, dimensions: $1 \times 1 \times C$). This feature map then undergoes a 1×1 convolution operation to generate a channel attention feature map (Feature, dimensions: $1 \times 1 \times C$). Subsequently, the channel attention feature map F' (dimensions: $H \times W \times C$) to obtain a feature map F' (dimensions: $H \times W \times C$). This process aimed to preserve crucial features while reducing noise and redundancy by suppressing irrelevant features.

In this study, ECANet was integrated into the YOLO v5s network, particularly during the upsampling and downsampling stages of the backbone and neck networks. By dynamically weighing the feature maps, ECANet enhances the model's ability to focus on maize tassels and minimizes background interference. This approach significantly improves the detection of missed tassels in complex field environments.



Enhancing tassel detection with a computationally efficient neck

Standard convolution (SC) is a common image-processing operation in convolutional neural network (CNN); and its core mechanism involves sliding convolutional kernels over the input images to extract features. During this process, each convolutional kernel performs element-wise multiplication with local regions of the input image and aggregates the results to generate output feature maps. This operation captures the spatial relationships within the input data, aiding in learning the local features of the image targets. Notably, the SC employed a parameter-sharing strategy when processing channel and spatial information, which resulted in a large number of parameters. Equations (1) and (2) describe the parameter count and computational complexity, respectively, of the SC.

$$P_{conv} = C_{in} \times C_{out} \times K \times K \tag{1}$$

$$S_{conv} = C_{in} \times C_{out} \times H \times W \times K \times K$$
(2)

Depthwise separable convolution (DSC) [34] is an operation used in convolutional neural networks to optimize computational efficiency. Figure 6 illustrates the structure. The core idea is to decompose the SC into two independent steps: depth- and point-wise convolutions. Each input channel is convolved independently to capture spatial information in the depth-wise convolution stage. Subsequently, in the point-wise convolution stage, a 1×1 convolutional kernel is used to linearly combine the output channels of the depth-wise convolution, thereby facilitating inter-channel information interaction and feature compression. This decomposition reduces the number of parameters and computational load, enhancing the model's lightweight and computational efficiency; the model is particularly suitable for DL tasks on resource-constrained or mobile devices. Equations (3) and (4) give the formulas for the parameter quantity and computational load, respectively, of the DSC.

$$P_{dpc} = C_{in} \times K \times K + C_{in} \times C_{out} \tag{3}$$



Fig. 6 Structure diagram of DSC

Equations (5) and (6) give the ratios of the numbers of parameters and computational costs between the SC and DSC, respectively,

$$Q_P = \frac{P_{conv}}{P_{dpc}} = \frac{1}{C_{out}} + \frac{1}{K^2}$$
(5)

$$Q_S = \frac{S_{conv}}{S_{dpc}} = \frac{1}{C_{out}} + \frac{1}{K^2} \tag{6}$$

where C_{in} is represented the number of input channels, C_{out} is the number of output channels, H is the height of the output feature map, W is the width of the output feature map, and K the convolutional kernel size.

In the improved YOLOv5 network, the C3 module in the neck was replaced with an optimized version (DP_ CSP) utilizing DSC. This change significantly reduced computational complexity and model parameters by decoupling spatial and channel-wise operations. The streamlined feature extraction maintained the model's ability to capture fine-grained details essential for tassel detection. This improvement not only enhanced computational efficiency and inference speed but also ensured that detection performance remained robust, making the model more effective for detailed identification tasks.

Enhanced detection with improved loss function for tassel localization

In object detection tasks, the choice of loss function is critical for measuring discrepancies between predicted and actual labels. One common evaluation metric is the intersection over union (IoU), which quantifies the overlap between predicted bounding boxes and ground truth boxes. In the YOLO v5s model, the complete intersection over union (CIoU) loss function [35] is used to optimize box positioning for accurate localization of objects.

