SA-YOLO FOR MAIZE SEEDLING ROW SEGMENTATION: TOWARD RELIABLE NAVIGATION LINE EXTRACTION IN AGRICULTURAL FIELDS
/
基于改进 YOLOv12n-seg 网络的苗期玉米作物行识别方法

Delin SHANG[1,2], Yu YANG[3], Xingchen QIAO[3], Zhiwei HU[3], Changchun FAN[3], Yuanwei JIANG[3], Ziyang QIN[3], Kuanyan ZHANG[3], Xiaole WANG[3], Tie ZHANG*[1,2], Zhenchao WU*[3], Gang ZHANG*[4]

[1] State Key Laboratory of Agricultural Equipment Technology, Beijing 100083, China;
[2] Chinese Academy of Agricultural Mechanization Sciences Group Co., Ltd., Beijing 100083, China;
[3] School of Engineering, Anhui Agricultural University, Anhui 230036, Hefei, China;
[4] School of Intelligent Manufacturing, Anhui Science and Technology University, Chuzhou 239000, China
Corresponding author: Tie Zhang; Zhenchao Wu; Gang Zhang
Tel: +8613691110166; +8613204655989; +8618756564355; E-mail: zhangtie1979@126.com; wuzhenchao@ahau.edu.cn; zhangg@ahstu.edu.cn
DOI: https://doi.org/10.35633/inmateh-79-109

***Keywords:** Crop row recognition; Maize seedlings; Deep learning model; Centreline extraction*

ABSTRACT
The seedling stage is a relatively short but critical period in the maize growth cycle, during which rapid weeding, pesticide application, and fertilization are required. Crop row centreline extraction technology based on computer vision can assist agricultural machinery in performing automated field operations, thereby improving work efficiency. Therefore, improving the accuracy of crop row centreline extraction imposes higher requirements on existing models and algorithms. In this study, a model named Synergistic Attention YOLO (SA-YOLO) was proposed. The feature extraction capability for crop rows was enhanced by introducing the spatial and channel synergistic attention module and designing the A2C2f module integrated with switchable atrous convolution. In addition, a crop topological skeleton feature point extraction strategy and a collaborative fitting method combining principal component analysis and least squares was integrated to accurately extract crop row centrelines from the skeleton feature points of crop pixels within the segmented region of interest. Comparative experiments showed that the SA-YOLO model achieved competitive performance, with an average precision at IoU threshold 0.5 (AP_{50}) of 99.1%, and a mask Intersection over Union (IoU_{mask}) of 83.8%, while maintaining a lightweight architecture with only 3.72 M parameters. The mean angle error (A_{mean}) of crop row centerline extraction was 0.590°, and the mean normalized lateral error (L_{mean}^{norm}) was 0.357%. This study provides a new approach for exploring crop row centerline extraction in agriculture and can further enrich the theoretical and technical foundation for visual navigation of agricultural robots.

摘要
苗期是玉米生长周期中相对较短但关键的时期，需要快速完成除草、施药、施肥等工作。基于计算机视觉的作物行中心线提取技术可以辅助农业机械进行自动化田间作业提高工作效率。因此，作物行中心线提取精度的提升对现有模型与算法提出了更高的要求。本研究提出了一种名为 SA-YOLO 的模型，通过引入 SCSA 注意力模块和设计的 A2C2f_SAConv 模块增强了对作物行特征的提取能力。此外，还集成了作物拓扑骨架特征点提取策略与 PCA-LSM 协同拟合方法，通过分割输出的 ROI 区域内作物像素的骨架特征点来精确提取作物行中心线。对比实验表明，SA-YOLO 模型实现了具有竞争力的性能，其 AP_{50} 为 99.1％，IoU_{mask} 为 83.8％，同时保持轻量级架构，参数仅为 3.72M。作物行中心线提取产生的平均角度误差为 0.590°，归一化横向误差为 0.357%。该研究为探索农业作物行中心线提取提供了新的思路，可以进一步丰富农业机器人视觉导航的理论和技术基础。

1. INTRODUCTION
As one of the most important food crops worldwide, maize plays a key role in global food security. Achieving efficient and intelligent maize production is an important pathway to ensure food supply capacity. In field management during the maize seedling stage *(Shahid et al., 2024)*, rapid and accurate recognition of maize row centrelines is an important foundation for autonomous navigation of agricultural machinery *(Liu et al., 2023; Liu et al., 2024b)*, precision agricultural operations *(Chamara et al., 2023; Huang et al., 2020)*, and crop phenotyping analysis.

Therefore, constructing an accurate and efficient maize row centreline recognition method through intelligent means is of great significance for improving maize production efficiency and sustainable development capability.

With the gradual introduction of sensor technology and agricultural Internet of Things technology into modern agriculture, the methods for crop row centreline extraction have become increasingly diversified. Current research mainly focuses on extracting centrelines from two types of data: crop images *(Song et al., 2023)* and point clouds *(Jiang et al., 2025)*. Among these, crop point cloud data are often acquired using sensors such as light detection and ranging (LiDAR) *(Liu et al., 2024a)* and ultrasonic sensors *(Bronson et al., 2021)*. Point clouds can directly reflect the row structure of crops in three-dimensional space, but the high complexity of acquisition and processing limits their widespread application in field operations requiring high real-time performance *(Luo et al., 2021; Fan et al., 2023)*. In contrast, RGB image-based methods have become the most commonly used data source for field row extraction due to their low acquisition cost and fast processing speed. However, relying solely on colour information makes it difficult to stably distinguish crops from soil under complex conditions such as illumination changes and weed interference, necessitating enhanced crop feature representation. Traditional methods enhance crop row features through operations such as excess green segmentation *(Ban et al., 2024)* and morphological operations *(Cai et al., 2026)* on RGB images. However, these methods are susceptible to field illumination, weeds, and other factors. In complex field environments, maize seedlings are short and sparse, occupy few pixels in the image, and lack sufficient morphological and textural information, making it difficult to effectively distinguish them from soil, residual film, and early-stage weeds *(Kim et al., 2023)*.

Recently, with the continuous development of deep learning technologies, many studies have improved the accuracy of crop row centrelines by continuously refining deep learning networks and post-processing fitting methods. In terms of deep learning networks, multi-scale feature extraction modules, attention mechanisms, or lightweight convolutions have been designed to improve segmentation accuracy while maintaining real-time performance *(Wang et al., 2023)*.

Lv et al. (2026) proposed a DAC feature fusion architecture based on YOLOv11s and adopted LAMP pruning optimization. The resulting YOLO-DP network achieved 93.8% mAP_{50} in complex orchard environments. In the post-processing fitting stage, depending on different annotation strategies, feature point extraction methods also vary. Existing annotation strategies are mainly divided into two categories: annotating specific parts of each crop plant and annotating each entire crop row as a whole. Current research on annotating specific parts of each crop plant has gradually refined from the whole plant to more morphologically stable parts such as plant cores or stems *(Diao et al., 2023; Li et al., 2023)*. This method, however, incurs high annotation cost and may lead to large errors in crop row centrelines due to missed or false detections. In contrast, the method of annotating an entire crop row as a connected region can directly guide the network to learn the row structure features of crops and significantly reduce annotation cost.

Some researchers have extracted feature points by methods such as vertical cutting of pixel strips to obtain centroids or centre points *(Han et al., 2026)*. Yet, this feature point extraction method requires high segmentation accuracy. Because the row spacing of maize seedlings is narrow, adjacent rows are prone to visual adhesion, and uneven emergence or high planting density may cause local staggered interruptions *(Gao et al., 2023)*, all of which can lead to deviation of the fitted centreline. Other researchers have combined deep learning algorithms with traditional feature enhancement methods to extract feature points.

