frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 01 December 2025
DOI 10.3389/fpls.2025.1706596

OPEN ACCESS

EDITED BY
Chengquan Zhou,
Zhejiang Academy of Agricultural Sciences, China

REVIEWED BY
Dawei Sun,
Zhejiang Academy of Agricultural Sciences, China
Jiandong Hu,
Henan Agricultural University, China
Kong Xiangyu,
South China Agricultural University, China

*CORRESPONDENCE
Long Qi
qilong@scau.edu.cn
Ruijun Ma
ruijunma@scau.edu.cn

RECEIVED 16 September 2025
REVISED 23 October 2025
ACCEPTED 17 November 2025
PUBLISHED 01 December 2025

CITATION
Feng B, He Q, Hu Y, Cai H, Luo D, Shen Z, Zhang B, Qi L and Ma R (2025) ALNet: towards real-time and accurate maize row detection via anchor-line network. *Front. Plant Sci.* 16:1706596. doi: 10.3389/fpls.2025.1706596

ALNet: towards real-time and accurate maize row detection via anchor-line network

Bofeng Feng[1], Qingliang He[1], Yun Hu[2], Hao Cai[1], Dongyu Luo[1], Zhiye Shen[1], Bob Zhang[3], Long Qi[4,5,6,7]* and Ruijun Ma[1]*

[1]College of Engineering, South China Agricultural University, Guangzhou, China, [2]School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China, [3]Department of Computer and Information Science, University of Macau, Macao, Macao SAR, China, [4]College of Water Conservancy and Civil Engineering, South China Agricultural University, Guangzhou, China, [5]Guangdong Engineering Technology Research Center of Rice Transplanting Mechanical Equipment, Guangzhou, China, [6]State Key Laboratory of Agricultural Equipment Technology, Guangzhou, China, [7]Department of Biosystems Engineering, University of Manitoba, Winnipeg, MB, Canada

Accurate and efficient crop row detection is essential for the visual navigation of agricultural machinery. However, existing deep learning–based methods often suffer from high computational costs, limited deployment capability on edge devices, and difficulty in maintaining both accuracy and speed. This study presents ALNet (Anchor-Line Network), a lightweight convolutional neural network tailored to the elongated geometry of maize rows. ALNet introduces an Anchor-Line mechanism to reformulate row detection as an end-to-end regression task, replacing pixel-wise convolutions with row-aligned kernel operations to reduce computation while preserving geometric continuity. An Attention-guided ROI Align module equipped with a Dual-Axis Extrusion Transformer (DAE-Former) is incorporated to capture global–local feature interactions and enhance robustness under challenging field conditions such as weed infestation, low light, and wind distortion. In addition, a Row IoU (RIoU) loss is designed to improve localization accuracy by aligning predicted and ground-truth row geometries more effectively. Experimental results on field-acquired maize datasets demonstrate that ALNet achieves an *mF*1 of 59.60 across IoU thresholds ($\geq$ 9.24 points higher than competing methods) and an inference speed of 161.26 FPS, with a computational cost of only 11.9 GFlops, demonstrating potential for real-time edge deployment. These advances establish ALNet as a practical and scalable solution for intelligent visual navigation in precision agriculture.

KEYWORDS

corn row detection, anchor-line, attention-guided ROI align, dual-axis extrusion transformer, precision agriculture

1 Introduction

1.1 Background

Corn, the largest global crop and a primary staple in China, demands highly effective field management to ensure sustainable production Yao et al. (2025). Precision management practices are critical for creating optimal growth conditions, mitigating losses from diseases, pests, and weeds, and improving both yield and quality while reducing operational costs ShaoKun (2017); Zhaosheng (2025). In recent years, machine vision systems have become pivotal tools in modern agriculture, enabling advanced perceptual capabilities for autonomous field operations Abbasi et al. (2022); Rabab et al. (2021).

Among these applications, corn row detection stands as a foundational visual perception task, directly influencing the efficiency and accuracy of agricultural machinery in field management Chen et al. (2025); Diao et al. (2023). However, achieving reliable corn row detection in real-world scenarios remains challenging due to several factors: (1) Illumination variability: Variable lighting, shadows, and insufficient illumination can degrade image quality. (2) Weed interference: Dense weed growth with similar visual characteristics to corn can cause misidentification Chen et al. (2025). (3) Seedling variation: Missing plants Xiangguang (2021) and seedlings at different growth stages reduce detection accuracy. (4) Image blur: Machinery vibration can blur images, impairing detection performance. (5) Wind effects: Strong winds can temporarily lodge corn plants, distorting their visual features.

1.2 Related work

Existing deep learning-based crop row detection methods can be categorized into three types based on the algorithm used: object detection-based methods, semantic segmentation-based methods, and instance segmentation-based methods.

1.2.1 Object detection-based methods

Current corn row detection methodologies typically employ a two-stage approach: plant localization followed by row aggregation. Hongbo (2024) developed a YOLOv8-G variant for corn seedling center detection, implementing affinity propagation clustering and least squares regression for row fitting. Lulu (2024) combined Faster R-CNN with Susan operator feature extraction, utilizing RANSAC algorithms to enhance row fitting precision. Gong and Zhuang (2024) improved YOLOv7-tiny through infrared imaging integration and ShuffleNet v1 optimization, incorporating Coordinate Attention mechanisms and EIOU loss functions to boost localization accuracy. While these approaches demonstrate technical merit, three fundamental limitations persist: (1) overhead imaging perspectives constrain practical field applicability, (2) computational complexity limits deployment on resource-constrained devices, and (3) processing speeds face challenges in meeting real-time agricultural operation requirements.

1.2.2 Semantic segmentation–based methods

Semantic segmentation approaches offer streamlined solutions for crop row detection by directly predicting row structures from pixel-level classifications. Liu et al. (2024) developed a canopy ROI segmentation method coupled with horizontal striping analysis and midpoint clustering for row extraction. Cao et al. (2022) enhanced the ENet architecture through residual connections, implementing an optimized RANSAC algorithm for precise navigation path identification. Ban et al. (2025) combined StemFormer-based segmentation with LiDAR point cloud processing to cluster maize stalk positions and fit row trajectories. Gan et al. (2025) demonstrated the effectiveness of semi-supervised learning through their Unimatch framework, achieving efficient rice seedling segmentation with reduced annotation requirements. While these end-to-end methods eliminate multi-stage processing bottlenecks, they face three persistent challenges: (1) blurred edge delineation in dense canopies, (2) feature confusion in overlapping plant regions, and (3) contextual information loss during high-resolution processing Ronneberger et al. (2015); Takikawa et al. (2019). These limitations ultimately constrain detection accuracy under complex field conditions.

1.2.3 Instance segmentation–based methods

Unlike semantic segmentation, methods based on instance segmentation typically employ a two-stage approach. For instance, they first identify the region where the instances are located, and then perform semantic segmentation within the detected bounding box, outputting each segmentation result as a separate instance. Wei et al. (2022) developed the RASCM model, reformulating row detection as optimal line-position selection through reinforced attention mechanisms. Bing (2021) proposed a perspective transformation pipeline that converts field images to bird's-eye views before segmenting row-parallelogram masks. Chen et al. (2025) designed a dual-path network architecture that concurrently performs semantic segmentation and feature embedding, enabling simultaneous rice row clustering and trajectory regression. Although such methods offer finer granularity and improved speed over pure semantic segmentation, they still face challenges:(1) Contextual fragmentation: Treating crop rows as isolated entities disregards field-scale spatial relationships. (2)Computational intensity: Dense pixel-wise predictions require 1.8-3.2× more operations than detection-only methods. (3)Accuracy-speed tradeoff: Current implementations struggle to maintain 70% mAP while achieving 100ms latency.

