RTrack-DL: A dynamic scene-oriented model for instance identification and continuous tracking of rice seedling rows

Ziqi Li[1†], Dongfang Li[1†], He Zheng[1], Weiwei Duan[1], Yicai Shan[2], Maohua Xiao[1], Ke Chen[1*]

(1. *College of Engineering, Nanjing Agricultural University, Nanjing 210031, China*;
2. *School of Electronic Engineering, Nanjing Xiaozhuang University, Nanjing 211171, China*)

Abstract: Visual navigation is critical for intelligent paddy field operations, where continuous and accurate tracking of rice seedling rows is essential. However, existing methods predominantly address static frame identification and lack robust solutions for continuous tracking in dynamic field environments, particularly under challenging conditions like missing seedlings and irregular planting patterns. To overcome these limitations, this study develops RTrack-DL, a novel dynamic scene-oriented model for instance-level identification and continuous tracking of rice seedling rows. The novelty lies in its dynamic framework that integrates a dedicated merge-and-deduplication strategy (IoU-Plus) to enhance line-fitting robustness against missing seedlings and irregular planting, and fuses the fitted lines with a multi-object tracker to maintain stable row identities and prevent track loss during field operations. The model employs DeepLabv3+ for pixel-level semantic segmentation of row centerlines. Subsequently, connected component analysis and line fitting are applied to obtain individual instances. To handle discontinuities caused by missing seedlings and irregular planting, a merge-and-deduplication strategy based on an IoU-Plus metric is introduced to eliminate redundant lines and enhance fitting robustness. A multi-object tracking module combining Kalman filtering and the Hungarian algorithm ensures consistent identity tracking across frames. The performance of RTrack-DL is evaluated through experiments under three illumination conditions—sunny, cloudy, and overcast—while considering scenarios involving missing seedlings and irregular planting. Results show optimal performance under cloudy conditions, with a fitting accuracy of 99.42% and an angular error of 0.49°, while sunny and overcast conditions yielded slightly reduced accuracy. In tracking tests, RTrack-DL achieved multiple object tracking accuracy scores of 93.5%, 92.1%, and 92.7% under cloudy, sunny, and overcast conditions, respectively, with real-time performance maintained at 15-20 FPS. The proposed framework demonstrates a practical solution for continuous tracking of dense rice seedling rows in dynamic field environments, providing a robust visual perception component for intelligent navigation systems in fully unmanned paddy field management.

Keywords: rice seedling row instance, seedling row line fitting, merge-and-deduplication strategy, continuous tracking, Kalman filtering, Hungarian algorithm

DOI: 10.25165/j.ijabe.20261904.10661

Citation: Li Z Q, Li D F, Zheng H, Duan W W, Shan Y C, Xiao M H, et al. RTrack-DL: A dynamic scene-oriented model for instance identification and continuous tracking of rice seedling rows. Int J Agric & Biol Eng, 2026; 19(4): 243–255.

1 Introduction

Rice is a critical foundation for global food security[1], and navigation technologies have been widely integrated into various stages of its production[2-4]. The commonly used Global Navigation Satellite System (GNSS), however, is susceptible to positioning inaccuracies and trajectory deviations exacerbated by the muddy conditions of paddy fields[5-7]. In contrast, visual navigation has emerged as a key enabling technology for precise unmanned operations due to its capability for local path correction[8-10], whose performance relies on stable and accurate perception of crop rows in the field[11]. While existing research has predominantly focused on the static identification of seedling rows[12,13], the muddy and unstable terrain in paddy fields challenges the assumption of static operation. Relying solely on frame-by-frame detection makes it difficult to establish precise inter-frame correspondences at the row level, leading to frequent target loss and identity discontinuity[14]. Therefore, introducing dynamic tracking is essential to ensure the temporal continuity of seedling rows across consecutive frames. Current research on dynamic tracking has been largely concentrated on structured road scenarios and has begun to be preliminarily applied to crop row tracking in dry fields[15]. However, research on dynamic visual tracking of rice seedling rows in the unique, unstructured paddy field environment— characterized by varying illumination, missing seedlings, and irregular row patterns—remains scarce. Consequently, research on dynamic scene-oriented identification and continuous tracking of rice seedling rows is crucial for ensuring stable and reliable autonomous operation of machinery in complex paddy field environments[16].

The primary tasks for the identification and continuous tracking

Received date: 2026-05-29 **Accepted date:** 2026-07-28

Biographies: Ziqi Li, M.S. student, research interest: agricultural machine vision, crop row perception and tracking, Email: liziqi@stu.njau.edu.cn; **Dongfang Li**, Lecturer, research interest: intelligent agricultural machinery and autonomous navigation of field robots, Email: dongfangli@njau.edu.cn; **He Zheng**, Ph.D. student, research interest: paddy-field robotic weeding, Email: zhenghe@stu.njau.edu.cn; **Weiwei Duan**, M.S. student, research interest: intelligent weeding technologies, Email: 2024112031@stu.njau.edu.cn; **Yicai Shan**, Professor, research interest: precision agricultural machinery and paddy-field weeding, Email: danyicai@njxzc.edu.cn; **Maohua Xiao**, Professor, research interest: intelligent weeding equipment and paddy-field machinery, Email: xiaomaohua@njau.edu.cn.

†The authors contributed equally to this work.

***Corresponding author: Ke Chen**, Lecturer, research interest: intelligent agricultural equipment, machine vision, field robotics. College of Engineering, Nanjing Agricultural University, Nanjing 210031, China. Tel: +86-18751909360, Email: ckyf@njau.edu.cn.

of rice seedling rows involve accurately extracting multiple rows from each frame[17]. Traditional methods, such as RGB threshold segmentation[18] and corner point detection[19], exhibit high sensitivity to threshold settings and geometric priors, making it challenging to precisely extract dense rice seedling rows. In recent years, deep learning-based semantic segmentation methods have been increasingly adopted for seedling row recognition. These approaches primarily consist of two steps: (ⅰ) pixel-level mask extraction of seedling rows[20-22], and (ⅱ) fitting centerlines to the extracted masks. Considerable research has been conducted on centerline fitting for seedling row masks. For instance, Li et al.[23] proposed an anchor point clustering algorithm based on dynamic search direction, Wu et al.[24] introduced an approximate B-spline fitting approach, He et al.[25] developed a prior-informed clustering method for centerline extraction, and Chen et al.[26] performed seedling row fitting within each instance using an instance segmentation network. While these methods have achieved progress under specific scenarios, they do not fully account for discontinuities in mask regions caused by missing seedlings and irregular planting patterns. Consequently, applying line fitting directly to such discontinuous masks in complex field environments can lead to fragmented or deviated row segments.

Moreover, existing methods for seedling row recognition are limited to providing instantaneous spatial information only[27-29], with little research dedicated to continuous visual tracking of seedling rows. These static approaches cannot establish spatial associations across multiple frames, making them ineffective in handling scenarios involving missing seedlings or irregular row patterns. Hernandez and Medeiros[30] proposed a multi-object tracking method using Vision Transformers for spatial association, which tracks bounding boxes of individual crops and estimates object trajectories by exploiting spatial correlations across adjacent frames. Similarly, Han et al.[31] introduced a multi-object tracking method based on graph optimization, which associates instance segmentation masks between consecutive frames by constructing a graph model to estimate trajectories. However, in densely planted paddy field environments with a large number of seedling rows and minimal inter-row spacing, relying solely on bounding boxes or mask-level associations leads to increased difficulty in data association, making it unsuitable for stable tracking of individual seedling rows and often resulting in target loss. Effective visual navigation in these scenarios requires simultaneous and continuous tracking of multiple seedling row instances to maintain a stable navigation path, which is beyond the capability of current association-based methods alone.

Therefore, achieving reliable identification and continuous tracking of dense rice seedling rows in dynamic scenarios remains a significant research gap. This gap is twofold: (ⅰ) the lack of robust line-fitting methods resilient to pervasive paddy field disturbances such as missing seedlings and irregular rows, leading to fragmented and deviated row representations[32]; and (ⅱ) the absence of a visual tracking framework capable of leveraging line features to maintain temporal consistency of seedling rows across adjacent frames[33]. These gaps translate into concrete technical challenges. First, the prevalent issue of missing seedlings[34] causes under-segmentation of rice seedling rows, resulting in disconnected mask clusters for a single physical row. Applying line fitting to such clusters yields multiple misaligned segments, failing to provide a unified and stable row structure. Second, in dense and visually similar rows, tracking methods reliant on bounding boxes or generic appearance features struggle to preserve continuity across frames[35], leading to track loss. Furthermore, operational viability demands a balance between accuracy and efficiency to meet real-time requirements[36].

To this end, this study proposes RTrack-DL, a dynamic scene-oriented model for instance identification and continuous tracking of rice seedling rows: (ⅰ) It establishes a dynamic framework that achieves instance-level row identification through semantic segmentation combined with connected component analysis and line fitting, and seamlessly bridges it with continuous multi-object tracking. (ⅱ) It innovatively designs and introduces a merge-and-deduplication strategy, guided by a composite IoU-Plus metric, to effectively handle line fragmentation and deviation caused by missing seedlings and irregular planting patterns, thereby enhancing fitting robustness. (ⅲ) It accomplishes the integration of fitted seedling row lines with a multi-object tracker (combining Kalman filtering and the Hungarian algorithm) to achieve persistent identity association across frames and prevent track loss during field operations. Experimental results demonstrate that the proposed method achieves high fitting accuracy and stable tracking performance under various common daylight conditions while maintaining real-time processing capabilities.

The remainder of this paper is organized as follows: Section 2 presents the detailed architecture and components of the proposed RTrack-DL model for rice seedling row instance identification and continuous tracking. Section 3 provides experimental results and discussions. Finally, Section 4 concludes the study with summary remarks.