For example, the Excess Green operator and Otsu method are used to segment crops and soil background in the region of interest (ROI), and then Features from Accelerated Segment Test (FAST) corner detection is used to extract crop feature points *(Yang et al., 2023)*. Nevertheless, the FAST corner detection method *(Chen et al., 2021)* mainly targets edges and texture-rich areas and does not act on the crop main stem, which may eventually cause deviation and instability in feature point extraction. In the stage of centreline fitting based on extracted feature points, studies have used fitting algorithms such as the Least Squares Method (LSM) *(Luo et al., 2025)*, Hough Transform (HT) *(Chen et al., 2020)*, and Random Sample Consensus (RANSAC) *(Jin et al., 2025)* to extract crop centrelines. However, the LSM algorithm is sensitive to outliers; the HT algorithm has high computational complexity and introduces errors due to discretization of the parameter space; and the RANSAC algorithm suffers from fluctuating fitting results due to the instability of iterative random sampling and high computational overhead, with centreline errors mostly exceeding 3°, making it difficult to meet the high accuracy requirements of agricultural machinery navigation.

Therefore, the specific contributions of this study to address the above main problems are summarized as follows:

(a) This study provides an open-source Corn Seedling Row (CSR) dataset, which offers a data foundation for related agricultural navigation research. The dataset can be publicly accessed at https://github.com/Possibility007/Corn-Seedling-Row.

(b) This study develops an open-source deep learning model, Synergistic Attention YOLO (SA-YOLO), which can accurately and efficiently segment crop rows, providing a lightweight perception solution for autonomous visual navigation of agricultural robots.

(c) This study proposes a principal component analysis and least squares method (PCA-LSM) collaborative optimization method based on crop topological skeleton feature points for crop row centreline extraction.

(d) The code is publicly available at https://github.com/yyyyyyu-git/SA-YOLO.

2. MATERIALS AND METHODS

The proposed method consists of five main processes. First, an RGB camera was used to collect maize seedling images. The rows were annotated to establish a maize seedling row dataset. Maize seedling rows were segmented based on the SA-YOLO model. ROI extraction and Jet mapping were used to enhance maize seedling features. Finally, based on the extracted skeleton feature points, principal component analysis (PCA) was used to determine the main direction, and then LSM was used to recognize the row centrelines. A schematic overview of the method is shown in Fig. 1.

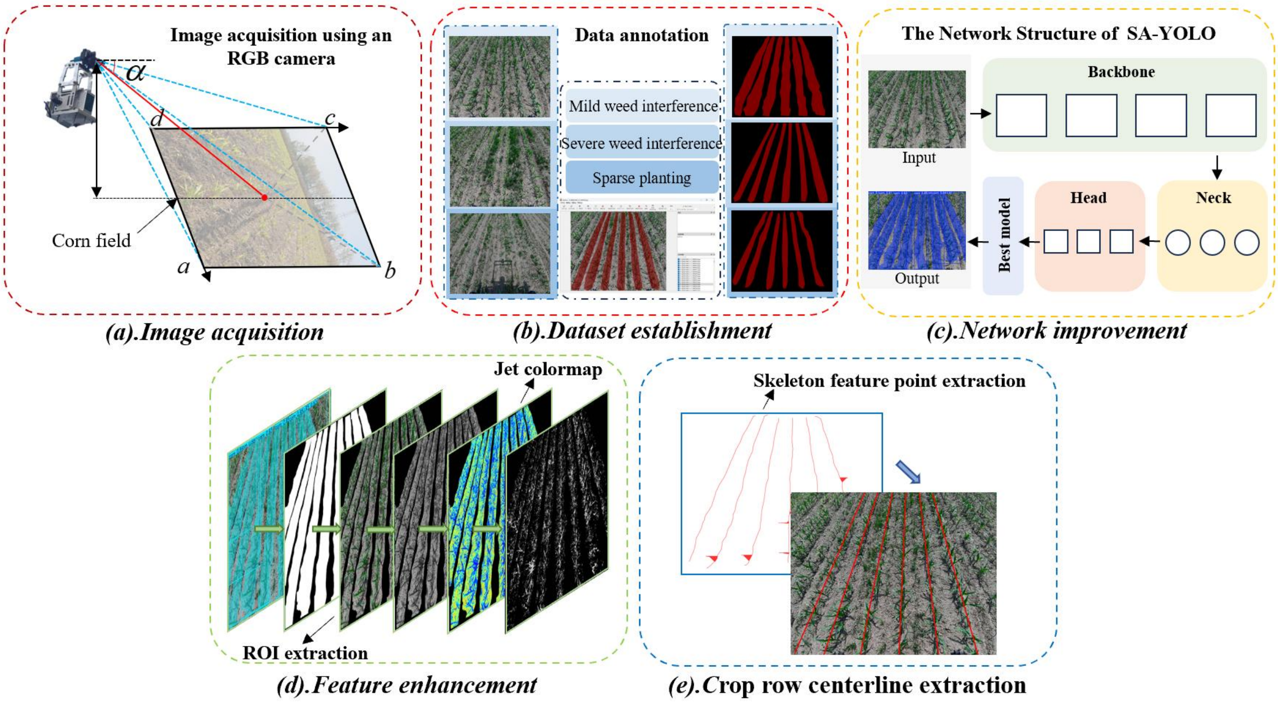


Fig. 1 - Schematic diagram of the overall workflow

2.1 Field image acquisition of maize seedlings

The maize seedling images used in this study were captured at the WanDong Comprehensive Experimental Station of Anhui Agricultural University in Mingguang County, Chuzhou City, Anhui Province (latitude 32°45′20.6″ N, longitude 118°6′55.9″ E, altitude 46 m), from July 5 to July 10, 2025. The image acquisition device was an Ailiguang ALI2312-SCH-SABA-IP67 monocular RGB camera, capturing images at 1920×1536 pixels. The camera was installed at a vertical height of 2.5 m above the ground, with a tilt angle of 50°. A total of 941 original images were collected, all saved in JPG format. The image acquisition scene and method are shown in Fig. 2.

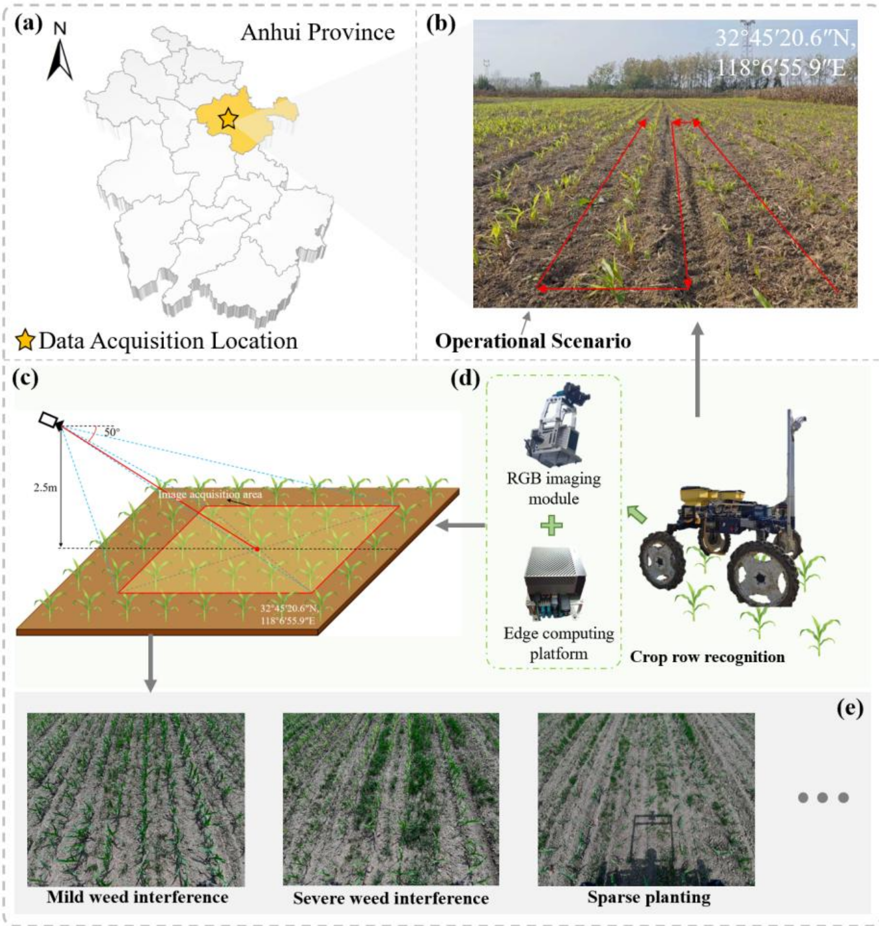


Fig. 2 - Schematic diagram of image acquisition location and method
(a) Data acquisition location; (b) Data acquisition scene; (c) Data acquisition method; (d) Data acquisition device; (b) (e) Example of acquired image.