1.3 Limitations and motivation

Despite considerable progress, most deep learning–based crop row detection methods face four major limitations: (1) Loss of holistic context–Treating crops and rows as independent objects reduces accuracy. (2) High computational cost–Pixel-wise prediction is expensive and slows inference. (3) Poor deployability–Heavy models consume excessive memory and are unsuitable for edge devices. (4) Speed–accuracy trade-off–Achieving both remains a challenge for field-scale application.

To address these issues, we propose ALNet (Anchor-Line Network), a deep learning corn row detection network based on an Anchor-Line strategy. This approach departs from conventional pixel-wise convolution by aligning convolutional kernels along the row direction, thereby lowering computational costs. A multi-scale feature extraction module combines deep semantic and shallow spatial cues, improving accuracy, while an Attention-guided ROI Align module recovers global context lost during Anchor-Line convolution. The network regresses entire maize rows as unified targets through a dedicated loss function.

The main contributions of this work are as follows:

- Anchor-Line design: tailored to maize row geometry, enabling end-to-end row extraction.
- Holistic target modeling: entire maize rows are detected and regressed as single entities.
- Edge deployment potential: a computationally efficient framework showing potential for edge device integration.
- Balanced performance: achieves both high accuracy and fast inference for practical production needs.

2 Materials and methods

As shown in Figure 1 the general workflow of our method mainly includes two steps: First, maize images at the 3–5 leaf stage with unfolded leaves are collected and annotated to create a dataset for model training and testing. Second, the dataset is used to train the proposed network. Based on the learned features, the distribution of Anchor-Lines is optimized according to the shape and positional characteristics of maize rows, and an end-to-end detection model is established.

2.1 Dataset acquisition

As shown in Figure 2, the image data used in this study were collected from mechanically sown maize fields at the Jilin University Agricultural Experiment Base, located in Lvyuan District(43° 49.02'N, 125°24.39'E), Changchun City, Jilin Province. Data collection was conducted in June 2024 and June 2025, when the maize was at the 3- to 5-leaf stage. At this growth stage maize roots begin to develop and weeds start to emerge. Maize leaves secrete substantial amounts of corn ketone, which can effectively reduce the toxicity of herbicides. Meanwhile, weeds are typically at the 2- to 4-leaf stage, during which their physiological activity is high and cell division is vigorous, rendering them more susceptible to herbicides. The efficiency of herbicide absorption and translocation is also higher at this stage, enabling more effective weed suppression Agathokleous (2018); Bing (2021).

The images were acquired with an RGB camera mounted on a handheld gimbal, producing frames at a resolution of 1280×720. To facilitate practical deployment of the algorithm, three shooting angles were selected: 60°, 45°, and 30°; here the angle denotes the inclination between the camera optical axis and the horizontal plane. In total, 6000 maize images were recorded. The collected images cover various maize growth states and a wide range of

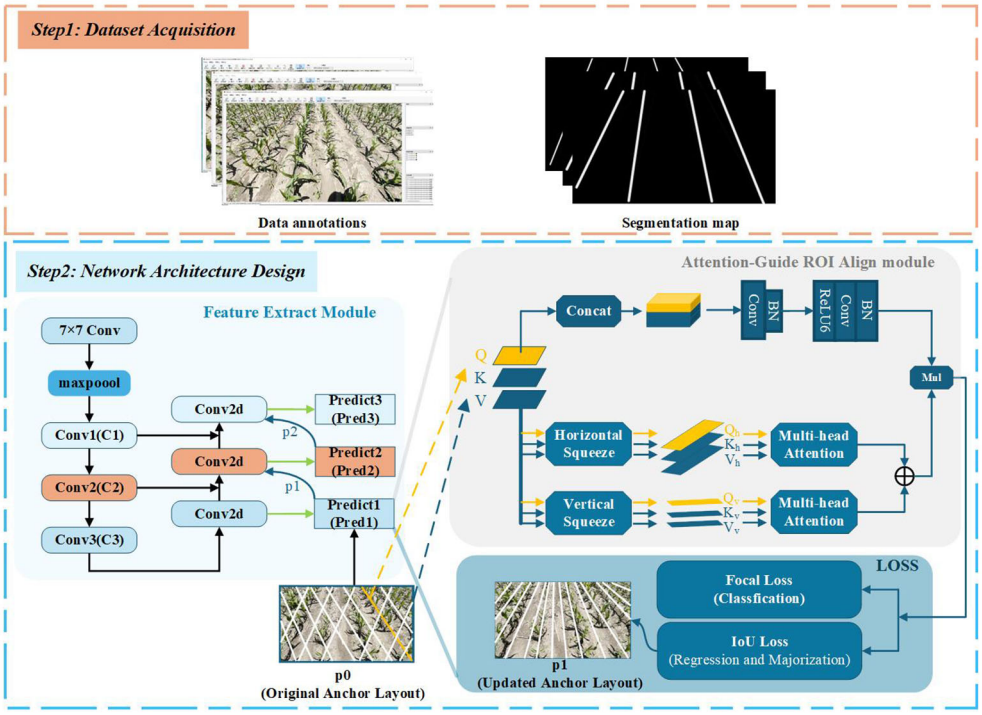

FIGURE 1
Flow diagram of the proposed crop row detection method.

10.3389/fpls.2025.1706596

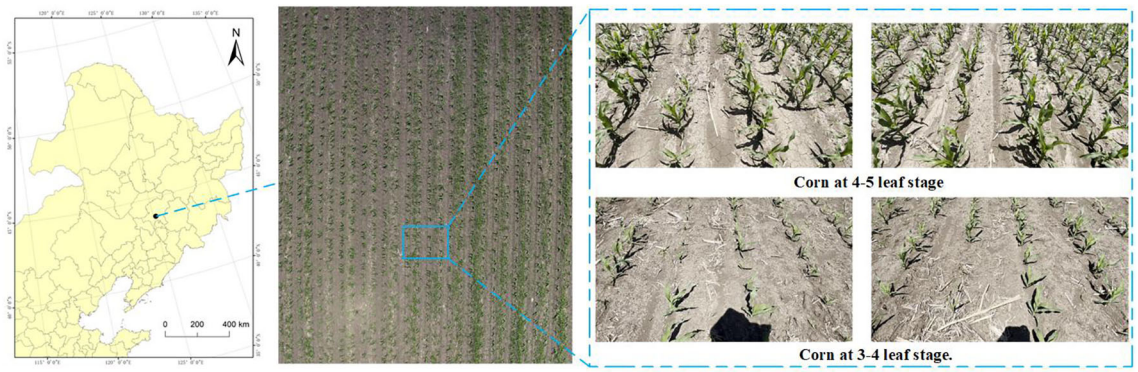

FIGURE 2
Data collection locations and sample images of corn rows.

complex field conditions, including dense weed coverage, missing seedlings, strong and low lighting, motion blur caused by agricultural machinery, and temporary lodging of maize due to strong winds. Several representative samples are shown in Figure 3.

All images were subsequently annotated using the open-source tool "Labelme". Because the entire maize row is treated as the recognition target, annotations were made with polylines that trace each maize row.