2 Materials and methods

In this study, we propose RTrack-DL, a dynamic scene-oriented model for instance identification and continuous tracking of rice seedling rows. This framework is specifically designed to meet the core perceptual requirements for visual navigation in paddy fields, which demands not only static detection but also: the translation of low-level pixels into stable, actionable geometric primitives (lines) for control; robustness of these primitives against field irregularities like missing seedlings; and temporal consistency of primitives across frames to ensure smooth guidance. To fulfill these requirements, the central focus of this work lies in the novel integration of a robust line-fitting module with a multi-object tracker, creating a coherent perception-to-tracking pipeline. The overall framework is illustrated in Figure 1.

The proposed method consists of four main stages, each addressing a specific need of the navigation pipeline. First, images are collected and annotated to construct the dataset. Next, the DeepLabv3+ semantic segmentation model[37] is employed to extract pixel-level masks of seedling row centerlines, providing the foundational visual perception. Third, to transform these masks into navigation-ready geometric features, connected component analysis identifies individual row instances. Critically, to satisfy the requirement for robust and unique line representations per physical row—even when masks are fragmented—a dedicated merge-and-deduplication strategy is applied during line fitting. This ensures each row instance yields a single, reliable line. Finally, to meet the necessity for temporal stability in the guidance signal, multi-object continuous tracking is achieved by integrating Kalman filtering[38-39] and the Hungarian algorithm[40-41]. This tracker maintains consistent identities for the fitted lines across frames, preventing navigation errors due to track loss. The RTrack-DL model is thus designed as an integrated solution, bridging robust instance-level perception with stable temporal association for deployment on autonomous paddy field machinery.

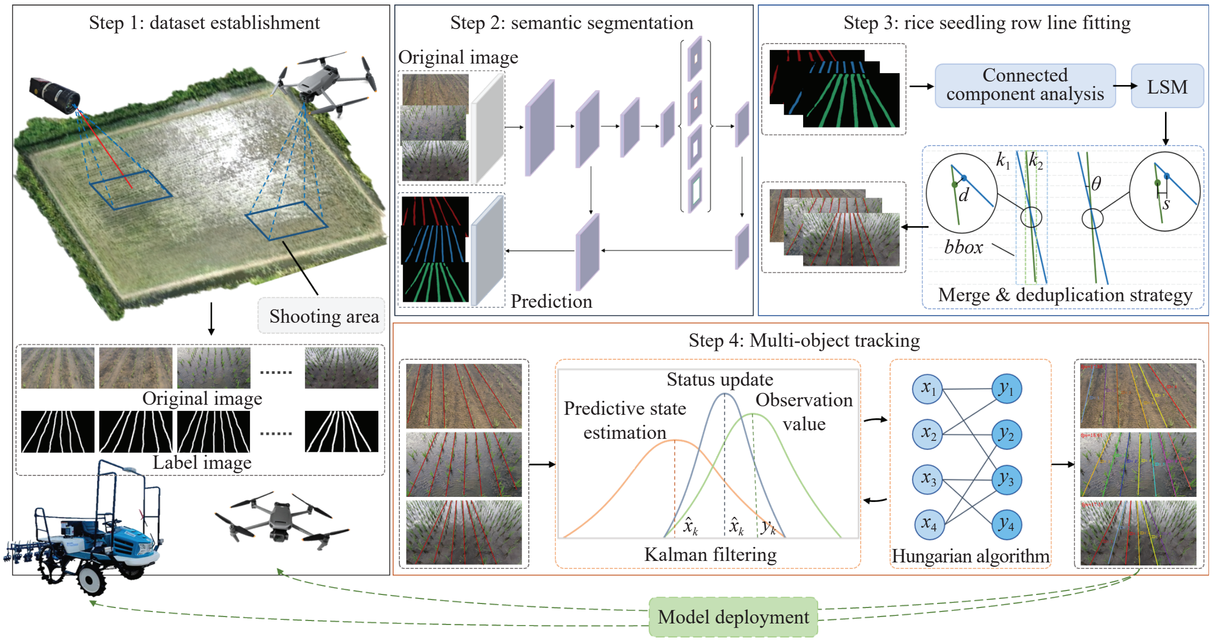

Figure 1 Flow diagram of the proposed rice seedling row tracking method

2.1 Dataset establishment

The dataset used in this study was collected on June 22 (8 d after transplanting), June 24 (10 d), June 28 (14 d), July 2 (18 d), July 4 (20 d), and July 7 (23 d) of 2024. Video data were captured using a DJI Mavic 3 Pro drone from the DJI Mavic series. The collection site was the rice cultivation demonstration base at Taihe Farm, Jiangning District, Nanjing, Jiangsu Province, China. The original videos were recorded in MP4 format with a resolution of 1920×1080 pixels and a frame rate of 25 fps. Individual video frames were extracted and saved in JPG format for subsequent processing. To simulate the visual configuration of an onboard camera mounted on paddy field machinery, the drone was flown at a height of 0.9 m with a camera inclination angle between 30° and 45° relative to the horizontal plane. This setting ensured that the projected characteristics of the rice seedling rows were consistent with those observed from an actual vehicle-mounted viewpoint. A constant flight speed was maintained during data collection to ensure high-quality video capture. The weather conditions during data collection were as follows: sunny (June 22 and June 28), cloudy (June 24 and July 4), and overcast (July 2 and July 7), corresponding to strong, soft, and low illumination levels, respectively. These three conditions—sunny, cloudy, and overcast—were selected as they represent the most common and operationally relevant daylight scenarios during typical field operation hours (e.g., 9 AM to 4 PM) for autonomous agricultural machinery. Across all conditions, the images contained missing seedlings and irregular rice seedling row patterns. For each weather condition, 1200 valid image frames were extracted, resulting in a total of 3600 images. The dataset of 3600 images was randomly split into training (70%, 2520 images), validation (15%, 540 images), and testing (15%, 540 images) sets, ensuring no temporal overlap between sets. Image annotation was performed by two graduate students with expertise in agricultural robotics using the Label Studio platform. The centerline of each visible seedling row was manually annotated as a polyline. Disagreements were resolved through discussion with a senior researcher. Inter-rater reliability (measured by average endpoint distance) was found to be below 5 pixels. Representative original images and corresponding label images of rice seedling rows collected under the three weather conditions are shown in Figure 2.

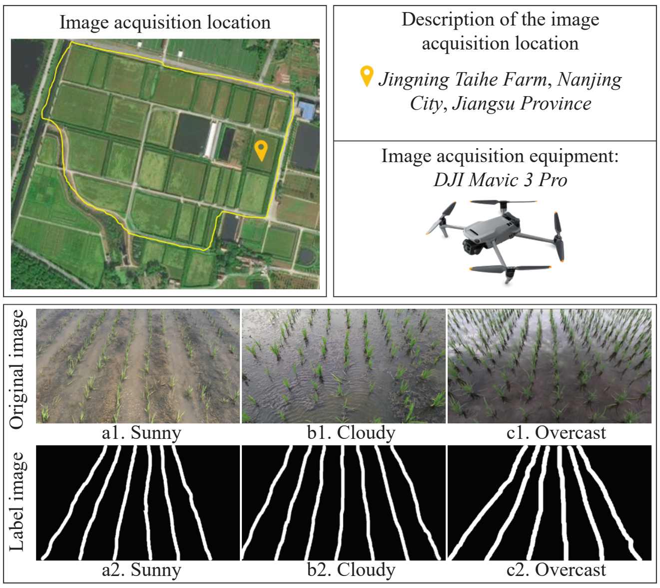

Figure 2 Rice seedling row original images and corresponding label images under different weather conditions

2.2 Architecture overview of RTrack-DL

The RTrack-DL model is built upon the DeepLabv3+ architecture, with the integration of a seedling row line-fitting module and a multi-object tracking module. In the line-fitting module, a merge-and-deduplication strategy is introduced to improve fitting precision, ensuring that only one line is fitted for each rice seedling row. The multi-object tracking module combines Kalman filtering with the Hungarian algorithm to achieve stable multi-object tracking of rice seedling rows, supporting fully autonomous operation in paddy field management. An overview of the proposed model is illustrated in Figure 3, which highlights the three core components of the RTrack-DL framework:

ⅰ) Feature Extraction Module. This module operates as the semantic segmentation front of the RTrack-DL model. The original image is first processed by the backbone network MobileNetV2 to extract multi-scale feature representations. Shallow features are passed to the decoder, while deep features are fed into the Atrous Spatial Pyramid Pooling (ASPP) structure. ASPP applies dilated convolutions and average pooling with different dilation rates to capture multi-scale context. The resulting feature maps are fused and refined through a 1×1 convolution to adjust the channel dimensions. In the decoder, the deep features from the encoder are upsampled by a factor of 4 and fused with the shallow features extracted from the backbone. After a 3×3 convolution and another 4×fold upsampling, the final segmentation map is produced, representing the pixel-level locations of rice seedling rows in the input image.

ii) Rice Seedling Row Line-Fitting Module. This module is designed to fit each segmented rice seedling row instance into a straight line. Connected component analysis is performed on the segmented mask to extract valid pixel regions to determine the position of each seedling row accurately. A least squares method is then applied to fit a straight line for each instance. Robust line fitting results are ensured by employing a merge-and-deduplication strategy to eliminate redundant lines generated during the fitting process, thereby ensuring that each rice seedling row instance is represented by a single optimal line.

iii) Multi-object Tracking Module. This module performs continuous tracking of rice seedling row instances across video frames by integrating Kalman filtering with the Hungarian algorithm. The Kalman filtering predicts the position of each seedling row line in the next frame, and a cost matrix is constructed on the basis of the distance and slope between the predicted lines and the newly fitted ones. The Hungarian algorithm is then used to determine the globally optimal one-to-one matching, minimizing the total cost. When a match is successful, the Kalman filtering is updated; if not, unmatched instances are discarded once their missed frame count exceeds a predefined threshold. This framework enables stable multi-object association and continuous tracking of seedling rows throughout the video sequence.