However, to improve the model's accuracy and robustness, especially in detecting maize tassels, this study replaces CIoU with SCYLLA-IoU (SIoU) [36]. SIoU optimizes spatial relationships between predicted and ground-truth bounding boxes by considering positional overlap, distance, aspect ratio, and scale similarity. Unlike traditional IoU-based functions, SIoU incorporates a penalty term that addresses centroid distance and geometric mismatches, such as aspect ratio and scale differences. This makes SIoU more effective for detecting tassels that exhibit varied scales and aspect ratios. The SIoU loss is computed as follows:

$$L_{SIoU} = 1 - IoU + P \tag{7}$$

Among them:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{8}$$

$$P = \alpha P_{distance} + \beta P_{aspect} + \gamma P_{scale} \tag{9}$$

In the above formula, L_{SloU} represents the SIoU loss value, while IoU measures the intersection over union between the predicted bounding box (A) and the ground truth bounding box (B), To address geometric mismatches, the penalty term(P) incorporates three components: $P_{distance}$, P_{aspect} , and P_{scale} , weighted by α , β , and γ , respectively. $P_{distance}$ penalizes the euclidean distance between the centroids of the predicted and ground truth boxes, P_{aspect} penalizes discrepancies in aspect ratios, and P_{scale} penalizes differences in scale (size).

MT-YOLO: an improved YOLOv5s model for detecting missed tassels

Based on the YOLOv5s model, this study proposes MT-YOLO, a customized framework designed for detecting missed tassels in hybrid maize seed production. Improvements were made using the methods mentioned in Sects. 3.2.2, 3.2.3, and 3.2.4 to optimize the model, integrating three key advancements to address challenges in accuracy and computational efficiency. First, the ECANet attention mechanism is introduced, refining the model's ability to extract and emphasize subtle, crucial features from complex images. This mechanism enables better detection of tassels, even when they are occluded or difficult to distinguish. To further reduce computational complexity, SCs in the C3 module were replaced with DSC, which separate operations on each input channel, significantly reducing the number of parameters while maintaining speed and efficiency. Finally, the use of the SIOU loss function instead of CIoU enhances the model's robustness and accuracy in localizing tassels, especially in densely planted fields. These combined enhancements result in a more efficient, accurate model optimized for precise and real-time detection of missed tassels, meeting the demands of hybrid maize seed production. Figure 7 illustrates the architecture of the proposed MT-YOLO model.

Network model evaluation index

This study evaluated object detection models using key performance metrics, including precision (P), recall (R), average precision (AP), F1-score, model size, parameters, detection speed, and FLOPs. Precision measures the likelihood of correctly detecting a target, while



Fig. 7 Structure diagram of MT-YOLO

recall indicates the model's ability to detect all targets. AP reflects the model's accuracy for a specific class, calculated as the area under the precision-recall curve, typically at an IoU of 0.5. The F1-score, as the harmonic mean of P and R, provides a balanced measure of performance. Model size denotes the storage space required, and parameters represent the number of weights and biases in the model, indicating its complexity. Detection speed, measured in frames per second (fps), evaluates the model's processing efficiency. FLOPs (floating-point operations per second) quantify computational complexity, reflecting the model's resource requirements during operation. Together, these metrics provide a comprehensive evaluation of model accuracy, efficiency, and speed.

Model training environment

In this study, all neural network models were trained and tested on a Windows 11 operating system computer, using Python 3.7 and PyTorch 1.13 for the DL framework construction. Model training used an NVIDIA GeForce RTX 3090 GPU and an Intel Core i9-10940X CPU. A series of uniform model parameter settings were adopted to mitigate overfitting in maize seed production target detection tasks. The target confidence threshold, initial learning rate, weight decay coefficient, and batch size were set as 0.5, 0.001, 0.0005, and 16, respectively. Furthermore, the number of training iterations was set to 300. Table 1 presents the relevant configurations of the training environment.

Model experiments and field test design

Comparative experiments of different attention mechanisms

This study introduced an efficient attention mechanism module, ECANet, which was integrated into the YOLO v5s network to enhance the model's performance in detecting maize tassels in complex maize seed production environments. To verify the impact of adding ECANet to the detection performance of the model, this study compared it with three mainstream attention mechanisms: the SENet [37], convolutional block attention module network (CBAMNet) [38], and coordinate attention network (CANet) [39]. The four attention

Training environment configuration	Name	Detailed description
Hardware	GPU	NVIDIA GeForce RTX 3090
	CPU	Intel Core i9-10940X
Software	Operating system	Windows 11
	Programming language version	Python 3.7
	Deep learning framework	PyTorch 1.13
	NVIDIA CUDA toolkit	CUDA 11.7
Network	Epoch	300
	Batch size	16

Table 1 The relevant configurations of the training environment

mechanisms were uniformly evaluated on the maize tassel images dataset constructed in this study.