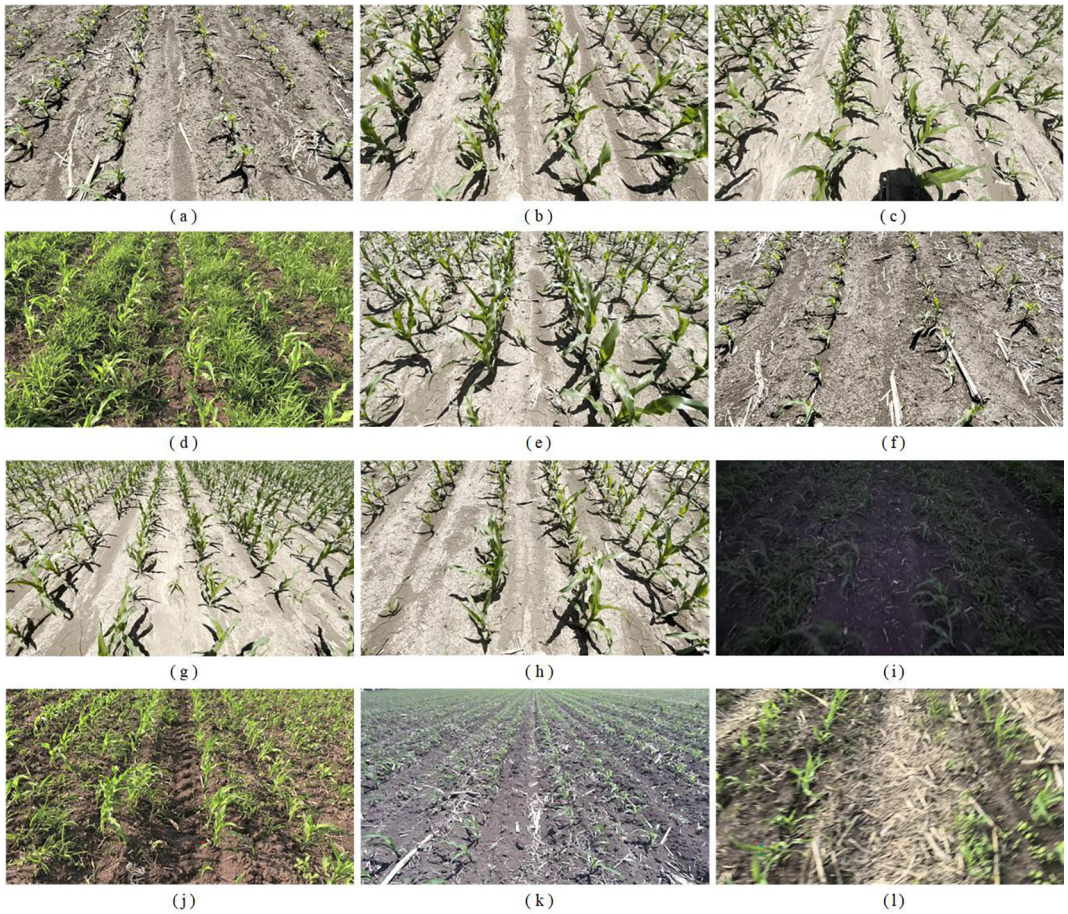

FIGURE 3
Some representative sample images of corn rows at the 3–5 leaf stage, including: **(a)** 3–4 leaf stage; **(b)** 4–5 leaf stage; **(c)** front-lit perspective; **(d)** weed-infested conditions; **(e)** imaging angle of 45°; **(f)** imaging angle of 60°; **(g)** imaging angle of 30°; **(h)** seedling gaps; **(i)** low-light conditions; **(j)** high-light conditions; **(k)** low-pixel images; **(l)** blurred images.

2.2 Architecture overview

An overview of ALNet is depicted in Figure 4. The network mainly consists of three components: the Feature Extract module, the Attention-guide ROI Align module, and the RIoU loss. Specifically, the network accepts RGB images and outputs lines in different colors, each corresponding to a distinct maize row. As noted above, we treat the entire maize row as the detection target. In conventional object detection, rectangular bounding boxes are commonly used to represent targets Liu et al. (2016); Redmon et al. (2016); Ren et al. (2015); however, elongated maize rows are poorly represented by bounding boxes. Taking into account the prior knowledge of corn row shapes, we adopt an Anchor-Line based formulation that casts maize-row detection as a combined classification and localization problem. An end-to-end network architecture is thus designed to achieve efficient and accurate row detection. In the following subsections we describe each component in detail.

2.2.1 Feature extract module

The Feature Extract module is responsible for extracting both deep semantic and shallow positional information from the input image, and for enabling flexible positioning of Anchor-Lines. Both accurate feature extraction and proper Anchor-Line placement are critical for high detection accuracy. To this end, we employed a Feature Pyramid Network (FPN) for multi-scale feature fusion He et al. (2016). Meanwhile, the detection head incorporated the Attention-guide ROI Align module and the RIoU loss (described below) to iteratively refine Anchor-Line positions so that they progressively aligned with the ground-truth maize rows. To mitigate gradient explosion and keep the model lightweight, ResNet-18 is adopted as the backbone Lin et al. (2017).

The specific workflow of the Feature Extract module is as follows. First, three convolutional layers are applied to the input image to extract multi-scale feature maps. The top-level feature map, C_3, is passed through convolutional layer L_1. An initial set of Anchor-Lines (p0) is introduced, and a convolutional kernel is applied sequentially along each Anchor-Line. The resulting feature maps are forwarded to the detection head Pred1. After Pred1 processes these features, an updated Anchor-Line arrangement (p1), which provides a more accurate estimate of the true maize row locations, is sent to L_2. Concurrently, feature map C_2 is fused with the convolved C_3 (from L_1), and this fused feature set is input to L_2. In L_2 the same convolution operation is performed along each Anchor-Line in p1. Finally, using the twice-updated Anchor-Line arrangement (p2) from L_2, convolution is performed on the multi-scale fused feature map in L_3; the result is fed into Pred3, which regresses the maize row lines.

2.2.2 Attention-guide ROI align module

In practical field conditions, maize rows may be occluded or blurred due to complex scenarios such as extreme lighting, numerous weeds, or missing seedlings. Under such circumstances local visual cues are crucial to verify the presence of maize rows; therefore, relying solely on contextual information extracted by the FPN can be insufficient Lamping et al. (2025). To improve robustness, we propose the Attention-guide ROI Align module to better exploit long-range dependencies and aggregate richer

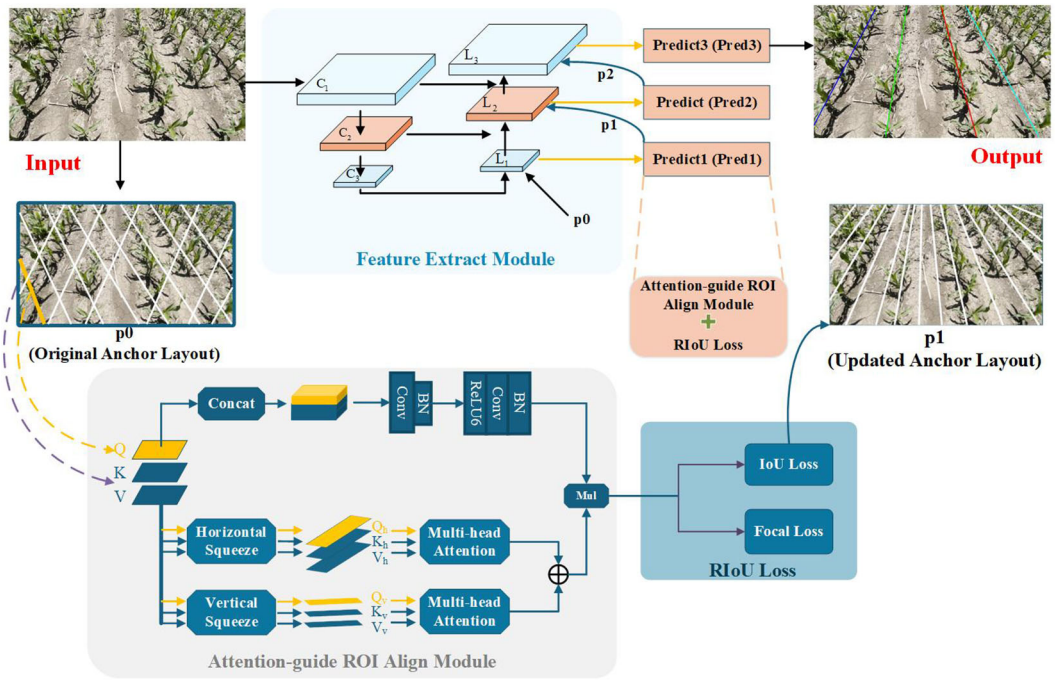

FIGURE 4
Overview of the proposed ALNet.