2.2 Image Annotation and Data Preprocessing

The model in this study is an instance segmentation model for maize seedling rows. Labelme was used for image annotation, and each maize row was annotated separately. Each maize row was annotated as an independent instance, with its mask covering an entire row of maize seedlings, forming a continuous management unit. The generated .json files were converted into .txt files required for model training. After annotation, the image dataset was split for better model training. The image dataset contained a total of 941 images, which were divided into training, validation, and test sets at a ratio of 7:1:2. An annotation example for the seedling maize dataset is shown in Fig. 3, including multiple scenes such as mild weed interference, severe weed interference, and sparse planting.

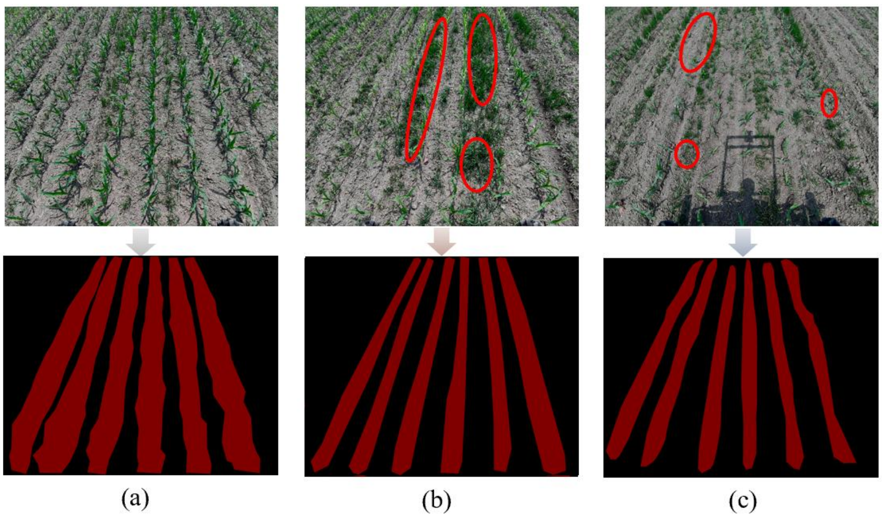


Fig. 3 - Example of dataset annotation
(a) Mild weed interference scene; (b) Severe weed interference scene; (c) Sparse planting scene.

2.3 SA-YOLO model

To address the problems of small plants, inter-row adhesion, and complex background in maize seedling row segmentation, an improved model named SA-YOLO was proposed based on YOLOv12n-seg. The SCSA attention module was introduced into the fast pooling layer of the backbone network to enhance the perception of row structure, and the designed A2C2f_SAConv module was introduced into the neck network to improve feature extraction capability, thereby improving segmentation accuracy while maintaining lightweight characteristics. The structure of the improved SA-YOLO is shown in Fig. 4.

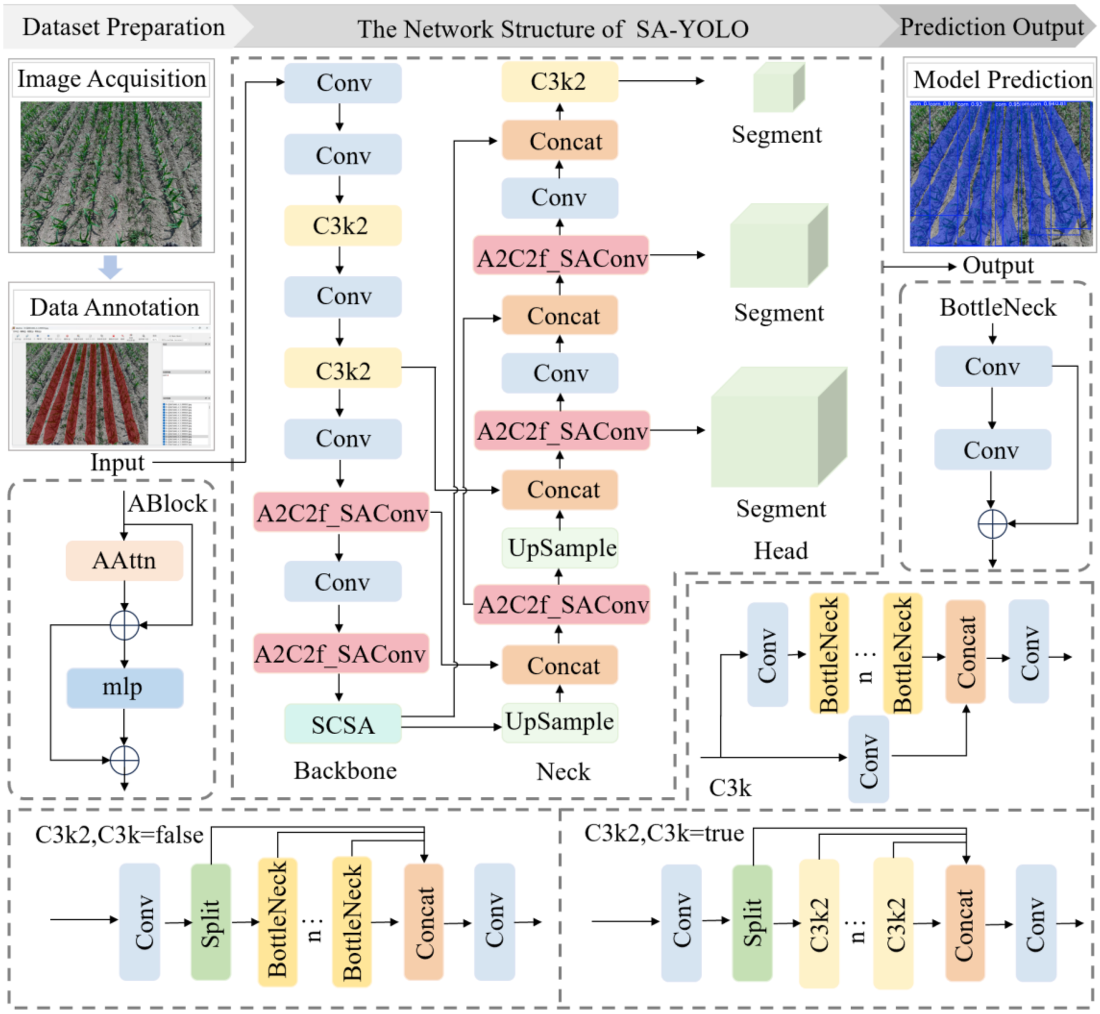


Fig. 4 - Structure of the SA-YOLO model

2.3.1 YOLOv12n-seg model

YOLOv12n-seg is a lightweight instance segmentation version of the YOLOv12 series. Its structure mainly consists of four parts: the input layer, attention-enhanced backbone, multi-scale fusion neck, and segmentation detection head. The input layer preprocesses the image to improve model robustness to different sizes and lighting conditions. The backbone network adopts the region attention module (A^2) and the residual efficient layer aggregation network (R-ELAN), reducing computational complexity while enhancing feature extraction capability. The neck network fuses multi-scale semantic and detail information through a bidirectional feature pyramid structure, improving model convergence and feature representation. The segmentation detection head integrates object localization, classification, and pixel-level segmentation to output instance bounding boxes and precise masks. Although this baseline model performs well in general scenes, in complex field environments, maize seedling row segmentation still faces challenges such as detail loss and mis-segmentation in adhesive regions, necessitating targeted improvements.