Ablation experiment

This study validates the effectiveness of the improved strategy using ablation experiments. This strategy included the introduction of the ECANet attention mechanism, the replacement of the SC with a DSC and change of the loss function. The experiments were divided into four groups:

- (1) The original YOLO v5s model.
- (2) The model based on the original YOLOv5s model, replacing the SC with DSC
- (3) The model in Group 2 that incorporates the ECANet module.
- (4) The model in Group 3 that replaced CIoU with SIoU, i.e., the MT-YOLO model

Field test

To validate the performance of the proposed MT-YOLO model for detecting missed tassels in hybrid maize fields, field tests were conducted using a DJI Phantom 4 RTK UAV (Fig. 8). The UAV was employed to capture high-resolution aerial images of the test plots. These images were subsequently analyzed using the MT-YOLO model to identify missed male tassels and determine their precise locations.

Four test plots (S1, S2, S3, and S4) were selected for evaluation, each representing a different hybrid maize female parent variety. Each plot covered an area of 663 m^2 , with a row spacing of 45 cm and a plant spacing of 20 cm. Manual bulk detasseling was carried out in all plots prior to the field tests, leaving behind a number of missed tassels to be detected during the study.

The UAV flew at a height of 3 m above the maize canopy, with a flight speed of 0.5 m/s, to ensure the capture of detailed and clear images. These images were processed offline using the MT-YOLO model, which identified missed tassels and generated their georeferenced locations based on RTK positioning data from the UAV.

Ground inspections were conducted to validate the tassel locations detected by the MT-YOLO model. Personnel equipped with GNSS differential positioning equipment (LTK-9980) marked the actual locations of missed tassels, which were then compared with the model's detection results. To assess the model's performance, the missed rate (M) was calculated, representing the proportion of missed tassels not detected by the model. Each test plot was inspected three times to ensure comprehensive detection accuracy, and the results from the MT-YOLO model were compared with manual inspections to provide valuable insights into its effectiveness in detecting missed tassels.

$$M = \frac{N - N_R}{N} \times 100\% \tag{10}$$

where N represents the total number of missed tassels identified through human inspections, and N_R denotes the number of missed tassels correctly detected by the MT-YOLO model. This metric quantifies the proportion of missed tassels that were not detected by the model after the detection process, providing a clear assessment of its effectiveness in hybrid maize seed production fields.

Results and discussion

Comparative experimental results of different attention mechanisms

Based on the updated data in Table 2, integrating attention mechanisms into YOLO v5s consistently improved its detection performance. Among the models, YOLO v5s-ECANet delivered the best results, achieving an AP of 92.3%, Precision of 89.2%, and Recall of 91.7%. Compared to the baseline YOLO v5s (AP: 91.8%, Precision: 88.4%, Recall: 90.5%), these improvements highlight



Fig. 8 Field test

 Table 2
 Comparisons of the four attention mechanism modules

Group number	Model	AP/%	Precision/%	Recall/%
1	YOLO v5s	91.8	88.4	90.5
2	YOLO v5s-CANet	92.1	88.7	90.4
3	YOLO v5s-SENet	92.3	89.0	90.9
4	YOLO v5s-CBAMNet	92.0	88.8	90.8
5	YOLO v5s-ECANet	92.3	89.2	91.7

ECANet's effectiveness in enhancing detection accuracy and reliability.

When compared to other attention mechanisms, ECANet outperformed YOLO v5s-CANet, YOLO v5s-SENet, and YOLO v5s-CBAMNet across all metrics. Its ability to achieve a strong balance between precision and recall makes it particularly well-suited for reducing missed detections and false positives in maize tassel identification. ECANet's superior performance is attributed to its efficient channel-wise attention mechanism, which strengthens the most relevant feature channels while maintaining computational efficiency. These results establish ECANet as the most effective choice for integration into the YOLO v5s network, ensuring a robust and precise detection framework for practical applications.