contextual information for learning maize-row features. Concretely, convolutional operations aligned with the preceding Anchor-Line are added so that each pixel along an Anchor-Line can aggregate information from its neighborhood and thus enhance the representation of occluded or blurred regions. In addition, correlations between local features and the global feature map are established, allowing richer context to augment local feature representations and improve discrimination under challenging conditions. To ensure the Attention-guide ROI Align module remains lightweight enough for edge-device deployment, we replace the standard Vision Transformer attention Dosovitskiy et al. (2020); Liang et al. (2025a); Xie et al. (2021); Zheng et al. (2021) with the Dual-Axis Extrusion Transformer (DAE-Former). The implementation is as follows:

$$\begin{aligned} q_{(h)} &= \tfrac{1}{W}\left(q^{\rightarrow(C_{qk},H,W)} I_W\right)^{\rightarrow(H,C_{qk})}, \\ q_{(v)} &= \tfrac{1}{H}\left(q^{\rightarrow(C_{qk},W,H)} I_H\right)^{\rightarrow(W,C_{qk})}. \end{aligned} \tag{1}$$

As illustrated in the lower half of the Attention-guide ROI Align module in Figure 5, the query vector q is extracted from a single Anchor Line using a weight matrix $W_q^{(s)} \in \mathbb{R}^{C_{qk}\times C}$, while the key k and the value v are extracted from the future map using a weight matrix $W_k^{(s)} \in \mathbb{R}^{C_{qk}\times C}$, $W_v^{(s)} \in \mathbb{R}^{C_v\times C}$. According to the upper formula in Equation 1, we first implement horizontal squeeze by averaging the queried feature map along the horizontal axis. Similarly, the lower formula in Equation 1 represents vertical squeeze through vertical direction averaging. Here, $Z^{\rightarrow(*)}$ denotes permuting the dimension of tensor Z as given, and I is an all-ones vector. The squeeze operation applied to q is simultaneously replicated on k and v, the resulting projected tensors satisfy: $q_{(h)}$, $k_{(h)}, v_{(h)} \in \mathbb{R}^{H\times C_{qk}}, q_{(v)}, k_{(v)}, v_{(v)} \in \mathbb{R}^{W\times C_{qk}}$.

Constraining feature interactions to a single axis substantially reduces computation. The per-position output is given by Equation 2:

$$y_{(i,j)} = \sum_{p=1}^{H} \mathrm{softmax}_p(q_{(h)i}^{\mathrm{T}} k_{(h)p}) v_{(h)p} + \sum_{p=1}^{W} \mathrm{softmax}_p(q_{(v)j}^{\mathrm{T}} k_{(v)p}) v_{(v)p}\,. \tag{2}$$

To mitigate possible contextual information loss introduced by biaxial compression, we add an auxiliary convolutional kernel to enhance local spatial details (upper part of the Attention-guide ROI Align module in Figure 5. Using weight matrices $W_q^{(e)}, W_k^{(e)} \in \mathbb{R}^{C_{qk}\times C}$ and $W_v^{(e)} \in \mathbb{R}^{C_v\times C}$, we extract a query q_1 from a single Anchor Line and the corresponding key k_1 and value v_1 from the feature map. These are concatenated along the channel dimension (with broadcasting as required) and processed by a 3×3 convolution to capture local detail. A subsequent linear projection — followed by batch normalization and an activation — reduces the concatenated channel size $(2C_{qk} + C_v)$ back to C, producing detail-enhancement weights. Finally, this enhanced feature is fused with the biaxial-attention output.

2.2.3 RIoU loss

As illustrated in Figure 6, each maize row is represented by a sequence of equally spaced 2D points:

$$G = \{(x_1, y_1), \cdots, (x_N, y_N)\}, \tag{3}$$

In Equation 3, the y-coordinates are sampled vertically at uniform intervals, i.e., $y_i = \frac{H}{N-1} \times i$, and H denotes the image height. Inspired by Liang et al. (2025b); Rezatofighi et al. (2019); Zheng et al. (2022), we adopt an Intersection-over-Union (IoU) based loss to measure Row IoU (RIoU), treating each maize row as an entity for regression. For each sampled position we consider the predicted point x_i^p and the corresponding ground-truth point x_i^g, and extend both horizontally by a half-length e. The IoU between the two extended line segments is then:

$$IoU = \frac{d_i^O}{d_i^u} = \frac{\min\ (x_i^p + e, x_i^g + e) - \max\ (x_i^p - e, x_i^g - e)}{\max\ (x_i^p + e, x_i^g + e) - \min\ (x_i^p - e, x_i^g - e)}, \tag{4}$$

In Equation 4, $[x_i^p - e, x_i^p + e]$ and $[x_i^g - e, x_i^g + e]$ are the extended prediction and ground-truth segments, respectively. Note that the numerator may become negative when segments do not overlap, which effectively expands the optimization space.

Treating the row as the aggregation of these sampled positions, we discretize the integral IoU and define RIoU as in Equation 5:

$$RIoU = \frac{\sum_{i=1}^{N} d_i^O}{\sum_{i=1}^{N} d_i^u}, \tag{5}$$

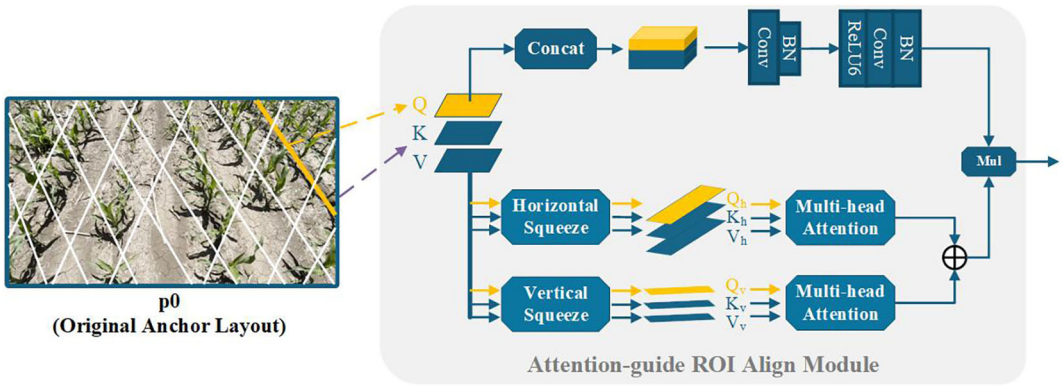

FIGURE 5
Overview of the attention-guide ROI align module.

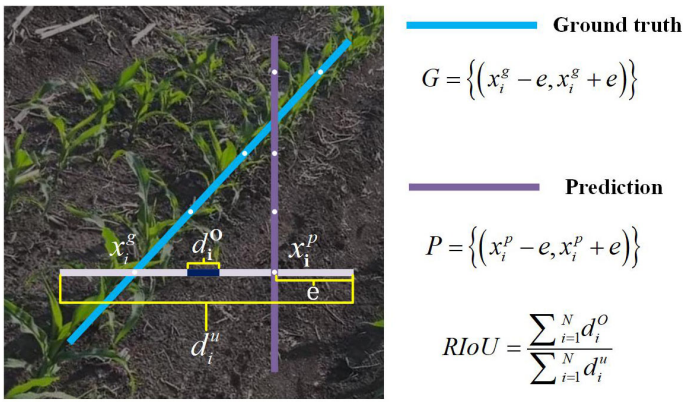

FIGURE 6
Illustration of Row IoU. Row IoU (Intersection over Union) is calculated by integrating the IoU of the extended segment at sampled positions x_i.

and the corresponding loss given in Equation 6:

$$L_{RIoU} = 1 - RIoU, \tag{6}$$

where $-1 \leqslant RIoU \leqslant 1$. RIoU attains 1 when the predicted and ground-truth extended segments are perfectly aligned, and approaches -1 as they move far apart.