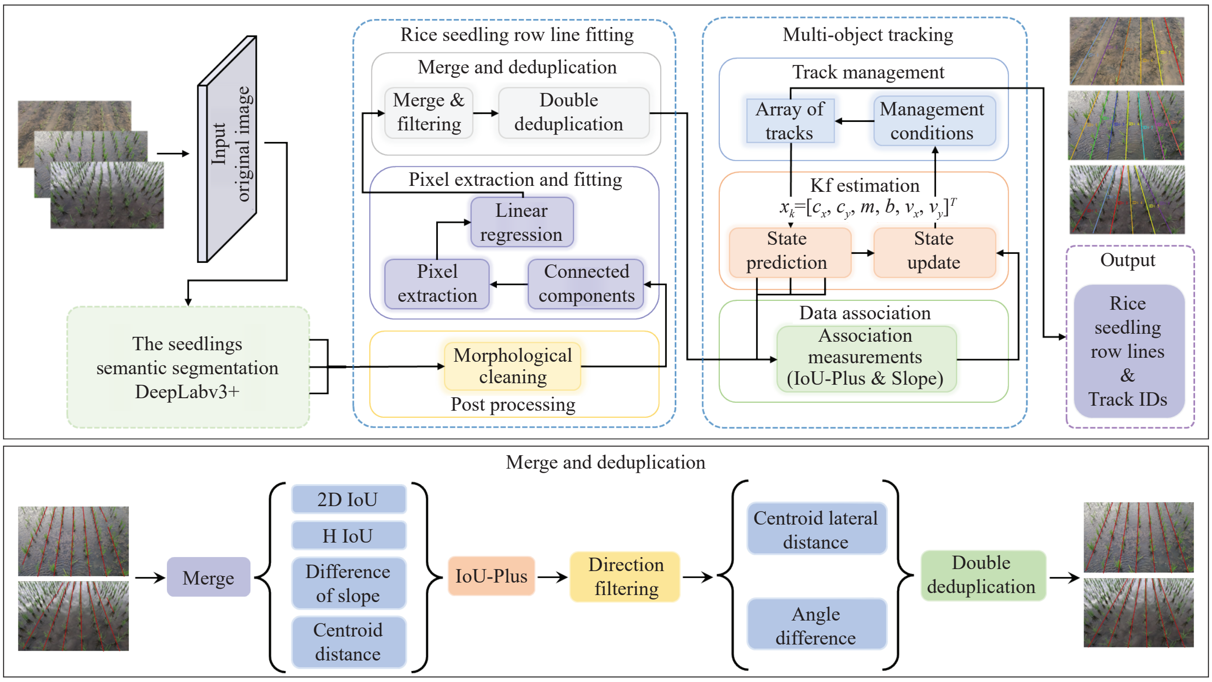

Figure 3 Framework of the proposed RTrack-DL model

2.3 DeepLabv3+ model

DeepLabv3+ is a widely used high-accuracy model for semantic segmentation. It adopts an encoder–decoder architecture combined with ASPP to fuse shallow and deep features, enabling pixel-level segmentation of complex visual scenes. The overall architecture of DeepLabv3+ is shown in Figure 4. In this study, a lightweight backbone—MobileNetV2[42]—is adopted to reduce the number of model parameters and enhance computational efficiency. This design ensures that the RTrack-DL model maintains real-time tracking performance without compromising segmentation accuracy. DeepLabv3+ with a MobileNetV2 backbone is selected as the segmentation model in this study due to its demonstrated balance between segmentation accuracy and computational efficiency, which is crucial for real-time agricultural applications. While newer architectures (e.g., SegFormer, Mask2Former) may offer marginal gains in certain metrics, they often incur higher computational costs[43]. For the specific task of segmenting dense, thin, linear structures such as rice seedling row centerlines, the Atrous Spatial Pyramid Pooling (ASPP) module in DeepLabv3+ effectively captures multi-scale contextual information without introducing excessive latency, making it a suitable choice for our dynamic field scenario.

2.4 Rice seedling row line fitting

The rice seedling row line-fitting process consists of four main steps: post processing, pixel extraction, line fitting, and the merge-and-deduplication strategy. The overall structure of this module is illustrated in Figure 5. In this study, morphological operations are first applied to the semantic segmentation results to obtain denoised binary segmentation images of rice seedling rows. Connected component analysis is then performed on the seedling row masks to cluster pixels into independent instance regions. For each instance, pixel coordinates are extracted for subsequent line fitting. Given the set of extracted pixel coordinates $Q_n(x_i, y_i)$, the fitted centerline is assumed to follow a linear form. The estimated parameters of the line are denoted by m and b, and the centroid of the seedling row instance is denoted by $(\bar{x}, \bar{y})$. The resulting line equation is expressed as in Equation (1):

$$y_i = mx_i + b \quad (i = 1,2,\ldots n) \tag{1}$$

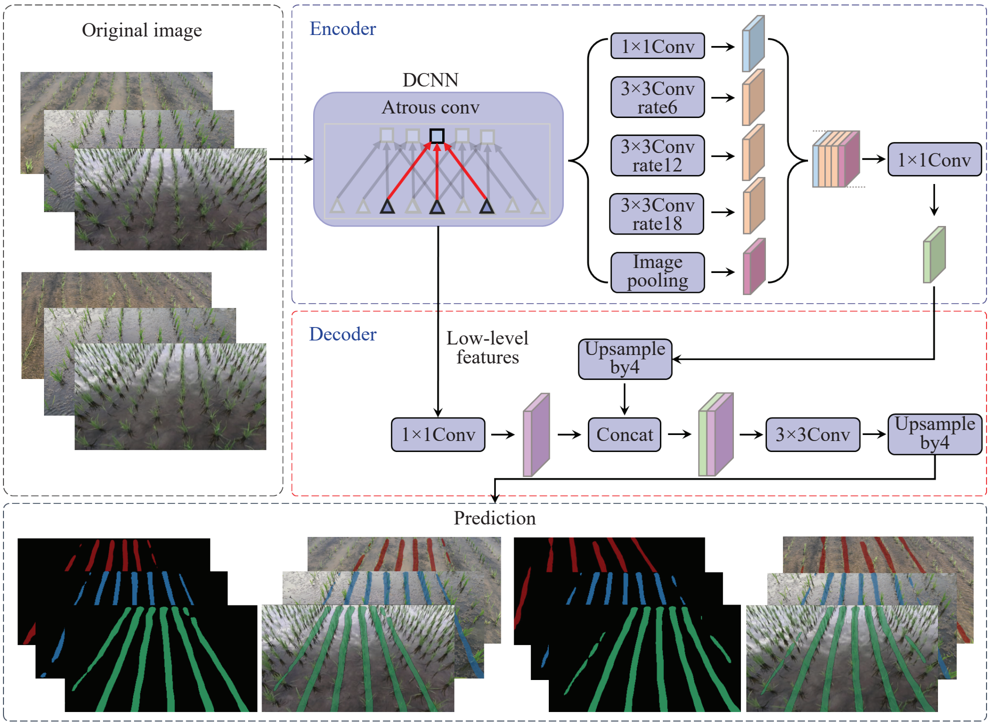

Figure 4 Architecture of the DeepLabv3+ model

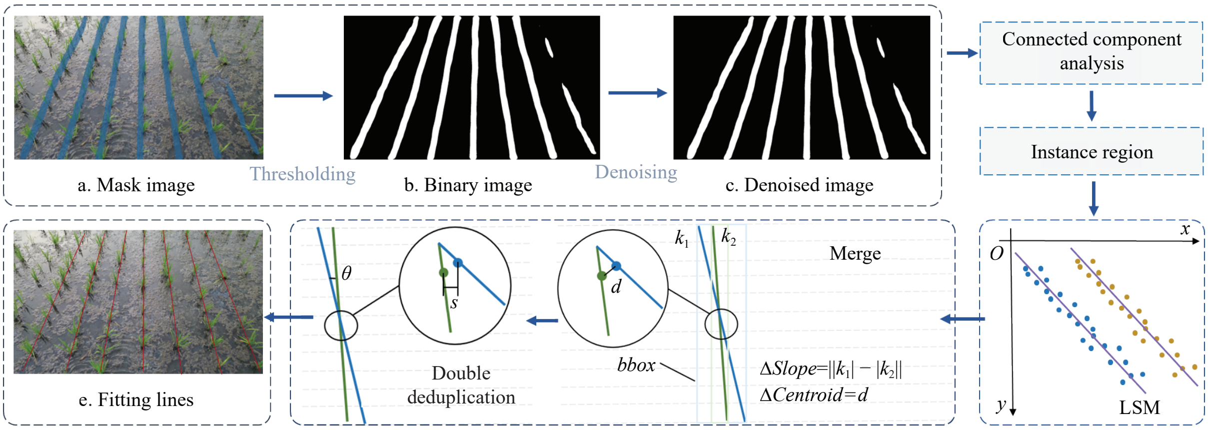

Figure 5 Framework of the rice seedling row line-fitting module

where,

$$\begin{cases} m = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \\ b = \bar{y} - m\bar{x} \end{cases} \tag{2}$$

$$\begin{cases} \bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i \\ \bar{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i \end{cases} \tag{3}$$

On the basis of Equations (2) and (3), the optimal combination of m and b can be determined using the least squares method[44], which is employed in this study to fit the centerline of each rice seedling row instance.