2.3.2 SCSA attention module

To enhance the model's perception of the spatial layout and key features of seedling rows, the Spatial and Channel Synergistic Attention (SCSA) module *(Si et al.,2025)* was introduced after the fast pooling layer of the YOLOv12n-seg backbone network. SCSA consists of a Shared Multi-Semantic Spatial Attention (SMSA) module and a Progressive Channel-wise Self-Attention (PCSA) module connected in series. Its structure is shown in Fig. 5.

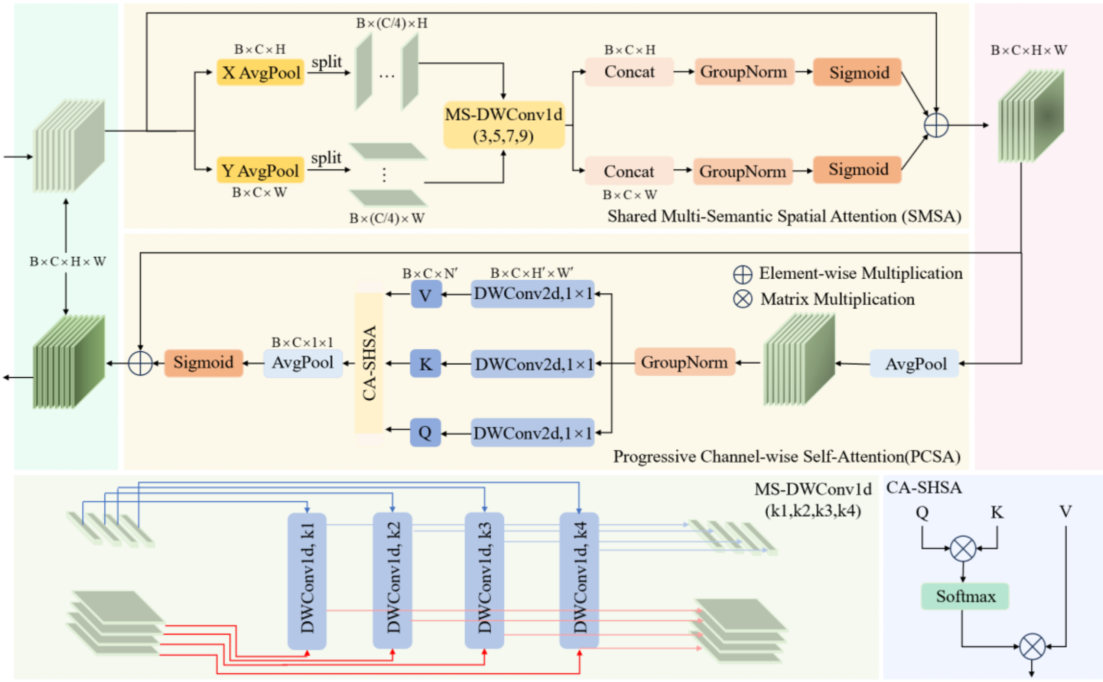


Fig. 5 - Structure of the SCSA attention mechanism module

The SMSA module first performs global average pooling on the input feature map along the height and width dimensions, obtaining two one-dimensional feature sequences. Each sequence is then evenly divided into K sub-features to extract spatial information at different semantic levels in parallel. For each sub-feature, one-dimensional depthwise separable convolutions with kernel sizes of 3, 5, 7, and 9 are used to capture multi-scale spatial context. The height and width dimension features are then aligned through shared convolution, and after concatenating the sub-features, group normalization and Sigmoid activation are applied to generate spatial attention weights. The calculation formulas are shown in (1)~(3).

$$\mathrm{Attn_H}=\sigma\left(\mathrm{GH_K^N}\left(\mathrm{Concat}\left(\mathrm{H_1^{\tilde{X}},H_2^{\tilde{X}},\cdots,H_K^{\tilde{X}}}\right)\right)\right) \tag{1}$$

$$\mathrm{Attn_W}=\sigma\left(\mathrm{GW_K^N}\left(\mathrm{Concat}\left(\mathrm{W_1^{\tilde{X}},W_2^{\tilde{X}},\cdots,W_K^{\tilde{X}}}\right)\right)\right) \tag{2}$$

$$\mathrm{SWSA(X)=X_s=Attn_H\times Attn_W\times X} \tag{3}$$

where:

$\sigma(\cdot)$ denotes Sigmoid normalization, and $GH_K^N(\cdot)$ and $GW_K^N(\cdot)$ denote group normalization along the H and W dimensions with K groups, respectively.

The PCSA module further performs channel-wise self-attention enhancement on the spatial features output by SMSA. First, progressive average pooling is applied to X_S to reduce resolution while preserving key spatial structure. Then, three independent two-dimensional depthwise convolutions are used to generate Query (Q), Key (K), and Value (V). Single-head self-attention is performed in the channel domain to model inter-channel dependencies. Finally, global pooling and Sigmoid activation are applied to generate channel attention weights, which are multiplied by X_S to obtain the final output. The calculation formulas are shown in (4)~(7).

$$\mathrm{X_p=Pool_{(7,7)}^{(H,W)\rightarrow(H',W')}(X_s)} \tag{4}$$

$$\mathrm{Q=F_{proj}^{Q}(X_p),K=F_{proj}^{K}(X_p),V=F_{proj}^{V}(X_p)} \tag{5}$$

$$\mathrm{X_{attn}=Attn(Q,K,V)=Softmax}\left(\frac{\mathrm{QK^T}}{\sqrt{\mathrm{C}}}\right)\mathrm{V} \tag{6}$$

$$\mathrm{PCSA(X_s)=X_c=X_s\times\sigma\left(Pool_{(H',W')}^{(H',W')\rightarrow(1,1)}(X_{attn})\right)} \tag{7}$$

where: $Pool_{(k,k)}^{(H,W)\to(H',W')}(\cdot)$ denotes a pooling operation with kernel size $k{\times}k$ that resizes the resolution from (H,W) to (H',W'). $Fproj(\cdot)$ denotes the linear projections for generating query, key, and value.

SMSA effectively captures multi-semantic spatial information of seedlings in complex field environments through multi-scale convolution, while PCSA enhances feature consistency and suppresses redundancy through progressive channel interaction. The SCSA module significantly improves the model's representation ability of fine seedling features and their spatial row distribution while maintaining low computational overhead, providing more robust feature input for subsequent segmentation tasks.

2.3.3 A2C2f_SAConv module

To address the limitation of standard convolution with a fixed receptive field that cannot adaptively match features at different scales, the A2C2f module in the network was replaced with an improved module incorporating Switchable Atrous Convolution (SAConv) *(Qiao et al.,2021)*, named A2C2f with Switchable Atrous Convolution (A2C2f_SAConv). Its structure is shown in Fig. 6b. SAConv dynamically adjusts the dilation rate and incorporates a context-aware mechanism, enabling the network to optimize multi-scale feature fusion.

