Computers and Electronics in Agriculture 239 (2025) 111041

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Tassel counting of individual ridge from UAV RGB imagery based on YOLOv8m with deep SORT and double-step Otsu thresholding algorithm by filtering abnormal IDs for maize breeding

Man Xia [a], Leilei He [a], Yu Gu [a], Xiaojuan Liu [a], Bryan Gilbert Murengami [a], Lamin L. Janneh [e], Rui Li [a], Yan Long [a,*], Anastasia Grecheneva [f], Longsheng Fu [a,b,c,d,**]

[a] College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
[b] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China
[c] Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling, Shaanxi 712100, China
[d] Shaanxi Agricultural Equipment Engineering Technology Research Center, Yangling, Shaanxi 712100, China
[e] School of Information Communication and Technology, University of The Gambia, Banjul, Kanifing 3530, Gambia
[f] Russian State Agrarian University - Moscow Timiryazev Agricultural Academy, Moscow 127434, Russia

ARTICLE INFO

Keywords:
YOLOv8m
Maize tassels counting
Individual ridge
Double-step Otsu thresholding
Filter abnormal IDs

ABSTRACT

Accurate maize tassel counting on individual ridge plays a critical role in advancing maize breeding programs by providing insights into crop adaptability and agronomic performance. The current manual method of processing male inflorescences has low accuracy and high labor intensity. Moreover, the existing algorithms cannot be directly applied to the counting of an individual ridge maize tassels. An automated system with UAV-based RGB imagery is presented for maize tassel counting. The object detection model was developed based on You Only Look Once version 8-medium (YOLOv8m), which detects tassels. A double-step Otsu threshold algorithm (DSOTSUTA) was designed to extract individual maize tassel ridge, which eliminated the interference of two adjacent ridges on both sides. Individual ridge tassel counting was implemented by Deep learning based Simple Online and Realtime Tracking (Deep SORT). It assigned identifiers (IDs) to each tassel and filtered abnormal IDs by analyzing the displacement increments of IDs in consecutive frames eliminating errors caused by ID switching. The object detection model achieved a mean Average Precision (*mAP*) of 91.6 %. The DSOTSUTA was tested on 5340 images and effectively extracted individual maize tassel ridges. The system achieved a root mean square error (*RMSE*) of 22.14 tassels per video, a mean absolute percentage error (*MAPE*) of 8.43 %, and an accuracy of 92.23 %, signifying a mean absolute percentage error (*MAPE*) of 8.43 % between the predicted tassel counts and ground truth observations. These results indicate that this automated system has the ability to enhance the accuracy of individual ridge maize tassel counting in breeding programs.

1. Introduction

Counting of maize tassels for each individual ridge aids breeders in developing superior varieties. As a kind of monoecious crop, the tassels of maize grow on the top of plant, whose traits are important indicators for breeding evaluation, especially the number of tassels in a ridge-based planting unit (He et al., 2023; Li et al., 2022). Additionally, the number of tassels in different regions, varieties, and cultivation conditions could reveal the adaptability and performance of maize in different environments, providing more evidence for the selection of superior varieties (Cooper and Messina, 2023; Hoopes et al., 2019; Rizzo et al., 2022). However, breeders usually rely on manual visual counting of tassels from sampling areas to evaluate the overall quantity in the field, which is laborious, error-prone, costly, and inefficient. Furthermore, manual tassel counting usually requires crossing fields, which may cause damage to maize plants and disrupt breeding plots (Karami et al., 2021; Liu et al., 2020; Zou et al., 2020). Thus, there is an urgent need for automated, precise counting of maize tassels per ridge to overcome these practical limitations.

Efficient information gathering on maize development is preferable

* Corresponding author.
** Corresponding author.
E-mail addresses: longyan@nwafu.edu.cn (Y. Long), fulsh@nwafu.edu.cn (L. Fu).

https://doi.org/10.1016/j.compag.2025.111041
Received 12 February 2025; Received in revised form 12 September 2025; Accepted 27 September 2025
Available online 15 October 2025
0168-1699/© 2025 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

to manual tallying of maize tassels in the field, through the utilization of advanced data collection methods. Lu et al. (2017) proposed a visual-based online monitoring system for maize image acquisition and growth analysis, which is restricted to fixed-field installation and only achieved tassel counting in specific imaging areas. Hasan et al. (2018) developed a vehicle-mounted mobile imaging system for field wheat breeding analysis, which realized spikes counting in continuous images of the breeding block, but there was a risk to damage crops during vehicle movement. A non-invasive, ridge-specific data collection approach is thus critical for large-area maize breeding management, where precise separation of individual ridges remains a key challenge.

As a flexible data acquisition platform, unmanned aerial vehicles (UAVs) are not restricted by terrain and can quickly cover large areas. UAVs have been used for rapid nondestructive growth assessment of maize combined with computer vision technology. Liu et al. (2022) collected images of maize at different stages from a UAV with a height of 20 m, which achieved a *mAP* of 44.7 % in small-size tassel detection by improving YOLOv5. Zhang et al. (2023) improved YOLOv4 to detect densely distributed maize tassels, which reached an accuracy of 95.11 % on UAV-collected images with a height of 7 m. Jia et al., (2024b) utilized a UAV for maize image collection at two flight heights (5 m and 10 m) covering the entire tasseling period, which obtained an accuracy of 96 % in tassel detection by improving YOLOv5 based on coordinate-attention mechanism. However, these studies primarily focused on tassel detection in sampled image regions rather than continuous counting across an entire planting ridge, leaving unresolved the challenge of double-counting due to overlapping UAV imagery and the need to distinguish tassels belonging to the target ridge from adjacent ones. Moreover, practical application of UAV-based tassel counting systems is hindered by several technical challenges, including occlusion caused by dense plant canopies, adjacent-ridge interference due to close planting patterns, ID-switching in object tracking when tassels move or overlap, wind-induced motion blur of tassels, and scale variation resulting from inconsistent flight altitudes.

Compared to counting objects in images, tracking based on position correlation of objects in continuous video frames has been proven to reduce double counting effectively. As a robust and high-performance multi-object tracking algorithm, Deep learning based Simple Online and Realtime Tracking (Deep SORT) has been widely applied in field crops applications (Bewley et al., 2016; Liang et al., 2022; Lin et al., 2022; Wojke et al., 2017). Yang et al. (2022) developed a tracking-by-detection strategy based on modified CenterNet and Deep SORT for cotton seedling counting, which achieved a total bias of 8 in 20 video clips with a ground truth of 166. Li et al. (2024) proposed a cross-line partition wheat ear counting algorithm based on Deep SORT and improved YOLOv7, which achieved a counting accuracy of over 94 % in videos of different speeds. While object tracking has demonstrated superior counting accuracy on a variety of field crops (Fernandez-Gallego et al., 2018; Madec et al., 2019), further processing is required to distinguish which of the tassels appearing in collected video frames belong to the target ridge. Jia et al., (2024b) utilized an unmanned aerial vehicle (UAV) to collect maize images at two flight altitudes (5 meters and 10 meters) throughout the entire tasseling period. This approach achieved a detection accuracy of 96% in tassels by improving YOLOv5 based on the coordinate-attention mechanism. Influenced by flight altitude and different breeding patterns, UAV-captured maize videos often show multiple ridges rather than a single ridge (Jia et al., 2024a; Li et al., 2023; Pauw, 2013).

As cross-pollination needs to be controlled during maize breeding, different varieties of maize are usually planted on separate ridges. Influenced by flight altitude and different breeding patterns, UAV-captured maize videos often show multiple ridges rather than a single ridge (Jia et al., 2024a; Li et al., 2023; Pauw, 2013). It is necessary to determine the bilateral boundaries of the target ridge and reject tassels from others during counting of maize tassels by object-tracking (Guo et al., 2021). In maize breeding fields, there is usually a ditch with a fixed length (0.4 m ~ 0.6 m) between neighboring ridges for separating planting ridges (Li et al., 2020), as shown in Fig. 1. Therefore, the position of ditches from video frames could be applied for target ridge extraction, excluding the outside tassels.

An automatic system for counting tassels in individual ridge maize was developed for the breeding plots. The specific research objectives include: (1) constructing an image dataset of maize tassels and employing state-of-the-art deep learning models to detect maize tassels in RGB images; (2) developing a new algorithm to extract individual ridge tassels; (3) utilizing multi-object tracking to track tassels in video frames; (3) filtering abnormal IDs due to ID switching in consecutive frames; and (5) evaluating the performance of the proposed system to count maize tassels in individual ridge.

2. Materials and methods

2.1. Dataset preparation

2.1.1. Data collection

Original video data of maize tassels was collected from a breeding plot in Yangling, Shaanxi Province, China (34°17′ N, 108°0′ E). The breeding plot is configured with a plant spacing of 10 cm, a ditch width of 90 cm, and a ridge width of 100 cm. UAV video data was collected in August 2023 using a Zenmuse P1 camera mounted on an industrial-grade UAV (DJI M300 RTK). The camera was equipped with a full-frame sensor (35.9 mm × 24 mm), a 35 mm lens, an aperture range of f/2.8 to f/16, a field of view (FoV) of 63.4°, and a shutter speed range of 1/8000 to 1 s. Video data was collected under clear and windless conditions (wind speed < 0.5 m/s) from 14:00 to 15:00, flying at an altitude of 16 m. The shutter speed was set at 1/1000 s to minimize any potential tassel movement between frames. The camera was set to an aperture of f/5, an ISO of 800, and a shutter speed of 1/1000 s. A total of 14 videos containing approximately 7,000 maize tassels were saved in MOV format, with a resolution of 3840 × 2160 pixels and a frame rate of 30 frames per second. A sample video frame captured from the maize breeding plot is presented in Fig. 2. The video frame contains individual ridge that needs to be counted in this study, the adjacent ridges on both sides that interfere with the individual ridge count, and the ditch between the ridges. Currently, the UAV flight direction needs to be parallel to the ridges to ensure optimal image capture.