2.3 Model training

During training, following the strategy in Ge et al. (2021), one or more predicted rows are dynamically assigned as positive samples for each ground-truth maize row. The assignment cost is defined as shown in Equation 7:

$$\begin{aligned} \mathcal{C}_{assign} &= \omega_{sim}\mathcal{C}_{sim} + \omega_{cls}\mathcal{C}_{cls}, \\ \mathcal{C}_{sim} &= (\mathcal{C}_{dis} \cdot \mathcal{C}_{xy} \cdot \mathcal{C}_{theta})^2 . \end{aligned} \tag{7}$$

where $\mathcal{C}_{cls}$ denotes the focal classification cost between prediction and ground truth. The similarity cost $\mathcal{C}_{sim}$ consists of three normalized components (all scaled to $[0, 1]$) — $\mathcal{C}_{dis}$ represents the average pixel distance of all valid line points, $\mathcal{C}_{xy}$ is the distance between the start-point coordinates, and $\mathcal{C}_{theta}$ is the angular difference θ between the predicted and ground-truth lines. The scalars ω_{sim} and ω_{cls} weight the respective costs.

The total training loss comprises classification and regression terms; regression losses are applied only to the assigned (positive) samples:

$$L_{total} = \omega_{cls}L_{cls} + \omega_{xytl}L_{xytl} + \omega_{RIoU}L_{RIoU} . \tag{8}$$

In Equation 8, L_{xytl} is the regression loss for the start-point coordinates, angle θ and the row length, implemented with the smooth-l_1 loss. L_{xytl} is the focal classification loss between predicted class scores and ground truth.

All experiments were performed on a machine with an Intel(R) Core(TM) i5–12500 CPU and an NVIDIA GeForce RTX 3060 (12GB) GPU. We used ResNet-18 with FPN as the backbone. The original input images (1280×720) were resized to C×W×H = 3×800×320 before training, the parameter count is 11.75M, the measured peak GPU memory usage during training was 9,878 MB. Optimization was performed with AdamW with an initial learning rate of 0.0001. A cosine decay learning-rate schedule (power factor 0.9) was used. Training ran for 70 epochs with a batch size of 40.

3 Results and discussion

3.1 Evaluation metric

We adopted the *F*1-measure as the principal evaluation metric. Intersection-over-Union (IoU) is computed between predicted rows and ground-truth rows as follows.

Predicted and ground-truth row lines are thickened to 50 pixels so that the enlarged masks closely cover the actual maize rows; IoU is then calculated between these thickened masks. A predicted row is counted as a true positive (TP) if its IoU with a ground-truth row exceeds a specified threshold. To compare localization performance more precisely across methods, we report *F*1 at multiple IoU thresholds and their mean (*mF*1):

$$\begin{aligned} F1 &= \tfrac{2\times Precision\times Recall}{Precision+Recall}, \\ mF1 &= (F1@50 + F1@55 + \ldots + F1@95)/10, \end{aligned} \tag{9}$$

In Equation 9, $F1_{@}0.50$, $F1_{@}0.55$,…, $F1_{@}0.95$ refer to the model's *F*1 scores computed at IoU thresholds of 0.5, 0.55,…, 0.95. When the IoU threshold is 0.5, predicted corn rows are considered to reasonably represent actual corn rows in real-world scenarios. However, to precisely quantify the model's detection accuracy and improve the ability to distinguish between different models, *F*1 scores at higher IoU thresholds are also included in the evaluation metrics.

3.2 Ablation study

To assess the contribution of key components in ALNet, we conducted ablation experiments on the same dataset using a

ResNet-18 backbone. We progressively evaluated the RIoU loss, Feature Pyramid Network (FPN), Attention-guide ROI Align module, and DAE-Fomer. Results are summarized in Table 1.

Efficacy of the RIoU loss Incorporating the RIoU loss into the baseline model (second row in Table 1) increased the *mF*1 score from 51.90 to 54.50 (approximately +5%), with only a slight drop in FPS to 185.72 (less than 3%). When the RIoU loss weight was replaced by the optimal regression weight of the traditional smooth-l_1 loss Yu et al. (2016), the *mF*1 score declined, confirming that RIoU loss offers greater stability and superior performance—particularly under stricter IoU thresholds such as *F*1@85 and *F*1@90. Visual comparison of weight distribution heatmaps under the two loss configurations (Figure 7) further validates these findings. While smooth-l_1 captures the spatial representation of corn rows, the RIoU loss produces heatmaps with greater spatial smoothness and markedly higher color contrast, improving target–background separation.

This performance gain stems from RIoU's continuous IoU-based optimization objective and favorable gradient properties. Smooth-l_1, due to its linear penalty, may cause gradient oscillations when coordinate deviations are small. In contrast, RIoU yields smoother, more stable gradients and avoids stagnation when predicted and ground-truth boxes do not overlap. It also dynamically adjusts loss weights to mitigate over-sensitivity to local errors, improving convergence efficiency and representation accuracy. In summary, both quantitative results and heatmap visualizations confirm the superiority of RIoU loss in localization accuracy, gradient stability, and convergence behavior.

Contributions of the FPN The Feature Pyramid Network (FPN) effectively fuses deep semantic information with shallow positional cues. As Figure 8 illustrates, attention weight heatmaps demonstrate FPN's role in improving corn row localization accuracy and feature representation. Without FPN, heatmaps exhibit weak attention weight distributions: though centered on corn rows, response intensity is low. With FPN enabled, weights concentrate more intensely in corn row regions with heightened color saturation, indicating dual improvements in detection sensitivity and localization accuracy.

Necessity of the Attention-guide ROI Align module To validate the role of the Attention-guided ROI Align module in enhancing robustness and accuracy, we analyzed weight distribution heatmaps (Figure 9). As can be seen, models without this module exhibit scattered heatmap color distributions, reflecting insufficient focus on maize rows when background interference (e.g., weeds) is present. This leads to reduced feature extraction accuracy and diminished robustness under challenging conditions such as dense weeds or low light. In contrast, even under ideal conditions (clear background, sufficient lighting), the module produces more concentrated and vivid heatmap colors, confirming improved feature localization. The module provides two principal advantages: (*i*) enhanced semantic feature capture via efficient global context integration; (*ii*) achieved robust and accurate maize row localization under diverse environmental disturbances.

Effectiveness of the DAE-Fomer Replacing DAE-Former with a traditional Vision Transformer confirmed that the Dual-Axis Extrusion design achieves an optimal trade-off between detection accuracy and speed (fourth row in Table 1). This design overcomes the speed bottleneck of standard Transformers while maintaining high accuracy.

Summary, the ablation results demonstrate that: (*i*) RIoU loss substantially improves localization precision; (*ii*) FPN and the Attention-guided ROI Align module jointly enhance feature representation and robustness; and (*iii*) DAE-Former's dual-axis design ensures high-speed, high-accuracy performance, validating its necessity in the corn row detection task.