The core idea of SAConv is to extend the traditional static atrous convolution into a dynamic, spatially adaptive feature extraction mechanism. Its structure is shown in Fig. 6a. It employs multiple convolution branches with different dilation rates in parallel and dynamically fuses the outputs of each branch through a lightweight, pixel-wise learned switch function, allowing each spatial position to autonomously select the most suitable receptive field based on local context. In addition, SAConv integrates global context preprocessing and postprocessing modules, which perform global average pooling and convolution to calibrate features globally, further improving feature consistency and discriminability. The calculation formulas are shown in (8) and (9).

$$\mathrm{S(x)=\sigma\left(Conv_{1\times1}\left(AvgPool_{5\times5}(x)\right)\right)} \tag{8}$$

$$\mathrm{y=S(x)\cdot Conv(x,w,1)+\left(1\text{-}S(x)\right)\cdot Conv(x,w+\Delta w,r)} \tag{9}$$

where:

The switch function *S(x)* is generated by a 5×5 average pooling layer followed by a 1×1 convolutional layer and normalized by the activation function σ. Conv(x,w,r) denotes the convolution operation with weight *w*, hyperparameter *r*, and trainable weight Δw.

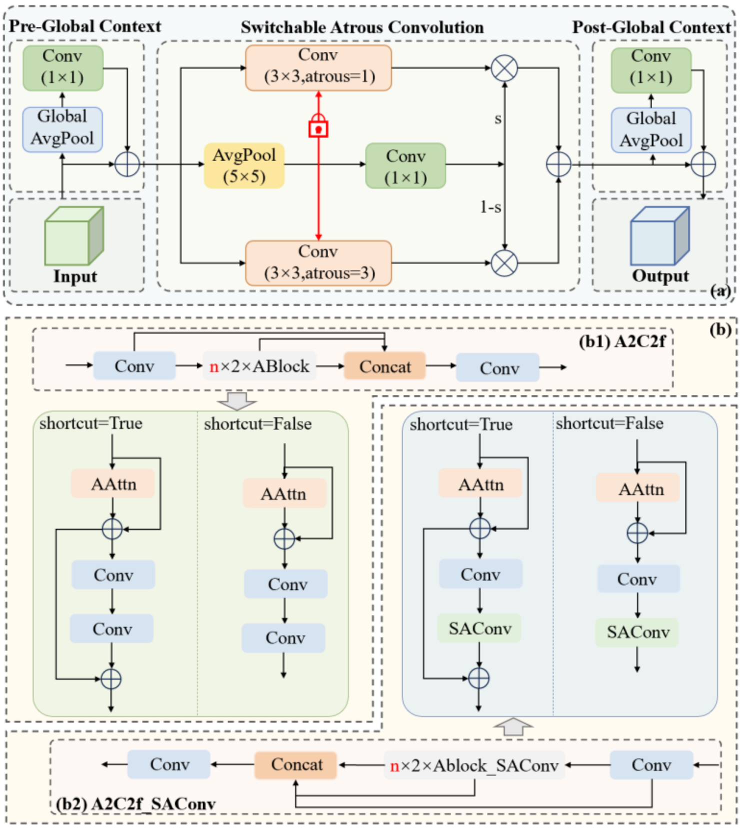


Fig. 6 - Structure of the SAConv module

(a) SAConv structure; (b) Difference between the A2C2f module and the A2C2f_SAConv module.

2.4 Crop row centreline fitting

In the crop row centreline fitting process, contour detection is a key factor affecting the accuracy of the fitted centreline. In this study, the segmentation results were used to extract the ROI region, and the skeleton feature points were extracted from the binary image of crops within the ROI region. Finally, PCA was used to determine the main direction, and LSM was used to fit these skeleton feature points to obtain the centrelines of maize seedling rows.

2.4.1 Crop row skeleton feature point extraction

The improved SA-YOLO model was used to infer the positions of maize seedling rows through instance segmentation. Masks were generated from the inferred segmentation regions, and bitwise operations were performed with the original image to extract the ROI region *(Luz et al.,2024)*, which eliminated a large amount of interference information and saved computational cost. The extraction process is shown in Fig. 7. After the ROI region was extracted, grayscale processing was applied to the image *(Rodríguez et al., 2021)*. However, because the colour features of maize seedlings and ground soil are similar in grayscale images, direct binarization made it difficult to effectively segment maize. Pseudo-colour processing technology *(Bayareh et al.,2021)* significantly enhances image distinguishability by assigning different colour maps to different gray values; this method has been widely used in fields such as medical imaging and meteorological analysis. Among them, the Jet mapping, as a typical pseudo-colour algorithm, is often used to highlight detailed features in images due to its high colour contrast. After Jet mapping, the visual features of the maize seedling regions were significantly enhanced, creating favourable conditions for subsequent processing. After maize seedling feature enhancement, binarization was applied to the image *(Surulivelu et al.,2026)*. Skeleton feature points were extracted from the binarized image. A morphological skeletonization method was used to iteratively delete edge pixels while preserving region connectivity, gradually thinning the crop rows into a single-pixel-wide topological skeleton. This process not only extracted the central feature points of the crop rows but also fully preserved their overall directional trend. The resulting skeleton feature point set was continuous, dense, and geometrically meaningful, faithfully reflecting the morphological characteristics of the crop rows.

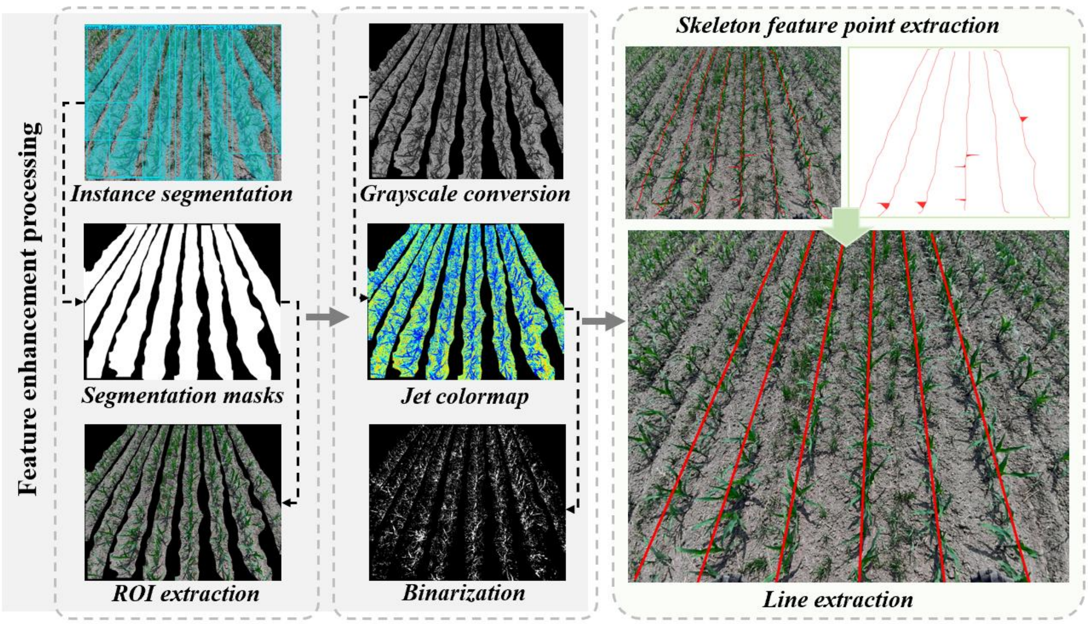


Fig. 7 - Flowchart of crop row centreline fitting

2.4.2 Crop row centreline fitting

PCA robustly estimates the dominant direction of a point set, which corresponds to the eigenvector associated with the largest eigenvalue. However, because the fitted straight line is determined only by the direction of the first principal component without optimizing errors in specific directions, it remains sensitive to outliers. Using LSM alone, although it minimizes the error along a preset independent variable direction, if that direction does not match the actual deviation direction of interest, the resulting line fails to accurately reflect the geometric characteristics of the data. Moreover, its robustness to outliers is poor, and a few outliers can cause the fitting result to deviate significantly from the true trend.