2.1.2. Dataset building

The dataset comprises both video data and image data. The image data was utilized for training and testing detection model, while the video data is employed to test the performance of the counting system. The original videos were evenly divided into experimental videos (serial numbers are test videos 1 to 7) and detecting videos, with each part containing 7 videos. The process of data preparation is shown in Fig. 3, a total of 600 images were extracted from the detecting videos. These images were then randomly divided into training and testing images, constituting 80 % (480 images) and 20 % (120 images), which were used to train and test detection model, respectively.

To improve the generalization ability of the deep learning model, after image annotation, data augmentation was carried out on the training dataset. The images were annotated using LabelImg (https://github.com/tzutalin/labelImg), with tassels being annotated with rectangular boxes and categorized as 'tassel' (as shown in Fig. 4). Data augmentation encompassed brightness adjustments, noise blurring, and image mirroring in both horizontal and vertical directions. As a result, the dataset was expanded to a total of 4,800 images. To facilitate complete replication and independent verification, experimental image sequences and codes have been stored in a public repository(https://github.com/fu3lab/tassel_individual_ridge_counting).

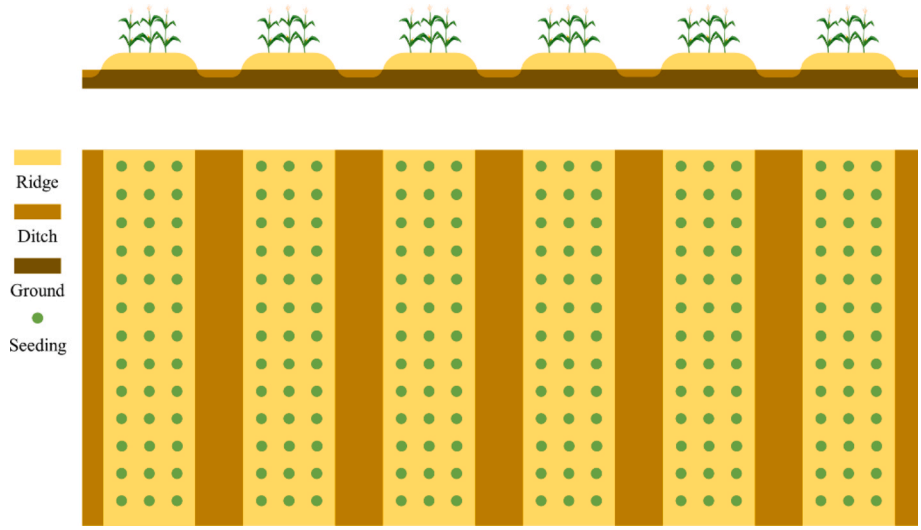


Fig. 1. Example of a maize breeding plot composition. Ridges, ditches, ground, and seeding are in light brown, brown, dark brown, and green circles, respectively.



Fig. 2. An example of maize tassels video frame obtained from breeding plot.

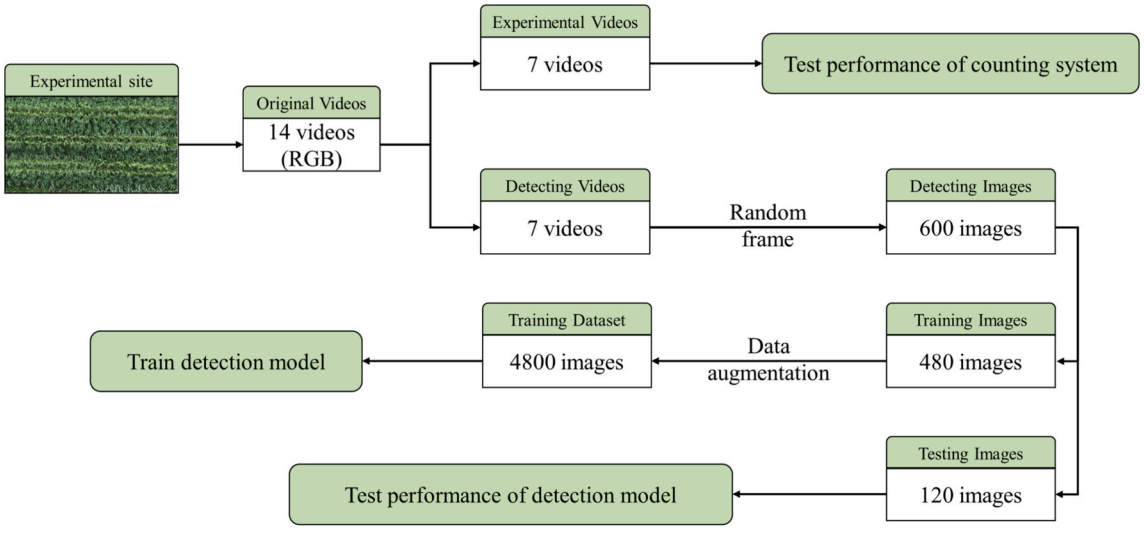


Fig. 3. Process of data preparation.

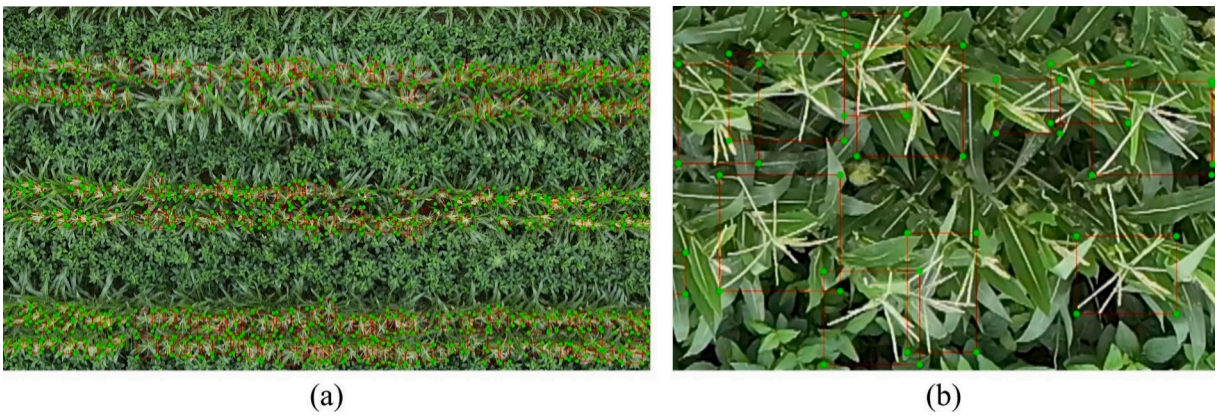


Fig. 4. Annotated example of maize tassels. (a) Example of maize tassels annotation in the entire image; (b) Example of detailed maize tassel annotation.

2.2. Maize tassels detection

2.2.1. YOLOv8m model

YOLOv8m has demonstrated strong potential in the field of object detection in machine vision (Varghese and Sambath, 2024). It features a CSPDarknet53 backbone with Cross-Stage Partial connections for better gradient flow and reduced computation. The dense prediction head directly estimates bounding boxes and class probabilities, leveraging both high-level and low-level features for precise detection. Given its outstanding performance metrics, this research specifically employed YOLOv8m to detect maize tassels. The Network structure of YOLOv8m is shown in Fig. 5.

2.2.2. Model training

Experiments were performed using the PyTorch framework, version 2.0.1, conducted on a desktop computer equipped with an Intel Xeon Gold 5128 32-core processor (2.3 GHz), dual Nvidia GeForce GTX 4090 24G GPUs, and 128 GB of memory, running a 64-bit Windows 10 system. The utilized software toolkit comprised CUDA 12.1, cuDNN 8.8.1, Python 3.9, and OpenCV 4.8. YOLOv8m was applied for training tassel detection on the PyTorch framework with a network input size of 1280 $\times$ 1280 pixels and a batch size of 4. Stochastic gradient descent was used for training and included a weight decay of 0.005. The initial learning rate was set to 0.01, and the training was conducted for 400 epochs.

2.3. Individual ridge extraction

Accurately extracting individual ridge is essential for the individual maize tassel ridge counting system. As shown in Fig. 2, maize plants were regularly arranged with a certain spacing within the same ridge and a ditch spacing between different ridges. This ditch spacing serves as a natural delimiter that can be used to isolate individual ridge tassels. In this work the separation of individual tassels was achieved by leveraging the peak-valley feature on a histogram. In this study, a *y*-coordinate system was established for extracting individual ridge. Each tassel was assigned a *y*-coordinate according to its vertical position within video frame, facilitating the creation of a histogram to show the tassel distribution along the *y*-axis.