3.3 Comparison with other methods

We compared ALNet with several mainstream detection and segmentation methods for maize-row detection. Quantitative results are reported in Table 2. Overall, ALNet attains a favorable trade-off between detection accuracy and inference speed: its *F*1 scores across multiple IoU thresholds exceed those of competing models, indicating a clear overall performance advantage. Further inspection of localization accuracy and real-time metrics shows that ALNet not only locates maize rows with high precision but also achieves a substantial improvement in inference speed.

Specifically, relative to the semantic-segmentation method TP-ISN (DeepLabV3+), ALNet increases the average *F*1 by 9.24 percentage points while delivering a 245% improvement in FPS, demonstrating superior performance in both accuracy and speed. Compared with RASCM (ResNet-18 backbone), ALNet incurs only a 19% reduction in FPS but improves *mF*1 by 51.73 percentage

TABLE 1 Effects of each component in our method.

RIoU	Attention-guide ROI align	DAE-fomer	FPN	*mF*1	*F*1@55	*F*1@65	*F*1@85	FPS
				51.90	78.37	60.32	19.54	193
✓				54.50	90.20	79.12	27.72	185.72
✓			✓	55.72	91.51	79.58	28.25	165.60
*	✓	✓	✓	56.74	90.68	79.61	31.68	163.22
✓	✓	#	✓	60.63	92.34	81.60	34.14	54.60
✓	✓	✓	✓	59.60	91.58	80.84	33.72	161.26

"*" Replace this module with Smooth-l1, "#" replace this module with a traditional Vision Transformer, as detailed below.

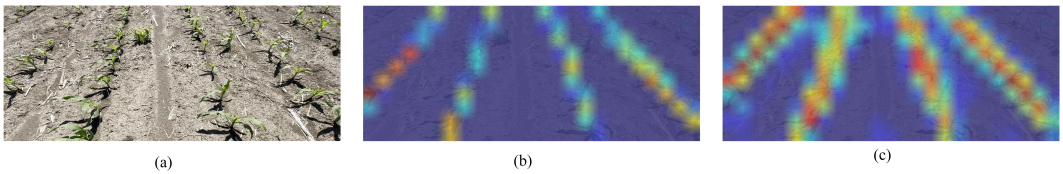

FIGURE 7
Illustration of attention weight. The red regions correspond to high score in attention weight. **(a)** represents the input image to the model; **(b)** depicts the weight heatmap of the model with smooth-l_1 loss; and **(c)** depicts the weight heatmap of the model with RIoU loss.

points. These results indicate that ALNet more accurately regresses maize-row positions while retaining real-time capability, thus achieving a favorable balance between precision and throughput.

Figure 10 presents a qualitative comparison on challenging maize-row scenes. Traditional object detection frameworks, such as the YOLO series and DEIM, produce discrete bounding-box outputs that inadequately capture the continuous, strip-like geometry of maize rows. This limitation compromises geometric coherence and results in irregular edge delineation. Furthermore, the two-stage fitting approach—first detecting individual corn plants and subsequently fitting the corn seedling line—incurs substantial.

FPS degradation for these models. Instance-segmentation approaches (e.g., RASCM) attempt contour-level representation but are limited by segmentation granularity, which hinders faithful modeling of contiguous row structures. TP-ISN (DeepLabV3+), although better at preserving continuity through semantic segmentation, is constrained by its dual-branch design that relies on traditional image-processing modules (e.g., morphological operations and handcrafted post-processing) in the embedding branch; these modules limit computational efficiency and the representational capacity of learned features, thereby impairing both accuracy and real-time performance. In contrast, ALNet accurately captures the continuity and geometric characteristics of maize rows under complex conditions, producing smooth, coherent predictions that demonstrate its robustness and practical effectiveness.

Figure 11 shows training curves: ALNet reaches an $mF1$ of 59.60 at the seventieth training epoch, whereas competing models attain $mF1$ values in the 30–50 range under the same or larger number of epochs. This gap indicates that ALNet converges more rapidly, a consequence of the proposed architectural and optimization strategies that enable faster adaptation to the task. Under limited training resources, ALNet still achieves superior performance, validating the design philosophy of balancing training efficiency and model capability for resource-constrained deployments. In addition, the training process of ALNet is more stable and less sensitive to hyperparameters (e.g., learning rate, weight initialization), which reduces tuning effort and mitigates overfitting risk—further evidence of improved training robustness.

4 Limitation and future work

Our network has advanced the balance between detection accuracy and speed for maize-row detection and provides real-time, precise positional information that can support automated field-management tasks. Nevertheless, several limitations remain.

First, experiments to date have been conducted using field-collected image datasets and have not yet been validated through deployment in operational field systems. Second, owing to the seasonal growth dynamics of maize, the temporal coverage and scene diversity of the current dataset require further expansion to validate robustness across all growth stages and agricultural conditions. Thirdly, the dataset is limited to a single geographic region and lacks validation across diverse environments and corn varieties. Consequently, relying solely on this dataset may exhibit reduced detection performance or potential failure in identifying

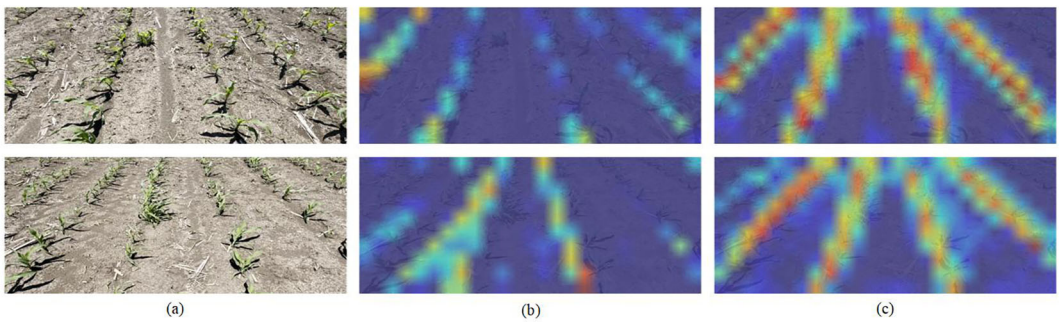

FIGURE 8
Illustration of attention weight. The red regions correspond to high score in attention weight. **(a)** represents the input image to the model; **(b)** depicts the weight heatmap of the model without FPN; and **(c)** depicts the weight heatmap of the model with FPN.

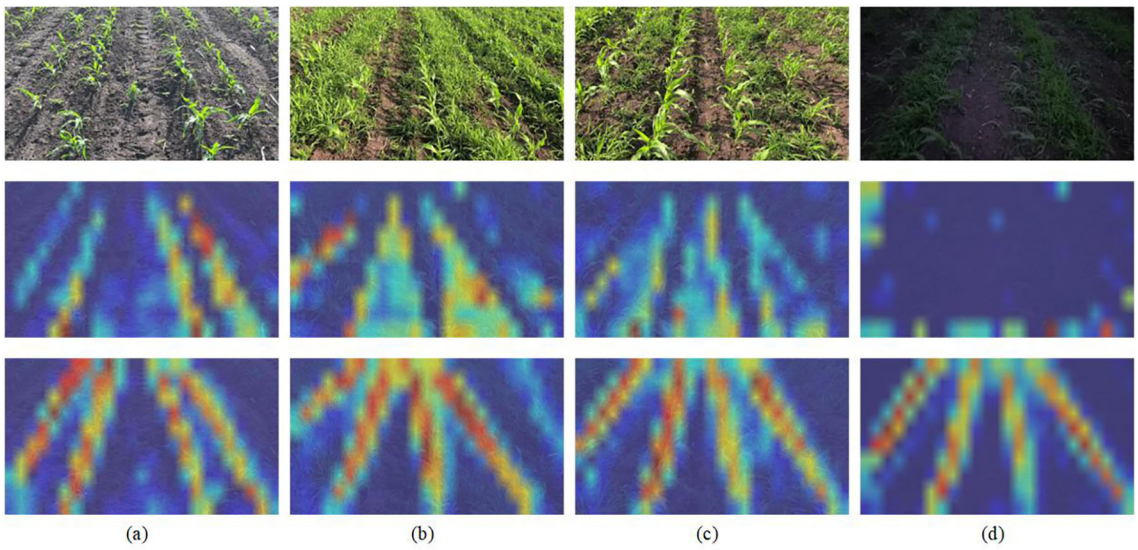

FIGURE 9
Illustration of attention weight. The red regions correspond to high score in attention weight. The first row represents the input image; the second row depicts the model without Attention-guide ROI Align; and the third row depicts the model with Attention-guide ROI Align. **(a)** depicts an ideal detection environment, while **(b, c)** depict weed-infested detection environments, and **(d)** depicts a detection environment with low-light and strong-wind conditions.

corn under different regional conditions or cultivar variations. Fourthly, while the study specifically targets corn detection during the 3–5 leaf stage, there remains a risk of performance degradation or detection failure when applied to corn at other growth stages due to the absence of cross-stage generalization validation in the current methodology.