However, the mask regions of some seedling rows may become discontinuous during the segmentation process because of missing seedlings and irregular planting patterns. As a result, multiple line segments may be fitted to the same seedling row instance. To mitigate this issue, we apply a merge-and-deduplication strategy after the fitting stage. This strategy effectively reduces the occurrence of multiple lines being assigned to a single instance, thereby enhancing the robustness and accuracy of the overall fitting results.

In the proposed strategy, line merging is performed by expanding the bounding boxes (bbox) of the fitted lines. A multi-dimensional metric named IoU-Plus is introduced to determine whether two fitted lines belong to the same rice seedling row instance. This metric incorporates four criteria: 2D intersection-over-union (2D IoU), horizontal IoU (H IoU), slope difference (ΔSlope), and centroid distance (Δ Centroid). Specifically, 2D IoU is used to evaluate the spatial overlap ratio between the bounding boxes of two fitted lines, as defined in Equation (4). When 2D IoU equals zero—indicating no intersection—H IoU is employed to assess the degree of horizontal overlap between the line segments, as defined in Equation (5). In addition to these geometric criteria, ΔSlope and ΔCentroid are applied as further constraints to determine whether the lines should be considered part of the same seedling row. By combining these four factors, the strategy effectively merges multiple lines fitted to discontinuous mask regions of the same rice seedling row instance into a single unified line.

$$\mathrm{IoU}_{ij}=\frac{\max(0,\min(x_{Ri},x_{Rj})-\max(x_{Li},x_{Lj}))\cdot\max(0,\min(y_{Bi},y_{Bj})-\max(y_{Ui},y_{Uj}))}{w_ih_i+w_jh_j-\max(0,\min(x_{Ri},x_{Rj})-\max(x_{Li},x_{Lj}))\cdot\max(0,\min(y_{Bi},y_{Bj})-\max(y_{Ui},y_{Uj}))} \quad (4)$$

$$\mathrm{H\ IoU}_{ij}=\frac{\max(0,\min(x_{Ri},x_{Rj})-\max(x_{Li},x_{Lj}))}{w_i+w_j-\max(0,\min(x_{Ri},x_{Rj})-\max(x_{Li},x_{Lj}))} \quad (5)$$

where,

$$x_{Lk}=x_k,\quad x_{Rk}=x_k+w_k \quad (6)$$

$$y_{Uk}=y_k,\quad y_{Bk}=y_k+h_k \quad (7)$$

Equations (6) and (7) are used to compute the 2D IoU and H IoU, respectively. The coordinates in the equations correspond to the expanded bounding boxes of the fitted lines, denoted as $B_k=[x_k,\ y_k,\ w_k,\ h_k]$, where x_{Lk} and x_{Rk} represent the left and right x-coordinates, and y_{Uk} and y_{Bk} represent the upper and lower y-coordinates of the bounding box, respectively.

Following the merging step, a direction-based filtering process is applied to eliminate horizontally aligned noise lines that do not contribute to actual seedling row structures. To ensure that only one optimal fitted line is retained for each rice seedling row instance, we conduct a second-stage deduplication on the basis of centroid horizontal distance and angular difference. Two fitted lines are considered to belong to the same seedling row if the horizontal centroid distance and angular deviation fall below predefined thresholds. Among all candidates satisfying this condition, the line with the largest instance region coverage is selected as the final fitted line. This two-stage refinement considerably reduces the probability of identity switching and false association during cross-frame tracking, enhancing the temporal consistency of the tracking results. The computational complexity of the proposed IoU-Plus merging strategy scales linearly with the number of initially fitted lines, making it efficient for typical field scenarios. In our experimental dataset, which featured an average of 6.13 rows per frame, the strategy effectively reduced false positive lines by over 95% (as shown in Table 1). Its reliance on geometric constraints—such as slope consistency, centroid proximity, and bounding box overlap—suggests that the method would remain effective as long as a minimum inter-row spacing is maintained.

The proposed merge-and-deduplication strategy is designed to handle discontinuities in seedling row masks caused by missing seedlings. It effectively integrates disconnected segments as long as they maintain geometric coherence, characterized by similar slope and spatial proximity (as quantified by the IoU-Plus metrics). Our dataset includes a variety of gap types, from common sparse gaps to localized large gaps (e.g., caused by bird damage). The strategy demonstrates robustness in these scenarios. However, it operates under the assumption that discontinuities are local. In cases of very large gaps (e.g., exceeding 50 cm in the field) that break the geometric coherence of a row, the strategy may fail to associate the disjointed segments into a single line. Such global row breaks represent a boundary condition for the current instance identification module, as the problem scope extends into inferring the overall row structure, which falls beyond its current perceptual objective. For such global row breaks, the problem may extend beyond instance identification and require integration with higher-level path planning algorithms to infer the overall row structure.

2.5 Rice seedling row multi-object tracking

The multi-object tracking module maintains consistent identities of rice seedling rows across frames. While deep learning trackers excel in handling occlusions for generic objects, they depend on rich appearance features and substantial computational resources[45]. By contrast, tracking rice seedling rows—which are geometrically structured lines with minimal appearance variation—favors the classical combination of Kalman filtering and the Hungarian algorithm. The Kalman filtering offers a computationally efficient and interpretable motion model for predicting line states (centroid, slope), while the Hungarian algorithm solves the global assignment problem using a cost matrix based on geometric consistency (e.g., IoU, centroid distance, slope difference). This pipeline is well-suited to the structured paddy field environment and supports real-time performance on embedded systems. Nevertheless, its reliance on motion and geometric data alone can limit performance under prolonged, heavy occlusion, where appearance or long-term temporal cues could provide complementary robustness.

2.5.1 Kalman filtering

The multi-object tracking module is designed to maintain consistent identities of rice seedling rows across video frames. While recent deep learning-based multi-object tracking (MOT) frameworks (e.g., ByteTrack) have shown strong performance in generic object tracking, they are predominantly designed for bounding box-based state representations[46]. Adapting them to effectively track the state of line segments, as required in our application, would necessitate non-trivial architectural redesign. In contrast, the combination of Kalman filtering and the Hungarian algorithm offers high interpretability, low computational cost, and has proven sufficient performance in structured agricultural scenes like paddy fields, where motion is relatively predictable[47]. Therefore, this classical paradigm was adopted to meet the real-time and robustness requirements of our system. The Kalman filtering is a widely used prediction algorithm in multi-object tracking tasks. It provides a simple yet effective motion model for predicting line states (centroid, slope) frame-to-frame, while the Hungarian algorithm efficiently solves the optimal assignment problem given the geometric cost matrix, making this combination suitable for real-time tracking of structured objects like rows. It leverages prior and posterior estimation errors to predict and update the system's covariance matrix iteratively. By continuously observing and processing measurements in a sequence, the Kalman filtering can accurately estimate the system's state variables and stably output the parameters of each seedling row line. The algorithm is computationally efficient and capable of real-time, online parameter updates. The Kalman filtering consists of two main stages: prediction and update. Under a constant velocity motion model, the filter first predicts the current system state using the previous state and the state transition matrix. It then updates this prediction by comparing the estimated value with the newly observed measurement, thereby refining the output and producing the optimal estimates of the seedling row line's centroid and parameters. The mathematical formulation of the prediction and update processes is presented as follows:

$$x_k^-=Ax_{k-1} \quad (8)$$

$$P_k^-=AP_{k-1}A^{\mathrm{T}}+Q \quad (9)$$

$$K_k=P_k^-H^{\mathrm{T}}(HP_k^-H^{\mathrm{T}}+R)^{-1} \quad (10)$$

$$x_k=x_k^-+K_K(z_k-Hx_k^-) \quad (11)$$

$$P_k = (I - K_k H)P_k^- \tag{12}$$

where, x_k^- and x_k represent the predicted and updated state vectors at time step k, respectively; x_{k-1} represents the state vector at time step $k-1$; H is the observation matrix that maps the system state to the measurement space; P_k^- and P_k denote the predicted and updated covariance matrices of the estimation error at time step k, respectively, while P_{k-1} denotes the covariance matrix at time step $k-1$; Q and R denote the process noise covariance and the observation noise covariance matrices, respectively; K_k is the Kalman gain matrix; z_k is the observation vector at time k; and I denotes the identity matrix.

The matrix A is the state transition matrix used to project the previous state forward in time, as defined in Equation (13):

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{13}$$

2.5.2 Hungarian algorithm

The Hungarian algorithm is a combinatorial optimization method that solves assignment problems in polynomial time. Its core principle is to find the optimal matching in a bipartite graph by iteratively searching for augmenting paths. In this study, the Hungarian algorithm is applied during the tracking stage to solve the inter frame multi-object association problem. Specifically, rice seedling row lines detected in each frame are matched with those from the previous frame to enable continuous tracking of seedling row instances over time. Three factors are considered to construct the cost matrix for matching: the intersection-over-union (IoU) between bounding boxes, the normalized centroid distance, and the slope difference. The calculation formula is provided in Equation (14). The Hungarian algorithm is then used to find the globally optimal one-to-one matching in the cost matrix, ensuring that each rice seedling row line maintains a consistent ID across frames while minimizing the total matching cost.

$$C_{ij} = \beta_1(1 - \mathrm{IoU}(B_i, B_j')) + \beta_2 \frac{\|\mu_i - \mu_j'\|}{W} + \beta_3 |m_i - m_j'| \tag{14}$$

where, β_1, β_2, β_3 represent the weights assigned to IoU, normalized centroid distance, and slope difference, respectively. Among them, β_1=0.5, β_2=0.3, β_3=0.2, and these values were determined through grid search on the validation set to maximize MOTA; B_i and B_j' denote the bounding boxes of the predicted line from the tracker and the newly fitted line; μ_i and μ_j' are the centroids of the tracker line and the newly fitted line, respectively; W is the width of the image frame; and m_i and m_j' are the slopes of the tracker line and the newly fitted line, respectively.