Ablation experiment results

Table 3 illustrates the ablation experiment results for the YOLO v5s model. Group 2, with the addition of DSC, reduced parameters by 14.55%, improved FLOPs efficiency by 13.29%, and achieved an AP of 91.3%. Group 3 introduced ECA, further improving the AP to 92.9% while maintaining a 12.03% improvement in FLOPs efficiency compared to the baseline. Finally, Group 4

Group number	YOLO v5s	DSC	ECA	SloU	AP/%	Parameters	FLOPs/G
1		/	1	/	91.8	7,012,822	15.8
2	$\sqrt{}$	$\sqrt{}$	/	/	91.3	5,992,278	13.7
3		$\sqrt{}$		/	92.9	5,993,139	13.9
4		$\sqrt{}$	$\sqrt{}$		93.1	5,993,139	13.9

Table 3 Results of the ablation experiment

" $\sqrt{"}$ means that the improvement factor was used, and "/" means that the improvement factor was not used

replaced CIoU with SIoU, resulting in the MT-YOLO model, which achieved the highest AP of 93.1%, with a 14.54% reduction in parameters and a 12.03% improvement in FLOPs efficiency compared to the baseline.

These results demonstrate a progressive enhancement in both accuracy and computational efficiency. The integration of DSC significantly reduced parameters while maintaining reasonable performance, and the addition of ECA effectively enhanced feature representation, further improving the detection accuracy. Replacing CIoU with SIoU in the final model provided a refined loss calculation, boosting AP while preserving computational efficiency, making MT-YOLO the optimal configuration in terms of both accuracy and resource utilization.

Comparing the performance of the MT-YOLO model with other models

The performance comparison in Table 4 highlights the advantages of the MT-YOLO model across a range of metrics when compared to Faster R-CNN, SSD, YOLO v5n, YOLO v5s, YOLO v7, and YOLO v8s. MT-YOLO achieved the highest AP (93.1%), Precision (93.3%), Recall (91.6%), and F1-score (92.4%) among all evaluated models. These results demonstrate MT-YOLO's enhanced capability to accurately detect maize tassels, particularly in challenging scenarios such as detecting tassels of various shapes, tassels affected by uneven lighting or shadows under strong light, and small target tassels in complex backgrounds.

Compared to YOLO v5s, MT-YOLO showed a 1.3% improvement in AP, a 4.9% increase in Precision, a 1.1% enhancement in Recall, and a 3.0% boost in F1-score, highlighting its balanced optimization of detection accuracy. Notably, MT-YOLO's detection speed of 124 fps remains competitive, surpassing most models except for YOLO v5n, which prioritizes speed by utilizing fewer parameters at the expense of accuracy. MT-YOLO's robust detection framework leverages architectural improvements like ECANet and SIOU to ensure better feature representation and loss calculation, enhancing detection consistency across varying tassel characteristics and environmental conditions.

Impact of sunlight and shadows on tassel detection performance

Figure 9 demonstrates the models' detection performance under sunlight and shadow interference. MT-YOLO shows a marked improvement, effectively reducing missed detections (yellow boxes) in areas with strong sunlight or deep shadows. YOLO v5s, however, struggles with such conditions, frequently failing to detect tassels in regions of intense light or darkness. This limitation is likely due to its inability to extract sufficient features under extreme lighting variations. MT-YOLO, with its improved architectural components such as ECANet and SIoU, captures richer feature representations, enabling accurate detections even in challenging lighting environments.

Models	AP/%	Precision/%	Recall/%	F1-score/%	Model size/MB	Detection speed/ fps
Faster R-CNN	80.1	79.5	75.0	77.2	108	47
SSD	77.5	73.2	80.4	76.6	90.6	65
YOLO v5n	90.3	88.0	86.6	87.3	3.9	132
YOLO v5s	91.8	88.4	90.5	89.4	14.4	118
YOLO v7	92.4	90.1	87.5	88.8	74.8	56
YOLO v8s	91.5	91.4	89.8	90.6	22.5	113
MT-YOLO	93.1	93.3	91.6	92.4	12.4	124

Table 4 Performance comparison of the four detection models



YOLO v5s

MT-YOLO

Sunlight interference

Fig. 9 Detection performance of tassels affected by sunlight and shadows

Detection performance for tassels with complex shapes

Figure 10 compares the detection performance for tassels with complex characteristics. For hollows caused by detasseling, YOLO v5s often misclassifies the gaps in maize plants as tassels, leading to false detections (green boxes), due to its limited ability to differentiate structural gaps. MT-YOLO effectively minimizes these errors by leveraging enhanced feature extraction. For incomplete target shapes along edges, YOLO v5s struggles with both missed and false detections, failing to recognize partial features near boundaries. MT-YOLO addresses this issue with improved attention mechanisms, ensuring better accuracy. In cases of occluded tassels, YOLO v5s frequently misses partially visible targets or misclassifies fragments, while MT-YOLO achieves more consistent detections through superior spatial and contextual understanding.