Therefore, this study proposed a collaborative PCA-LSM method. The skeleton feature points were used to fit a straight line that best represents the crop rows. First, PCA was used to determine the main direction of the point set. Then, based on the relative inclination angle between this main direction and the coordinate axes, an appropriate LSM form was selected. This method combines the directional robustness of PCA with the precise fitting capability of LSM, providing better tolerance to outliers and enabling extraction of crop row centrelines that conform to the overall distribution of the data. A schematic comparison of the three methods is shown in Fig. 8.

The calculation formula for the PCA principal direction is shown in Equation (10):

$$C=\frac{1}{N-1}\sum_{i=1}^{N}(p_i-\bar{p})(p_i-\bar{p})^T \tag{10}$$

where: $\bar{P}$ is the centroid of the point set, $P_i=(x_i,y_i)$, and the principal direction vector is denoted as $v=(v_x,v_y)$.

The method for selecting the appropriate least squares form is shown in Equation (11):

$$S_{\in}^2=\begin{cases}\min\limits_{k,b}\sum_{i=1}^{N}\left(x_i-(ky_i+b)\right)^2, & \text{if } |v_y|>|v_x| \\ \min\limits_{m,b}\sum_{i=1}^{N}\left(y_i-(mx_i+b)\right)^2, & \text{if } |v_y|<|v_x|\end{cases} \tag{11}$$

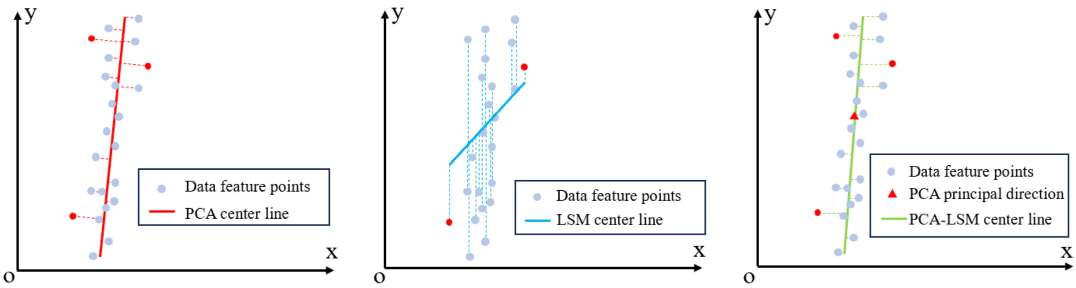


Fig. 8 - Schematic diagrams of three methods

(a) Schematic diagram of the PCA method; (b) Schematic diagram of the LSM method; (c) Schematic diagram of the PCA-LSM method.

3. RESULTS AND DISCUSSION

3.1 Training environment and evaluation metrics

3.1.1 Parameter settings and model training

The model training platform in this experiment used an NVIDIA GeForce RTX 4050 GPU, a 13th Gen Intel(R) Core(TM) i7-13620H processor, 6 GB of video memory, and the Windows 11 operating system. The programming language was Python 3.12, the software platform was PyCharm 2023.2.4, the PyTorch framework version was 2.0.0, and the CUDA version was 11.8. Stochastic gradient descent was used for training. The image resolution was set to 640×640 pixels, the momentum was 0.937, and the weight decay coefficient was 0.0005. The initial learning rate was set to 0.01, and the number of training epochs was 100.

3.1.2 Segmentation evaluation

To evaluate the performance of the improved network for maize seedling row segmentation in field scenes, metrics at the mask level, including Precision (P), Recall (R), Average Precision (AP_{50}), and mask Intersection over Union (IoU_{mask}), were adopted to assess model performance. Among them, AP_{50} is the area under the precision-recall curve, which comprehensively measures the model's detection capability for a single class and serves as the core metric for evaluating overall network performance. The calculation formulas for the specific parameters are shown in Equations (12)~(15).

$$P=\frac{TP}{TP+FP} \tag{12}$$

$$R=\frac{TP}{TP+FN} \tag{13}$$

$$AP_i=\int_0^1 P_i(R_i)dR_i \tag{14}$$

$$IoU_{mask}=\frac{Area(Mask_{predicted}\cap Mask_{ground\_truth})}{Area(Mask_{predicted}\cup Mask_{ground\_truth})} \tag{15}$$

where:

TP (True Positive), FP (False Positive), and FN (False Negative) in Equations (12)~(14) are defined at the instance level: TP is the number of correctly segmented maize seedling row instances, FP is the number of incorrectly segmented instances, and FN is the number of missed instances. Equation (15) computes the pixel-level mask IoU for a single matched instance using mask areas.

3.1.3 Centreline extraction evaluation

To quantify the accuracy of the extracted crop row centrelines, evaluation metrics based on the lateral error and angle error between the predicted centreline and the ground truth centreline were constructed. The test image set was defined as S_{images}, the total number of images was $k=|S_{images}|$, i=1, 2, …, k. The mean normalized lateral error (L_{mean}^{norm}), the standard deviation of lateral error (L_{std}^{norm}), the mean angle error (A_{mean}), and the standard deviation of angle error (A_{std}) were used to reflect the average deviation level of the predicted navigation line relative to the ground truth centerline across all images. Smaller values indicate higher overall fitting accuracy. The calculation principle of this method is shown in Fig. 9. The formulas for comprehensive evaluation of navigation line accuracy and stability are shown in Equations (16)~(19).

$$\mathrm{L}_{\mathrm{mean}}^{\mathrm{norm}}=\frac{1}{\mathrm{k}}\sum_{\mathrm{s}\in\mathrm{S}_{\mathrm{images}}}\mathrm{L}_{\mathrm{s}}^{\mathrm{norm}} \tag{16}$$

$$\mathrm{L}_{\mathrm{std}}^{\mathrm{norm}}=\sqrt{\frac{1}{\mathrm{k}}\sum_{\mathrm{s}\in\mathrm{S}_{\mathrm{images}}}(\mathrm{L}_{\mathrm{S}}^{\mathrm{norm}}\text{-}\mathrm{L}_{\mathrm{mean}}^{\mathrm{norm}})^2} \tag{17}$$

$$\mathrm{A}_{\mathrm{mean}}=\frac{1}{\mathrm{k}}\sum_{\mathrm{i}=1}^{\mathrm{k}}\theta_{\mathrm{i}} \tag{18}$$

$$\mathrm{A}_{\mathrm{std}}=\sqrt{\frac{1}{\mathrm{k}\text{-}1}\sum_{\mathrm{i}=1}^{\mathrm{k}}(\theta_{\mathrm{i}}\text{-}\mathrm{A}_{\mathrm{mean}})^2} \tag{19}$$