The multi-object tracking module uses the Kalman filtering to predict and update the state of each rice seedling row line continuously, ensuring stable and accurate outputs of its centroid and parameters. To manage the state of each tracked line, the module defines three tracking states: initial, active, and inactive. The structure of this tracking state machine is illustrated in Figure 6. In the initial state, a newly fitted seedling row line is added to the tracker and prepared for association. Once a new detection is successfully matched with a predicted tracker line, the corresponding state is updated, and the line enters the active state. The Kalman filtering continuously outputs predictions, and successful matches update the state and retain the same instance ID. Seedling row lines in the active state can be tracked stably across frames using ID and color. If a seedling row line fails to be matched for three consecutive frames, it transitions to the inactive state and is removed from tracking once the maximum number of missed frames is exceeded. This mechanism ensures that even in the presence of missed detections or false positives, the tracker maintains correct outputs for a short period and reestablishes stable tracking when new detections appear in subsequent frames. By integrating Kalman filtering for prediction and the Hungarian algorithm for optimal matching, the proposed module ensures accurate and temporally consistent tracking of seedling row lines fitted from pixel-level instance masks. It supports robust multi-row tracking under dynamic scenarios, maintaining stable identity association and parameter continuity. This mechanism of discarding tracks after a short, fixed number of missed detections (three frames in our implementation) effectively prevents long-term drift of tracker identities by promptly cleaning up lost tracks. It represents a design choice that prioritizes responsiveness and robustness against false positives in the short term. However, this strategy also precludes long-term re-identification (re-ID), as a seedling row that re-enters the field of view after an extended occlusion will be assigned a new identity. This limitation is expected to occur under prolonged occlusion, severe segmentation failure, abrupt camera motion, or when a seedling row temporarily leaves and later re-enters the field of view. To mitigate this limitation in future versions, integrating more sophisticated re-association techniques, such as leveraging appearance embeddings from the segmentation backbone or employing trajectory forecasting, is a high-priority development goal.

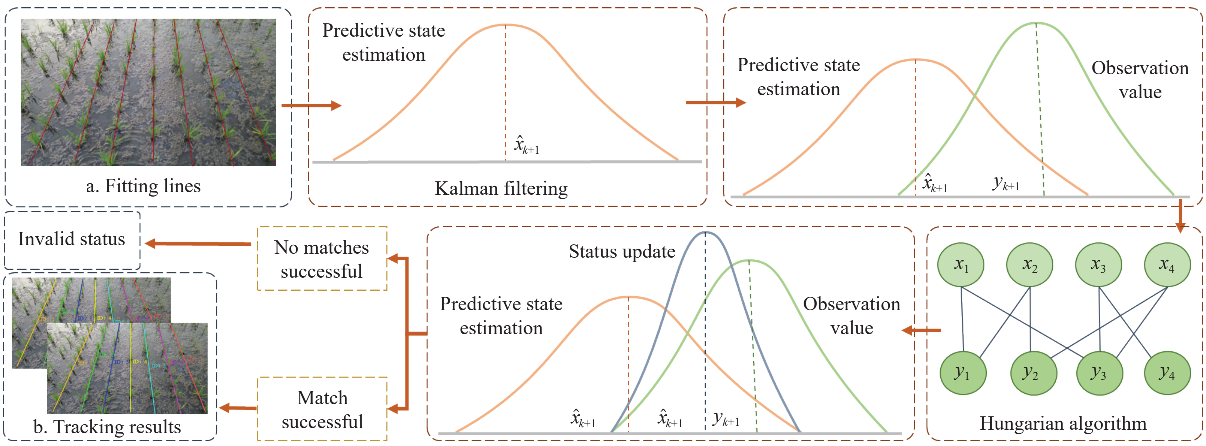

Figure 6 Framework of the multi-object tracking module for rice seedling rows

To mitigate identity drift, tracks are terminated after three consecutive unmatched frames, effectively removing “ghost” tracks and ensuring short-term stability. However, this prevents long-term re-identification: any seedling row lost due to temporary occlusion will be assigned a new ID upon reappearance, causing identity switches. This reflects a fundamental trade-off between short-term stability and long-term association in our current design.

3 Results and discussion

The proposed RTrack-DL model is a dynamic scene-oriented framework for instance identification and continuous tracking of rice seedling rows. It primarily consists of two key components: seedling row line fitting based on segmentation results and continuous multi-object tracking across video frames. Therefore, the performance of the model is evaluated from two perspectives: line fitting accuracy and multi-object tracking performance.

3.1 Experimental environment and DeepLabv3+ model training configuration

The experiments in this study were conducted on a workstation equipped with an NVIDIA GeForce RTX 3090 GPU, an AMD Ryzen 9 5950X 16-core processor, and 128 GB of RAM. The software environment was based on Python 3.8, PyTorch 1.12.1, and CUDA 11.8.0, running on Windows 10 (version 22H2). The resulting DeepLabv3+ model with MobileNetV2 backbone contains 5.81 million parameters, with its checkpoint file occupying approximately 22.4 MB of storage. The inference speed of the model was improved by adopting MobileNetV2 as the lightweight backbone for DeepLabv3+. The network was trained from scratch without using any pretrained weights. Initially, the backbone was trained through a classification task to obtain stable weight initialization, which was subsequently used for training the full segmentation model. During training, the initial learning rate was set to 0.007, and the batch size was set to 4. The weights were updated using stochastic gradient descent with a weight decay of 0.0001. The main loss function was dice loss, while cross entropy was used as the auxiliary loss function. The learning rate was decayed once every 10 epochs, and the model was trained for 300 epochs in total to ensure sufficient convergence. The optimal checkpoint obtained from training was used in subsequent modules as a stable and robust source of segmentation input, which improved the overall system performance in terms of accuracy and real-time capability.

3.2 Visualization of rice seedling row segmentation results

During the training of the DeepLabv3+ model, as the number of epochs increased, the loss value gradually stabilized and ultimately converged to below 0.21. The overall pixel accuracy reached 93.57%, indicating strong convergence performance. To evaluate segmentation performance across different conditions, we randomly selected two image samples from each weather scenario—sunny, cloudy, and overcast—for testing and visualization. Representative visual results are shown in Figure 7. As observed from the segmentation visualizations, the model demonstrates clear boundary delineation between seedling rows and the background, achieving accurate pixel-level identification. This result indicates that the model effectively captures high-level semantic features and long-range contextual information. Even in challenging cases with missing or irregular seedling rows, the model is still capable of producing mostly connected segmentation outputs. Only a few discontinuities appear at the image edges, where segmentation masks become fragmented due to boundary artifacts. To mitigate this issue during the line-fitting stage, we introduced a merge-and-deduplication strategy to consolidate line segments belonging to the same seedling row. As a result, these minor segmentation discontinuities did not affect the effectiveness of subsequent multi-object tracking.

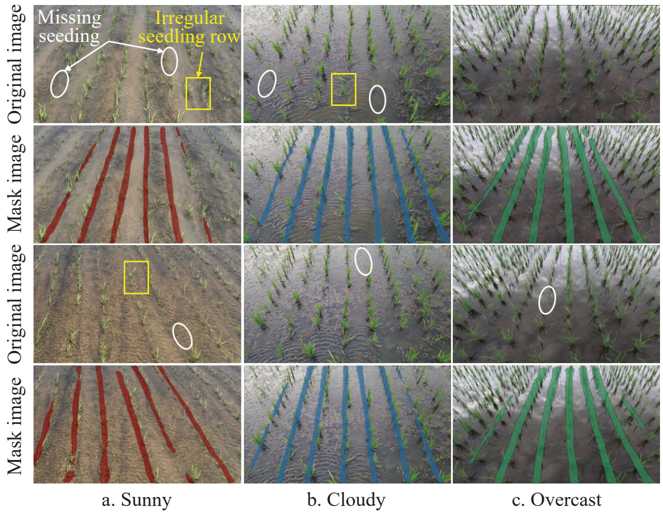

Figure 7 Rice seedling row segmentation mask images under different weather conditions

3.3 Evaluation of rice seedling row line-fitting performance

To evaluate the line-fitting performance of the proposed RTrack-DL model, we designed a comparative experiment by assessing the difference between automatically identified seedling row lines and manually annotated ground truth (GT) lines. Two evaluation metrics were adopted under different weather conditions: the rice seedling row identification rate and the angle error ($\Delta\theta$). A predicted seedling row line is deemed accurate if it satisfies two conditions: its Mean Perpendicular Error (MPE) must be less than 14 pixels (approximately one-fifth of the average row width), and the absolute angular deviation ($|\Delta\theta|$) from the ground truth must be less than 3°. The MPE threshold of 14 pixels corresponds to approximately 2.5 cm on the ground plane, given the average camera height of 0.9 m and image resolution. The angular error threshold of 3° is deemed sufficient for navigation purposes in relatively straight rows. This metric is calculated by averaging the vertical distances between corresponding points along the predicted and true lines at each image row. The angle error is computed by measuring the angular difference between the predicted line and the GT line, which is typically expressed in the general form $Ax+By+C=0$. MPE and angle error are jointly considered to determine whether a seedling row line is correctly fitted. The fitting accuracy is evaluated by computing these values over N discrete points on the predicted seedling row line. The detailed formulas are presented in Equations (15) and (16):

$$\mathrm{MPE}=\frac{1}{N}\sum_{i=1}^{N}\frac{|Ax_i+By_i+C|}{\sqrt{A^2+B^2}} \tag{15}$$

$$\Delta\theta=\left|\tan^{-1}\left(\frac{y_2-y_1}{x_2-x_1}\right)-\tan^{-1}\left(\frac{y_2'-y_1'}{x_2'-x_1'}\right)\right|\cdot\frac{180}{\pi} \tag{16}$$

where, (x_1,y_1), (x_2,y_2) are the pixel coordinates of the two endpoints of the recognized rice seedling row line; (x_1',y_1') and (x_2',y_2') are the pixel coordinates of the two endpoints of the real seedling row line, respectively.