To address these limitations, we propose the following technical directions for future work:

1. Deploy the algorithm on mobile embedded platforms to evaluate real-world inference performance and energy constraints.
2. Integrate the perception pipeline with autonomous-control modules for agricultural robots (e.g., path tracking, trajectory planning, and obstacle avoidance) to enable stable and reliable visual navigation in-field.
3. Extend training data and model generalization to additional crops with more complex planting patterns (e.g., wheat and rice) to improve multi-crop adaptability and broaden practical applicability in diverse agricultural scenarios.
4. Enrich the dataset to cover more growth stages, lighting conditions, and terrain variations, and perform field-deployment trials to validate end-to-end performance and robustness in real operational settings.

These directions will both test the practical applicability of ALNet and further enhance its generalization, efficiency, and suitability for real-world agricultural deployments.

5 Conclusions

This study presents ALNet (Anchor-Line Network), a lightweight convolutional neural network tailored for the elongated geometry of maize rows. By leveraging the Anchor-Line

TABLE 2 Performance comparison of detection models.

Method	Backbone	Parameters	mF1	F1@50	F1@70	F1@90	FPS	GFLOPs
RASCM Wei et al. (2022)	ResNet18	6.23M	39.28	72.96	46.54	6.83	200	8.4
YOLO11-N Hongbo (2024)	Darknet	3.8M	35.21	70.08	43.96	3.08	7.04	21.5
TP-ISN Chen et al. (2025)	DeepLabV3+	4.41M	50.36	81.52	62.85	7.72	46.70	19.6
DEIM-N Huang et al. (2025)	HGNetv2	4M	33.65	69.43	43.78	2.29	3.54	7
ALNet (*ours*)	ResNet18	11.23M	59.60	94.58	72.92	12.44	161.26	11.9

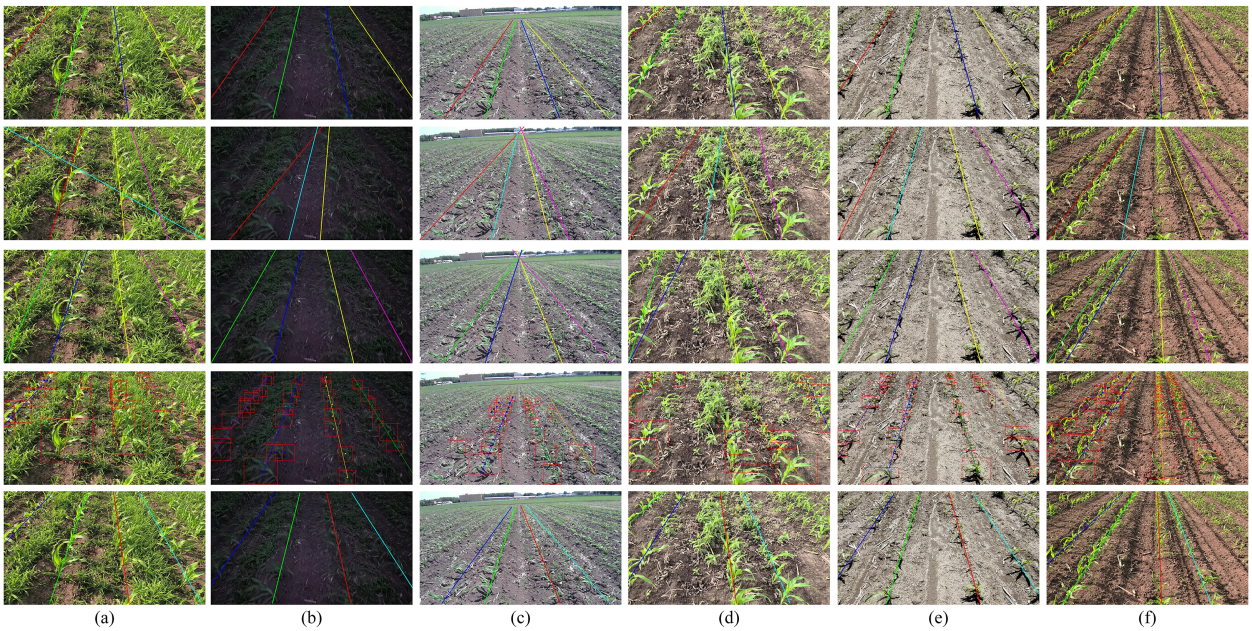

FIGURE 10
The visualization of different detection models. The first row presents the ground truth of corn rows, the second row shows the prediction maps from the RASCM model, the third row corresponds to the TP-ISN model, the fourth row displays the YOLO11 model's predictions, and the fifth row illustrates the ALNet (ours) model's predictions. **(a)** Weedy field with strong illumination; **(b)** low-light conditions with strong wind; **(c)** low-resolution scenario; **(d)** blurred scenario; **(e)** missing plant scenario; **(f)** overexposed illumination scenario.

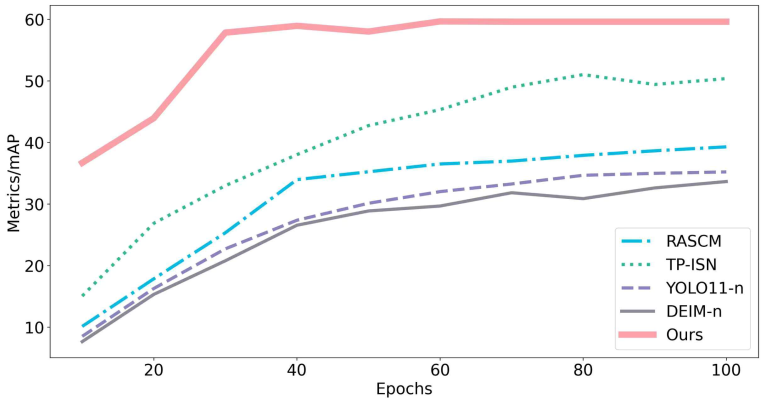

FIGURE 11
*mF*1 score *vs*. epochs for different models.

mechanism, ALNet reformulates row detection as an end-to-end regression task, avoiding the computational burden of pixel-wise prediction while preserving the holistic continuity of crop rows. The integration of an Attention-guided ROI Align module with a Dual-Axis Extrusion Transformer (DAE-Former) enables efficient global–local feature interaction, enhancing robustness under challenging field conditions such as weed infestation, low light, and wind distortion. Experimental results show that ALNet achieves a favorable accuracy–speed trade-off, with an *mF*1 of 59.60 across IoU thresholds and an inference speed of 161.26 FPS, both outperforming existing mainstream methods. Its lightweight design (11.9 GFlops) suggests potential for real-time edge deployment. These advances establish ALNet as a practical and scalable solution for visual navigation in precision agriculture.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