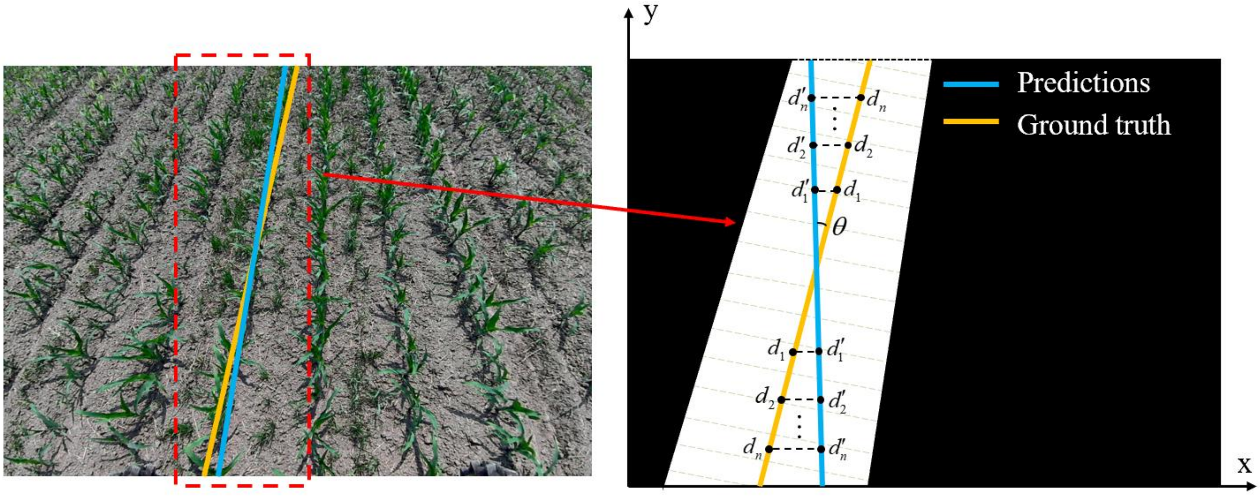


Fig. 9 - Schematic diagram of centreline extraction accuracy evaluation

3.2 Comparison of different segmentation models

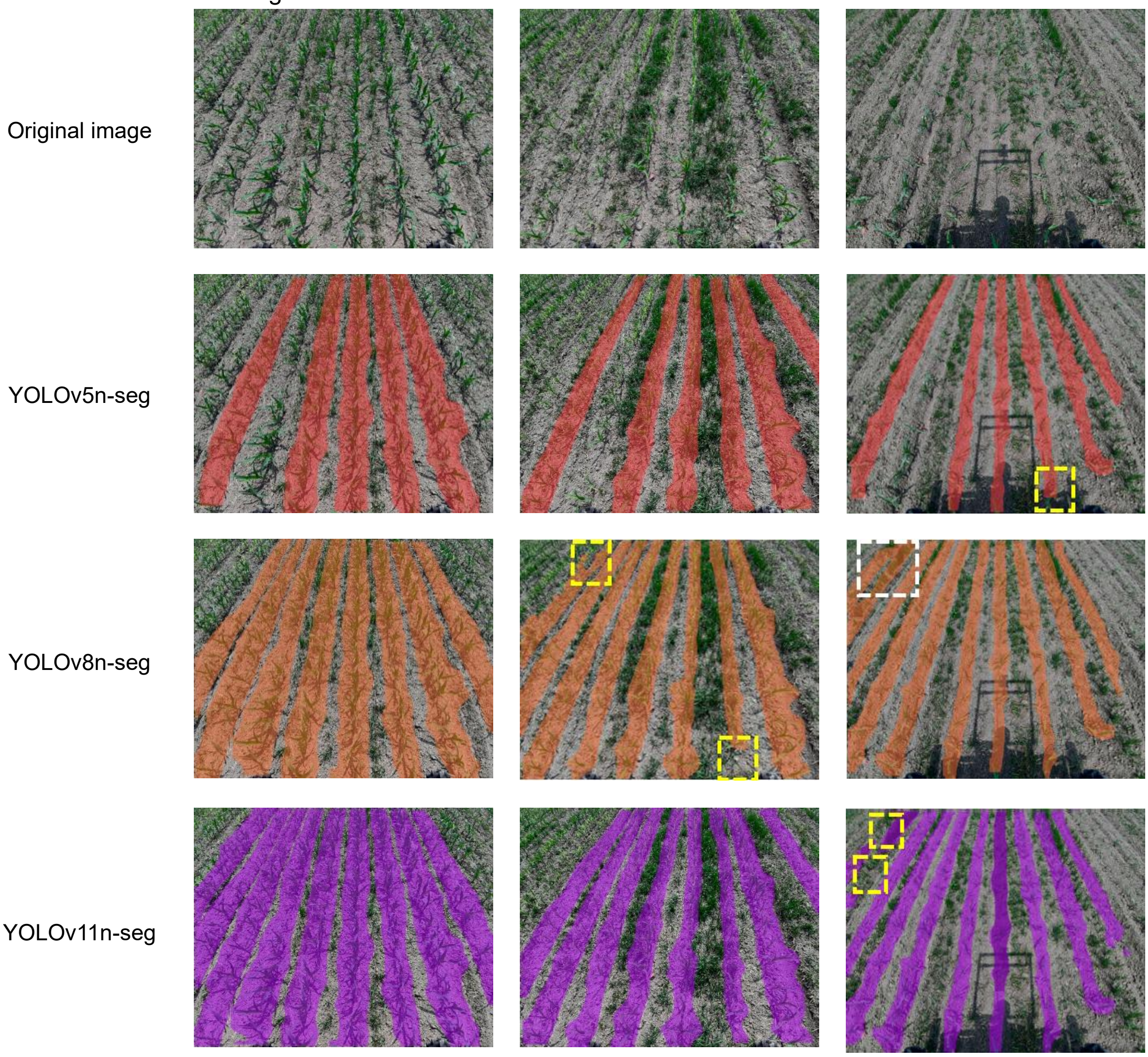
To validate the performance of the SA-YOLO model, comparative experiments were conducted against YOLOv5n-seg, YOLOv8n-seg, YOLOv11n-seg, and the baseline YOLOv12n-seg in a complex maize field environment at the seedling stage. The results are shown in Table 1. The SA-YOLO model achieved the best segmentation accuracy, with an AP_{50} of 99.1% and an IoU_{mask} of 83.8%, which was 1.5% higher than the suboptimal YOLOv8n-seg model. It also maintained leading performance in both precision and recall, indicating that the model can more completely and accurately identify maize row regions. From a lightweight perspective, although YOLOv5n-seg had the smallest number of parameters, its AP_{50} was relatively low. The YOLOv11n-seg model achieves the same *P* and AP_{50} as the SA-YOLO model, but its IoU_{mask} is 2.0 percentage points lower. Under a lightweight configuration with only 3.72 M parameters and 8.9 Giga Floating Point Operations (GFLOPs) of computation, SA-YOLO achieved an inference speed of 161.29 Frames Per Second (FPS), which was an improvement over the baseline YOLOv12n-seg's 153.85 FPS. Compared with the baseline YOLOv12n-seg, SA-YOLO increased AP_{50} by 0.3 percentage points and IoU_{mask} by 5.0 percentage points, validating the effectiveness of the proposed improved modules in enhancing feature extraction capability and segmentation accuracy.

Table 1

Comparison of detection performance of different network models

Model	P (%)	R (%)	AP_{50} (%)	AP_{50-95} (%)	IoU_{mask} (%)	FPS	Inference Times (ms)	Parameters (M)	GFLOPs	Model size (MB)
YOLOv5n-seg	96.5	94.7	98.5	65.2	82.2	117.65	8.5	**1.88**	**6.7**	**4.1**
YOLOv8n-seg	**96.6**	98.0	98.9	**66.8**	82.3	**232.56**	**4.3**	3.26	12.0	6.8
YOLOv11n-seg	96.5	**98.4**	**99.1**	65.2	81.8	185.19	5.4	2.83	10.2	6.0
YOLOv12n-seg	95.6	97.4	98.8	64.2	78.8	153.85	6.5	2.81	10.2	6.1
SA-YOLO	96.5	98.0	**99.1**	64.0	**83.8**	161.29	6.2	3.72	8.9	7.8

Maize seedling row images from different scenes, including mild weed interference, severe weed interference, and sparse planting, were randomly selected from the dataset for recognition experiments. As shown in Fig. 10, under the mild weed interference scene, the SA-YOLO model successfully segments 9 rows of maize seedlings, while other models all suffered from missed segmentation. Under the severe weed interference scene, YOLOv8n-seg had incomplete segmentation boundaries within the yellow box, affecting row continuity, and YOLOv12n-seg had false detection of weeds as maize rows within the red box. Under the sparse planting scene, the models exhibited blurred boundaries. YOLOv8n-seg and SA-YOLO showed some adhesion between rows within the white box. Across the three scenes, SA-YOLO successfully segmented the most rows of maize seedlings.