Field evaluation of missed tassel detection accuracy

Figure 11 illustrates the process of detecting missed tassels in maize fields. Table 5 presents the field test results, showing a mean missed rate of 9.1% for the detection system. The MT-YOLO model effectively identified most missed tassels, with variations across plots. Notably, in plot S1, the system not only detected all ground-truth missed tassels but also identified an additional missed tassel overlooked by human workers, demonstrating its potential for enhanced detection capabilities.

Analysis revealed that the primary cause of missed detections was related to the height disparity among maize plants. Dwarf maize plants, measuring between 1.4 m and 1.6 m, were more likely to be missed compared to taller plants (1.8 m to 2.0 m), due to the UAS's aerial perspective. These results indicate that the combination of UAVs and deep learning detection models offers significant potential for replacing manual inspections, enabling rapid and accurate detection of missed tassels in maize fields.

Limitations of the missed tassel detection model

The proposed MT-YOLO model for maize tassel detection has several limitations that need to be addressed to enhance its performance and adaptability. First, the dataset lacks diversity in terms of planting agronomy, maize varieties, and flight altitudes, limiting the model's ability to adapt to different field conditions. For example, maize



Incomplete target shapes due to occlusion

Hollows caused by detasseling Fig. 10 Detection performance of tassels with different shapes



along edges

Fig. 11 Field-missed tassel detection

Table 5 Field test results comparison of tassel detection accurate	acy
--	-----

Test plots	Total missed tassels	Total detected by	Correct detections	False detections	Missed rate/%
	(ground truth)	UAS			
S1	26	29	24	5	7.7
S2	17	19	15	4	11.8
S3	14	15	13	2	7.1
S4	20	19	18	1	10.0
Mean missed rate	/	/		/	9.1

varieties with large open leaves and varying tassel morphologies are underrepresented, which constrains the model's generalizability. Second, the dataset only annotates tassels under a single category ("detassel"), which fails to capture distinctions such as developmental stages, thereby limiting the precision of detasseling operations. Third, data collection is currently conducted at lower altitudes, restricting the photographed area and reducing inspection efficiency. Expanding data collection to higher altitudes would cover a larger area per flight, improving operational efficiency. Fourth, the current system lacks robust mechanisms for batch processing and precise correction of exported positional data, which are crucial for improving the accuracy of tassel localization and enabling efficient large-scale operations. Lastly, the UAS's top-down view faces significant challenges in detecting tassels on dwarf maize plants, as these tassels are often occluded by leaves or fall outside the optimal detection range, leading to higher omission rates. Addressing these challenges will require expanding and refining the dataset, improving data processing pipelines, and optimizing the model for diverse field conditions and operational requirements.

Conclusions

The MT-YOLO model presents a significant advancement in the automated detection of missed tassels during hybrid maize seed production. By addressing challenges such as small target detection, varying tassel morphologies, and environmental interferences, MT-YOLO outperforms existing models in both accuracy and speed. Its integration with UASs offers a practical approach to reducing labor dependency and enhancing detasseling efficiency. Field validations confirm its applicability in real-world scenarios, achieving a mean missed rate of 9.1%. Future work will focus on further refining the model to handle more complex field conditions and exploring its integration into broader agricultural automation systems.

To enhance the adaptability and performance of the MT-YOLO model for maize tassel detection, future research should focus on expanding the dataset to include diverse maize varieties, planting agronomy, and tassel morphologies, with annotations for developmental stages to improve generalizability and precision. Optimizing data collection through higher-altitude flights and multi-angle imaging can increase operational efficiency and mitigate challenges related to occlusions, particularly for dwarf maize plants. Developing robust data processing pipelines, including batch processing and positional correction mechanisms, will enable accurate tassel localization and efficient large-scale operations. Furthermore, refining the model to handle environmental complexities

such as occlusions, variable lighting, and dense planting configurations through advanced attention mechanisms and transfer learning is essential. Addressing these areas will strengthen the model's scalability and integration into automated agricultural systems, ensuring broader applicability and improved efficiency in hybrid maize seed production.