The established *y*-coordinate was a crucial parameter for differentiating tassels on different ridges, given the regular intra-ridge spacing and inter-ridge ditch spacing. Specific steps included determining the *y*-coordinate range for individual ridges and analyzing the *y*-coordinate histogram of all detected maize tassels within video frame, as shown in Fig. 6b. When the video frame contained only one ridge, the tassels could be directly counted without algorithmic extraction. For frames containing two ridges, an Otsu thresholding algorithm was applied to automatically determine the optimal separation threshold between ridges, effectively isolating individual ridge for counting. For frames containing three ridges, an adaptive three-peak threshold segmentation algorithm would segment the position of individual ridges. However, existing algorithms were unreliable for this task because of the uneven peak distribution in the histogram, complicating the development of an effective segmentation algorithm.

Reliable segmentation of three-peak histogram is an important task in extracting maize tassels in individual ridge. To achieve this, a double-step Otsu threshold algorithm (DSOTSUTA) was developed based on the distinct vertical distribution characteristics of maize tassels in ridges, which is influenced by the regular intra-ridge spacing and inter-ridge ditch spacing. First, the median of the *y*-coordinates for all detected tassels in the image were calculated and used to divide the three-peak histogram into two double-peak histograms. Each of these histograms was then processed by the Otsu threshold algorithm to determine the optimal threshold for segmentation. This algorithm enabled the accurate extraction of individual ridge tassels, as shown in Fig. 6d, where the area of individual ridge was highlighted by a specific ridge height R_i. Before the tracker processes each frame, the DSOTSUTA was used to extract the tassel regions of individual maize ridge, enabling the extraction of such regions in the video.

The steps of the DSOTSUTA are detailed as follows:

Step 1 (S1): Calculate the median N_M

First, calculate the median N_M of the *y*-coordinates from the set N_i of all detection boxes within video frame. This median will serve as a dividing line to split the data into two subsets.

$$N_M = median(N_i) \tag{1}$$

Step 2 (S2): Divide the set N_i into two subsets

Divide the set N_i into two subsets, N' and N'', based on the median N_M.

$$N' = \{N_i | N_i \leqslant N_M\} \tag{2}$$

$$N'' = \{N_i | N_i \geqslant N_M\} \tag{3}$$

Step 3 (S3): Apply the Otsu threshold algorithm

Apply the Otsu threshold algorithm to each subset to obtain two thresholds $k_1$and k_2. The Otsu threshold algorithm works by minimizing the intra-class variance while maximizing the inter-class variance.

For each subset N' and N'', the algorithm finds the optimal threshold k that maximizes the inter-class variance, which defined as:

$$\sigma_B^2(k_1) = \sum_{i=0}^{k_1} \frac{n_i}{n}[m_1(k_1) - \sum_{i=0}^{L-1} i\frac{n_i}{n}]^2 + (1 - \sum_{i=0}^{k_1} \frac{n_i}{n})[m_2(k_1) - \sum_{i=0}^{L-1} i\frac{n_i}{n}]^2 \tag{4}$$

where: n_i refers to the number of N' in the i-th class, n refers to the total number of N', $m(k_i)$ refers to the mean of the i-th class, μ refers to the global mean.

Step 4 (S4): Use thresholds to segment and extract individual ridge tassels

Use the thresholds k_1 and k_2 to segment the *y*-coordinates and extract

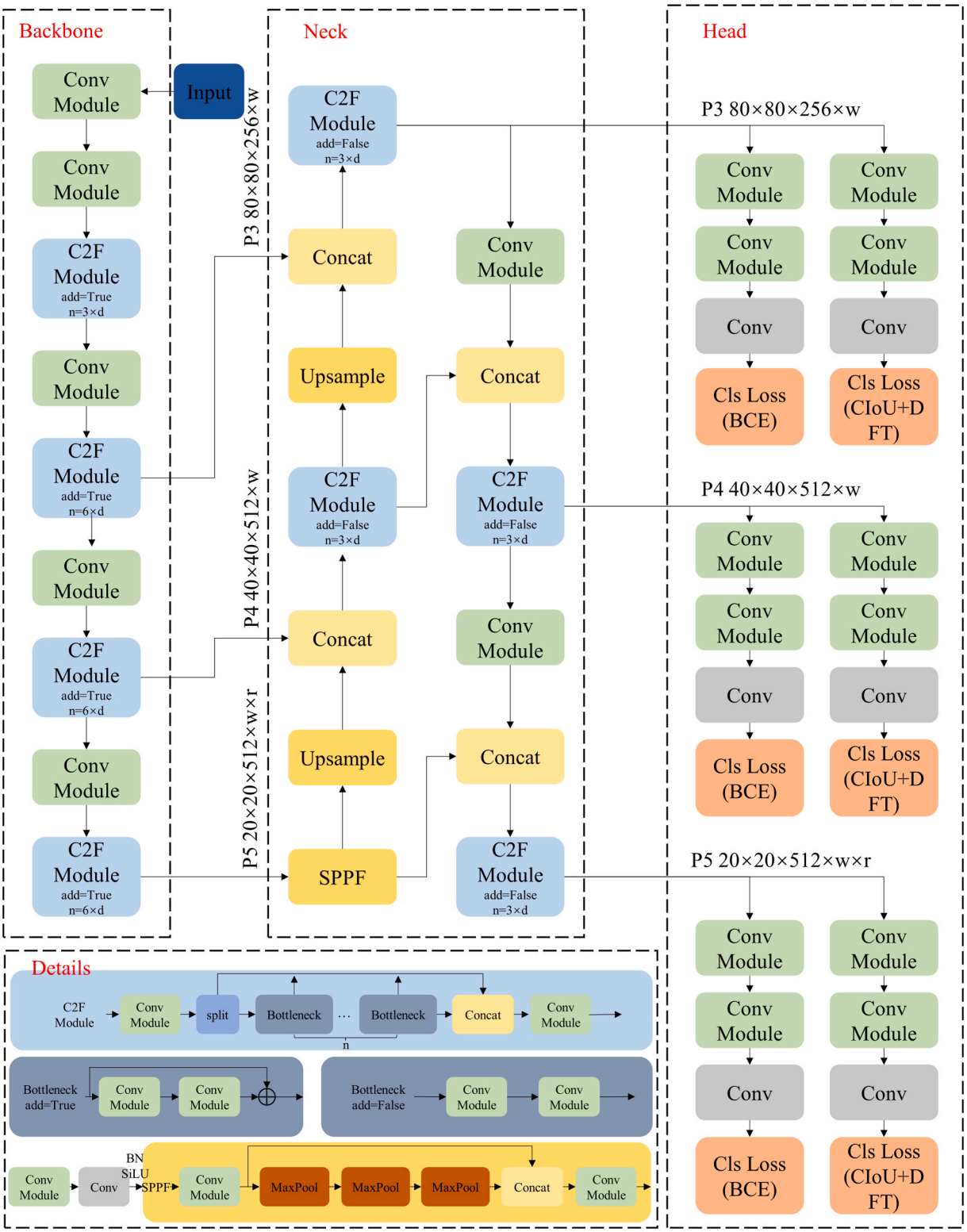


Fig. 5. The model architecture of YOLOv8m.

individual ridge tassels. The final y-coordinate set R_i for center points of the middle ridge can be obtained by intersecting subsets:

$$R_i = N_{\leqslant k_1} \cap N_{\leqslant k_2} \tag{5}$$

2.4. Tassel tracking and counting

2.4.1. Tassel tracking by Deep SORT

Detection alone was insufficient for accurate counting in the videos, so employing Deep SORT for tracking could mitigate the issue of repeated counting. Since our system is built based on Deep SORT, its tracking accuracy is inevitably limited by the performance of the