BF: Conceptualization, Data curation, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. QH: Investigation, Software, Validation, Writing – review & editing. YH: Data curation, Methodology, Resources, Validation, Visualization, Writing – review & editing. HC: Data curation, Resources, Software, Writing – review & editing. DL: Conceptualization, Data curation, Investigation, Writing – review & editing. ZS: Methodology, Project administration, Writing – review & editing. BZ: Software, Supervision, Validation, Writing – review & editing. LQ: Funding acquisition, Methodology, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. RM: Methodology, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This research was funded by the Specific University Discipline Construction Project (File No.2023B10564002).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer KX declared a shared affiliation with the author(s) RM, BF, QH, HC, DL, ZS, LQ to the handling editor at the time of review.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Abbasi, R., Martinez, P., and Ahmad, R. (2022). The digitization of agricultural industry – a systematic literature review on agriculture 4.0. *Smart Agric. Technol.* 2, 100042. doi: 10.1016/j.atech.2022.100042

Agathokleous, E. (2018). Environmental hormesis, a fundamental non-monotonic biological phenomenon with implications in ecotoxicology and environmental safety. *Ecotoxicology Environ. Saf.* 148, 1042–1053. doi: 10.1016/j.ecoenv.2017.12.003

Ban, C., Wang, L., Su, T., Chi, R., and Fu, G. (2025). Fusion of monocular camera and 3d lidar data for navigation line extraction under corn canopy. *Comput. Electron. Agric.* 232, 110124. doi: 10.1016/j.compag.2025.110124

Bing, W. (2021). The Acquisition of Weed Phenotype Information Based on Instance Segmentation and Development of Target Application System. Harbin, Heilongjiang Province, China: Northeast Agricultural University.

Cao, M., Tang, F., Ji, P., and Ma, F. (2022). Improved real-time semantic segmentation network model for crop vision navigation line detection. *Front. Plant Sci.* 13, 898131. doi: 10.3389/fpls.2022.898131

Chen, Z., Cai, Y., Liu, Y., Liang, Z., Chen, H., Ma, R., et al. (2025). Towards end-to-end rice row detection in paddy fields exploiting two-pathway instance segmentation. *Comput. Electron. Agric.* 231, 109963. doi: 10.1016/j.compag.2025.109963

Diao, Z., Guo, P., Zhang, B., Zhang, D., Yan, J., He, Z., et al. (2023). Navigation line extraction algorithm for corn spraying robot based on improved YOLOv8s network. *Comput. Electron. Agric.* 212, 108049. doi: 10.1016/j.compag.2023.108049

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. doi: 10.48550/arXiv.2010.11929

Gan, S., Yu, G., Wang, L., and Sun, L. (2025). Seedling row extraction on unmanned rice transplanter operating side based on semi-supervised semantic segmentation. *Comput. Electron. Agric.* 230, 109759. doi: 10.1016/j.compag.2024.109759

Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. (2021). YOLOX: Exceeding YOLO series in 2021. *arXiv preprint arXiv:2107.08430*. doi: 10.48550/arXiv.2107.08430

Gong, H., and Zhuang, W. (2024). An improved method for extracting inter-row navigation lines in nighttime maize crops using YOLOv7-tiny. *IEEE Access* 12, 27444–27455. doi: 10.1109/ACCESS.2024.3365555

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Las Vegas, USA: IEEE), 770–778.

Hongbo, L. (2024). Seedling stage corn line detection method based on improved yolov8. *Smart Agric.* 6, 72–84. doi: 10.12133/j.smartag.SA202408008

Huang, S., Lu, Z., Cun, X., Yu, Y., Zhou, X., and Shen, X. (2025). "DEIM: DETR with improved matching for fast convergence," in *Proceedings of the Computer Vision and Pattern Recognition Conference*. (Nashville, Tennessee, USA: IEEE), 15162–15171.

Lamping, C., Kootstra, G., and Derks, M. (2025). Transformer-based similarity learning for re-identification of chickens. *Smart Agric. Technol.* 11, 100945. doi: 10.1016/j.atech.2025.100945

Liang, B., Hu, L., Liu, G., Hu, P., Xu, S., and Jie, B. (2025a). YOLOv9-GSSA model for efficient soybean seedlings and weeds detection. *Smart Agric. Technol.* 12, 101134. doi: 10.1016/j.atech.2025.101134

Liang, X., Xiang, J., Qin, S., Xiao, Y., Chen, L., Zou, D., et al. (2025b). Small target detection algorithm based on sahi-improved-YOLOv8 for UAV imagery: A case study of tree pit detection. *Smart Agric. Technol.* 12, 101181. doi: 10.1016/j.atech.2025.101181

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Honolulu, Hawaii, United States: IEEE), 2117–2125.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). "SSD: Single shot multibox detector," in *European Conference on Computer Vision* (Amsterdam, Netherlands: Springer), 21–37.

Liu, Y., Guo, Y., Wang, X., Yang, Y., Zhang, J., An, D., et al. (2024). Crop root rows detection based on crop canopy image. *Agriculture* 14, 969. doi: 10.3390/agriculture14070969

Lulu, X. (2024). *Extraction of Row Centerline at the Early Stage of Corn Growth Based on UAV Images*. (Chinese Master's Theses Full-text Database). (Chengdu, Sichuan, China: University of Electronic Science and Technology of China).

Rabab, S., Badenhorst, P., Chen, Y.-P. P., and Daetwyler, H. D. (2021). A template-free machine vision-based crop row detection algorithm. *Precis. Agric.* 22, 124–153. doi: 10.1007/s11119-020-09732-4

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Las Vegas, USA: IEEE), 779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 28.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. (2019). "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Long Beach, USA: IEEE), 658–666.

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention* (Munich, Germany: Springer), 234–241.

ShaoKun, L. (2017). Advances and prospects of maize cultivation in China. *Scientia Agricultura Sin.* 50, 1941–1959. doi: 10.3864/j.issn.0578-1752.2017.11.001

Takikawa, T., Acuna, D., Jampani, V., and Fidler, S. (2019). "Gated-SCNN: Gated shape CNNs for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*. (Seoul, South Korea: IEEE), 5229–5238.

Wei, C., Li, H., Shi, J., Zhao, G., Feng, H., and Quan, L. (2022). Row anchor selection classification method for early-stage crop row-following. *Comput. Electron. Agric.* 192, 106577. doi: 10.1016/j.compag.2021.106577

Xiangguang, L. (2021). Extraction algorithm of the center line of maize row in case of plants lacking. *Trans. Chin. Soc. Agric. Eng.* 37, 203–210. doi: 10.11975/j.issn.1002-6819.2021.18.024

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* 34, 12077–12090.

Yao, C., Lv, D., Li, H., Fu, J., Li, C., Gao, X., et al. (2025). A real-time crop lodging recognition method for combine harvesters based on machine vision and modified DeepLab v3+. *Smart Agric. Technol.* 11, 100926. doi: 10.1016/j.atech.2025.100926

Yu, J., Jiang, Y., Wang, Z., Cao, Z., and Huang, T. (2016). "UnitBox: An advanced object detection network," in *Proceedings of the 24th ACM international conference on Multimedia*. (Amsterdam, Netherlands: ACM), 516–520.

Zhaosheng, L. (2025). Application and development trend of mechanized weeding technology for corn. *China Agric. Machinery Equip.*, 133–136.

Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., et al. (2021). "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (IEEE), 6881–6890.

Zheng, T., Zhao, S., Liu, Y., Liu, Z., and Cai, D. (2022). "SCALoss: Side and corner aligned loss for bounding box regression," in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 36. (Vancouver, Canada: AAAI), 3535–3543.