Experiments were conducted under three different weather conditions, with 460 images selected for each condition. The proposed line-fitting method was compared with two baseline approaches: the traditional K-Means method[48,49] and the recently proposed CAROLIF method[50]. In the K-Means method, path fitting was performed using the least squares algorithm. The statistical results of the comparative evaluation are summarized in Table 1. Under sunny conditions, the proposed method achieved a seedling row line-fitting accuracy of 98.34%, outperforming the K-Means method (79.37%) and the CAROLIF method (96.93%). Under overcast conditions, the fitting accuracy reached 98.92%, again exceeding that of K-Means (81.93%) and CAROLIF (97.03%). The highest fitting accuracy was achieved under cloudy conditions, reaching 99.42%, compared with 82.76% for K-Means and 97.41% for CAROLIF. The proposed method also resulted in a considerably lower average angle error of 0.49°, compared with 2.73° for K-Means and 1.86° for CAROLIF. The differences in fitting performance across weather conditions are primarily attributed to varying illumination levels. The superior performance under cloudy conditions is likely attributed to the soft, diffuse illumination which minimizes harsh shadows and specular reflections (common in sunny conditions) and provides better contrast than the low-light, low-contrast scenario of overcast conditions. By contrast, the performance differences among the methods are mainly due to their sensitivity to missing seedlings and irregular row patterns. The CAROLIF method is more susceptible to performance degradation under such conditions, followed by K-Means. Although illumination has some influence on the performance of the proposed method, its overall fitting accuracy remains consistently high across all scenarios.

In the K-Means method, discontinuities in seedling row instance regions often lead to the generation of redundant lines. Furthermore, the fitted lines show notable deviations from the GT and are highly susceptible to disturbances caused by missing seedlings and irregular row patterns. The CAROLIF method segments seedling rows based on color and clusters the extracted pixels before fitting lines using the RANSAC algorithm. Compared with K-Means, CAROLIF achieves remarkably better line-fitting accuracy. However, it still suffers from the issue of producing redundant lines because of its limited handling of disjointed seedling row instances. By contrast, the proposed method incorporates a merge-and-deduplication strategy that effectively mitigates the problem of overfitting multiple lines to the same seedling row. This incorporation ensures that only one line is fitted per instance, resulting in more accurate and consistent outputs. The seedling row line-fitting results of all three methods under different weather conditions are visually compared in Figure 8. It is important to note that the evaluated illumination conditions cover the primary daylight operational window. Performance under extreme illumination scenarios, such as dusk, night, or strong water glare, falls outside the current operational envelope of typical vision-based agricultural robots and remains an open boundary for the model's applicability.

Table 1 Comparison of seedling row line-fitting results under different methods and weather conditions

Methods	Scenarios	Line-fitting accuracy/%	Max	Mean	SD	RMSE
			Angle error/(°)			
K-Means + LSM	Sunny	79.37	14.17	2.96	3.07	3.90
	Cloudy	82.76	13.94	2.73	2.80	3.79
	Overcast	81.93	14.01	2.87	2.96	3.87
CAROLIF	Sunny	96.93	11.12	2.04	2.11	2.54
	Cloudy	97.41	10.71	1.86	1.89	2.03
	Overcast	97.03	11.06	1.93	2.05	2.11
Ours	Sunny	98.34	7.73	0. 67	0.64	0.95
	Cloudy	99.42	7.58	0.49	0.51	0.71
	Overcast	98.92	7.61	0.61	0.57	0.84

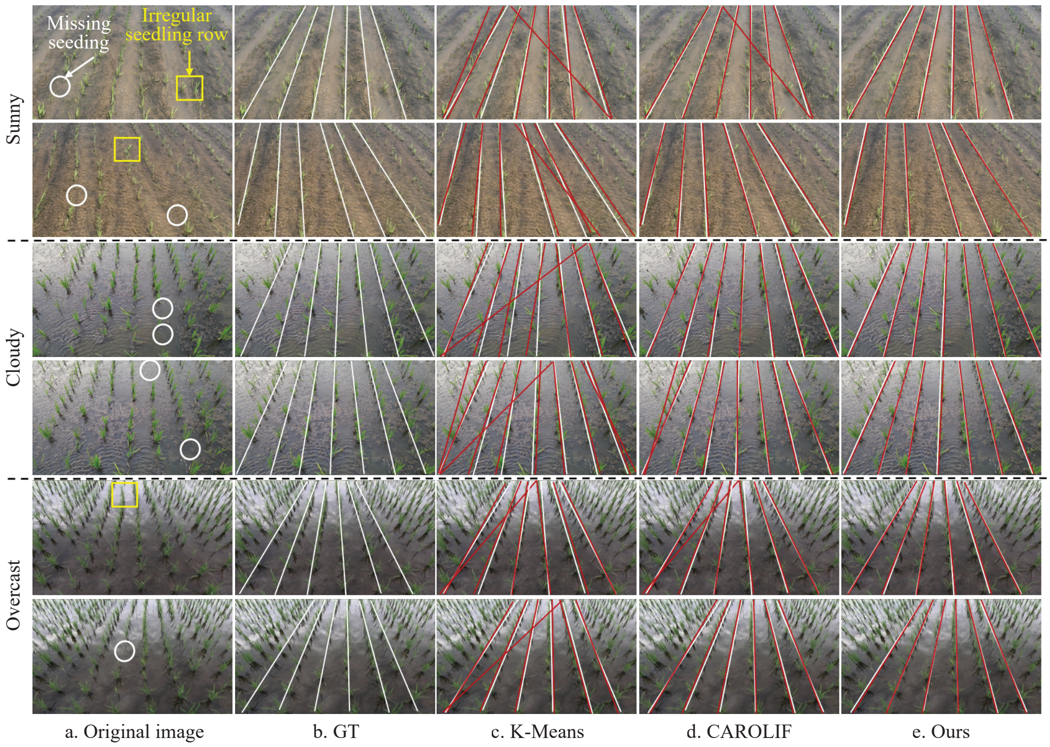

Figure 8 Comparison of seedling row line-fitting results using different methods

3.4 Ablation study on rice seedling row line fitting

An ablation study was conducted to investigate further the effectiveness of line fitting in the proposed RTrack-DL model. The experimental results are summarized in Table 2. A total of 465 frames were selected for the experiment, including 155 frames for each of the three weather conditions. Among these, 93 frames contained five seedling rows, 217 frames contained six rows, and 155 frames contained seven rows, resulting in an average of 6.13 fitted lines per frame. Given the missing seedlings during segmentation, certain seedling row mask regions—particularly those located near the left and right edges of the image—appeared discontinuous. This discontinuity often led to the generation of multiple fitted lines for a single seedling row instance. To address this issue and reduce the impact of irregular seedling row layouts, we applied a merge-and-deduplication strategy after line fitting. This strategy considerably reduced the number of redundant lines fitted to the same instance. A key component of this strategy is the IOU-Plus merging module, which jointly considers four constraints: 2D IoU, H IoU, ΔSlope, and Δ Centroid. Additionally, a second-stage deduplication process was applied to ensure that only the most optimal line is retained for each seedling row instance, further enhancing the robustness of the fitting results and complementing the IoU-Plus module.

In Table 2, the metric AvgLines represents the average number of fitted lines per frame; values closer to the true number of seedling rows indicate better performance. FP refers to the number of redundant lines fitted due to discontinuities in instance regions. FAF denotes the average number of false fitted lines per frame, while Frag counts the number of fitting failures caused by broken or irregular seedling row masks. After incorporating the merge-and-deduplication strategy into the line-fitting module, the fitting performance of the model improved remarkably. The number of redundant lines caused by missing seedlings and irregular row patterns was substantially reduced. The AvgLines value decreased to 6.14, reflecting a 4.66% reduction compared with the baseline without the strategy, and is nearly identical to the true value of 6.13, with only a 0.16% error. In addition, FAF dropped from 30.54% to 1.49%. This result demonstrates the strategy's effectiveness in selecting the optimal fitted centerline by applying multilevel constraints on candidate lines within the same rice seedling row instance. The visual comparison of ablation results is shown in Figure 9. The ablation study demonstrates consistent and substantial performance gains with the incremental addition of each component in the merge-and-deduplication strategy. It is important to note that the evaluation was conducted on a fixed test set using deterministic algorithms. Therefore, while the improvements reported in Table 2 are clear and consistent across the evaluated samples, formal statistical significance testing (e.g., using t-tests and reporting confidence intervals) was not performed in this study. Incorporating such rigorous statistical validation will be a standard practice in our future comparative studies to further substantiate the findings.