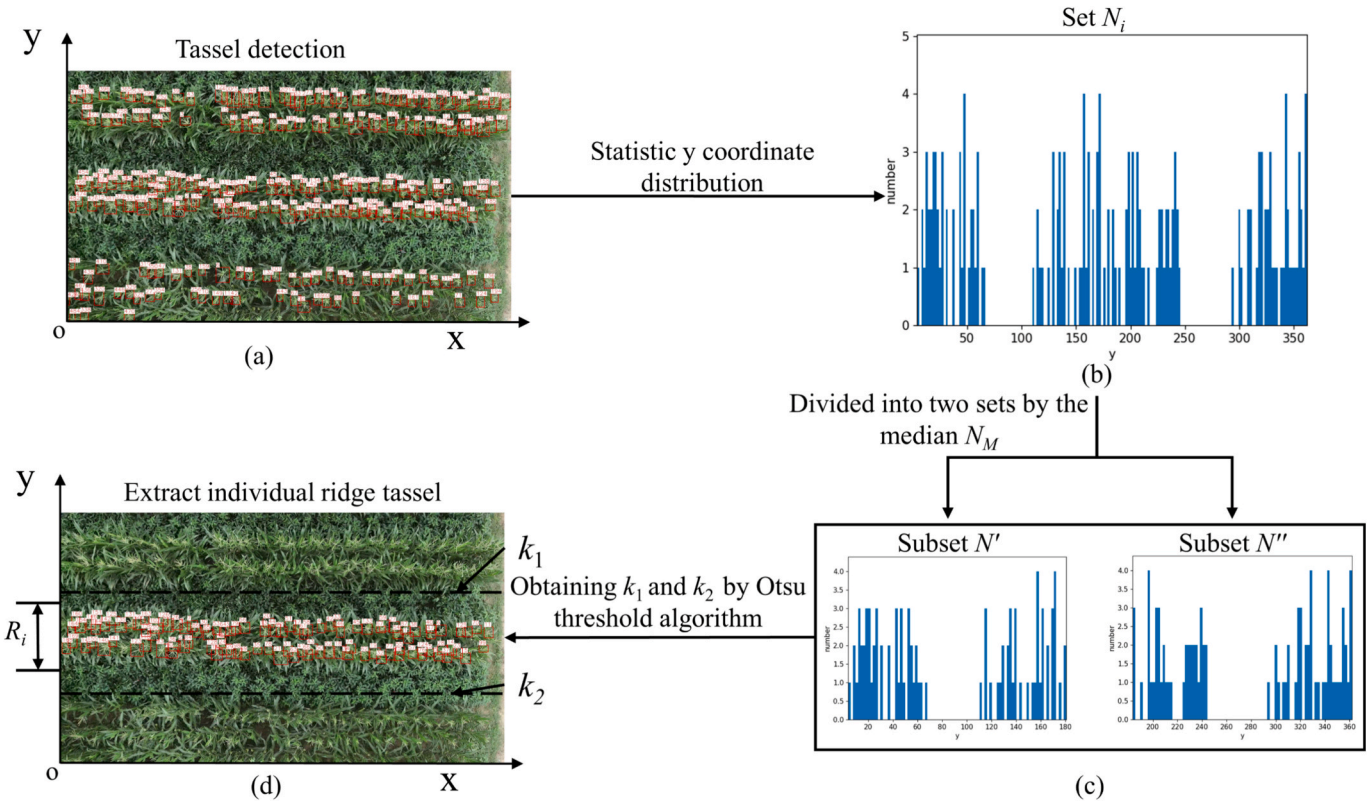


Fig. 6. Processing flow of DSOTSUTA for extraction of maize tassels on individual ridge. (a) Tassel detection result in video frame; (b) Statistical distribution of each tassel *y*-coordinate, N_i is a set of *y*-coordinates of all maize tassels; (c) Division into N' and N'' subsets by N_M; N_M is a median of N_i; (d) Extraction of individual ridge tassels area R_i.

detectors. Moreover, issues like occlusion, close spacing of tassels, or a large number of tassels within a frame can all lead to ID switches. During these switches, the algorithm assigns a new ID to a tassel that was previously being tracked, potentially causing inaccurate counting (Santos et al., 2024). To effectively overcome the problem of ID switches, it is necessary to incorporate new steps between frames to specifically filter such occurrences.

2.4.2. *Abnormal IDs judgment and tassel counting*

Filtering abnormal IDs caused by ID switching in Deep SORT is crucial for ensuring accurate tassel counting. ID switching is an unavoidable phenomenon in video tracking, but it could be judged whether ID switching occurs according to the displacement of maize tassel ID between consecutive frames. As shown in Fig. 7, the displacement x_1, x_2, x_3 and x_4 of normally tracked maize tassels IDs of 1, 2, 3 and 4 between the Frame 1 and Frame 2 were approximately equal in size, but the displacement x_5 and x_6 of IDs of 5 and 6 with ID switching were different from that of x_1, x_2, x_3 and x_4. Consequently, abnormal ID switching could be filtered by comparing the *x*-axis displacement increment of ID between consecutive frames.

Accurate calculation of tassel displacement increments between consecutive frames was a critical step in filtering abnormal tassel IDs. First, the YOLOv8 detection model was employed to identify tassels and their coordinates in each frame. The center coordinate of each tassel was calculated by averaging the corner coordinates of the detected tassel bounding box. Subsequently, the Deep SORT tracking model predicted tassel positions across frames and assigned unique IDs to each tassel. For each tracked tassel ID, the *x*-axis displacement increment (Δx_j) was computed as the difference between the *x*-coordinates of the center of the tassel in consecutive frames, as described in Eq. (6).

$$\Delta x_j = x_{t+1} - x_t \tag{6}$$

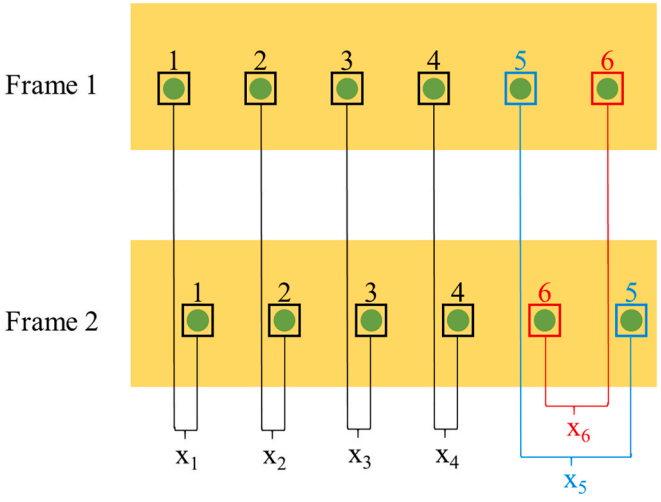


Fig. 7. Example of tassel ID switching and displacement in Deep SORT. Black indicates normal tracking, blue and red indicate ID switching. Green circles indicate maize tassels, and x_1-x_6 represent the consecutive displacements matched by Deep SORT.

where x_t and x_{t+1} refer to the center *x*-coordinates of the tassel in frame *t* and frame $t + 1$, respectively.

Finding abnormal increments from displacement increments was the core step of ID filtering. By recording the Δx_j of each tassel ID across consecutive frames and comparing them with statistical metrics such as mean and variance, displacement increments that deviate from the normal range can be identified. Such deviations typically indicate ID switches or other tracking anomalies. To further refine the abnormal

tassel IDs, a threshold-based criterion derived from statistical analysis was applied. Specifically, if Δx_j of a tassel ID deviate from the mean displacement increment (μ) by more than the standard deviation (σ), the ID was flagged as abnormal and excluded. The specific calculation formula was shown in Eq. (7).

$$\left|\Delta x_j - \mu\right| \geqslant \sigma \tag{7}$$

where μ refer to the mean Δx_j of all IDs, and σ refer to the standard deviation of these displacement increments. The count S, which represents the result of the proposed algorithm, was obtained by subtracting the abnormal IDs from the total number of IDs tracked by Deep SORT, as shown in Eq. (8).

$$S = Deep\,SORT_{IDs} - Abnormal_{IDs} \tag{8}$$

2.5. Performance evaluation

The performance of the maize tassel detection model was evaluated using precision (P), recall (R), and the Average Precision (AP), with their definitions provided in Eq. (9) to (11). Samples were classified into four types based on the combinations of true and predicted annotations: true positive (TP), false negative (FN), false positive (FP), and true negative (TN). The precision (P) and recall (R) were calculated as follows:

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$AP = \int_0^1 P(R)dR \tag{11}$$

In assessing the effectiveness of maize tassel detection and tracking, the following indicators were employed: mean absolute percentage error ($MAPE$), root mean square error ($RMSE$), and counting accuracy (P_c), as defined by Eq. (12)-(14).

$$RMSE = \sqrt{\frac{1}{n}\sum_{m=1}^{n}(S_m - G_m)} \tag{12}$$

$$MAPE = \frac{1}{n}\sum\nolimits_{m=1}^{n}\left|\frac{S_m - G_m}{G_m}\right| \times 100\% \tag{13}$$

$$P_c = \left(1 - \frac{|S - G|}{G}\right) \times 100\% \tag{14}$$

Where G_m refers to ground truth number of tassels in video m, S_m refers to the counted number of tassels by the proposed algorithm in video m, S and G refer to the number of the proposed algorithm count maize tassels and the ground truth number of tassels in video, respectively.

3. Results and discussion

3.1. Performance of object detection model

The object detection model achieved satisfactory results in detecting maize tassels. The performance metrics displayed in Table 1 demonstrate the effectiveness of the detection system on the maize tassel dataset. The test dataset, consisting of 60 images and totaling 18,067 ground truth annotated objects, yielded 15,694 TP, 1,519 FP, and 854 FN. These results correspond to a P of 91.6 % and a R of 93.2 %, with the AP closely aligning with the P. These results highlight the high accuracy of the system in detecting tassels.

Table 1
The detection results in the testing dataset.

Object	TP	FP	FN	TN	P/%	R/ %	AP/ %	Speed/(ms/ image)
Tassel	15,694	1519	854	0	91.6	93.2	91.6	48

In maize tassel detection, the high FP incidence is primarily caused by severe occlusion among densely packed tassels. Specifically, when overlapping tassels are mis-detected as additional instances, the bounding boxes become duplicated or displaced, leading to FP errors. When maize was densely planted, tassels often overlap, making it hard for detection algorithms to detect tassels accurately. This overlap was a key cause of FP, as the detection system cannot differentiate closely-positioned tassel overlaps (Shi et al., 2024). Fig. 8 shown FP examples in the test dataset. Red bounding boxes shows the detected tassels. In Fig. 8a, blue-circled overlapping tassels were successfully detected by the model under mild occlusion. In Fig. 8b and Fig. 8c, yellow-circled overlapping inflorescences highlight scenarios where the model faced difficulty in detecting tassels with severe overlap. Although the model has an AP of 91.6 %, it struggles with heavily overlapping tassels. The dense arrangement of tassels in the field worsens the occlusion problem, increasing FP likelihood. High-density planting, along with tassel size variability and irregular growth, adds to the detection complexity (W. Jia et al., 2024), making accurate and efficient tassel detection challenging.