Acknowledgements

The authors express their gratitude to the Intelligent Equipment Research Center, Beijing Academy of Agriculture and Forestry Sciences, and the Key Laboratory of Bionic Engineering, Ministry of Education, Jilin University, for their valuable comments and encouragement.

Author contributions

Jiangtao Qi: Conceptualization, Methodology, Resources, Funding acquisition, Writing – review & editing. Chenchen Ding: Writing – review & editing, Software, Methodology, Investigation, Formal analysis, Data curation. Ruirui Zhang: Methodology, Funding acquisition, Formal analysis, Conceptualization. Yuxin Xie: Resources, Investigation, Software, Data curation. Longlong Li: Investigation, Data curation. Weirong Zhang: Investigation, Supervision. Liping Chen: Methodology, Validation, Supervision, Resources, Funding acquisition.

Funding

This work was funded by the Key Research and Development Project of China (Grant No. 2023YFD2000403-2). The Science and Technology Project of Heilongjiang Province Subject (Grant No. 20212XJ05A0204). The Young and Middle-aged Technology Innovation and Entrepreneurship Outstanding Talents and Team Projects, Science and Technology Development Plan of Jilin Province (Grant No. 20230508032RC).

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable for that section.

Competing interests

The authors declare no competing interests.

Received: 24 November 2024 Accepted: 5 February 2025 Published online: 19 February 2025

References

- Food and Agriculture Organization of the United Nations. Agriculture Databases. https://www.fao.org/statistics/databases/en/. Accessed 24 Jan 2023
- 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. https://doi.org/10.1007/s12571-022-012887.
- Li H, Yue H, Li L, Su C. A comparative analysis of the hybrid maize (Zea mays L) seed quality in China from 2013 to 2018. Agron J. 2019. https:// doi.org/10.3390/agronomy9100625.
- Wright H. Commercial hybrid seed production. Hybridization of Crop Plants. 1980;165:161–76. https://doi.org/10.2135/1980.hybridizationof crops.c8.
- Jiang H, Jiang L, Wang Y, Zhan L. Directions and policy suggestions for China's seed industry development from the perspective of national food security. Xinjiang Norm Univ. 2022. https://doi.org/10.14100/j.cnki.65-1039/g4.20210811.001.
- Costa L, Carvalho I, Ferreira L, Szareski V. The effects of different mechanical detasseling methods on hybrid maize seed production. Genet Mol Res. 2019;18:18207. https://doi.org/10.4238/gmr18207.