Table 2 Ablation results of seedling row line fitting under different module configurations

Name		AvgLines	FP	FAF/%	Frag
A	Baseline	6.44	142	30.54	164
B1	Baseline +2D IoU+H IoU	6.36	107	23.01	127
B2	Baseline +2D IoU+H IoU +ΔSlope	6.26	57	12.26	83
B3	Baseline +2D IoU+H IoU +ΔCentroid	6.25	53	11.40	77
B4	Baseline+ΔSlope+ΔCentroid	6.19	26	5.59	38
C	Baseline + IoU-Plus	6.15	8	1.67	26
D	Baseline + IoU-Plus+ Double deduplication	6.14	7	1.49	19

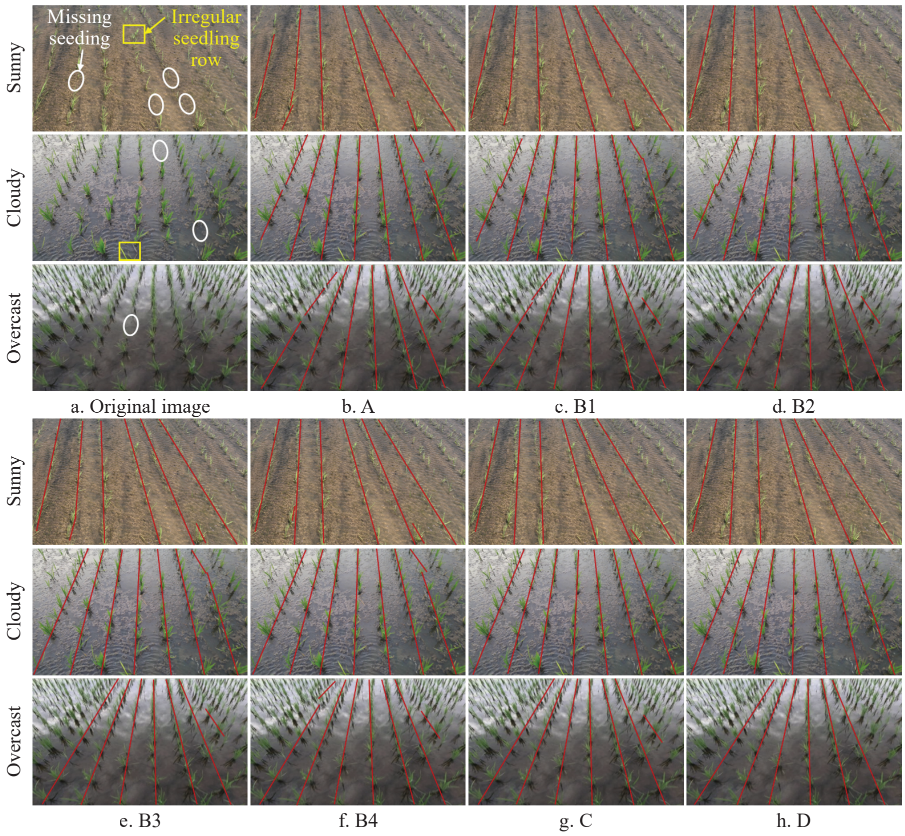

Figure 9 Visual comparison of rice seedling row line-fitting results in the ablation study

3.5 Evaluation of rice seedling row line tracking performance

To evaluate the continuous multi-object tracking performance of the RTrack-DL model, we selected three video sequences for testing, corresponding to sunny, cloudy, and overcast weather conditions. Each video had a duration of 90 s, comprising 2250 frames. The tracking results for each scenario are listed in Table 3.

Table 3 Experimental results of seedling row line tracking using the RTrack-DL model

Scenarios	MOTA/%	MOTP/%	IDF1/%	FAF/%	MT/%	ID sw	Frag	FPS
Sunny	92.1	92.6	92.4	1.67	93.6	191	237	17.96
Cloudy	93.5	94.7	94.8	1.43	94.3	117	138	18.72
Overcast	92.7	93.3	93.9	1.52	94.0	169	195	18.43

In Table 3, multiple-object tracking accuracy (MOTA) reflects the overall tracking accuracy in terms of correct identification of the number and attributes of rice seedling rows. Multiple-object tracking precision (MOTP) evaluates the spatial precision of the predicted seedling row lines. IDF1 measures the ability of the tracker to associate and maintain consistent IDs across frames correctly. MT indicates the number of GT seedling row trajectories that are successfully tracked for at least 80% of their lifespan. Higher values of MOTA, MOTP, IDF1, and MT indicate better tracking performance. By contrast, ML represents the number of GT seedling row trajectories that are mostly lost, meaning the object was not tracked for at least 20% of its lifetime. ID switches (ID sw) count the number of times an object's ID was reassigned during tracking. Lower values of ML and ID sw indicate better tracking stability. These eight metrics together reflect the overall ability of the algorithm to perform continuous and accurate multi-object tracking of seedling row lines. FPS (frames per second) represents the average inference speed and serves as an indicator of the model' s real-time performance.

The proposed RTrack-DL model achieved an average MOTA of 92.77%, with a maximum of 93.5% and a minimum of 92.1% across all video sequences. The MOTP remained consistently around 93%, and the IDF1 ranged from 92.4% to 94.8%. The average FAF was approximately 1.54%, indicating that the model provides high tracking accuracy and stability across varying conditions. Throughout the tracking process, the model maintained an FPS between 15-20 fps, demonstrating its computational efficiency and capability for real-time, continuous tracking. The visual results of multi-object tracking of rice seedling row lines are illustrated in Figure 10. While our tracking module combines established components (Kalman filtering and Hungarian algorithm), a direct comparison with state-of-the-art deep learning-based MOT frameworks (e.g., ByteTrack, OC-SORT) adapted to line states was not performed due to implementation complexity. This represents a potential avenue for future comparative studies.

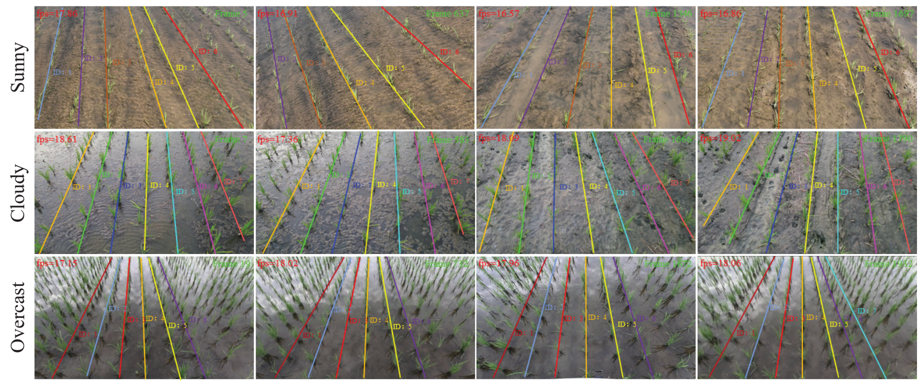

Figure 10 Multi-object tracking results of rice seedling row lines

The reported performance was achieved under conditions simulating stable, low-speed field operations. The drone was flown at a constant, slow speed (<1 m/s) and equipped with a gimbal for stabilization, which mitigated the effects of camera jitter and vibration. Consequently, the model's resilience to more severe motion artifacts or rapid perspective changes remains to be fully evaluated. Regarding real-time capability, the current frame rate of 15-20 fps is adequate for ground robots operating at similarly low speeds (e.g., ≤1 m/s), as it ensures sufficient frame overlap for stable tracking. However, for high-speed platforms such as unmanned aerial vehicles (UAVs) traveling faster than 3 m/s, this frame rate may be insufficient, and further optimization (e.g., model pruning or quantization) would be necessary to maintain tracking accuracy.

4 Conclusions

This study proposed RTrack-DL, a model for instance identification and continuous tracking of rice seedling rows in dynamic field scenarios. The model employs a DeepLabv3+-based semantic segmentation network to extract pixel-level centerlines of rice seedling rows, providing reliable masks for subsequent line fitting and tracking. To address segmentation discontinuities caused by missing seedlings and irregular planting patterns, a merge-and-deduplication strategy guided by the composite IoU-Plus metric is incorporated into the line-fitting module. This strategy consolidates multiple line segments belonging to the same rice seedling row instance into a single optimal line, thereby improving fitting robustness. Furthermore, the model combines Kalman filtering and the Hungarian algorithm to achieve continuous multi-object association and maintain stable rice seedling row identities across consecutive frames. Experimental results demonstrate that RTrack-DL achieves high line-fitting accuracy and stable continuous tracking under common daylight conditions while maintaining real-time performance. Overall, the proposed framework bridges static perception and dynamic tracking, providing reliable visual information for autonomous paddy field machinery.

The current framework still has limitations under extremely challenging illumination conditions, such as dusk and strong glare, and its computational efficiency on high-speed platforms requires further improvement. Future work will focus on integrating complementary sensing modalities, such as near-infrared and LiDAR, to enhance perceptual robustness under adverse conditions. Model quantization and hardware-aware pruning will also be investigated to improve processing speed. In addition, the generalizability of RTrack-DL will be evaluated across different field environments, crop growth stages, and geographic regions.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Young Scientists Fund) [52502530], the

Natural Science Foundation of Jiangsu Province (Youth Fund) [BK20251524], the Jiangsu Provincial Pilot Project for Integrated Research, Development, Manufacturing and Application of Agricultural Machinery [JSYTH2025-05], and the National Natural Science Foundation of China [32572220].

[References]

[1] Dadrasi A, Vačkářová D, Salmani F, Nguyen C T, Weinzettel J. Global rice land suitability and adaptation strategies under climate change. J. Environ. Manage., 2025; 394: 127630.

[2] Kotaniemi J, Känsäkoski N, Halbach E, Heikkilä T. A weeding robot for seedling removal. in: 20th IEEEASME Int. Conf. Mechatron. Embed. Syst. Appl. MESA, 2024; pp.1–8. DOI: 10.1109/MESA61532.2024.10704901.

[3] Chen D S, Wang Q L, Lin Y, Ma Z L, Sun L, Yu G H. Path tracking control of paddy field weeder integrated with satellite and visual methods. Comput. Electron. Agric, 2025; 234: 110257.

[4] Feng L Y, Xu C S, Tang H, Wei Z C, Guan X D, Xu J C, et al. Application of navigation technology in agricultural machinery: A review and prospects. Artif. Intell. Agric, 2026; 16(1): 94–123.

[5] Ju J Y, Chen G Q, Lv Z Y, Zhao M Y, Sun L, Wang Z T, et al. Design and experiment of an adaptive cruise weeding robot for paddy fields based on improved YOLOv5. Comput. Electron. Agric., 2024; 219: 108824.