3.2. Performance of DSOTSUTA

The DSOTSUTA exhibited high effectiveness in accurately extracting individual ridges. In a series of seven test videos comprising a total of 5340 frames, the algorithm successfully extracted individual maize ridges. Crucially, all segmentation thresholds were automatically derived from the frame-level distributions of tassel y-coordinates, eliminating the need for manual tuning and enabling robust adaptation to varying illumination, soil backgrounds, and other environmental conditions specific to the breeding plot. Fig. 9 shows the algorithm workflow and results. The original images of maize fields are displayed in Fig. 9a. Fig. 9b presented the detection results when the DSOTSUTA was not utilized. Histograms of the y-coordinates for all detected maize tassels are shown in Fig. 9c, providing a quantitative representation of tassel distribution across each frame. The histograms after the first step of the DSOTSUTA are presented in Fig. 9d, where initial segmentation isolated individual tassel ridges. The results of applying the DSOTSUTA are depicted in Fig. 9e, where the central maize tassel was successfully extracted, with disturbances from adjacent ridges effectively removed. The success of the algorithm can be attributed to its incorporation of maize planting characteristics and ridge distribution patterns. Maize was typically grown in evenly sized ridges with uniform spacing and parallel alignment, which enhanced the performance of the position-based DSOTSUTA. This alignment ensured that the extraction process was well-suited to the natural layout of maize fields.

The flight path of UAV video shooting and various interferences in the field will affect the extraction of individual ridges. Direction of the UAV video shooting is crucial, it has to be shot along the ridge, because it needs to remain parallel to the sequence of the maize ridge to avoid excessive tilt. Tilted video footage cannot not capture the spatial distribution characteristics of maize and maize tassel, which may result in the algorithm being unable to accurately extract individual ridges (Wu et al., 2022).

3.3. Performance of tassels tracking and counting

The proposed maize tassel counting system based on Deep SORT with filtering abnormal IDs effectively solves the problem of ID switching and achieves good performance on seven test videos. The verification was conducted in a test field used for maize breeding. The breeders counted each maize tassels separately to obtain the true value. This truth value will be directly compared with the proposed system

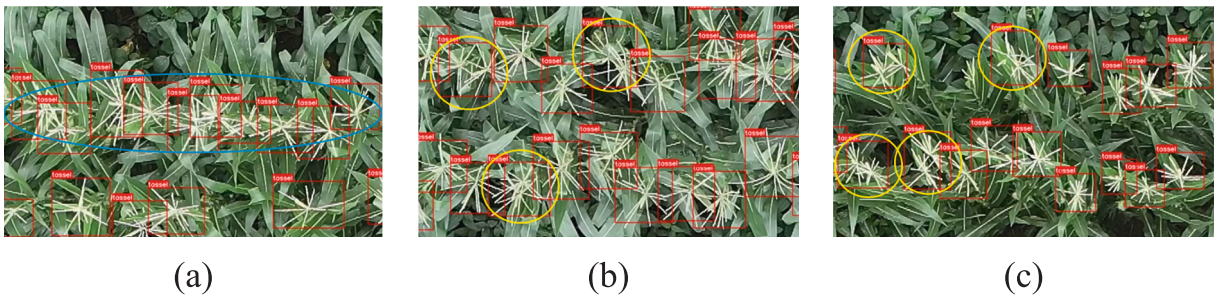


Fig. 8. Examples of *FP* in testing dataset, where the detected maize tassels were labeled with red bounding boxes. (a) Successfully detected overlapping maize tassels under occlusion within the blue circle; (b) and (c) Unsuccessfully detected overlapping maize inflorescence under severe occlusion indicated by a yellow circle.

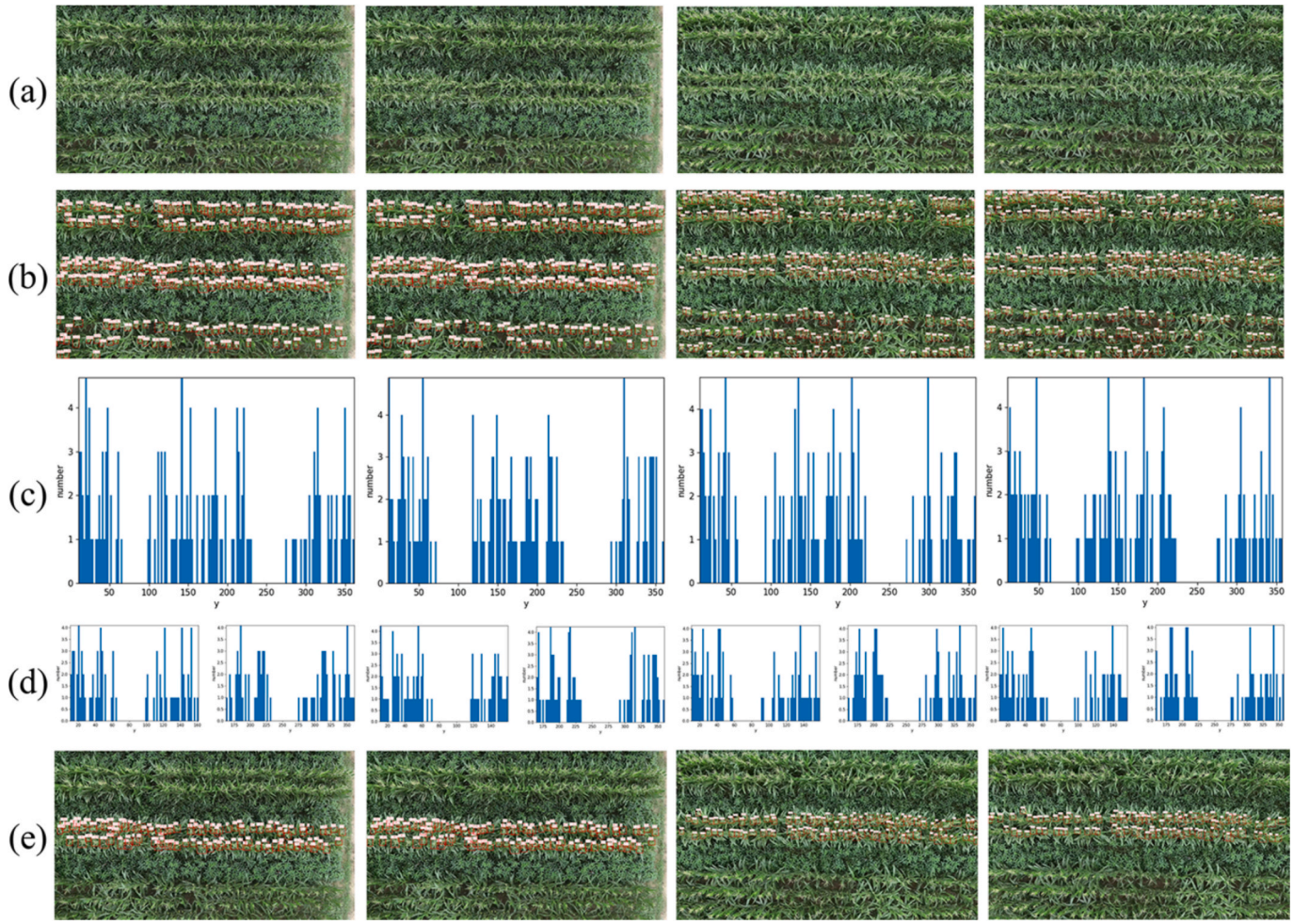


Fig. 9. Examples of DSOTSUTA extraction for an individual ridge maize tassel. (a) Original images. (b) Detection results by YOLOv8m. (c) Histograms of *y*-coordinates of detection bounding boxes based on detection results. (d) Histograms of *y*-coordinates segment by median N_M. (e) Individual ridge extraction results by DSOTSUTA.

Table 2
Maize tassel counts between ground truth, Deep SORT and Deep SORT with filter abnormal IDs and their accuracy in different videos.

Video number	1	2	3	4	5	6	7
Ground Truth	163	336	401	433	199	338	376
Deep SORT with filter abnormal IDs	151	303	390	512	202	371	412
Deep SORT with filter abnormal IDs Accuracy/%	92.64	90.18	97.26	84.57	98.51	91.16	91.26
Deep SORT	252	646	854	991	401	649	714
Deep SORT Accuracy/%	45.40	7.74	12.97	28.87	−1.51	7.99	10.11
Difference (Deep SORT with filter abnormal IDs –Ground Truth)	−12	−33	−11	79	3	33	36
Difference (Deep SORT with filter abnormal IDs –Ground Truth)/%	−7.36	−9.82	−2.74	18.25	1.51	9.76	9.57

count value to verify the effectiveness of the system. As shown in Table 2, this system performed extremely well in the testing of 7 videos, with an average P_c of 92.23 %, while the average P_c of Deep SORT was only 3.98 %. In practical breeding applications, breeders typically require a counting accuracy of around 90 % to ensure effective decision-making and resource allocation. The proposed system achieves a counting average P_c of 92.23 %, which meets this requirement and provides breeders with a reliable tool for their operations. The improvement was clearly evident in Fig. 11 and Table 2. Across seven test videos, Deep SORT achieved a mean P_c of only 3.98 %, with three videos showing negative accuracy (–12.97 % to –1.51 %), indicating systematic overcounting. In contrast, Deep SORT with filter abnormal IDs improved mean P_c to 92.23 %. The error reduction is further supported by *RMSE*: traditional Deep SORT had *RMSE* = 349.91 maize tassels per video, while the filtered version reduced this to 22.14 maize tassels per video. These results demonstrate both statistically and practically improvements. Notably, for real-time processing of the video streams captured by UAVs, this system requires a minimum of 6 GB GPU memory to ensure stable and efficient operation. All the experiments were conducted in a single, industry-standard breeding trial field. The maize planted in this field was arranged in a strictly parallel and equidistant ridge-like structure – this layout is widely adopted in breeding plans. This controllable and realistic environment provided a representative platform for testing: the average counting accuracy rate was 92.23 %, and the average error in the video was 22.14 tassels per video. This system proved its robustness under the conditions faced by breeders on a daily basis and was not affected by the complex variations introduced by different field designs.