- Zou Z, Wang J, Zhao Q, Li J, Cui L, Hong M, et al. Research status of seed corn emasculation technology and equipment. Agric Eng. 2020;10(7):19– 23. https://doi.org/10.3969/j.issn.2095-1795.2020.07.005.
- Chen M, Zhang R, Chen L, Tang Q, Xia L. Investigation on advances of unmanned aerial vehicle application research in agriculture and forestry. Smart Agric. 2021;3(3):22. https://doi.org/10.12133/j.smartag.2021.3.3. 202107-SA006.
- Han L, Yang G, Dai H, et al. Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. Plant Methods. 2019;15:1–19. https://doi.org/10.1186/s13007-019-0394-z.
- Hu P, Zhang R, Yang J, Chen L. Development status and key technologies of plant protection UAVs in China: a review. Drones. 2022;6(11):354. https://doi.org/10.3390/drones6110354.
- Zaman-Allah M, Vergara O, Araus JL, et al. Unmanned aerial platformbased multi-spectral imaging for field phenotyping of maize. Plant Methods. 2015;11:1–10. https://doi.org/10.1186/s13007-015-0078-2.
- Zhang R, Hewitt AJ, Chen L, Li L, Tang Q. Challenges and opportunities of unmanned aerial vehicles as a new tool for crop pest control. Pest Manage Sci. 2023;79(11):4123–31. https://doi.org/10.1002/ps.7683.
- Zhang R, Lian S, Li L, Zhang L, Zhang C, Chen L. Design and experiment of a binocular vision-based canopy volume extraction system for precision pesticide application by UAVs. Comput Electron Agric. 2023;213: 108197. https://doi.org/10.1016/j.compag.2023.108197.
- 14. Chen L, Zhang R, Ding C, et al. Method, system, and aerial unmanned detasseling machine for the synergistic management of maize seed production. Beijing: CN114402995B; 2023 Mar 24.
- Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. Proc IEEE Conf Comput Vision Pattern Recognit. 2016. https://doi.org/10.1109/CVPR.2016.91.
- Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. p. 7464–75. https://doi.org/10.48550/arXiv.2207.02696.
- Varghese R, Sambath M. YOLOV8: A novel object detection algorithm with enhanced performance and robustness. In: 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). IEEE; 2024. p. 1–6.https://doi.org/10.1109/ADICS58448. 2024.10533619.
- Khanam R, Hussain M. YOLOV11: an overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725. 2024. https://doi.org/10. 48550/arXiv.2410.17725.
- Mirhaji H, Soleymani M, Asakereh A, Mehdizadeh SA. Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions. Comput Electron Agric. 2021;191: 106533. https://doi.org/10.1016/j. compag.2021.106533.
- Zhang W, Huang H, Sun Y, Wu X. AgriPest-YOLO: A rapid light-trap agricultural pest detection method based on deep learning. Front Plant Sci. 2022;13:1079384. https://doi.org/10.3389/fpls.2022.1079384.
- Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. ECA-Net: efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition. 2020. p. 11534–42. https://doi.org/10.1109/CVPR42600.2020.01155.
- 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(2):338. https://doi. org/10.3390/rs12020338.
- Alzadjali A, Alali MH, Veeranampalayam Sivakumar AN, Deogun JS. Maize tassel detection from UAV imagery using deep learning. Front Rob Al. 2021;8: 600410. https://doi.org/10.3389/frobt.2021.600410.
- Liu W, Quijano K, Crawford MM. YOLOv5-Tassel: detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning. IEEE J Sel Top Appl Earth Obs Remote Sens. 2022;15:8085–94. https://doi.org/10. 1109/JTARS.2022.3206399.
- Song C, Zhang F, Li JS, Xie JY. Detection of maize tassels for UAV remote sensing image with an improved YOLOX model. J Integr Agric. 2023;22(6):1671–83. https://doi.org/10.1016/j.jia.2022.09.021.
- Liang Y, Chen Q, Dong C, Yang C. Application of deep-learning and UAV for field surveying corn tassel. Fujian J Agric Sci. 2022. https://doi.org/10. 19303/j.issn.1008-0384.2020.04.014.
- 27. 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(8):492. https://doi.org/10.3390/drones7080492.

- Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data. 2019;6(1):1–48. https://doi.org/10.1186/ s40537-019-0197-0.
- Zan X, Zhang X, Xing Z, Liu W, Zhang X, Su W, et al. Automatic detection of maize tassels from UAV images by combining random forest classifier and VGG16. Remote Sens. 2020;12(18):3049. https://doi.org/10.3390/ rs12183049.
- Glenn J. YOLOv5. Git code. 2020 https://github.com/ultralytics/yolov5. Accessed 24 Jan 2023
- Zhu X, Lyu S, Wang X, Zhao Q. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. p. 2778–88. https://doi.org/10.1109/ICCVW54120. 2021.00312.
- 32. Sapkota R, Karkee M. Comparing YOLO11 and YOLOv8 for instance segmentation of occluded and non-occluded immature green fruits in complex orchard environments. arXiv preprint arXiv:2410.19869. 2024. https://doi.org/10.48550/arXiv.2410.19869.
- Guo MH, Xu TX, Liu JJ, Lui ZN, Jiang PT, Mu TJ, et al. Attention mechanisms in computer vision: a survey. Comput Vis Media. 2022;8:331–68. https:// doi.org/10.1007/s41095-022-0271-y.
- Chollet F. deep learning with depth-wise separable convolutions. Proc IEEE Conf Comput Vision Pattern Recognit. 2017. https://doi.org/10.1109/ CVPR.2017.195.
- 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(8):8574–86. https://doi.org/10. 48550/arXiv.2005.03572.
- Gevorgyan Z. SloU loss: more powerful learning for bounding box regression. arXiv preprint arXiv:2205.12740. 2022. https://doi.org/10.48550/ arXiv.2205.12740.
- Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. p. 7132–41. https://doi.org/10.1109/CVPR.2018.00745
- Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV). 2018. p. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
- Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. p. 13713–22. https://doi.org/10.1109/ CVPR46437.2021.01350

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.