[6] Nunes F, Sousa F. Deep learning soft-decision GNSS multipath detection and mitigation. Sensors, 2024; 24(14): 4663.

[7] Zhang Y W, Liu M N, Chen D, Xu X M, Lu J B, Lai H R, et al. Development and testing of row‐controlled weeding intelligent robot for corn. J. Field Robotics, 2025; 42(3): 850–866.

[8] Han C Y, Wu W B, Luo X W, Li J H. Visual navigation and obstacle avoidance control for agricultural robots via LiDAR and camera. Remote Sensing, 2023; 15(22): 5402.

[9] Martini M, Ambrosio M, Navone A, Tuberga B, Chiaberge M. Enhancing visual autonomous navigation in row-based crops with effective synthetic data generation. Precis. Agric., 2024; 25: 2881–2902.

[10] Navone A, Martini M, Ambrosio M, Ostuni A, Angarano S, Chiaberge M. GPS-free autonomous navigation in cluttered tree rows with deep semantic segmentation. Robot Auton. Syst., 2025; 183: 104854.

[11] Wang S L, Zhang B K, Zhou J, Yang B, He Y. Time–jerk optimal trajectory planning for industrial robots with coupled interpolation function selection. J Field Robotics, 2024; 41(4): 917–941.

[12] Chakhvashvili E, Machwitz M, Antala M, Rozenstein O, Prikaziuk E, Schlerf M, et al. Crop stress detection from UAVs: best practices and lessons learned for exploiting sensor synergies. Precis. Agric., 2024; 25: 2614–2642.

[13] Shammi S, Diepeveen D, Zander S, Jones M G K, Sohel F. Early frost detection in wheat using machine learning from vertical temperature distributions. Comput. Electron. Agric, 2024; 221: 108950.

[14] Zhang Y X, Shen Y, Wang Q, Song C, Dai N, He B. A novel hybrid swarm intelligence algorithm for solving TSP and desired-path-based online obstacle avoidance strategy for AUV. Robot Auton. Syst, 2024; 177: 104678.

[15] Ariza-Sentís M, Vélez S, Martínez-Peña R, Baja H, Valente J. Object detection and tracking in Precision Farming : A systematic review. Comput. Electron. Agric, 2024; 219: 108757.

[16] Kang S, Fan J L, Ye Y K, Li C L, Du D D, Wang J. Maturity recognition and localisation of broccoli under occlusion based on RGB-D instance segmentation network. Biosyst Eng, 2025; 250: 270–284.

[17] Ma Z H, Yin C, Du X Q, Zhao L J, Lin L P, Zhang G F, Wu C Y. Rice row tracking control of crawler tractor based on the satellite and visual integrated navigation. Comput Electron Agric, 2022; 197: 106935.

[18] Liu W, Hu J P, Liu J X, Yue R C, Zhang T F, Yao M J, et al. Method for the navigation line recognition of the ridge without crops via machine vision. Int J Agric & Biol Eng, 2024; 17(2): 230–239.

[19] Li B Y, Xiao C, Wang L G, Wang Y Q, Lin Z P, Li M, et al. Dense nested attention network for infrared small target detection. IEEE Trans. Image Process, 2023; 32: 1745–1758.

[20] Yang Y, Zhou Y, Yue X, Zhang G, Wen X, Ma B, et al. Real-time detection of crop rows in maize fields based on autonomous extraction of ROI. Expert Syst. Appl, 2023; 213(Part A): 118826.

[21] Gong H L, Zhuang W D, Wang X. Improving the maize crop row navigation line recognition method of YOLOX. Front. Plant Sci, 2024; 15: 1338228.

[22] Mohialden Y M, Hussien N M, Salman S A, Alwahhab A B A, Ali M. Enhancing agriculture crop classification with deep learning. Babylon J Artif Intell, 2024; 2024: 20–26.

[23] Li D F, Li B L, Long S F, Feng H Q, Xi T, Kang S, et al. Rice seedling row detection based on morphological anchor points of rice stems. Biosyst Eng, 2023; 226: 71–85.

[24] Wu S L, Chen Z G, Bangura K, Jiang J, Ma X G, Li J Y, et al. A navigation method for paddy field management based on seedlings coordinate information. Comput. Electron. Agric, 2023; 215: 108436.

[25] He J, Dong W H, Tan Q N, Li J N, Song X W, Zhao R M. A variable-threshold segmentation method for rice row detection considering robot travelling prior information. Agriculture, 2025; 15(4): 413.

[26] Chen Z Y, Cai Y H, Liu Y X, Liang Z P, Chen H, Ma R J, et al. Towards end-to-end rice row detection in paddy fields exploiting two-pathway instance segmentation. Comput. Electron. Agric, 2025; 231: 109963.

[27] Zhang S, Wang Y Y, Zhu Z X, Li Z, Du Y F, Mao E R. Tractor path tracking control based on binocular vision. Inf. Process. Agric., 2018; 5(4): 422–432.

[28] Fu D B, Chen Z Y, Yao Z Q, Liang Z P, Cai Y H, Liu C, et al. Vision-based trajectory generation and tracking algorithm for maneuvering of a paddy field robot. Comput. Electron. Agric, 2024; 226: 109368.

[29] Tao H Q, Zhang R R, Zhang L H, Zhang D Z, Yi T C, Wu M Q. Tea harvest robot navigation path generation algorithm based on semantic segmentation using a visual sensor. Electronics, 2025; 14(5): 988.

[30] Hernandez B, Medeiros H. Multi-object tracking in agricultural applications using a vision transformer for spatial association. Comput Electron. Agric, 2024; 226: 109379.

[31] Han X D, Fang P C, Tian Y Y, Yu J H, Cai X H, Roggen D, et al. GRASPTrack: Geometry-reasoned association via segmentation and projection for multi-object tracking. arXiv: 2508.08117, 2025. DOI: 10.48550/arXiv.2508.08117.

[32] Li H W, Lai X D, Mo Y M, He D Q, Wu T. Pixel-wise navigation line extraction of cross-growth-stage seedlings in complex sugarcane fields and extension to corn and rice. Front. Plant Sci, 2025; 15: 1499896.

[33] Zhai W G, Li C C, Fei S P, Liu Y H, Ding F, Cheng Q, et al. CatBoost algorithm for estimating maize above-ground biomass using unmanned aerial vehicle-based multi-source sensor data and SPAD values. Comput. Electron. Agric, 2023; 214: 108306.

[34] Guo Z M, Geng Y H, Wang C, Xue Y, Sun D, Lou Z X, et al. InstaCropNet: An efficient Unet-based architecture for precise crop row detection in agricultural applications. Artif. Intell. Agric, 2024; 12: 85–96.

[35] Zhang X Y, Lu X Y, Zhang Z Q, Yang G F, He Y, Fang H. Simultaneous detection of reference lines in paddy fields using a machine vision-based framework. Comput. Electron. Agric, 2024; 221: 108923.

[36] Yu C H, Feng J Z, Zheng Z Z, Guo J P, Hu Y H. A lightweight SOD-YOLOv5n model-based winter jujube detection and counting method deployed on Android. Comput. Electron. Agric, 2024; 218: 108701.

[37] Chen L C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. 2018; DOI: 10.48550/arXiv.1802.02611.

[38] Kalman R E. A new approach to linear filtering and prediction problems. J. Basic Eng, 1960; 82: 35–45.

[39] Rapado-Rincón D, van Henten E J, Kootstra G. MinkSORT: A 3D deep feature extractor using sparse convolutions to improve 3D multi-object tracking in greenhouse tomato plants. Biosyst. Eng, 2023; 236: 193–200.

[40] Kuhn H W. The Hungarian method for the assignment problem. Nav. Res. Logist, 1955; 2: 83–97.

[41] Jang M, Hwang Y. Tomato pose estimation using the association of tomato body and sepal. Comput. Electron. Agric, 2024; 221: 108961.

[42] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. MobileNetV2: Inverted residuals and linear bottlenecks. in: 2018 IEEECVF Conf. Comput. Vis. Pattern Recognit, 2018; pp.4510–4520. DOI: 10.1109/CVPR.2018.00474.

[43] Pal A, Leite A C, From P J. A novel end-to-end vision-based architecture for agricultural human–robot collaboration in fruit picking operations. Robot. Auton. Syst, 2024; 172: 104567.

[44] Guerrero J M, Guijarro M, Montalvo M, Romeo J, Emmi L, Ribeiro A, et al. Automatic expert system based on images for accuracy crop row detection in maize fields. Expert Syst Appl, 2013; 40: 656–664.

[45] Chen J, Wu J, Wang Z, Qiang H, Cai G, Tan C, et al. Detecting ripe fruits under natural occlusion and illumination conditions. Comput. Electron. Agric, 2021; 190: 106450.

[46] Zhang Y A, Wang T C, Zhang X Y. MOTRv2: Bootstrapping end-to-end multi-object tracking by pretrained object detectors. arXiv: 2211.09791, 2023. DOI: 10.48550/arXiv.2211.09791.
[47] Zheng Z Y, Li J W, Qin L F. YOLO-BYTE: An efficient multi-object tracking algorithm for automatic monitoring of dairy cows. Comput. Electron. Agric., 2023; 209: 107857.
[48] Jiang G Q, Wang Z H, Liu H M. Automatic detection of crop rows based on multi-ROIs. Expert Syst Appl, 2015; 42(5): 2429–2441.
[49] Zhang S L, Ma Q L, Cheng S K, An D, Yang Z L, Ma B, et al. Crop row detection in the middle and late periods of maize under sheltering based on solid state LiDAR. Agriculture, 2022; 12(12): 2011.
[50] Khan M N, Rahi A, Rajendran V P, Al Hasan M, Anwar S. Real-time crop row detection using computer vision- application in agricultural robots. Front. Artif. Intell, 2024; 7: 1435686.