In contrast, Deep SORT with filtering abnormal IDs was much closer to ground truth, and the *RMSE* was reduced to 22.14 maize tassels per video, greatly minimizing the average error between the counting results and the true values. Furthermore, the *MAPE* is 8.43 %, indicating a strong correlation between the improved system and the actual data. This suggests that the system can closely align with the real maize tassel count in practical applications. Ma et al. (2024) reduced the ID switching problem by introducing an attention mechanism, which showed that constant ID switching could effectively improve the performance of the tracker. These results further confirm that maize tassel counting system based on Deep SORT with filtering abnormal ID effectively suppresses ID switching phenomenon during Deep SORT tracking process, providing a highly accurate reflection of the actual quantity.

To assess the agreement between the system counts and the ground truth, we utilized a Bland – Altman plot. As shown in Fig. 10, this plot illustrates the difference (proposed algorithm – Ground Truth) against the mean of Ground Truth and the proposed algorithm, with lines representing the mean difference and ± 1.96 standard deviations (SD). Bland – Altman plot: We have included the Bland – Altman plot (mean difference ± 1.96 SD) between the system counts and ground truth, which is now provided as Fig. 10. This plot visually demonstrates the agreement between our system's counts and the ground truth. Most data points lie within the limits defined by ± 1.96 SD, indicating a good level of agreement and suggesting that the system's counting results are consistent with the ground truth values.

While the maize tassel counting system based on Deep SORT and abnormal ID filtering performs well in practice, the traditional Deep SORT showed obvious drawbacks in counting. In Fig. 11, the results of Deep SORT were less than satisfactory. The underlying reason for this lies in the nature of the Deep SORT algorithm, which integrates motion and appearance data through a recursive Kalman filter and the Hungarian algorithm to link maize tassels between video frames for counting purposes. However, in practical applications, because the video was captured by a UAV, the maize tassels were relatively stationary, and there was minimal motion variation between consecutive frames. Furthermore, the differences in the appearance of maize tassels were subtle, which presents a challenge for the algorithm in associating targets based on motion and appearance data. As shown in Fig. 12a and Fig. 12b, Deep SORT experienced a noticeable ID switch between frames 36 and 37 in test video 2, which undoubtedly leads to duplicate counting. This was the primary reason why the Deep SORT overestimated the ground truth (An et al., 2024).

The algorithm of filtering abnormal IDs effectively addresses the problem of ID switching between video frames in Deep SORT. As shown

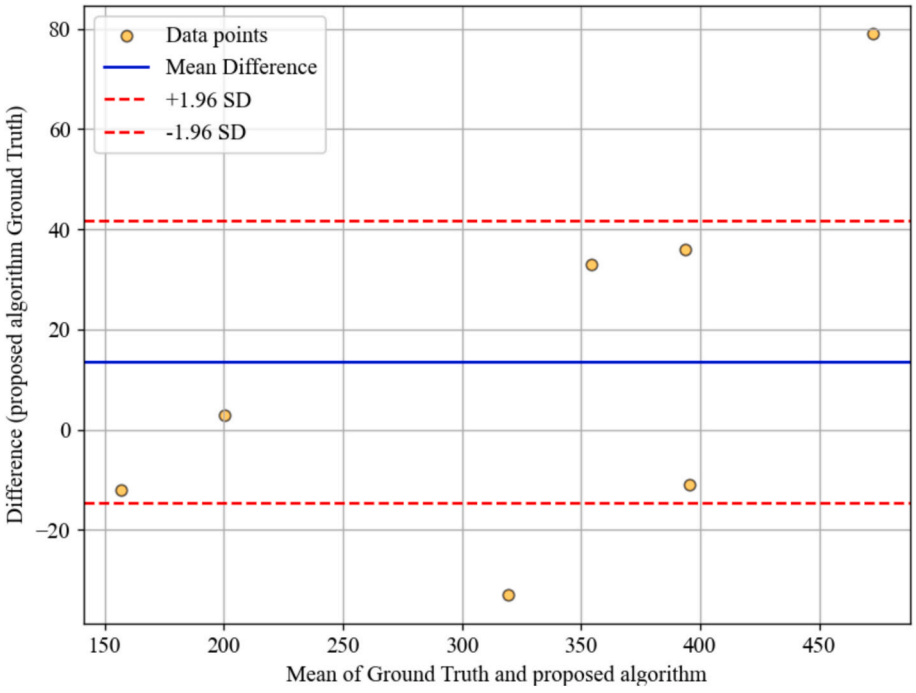


Fig. 10. The Bland-Altman plot between the truth value and the proposed method. The yellow circles represent the differences under the average values of different Ground Truth and proposed algorithm; The blue solid line represents the average value of these differences; The red dotted line indicates the range of ± 1.96 standard deviations (Standard Deviation, SD) of the average difference.

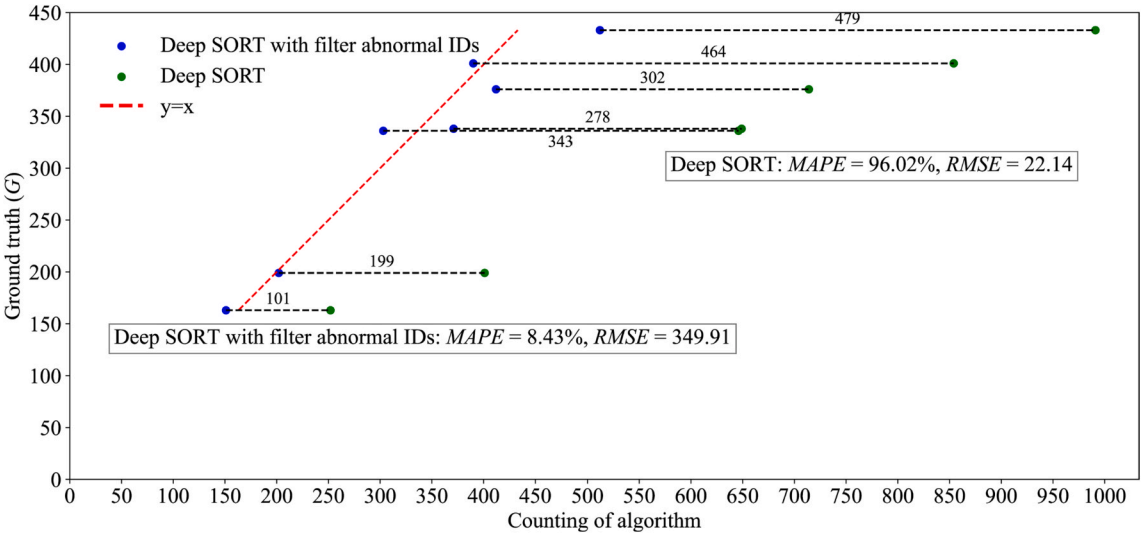


Fig. 11. Scatter plot of maize tassel counting algorithms. Red dashed line indicates the line of perfect agreement where the counting result by algorithm matches the ground truth (y = x); Black dashed line represents the count difference between Deep SORT and Deep SORT with filter abnormal IDs; Blue dots represents the counts obtained from Deep SORT with filter abnormal IDs; Green dots represents the counts obtained from Deep SORT.

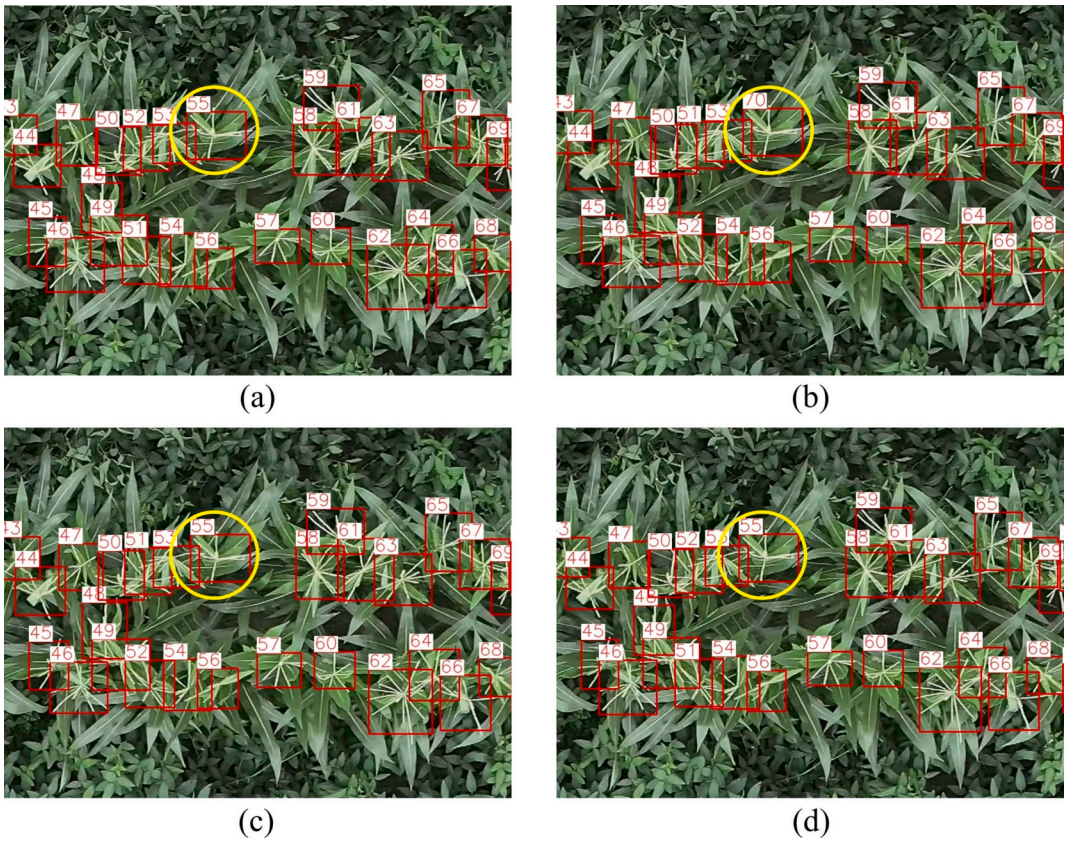


Fig. 12. Comparison of the effect of interframe maize tassels tracking and abnormal ID filtering. (a) Tracking results of maize tassels in frame 36 of test video 2 by Deep SORT algorithm; (b) Tracking results of maize tassels in frame 37 of test video 2 by Deep SORT algorithm; (c) Maize tassel tracking results in frame 36 of test video 2 based on Deep SORT with abnormal ID filtering algorithm; (d) Maize stalk tracking results in frame 37 of test video 2 based on Deep SORT and abnormal ID filtering algorithm; The red box marked the maize tassel, the number in the box was the tracking ID, and the yellow circle could observe the sign of an upcoming ID switch.

in Fig. 12a and Fig. 12b, prior to the improvement, noticeable ID switching occurred, leading to redundant counting (Zhang et al., 2021). In Fig. 12c and Fig. 12d, Deep SORT with filtering abnormal IDs algorithm effectively resolved ID switching problem between video frames. By comparing the σ of maize tassel IDs between consecutive frames with the μ of all maize tassels between consecutive frames, the system could filter out abnormal IDs, caused by ID switching. This made the tracking and counting of the maize tassels more reliable. Wu et al., (2023) also improved the counting accuracy by matching the displacement of counting objects between video frames, achieving a comparable or

potentially superior performance to our current method. Specifically, the method proposed by Wu et al., (2023) reduced the counting error rate from 12 % to 8 % and increased the overall counting accuracy from 88 % to 95 %. Additionally, Deep SORT with filtering abnormal IDs achieved an average P_c of 92.23 %, which further validates the potential of the filtering in improving counting performance.

When processing video with more frames, the maize tassel counting process based on the Deep SORT with filtering abnormal IDs exhibits noticeable performance issues. As shown in Fig. 11, as the number of maize tassels tracked by the Deep SORT algorithm increases, the error in the tracking and filtering abnormal IDs algorithm notably escalates. The overall success rate of maize tassel counting depends on the success rates of filtering abnormal IDs in consecutive frames, and they are in a multiplicative relationship. As the number of frames correspondingly rises, which means the number of operations for filtering abnormal IDs also increases. This is similar to a multi-step technological process (Xu et al., 2019). Assuming the multi-step technological process pass rate of each step is 99 %, as the number of steps increases, the overall success rate will be 0.99 raised to the corresponding power, and it will inevitably decrease. Similarly, the greater the number of video frames, the lower the overall success rate of counting. However, in practical maize breeding applications, the breeding plots are relatively small, and the number of tassels in each plot typically does not exceed 1000 plants. This limitation aids in mitigating the impact of increased errors in long videos on the overall counting accuracy.

In summary, the video-based tracking counting method, centered on multi-object tracking, has been applied in agriculture. This approach has shown promise in various applications, as demonstrated in recent studies (An et al., 2024; Gao et al., 2023; Tu et al., 2024). The individual ridge maize tassels counting system, which uses Deep SORT with filter abnormal IDs, is expected to be utilized in maize breeding plots. Building on these foundations, the present work makes three unique contributions. (1) proposed a counting system specifically designed for individual-ridge maize tassels. (2) By integrating DSOTSUTA and the filter abnormal IDs in Deep SORT, reduce adjacent-ridge interference and achieve 92.23 % mean P_c—an 84-percentage-point improvement over the unmodified Deep SORT baseline. (3) The system lowers breeding-expert workload by automating individual ridge maize tassels counting with an *RMSE* 22.14 tassels per video, demonstrating direct practical value in maize breeding plots. Beyond maize, the same motion-based Deep SORT framework—leveraging predictable ridge-parallel trajectories and adaptive ID filtering—can be readily transferred to other row crops such as wheat, soybean, sorghum, and even horticultural vegetables grown on ridges, offering a versatile, low-cost UAV solution for high-throughput, per-row plant counting across diverse agricultural systems. However, this study has several limitations that need to be acknowledged. First, the dataset was collected from a single location, so the performance of the system in tropical or sub-tropical maize growing regions remains unknown. Second, the flight direction was fixed parallel to the ridge, and the system's sensitivity to off-nadir angles greater than 15° has not been tested. Third, under dense canopy closure with more than 3 tassels per square meter, the displacement-increment assumption may be violated, affecting counting accuracy. Fourth, there is a lack of validation under night-time or low-light conditions, which is relevant as breeders sometimes conduct flights during dawn or dusk.

4. Conclusions

This study proposes a system for counting maize tassels on an individual ridge using YOLOv8 and Deep SORT. The proposed DSOTSUTA effectively filters out adjacent interference and extracts individual ridge tassels, obtaining the target area for accurate counting, which allowing UAV video data containing multiple ridge maize tassels to be used for individual ridge tassel counting. In the tested breeding trial field with strictly parallel and equidistant ridge layouts, YOLOv8m demonstrates excellent performance in field maize tassel detection. However, Deep SORT exhibits large errors in counting high-density tassels. Utilizing consecutive frame displacement to filter abnormal IDs, this process effectively reduces the counting errors typically caused by ID switching in Deep SORT, aligning the count results more closely with the ground truth of maize tassel counts. This study provides a methodology for individual ridge tassel counting, offering maize breeders an automated counting way. In the future, efforts will be directed towards adapting the system to mobile platforms to enhance accessibility for breeders in the field.

CRediT authorship contribution statement

Man Xia: Writing – original draft, Software, Data curation, Conceptualization. **Leilei He:** Writing – review & editing. **Yu Gu:** Software. **Xiaojuan Liu:** Formal analysis. **Bryan Gilbert Murengami:** Writing – review & editing. **Lamin L. Janneh:** Validation, Investigation. **Rui Li:** Funding acquisition. **Yan Long:** Project administration, Methodology. **Anastasia Grecheneva:** Supervision. **Longsheng Fu:** Writing – review & editing, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was partially supported by National Natural Science Foundation of China (32501786); National Foreign Expert Project, Ministry of Human Resources and Social Security, China (H20240238, Y20240046); Science and Technology Program of Yulin City, China (2023-CXY-183).

Data availability

Data is available at https://github.com/fu3lab/tassel_individual_ridge_counting.

References

An, Q., Cui, Y., Tong, W., Liu, Y., Zhao, B., Wei, L., 2024. StraTracker: a dynamic counting method for growing strawberries based on multi-target tracking. Comput. Electron. Agric. 227, 109564. https://doi.org/10.1016/j.compag.2024.109564.

Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B., 2016. Simple online and realtime tracking. Proc. - Int. Conf. Image Process. ICIP 2016-Augus, 3464–3468. DOI: 10.1109/ICIP.2016.7533003.

Cooper, M., Messina, C.D., 2023. Breeding crops for drought-affected environments and improved climate resilience. Plant Cell 35, 162–186. https://doi.org/10.1093/plcell/koac321.

Fernandez-Gallego, J.A., Kefauver, S.C., Gutiérrez, N.A., Nieto-Taladriz, M.T., Araus, J. L., 2018. Wheat ear counting in-field conditions: High throughput and low-cost approach using RGB images. Plant Methods 14, 1–12. https://doi.org/10.1186/s13007-018-0289-4.

Gao, C., He, L., Fang, W., Wu, Z., Jiang, H., Li, R., Fu, L., 2023. A novel pollination robot for kiwifruit flower based on preferential flowers selection and precisely target. Comput. Electron. Agric. 207, 107762. https://doi.org/10.1016/j.compag.2023.107762.

Guo, Y., Fu, Y.H., Chen, S., Robin Bryant, C., Li, X., Senthilnath, J., Sun, H., Wang, S., Wu, Z., de Beurs, K., 2021. Integrating spectral and textural information for identifying the tasseling date of summer maize using UAV based RGB images. Int. J. Appl. Earth Obs. Geoinf. 102, 102435. https://doi.org/10.1016/j.jag.2021.102435.

Hasan, M.M., Chopin, J.P., Laga, H., Miklavcic, S.J., 2018. Detection and analysis of wheat spikes using Convolutional Neural Networks. Plant Methods 14, 100. https://doi.org/10.1186/s13007-018-0366-8.

He, K., Zhao, Z., Ren, W., Chen, Z., Chen, L., Chen, F., Mi, G., Pan, Q., 2023. Mining genes regulating root system architecture in maize based on data integration analysis. Theor. Appl. Genet. 136, 1–15. https://doi.org/10.1007/s00122-023-04376-0.

Hoopes, G.M., Hamilton, J.P., Wood, J.C., Esteban, E., Pasha, A., Vaillancourt, B., Provart, N.J., Buell, C.R., 2019. An updated gene atlas for maize reveals organ-

specific and stress-induced genes. Plant J. 97, 1154–1167. https://doi.org/10.1111/tpj.14184.
Jia, W., Cao, K., Liu, M., Lu, Y., Ji, Z., Liu, G., Yin, X., Ge, X., 2024a. FCOS-EAM: an accurate segmentation method for overlapping green fruits. Comput. Electron. Agric. 226, 109392. https://doi.org/10.1016/j.compag.2024.109392.
Jia, Y., Fu, K., Lan, H., Wang, X., Su, Z., 2024b. Maize tassel detection with CA-YOLO for UAV images in complex field environments. Comput. Electron. Agric. 217, 108562. https://doi.org/10.1016/j.compag.2023.108562.
Karami, A., Quijano, K., Crawford, M., 2021. Advancing tassel detection and counting: annotation and algorithms. Remote Sens. 13, 2881. https://doi.org/10.3390/rs13152881.
Li, C., Binaghi, M., Pichon, V., Cannarozzi, G., Brandão de Freitas, L., Hanemian, M., Kuhlemeier, C., 2023. Tight genetic linkage of genes causing hybrid necrosis and pollinator isolation between young species. Nat. Plants 9, 420–432. https://doi.org/10.1038/s41477-023-01354-8.
Li, Y., Song, D., Dang, P., Wei, L., Qin, X., Siddique, K.H.M., 2020. Combined ditch buried straw return technology in a ridge–furrow plastic film mulch system: Implications for crop yield and soil organic matter dynamics. Soil Tillage Res. 199, 104596. https://doi.org/10.1016/j.still.2020.104596.
Li, Yong xiang, Lu, J., He, C., Wu, X., Cui, Y., Chen, L., Zhang, J., Xie, Y., An, Y., Liu, X., Zhen, S., Liu, Y., Li, C., Zhang, D., Shi, Y.S., Song, Y., Wang, J., Li, Yu, Wang, G., Fu, J., Wang, T., 2022. cis-Regulatory variation affecting gene expression contributes to the improvement of maize kernel size. Plant J. 1595–1608. DOI: 10.1111/tpj.15910.
Li, Z., Zhu, Y., Sui, S., Zhao, Y., Liu, P., Li, X., 2024. Real-time detection and counting of wheat ears based on improved YOLOv7. Comput. Electron. Agric. 218, 108670. https://doi.org/10.1016/j.compag.2024.108670.
Liang, C., Zhang, Z., Zhou, X., Li, B., Zhu, S., Hu, W., 2022. Rethinking the competition between Detection and ReID in Multiobject Tracking. IEEE Trans. Image Process. 31, 3182–3196. https://doi.org/10.1109/TIP.2022.3165376.
Lin, Y., Chen, T., Liu, S., Cai, Y., Shi, H., Zheng, D., Lan, Y., Yue, X., Zhang, L., 2022. Quick and accurate monitoring peanut seedlings emergence rate through UAV video and deep learning. Comput. Electron. Agric. 197, 106938. https://doi.org/10.1016/j.compag.2022.106938.
Liu, W., Quijano, K., Crawford, M.M., 2022. YOLOv5-Tassel: detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 8085–8094. https://doi.org/10.1109/JSTARS.2022.3206399.
Liu, Y., Cen, C., Che, Y., Ke, R., Ma, Y., Ma, Y., 2020. Detection of maize tassels from UAV RGB imagery with Faster R-CNN. Remote Sens. 12, 338. https://doi.org/10.3390/rs12020338.
Lu, H., Cao, Z., Xiao, Y., Zhuang, B., Shen, C., 2017. TasselNet: counting maize tassels in the wild via local counts regression network. Plant Methods 13, 79. https://doi.org/10.1186/s13007-017-0224-0.
Ma, S., Duan, S., Hou, Z., Yu, W., Pu, L., Zhao, X., 2024. Multi-object tracking algorithm based on interactive attention network and adaptive trajectory reconnection. Expert Syst. Appl. 249, 123581. https://doi.org/10.1016/j.eswa.2024.123581.
Madec, S., Jin, X., Lu, H., De Solan, B., Liu, S., Duyme, F., Heritier, E., Baret, F., 2019. Ear density estimation from high resolution RGB imagery using deep learning technique. Agric. For. Meteorol. 264, 225–234. https://doi.org/10.1016/j.agrformet.2018.10.013.
Pauw, A., 2013. Can pollination niches facilitate plant coexistence? Trends Ecol. Evol 28, 30–37. https://doi.org/10.1016/j.tree.2012.07.019.
Rizzo, G., Pablo Monzon, J., Tenorio, F.A., Howard, R., Cassman, K.G., Grassini, P., 2022. Climate and agronomy, not genetics, underpin recent maize yield gains in favorable environments. PNAS 119, e2113629119. https://doi.org/10.1073/pnas.2113629119.
Santos, T.T., de Souza, K.X.S., Camargo Neto, J., Koenigkan, L.V., Moreira, A.S., Ternes, S., 2024. Multiple orange detection and tracking with 3-D fruit relocalization and neural-net based yield regression in commercial sweet orange orchards. Comput. Electron. Agric. 224, 109199. https://doi.org/10.1016/j.compag.2024.109199.
Shi, H., Zhang, J., Lei, A., Wang, C., Xiao, Y., Wu, C., Wu, Q., Zhang, S., Xie, J., 2024. Enhancing detection accuracy of highly overlapping targets in agricultural imagery using IoA-SoftNMS algorithm across diverse image sizes. Comput. Electron. Agric. 227, 109475. https://doi.org/10.1016/j.compag.2024.109475.
Tu, S., Huang, Y., Liang, Y., Liu, H., Cai, Y., Lei, H., 2024. A passion fruit counting method based on the lightweight YOLOv5s and improved DeepSORT. Precis. Agric. 25, 1731–1750. https://doi.org/10.1007/s11119-024-10132-1.
Varghese, R., Sambath, M., 2024. YOLOv8: a novel object detection algorithm with enhanced performance and robustness. In: In: 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). IEEE, pp. 1–6.
Wojke, N., Bewley, A., Paulus, D., 2017. Simple online and realtime tracking with a deep association metric. Proc. - Int. Conf. Image Process. ICIP 2017-Septe, 3645–3649. DOI: 10.1109/ICIP.2017.8296962.
Wu, X., Li, W., Hong, D., Tao, R., Du, Q., 2022. Deep Learning for unmanned aerial vehicle-based object detection and tracking: a survey. IEEE Geosci. Remote Sens. Mag. 10, 91–124. https://doi.org/10.1109/MGRS.2021.3115137.
Wu, Z., Sun, X., Jiang, H., Gao, F., 2023. ScienceDirect twice matched fruit counting system : an automatic fruit counting pipeline in modern apple orchard using mutual and secondary matches. Biosyst. Eng. 234, 140–155. https://doi.org/10.1016/j.biosystemseng.2023.09.005.
Xu, J., Cao, Y., Zhang, Z., Hu, H., 2019. Spatial-temporal relation networks for multi-object tracking. Proc. IEEE Int. Conf. Comput. Vis. 2019-Octob, 3987–3997. DOI: 10.1109/ICCV.2019.00409.
Yang, H., Chang, F., Huang, Y., Xu, M., Zhao, Y., Ma, L., Su, H., 2022. Multi-object tracking using Deep SORT and modified CenterNet in cotton seedling counting. Comput. Electron. Agric. 202, 107–339. https://doi.org/10.1016/j.compag.2022.107339.
Zhang, X., Zhu, D., Wen, R., 2023. SwinT-YOLO: Detection of densely distributed maize tassels in remote sensing images. Comput. Electron. Agric. 210, 107905. https://doi.org/10.1016/j.compag.2023.107905.
Zhang, Y., Wang, C., Wang, X., Zeng, W., Liu, W., 2021. FairMOT: on the fairness of detection and re-identification in multiple object tracking. Int. J. Comput. Vis. 129, 3069–3087. https://doi.org/10.1007/s11263-021-01513-4.
Zou, H., Lu, H., Li, Y., Liu, L., Cao, Z., 2020. Maize tassels detection: a benchmark of the state of the art. Plant Methods 16, 108. https://doi.org/10.1186/s13007-020-00651-z.