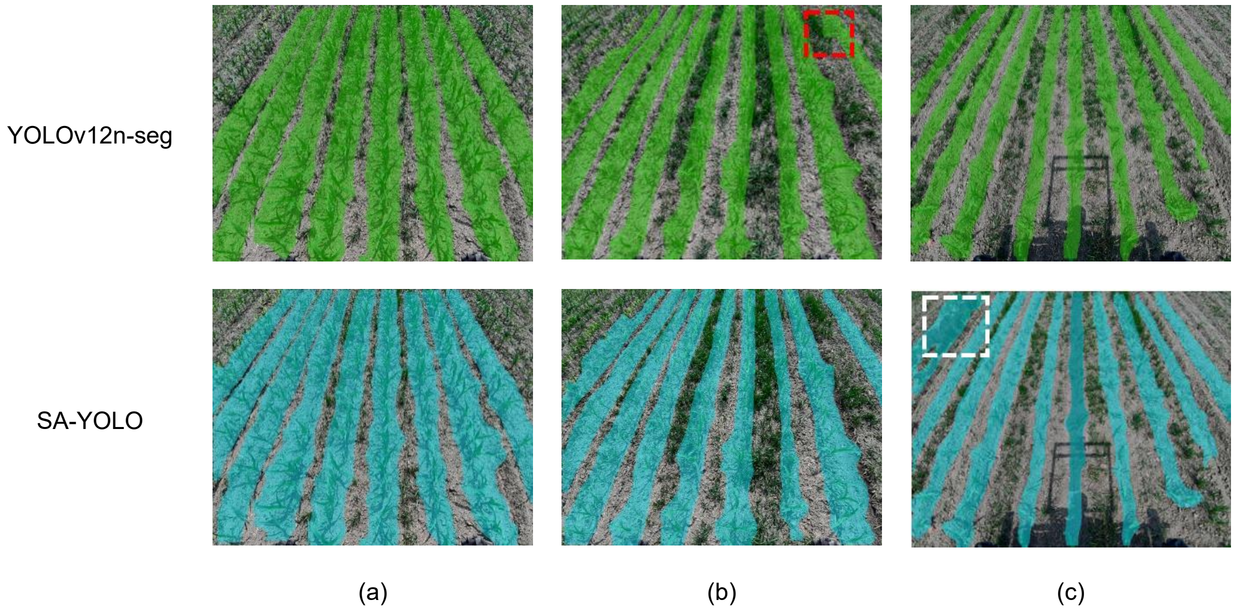


Fig. 10 - Segmentation results of five models under different scenes
(a) Segmentation results in mild weed interference scene; (b) Segmentation results in severe weed interference scene; (c) Segmentation results in sparse planting scene. The yellow boxes indicate incomplete segmentation regions, the white boxes indicate inter-row adhesion regions, and the red boxes indicate false segmentation of weeds.

Based on the above data, the improved SA-YOLO model in this study can accurately segment maize seedling rows in the above scenes, demonstrating that the introduction of the SCSA attention mechanism and the designed A2C2f_SAConv module can achieve better maize row segmentation accuracy in complex field environments, providing theoretical and technical support for the application of agricultural navigation systems *(Lopes et al., 2025)*.

3.3 Ablation experiments

To validate the effectiveness of each improved module, ablation experiments were conducted by sequentially introducing A2C2f_SAConv, SCSA and CBAM modules into the YOLOv12n-seg baseline model. The results are shown in Table 2. The baseline model achieved AP_{50}, AP_{50-95}, and IoU_{mask} of 98.8%, 64.2%, and 78.8%, respectively, on the test set. After introducing only the A2C2f_SAConv module, AP_{50} increased to 99.1%, IoU_{mask} increased by 5.0 percentage points to 83.8%, and the parameter count increased by only 0.91 M, indicating that this module effectively enhances multi-scale feature extraction capability through switchable atrous convolution while maintaining lightweight advantages *(Hu et al., 2025)*. Adding only the SCSA attention module achieved an AP_{50} of 98.9% and an IoU_{mask} of 81.6% with no increase in parameter count, validating the effectiveness of the SCSA spatial and channel synergistic attention mechanism. In contrast, introducing only the Convolutional Block Attention Module (CBAM) module gave an AP_{50} of 98.9% and an IoU_{mask} of 82.5%, with a slight increase in parameters *(Arthur et al., 2025; Guo et al., 2023)*.

Table 2

Ablation experiment results of YOLOv12n-seg

A2C2f_SA Conv	SCSA	CBAM	P (%)	R (%)	AP_{50} (%)	AP_{50-95} (%)	IoU_{mask} (%)	FPS	Inference Times (ms)	Parameters (M)	GFLOPs	Model size (MB)
×	×	×	95.6	97.4	98.8	64.2	78.8	153.85	6.5	**2.81**	10.2	**6.1**
√	×	×	96.3	**98.3**	**99.1**	64.6	**83.8**	178.57	5.6	3.72	8.9	7.7
×	√	×	95.9	98.1	98.9	64.6	81.6	163.93	6.1	**2.81**	10.2	**6.1**
×	×	√	95.6	98.0	98.9	**65.4**	82.5	**196.08**	**5.1**	2.83	9.6	**6.1**
√	√	×	96.5	98.0	**99.1**	64.0	**83.8**	161.29	6.2	3.72	8.9	7.8
√	×	√	**97.2**	96.8	99.0	64.9	83.5	181.82	5.5	3.73	**8.3**	7.8

Although optimization with a single module can improve model performance, more significant breakthroughs rely on synergy and combination among modules *(Liu et al.,2022; Shao et al., 2026; Gao et al., 2024; Meng et al., 2024)*. When both A2C2f_SAConv and SCSA were introduced, the model achieved the best performance, with AP_{50} of 99.1%, IoU_{mask} of 83.8%, *P* and *R* of 96.5% and 98.0%, respectively, and an inference speed of 161.29 FPS. However, introducing both A2C2f_SAConv and CBAM did not bring gains and reduced recall by 1.2 percentage points. Fig. 11 shows the performance curves of the ablation experiments, intuitively demonstrating the improvement in segmentation performance before and after the improvements.

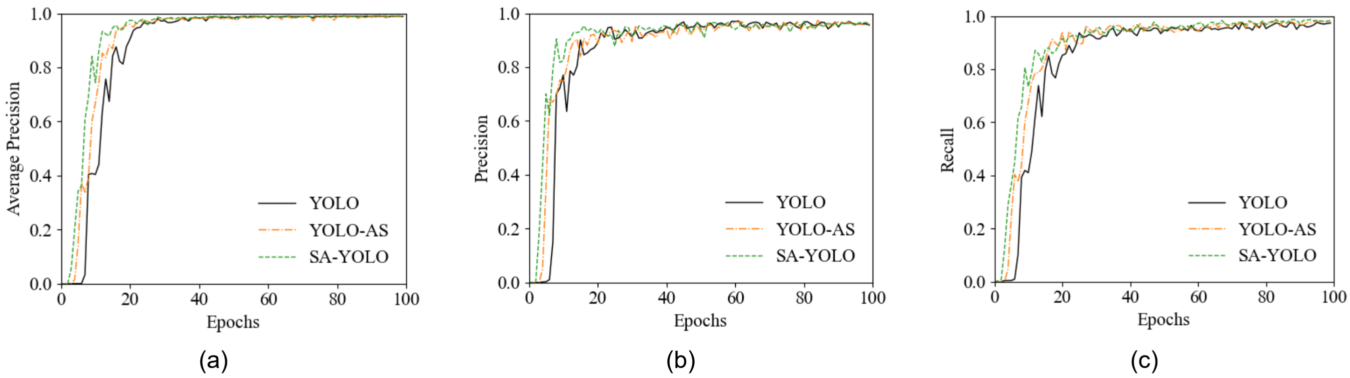


Fig. 11 - Performance curve comparison of improved algorithms
(a) Comparison of average precision curves of YOLOv12; (b) Comparison of precision curves of YOLOv12; (c) Comparison of recall curves of YOLOv12. YOLO refers to the YOLOv12n-seg model, and YOLO-AS refers to YOLOv12n-seg with the A2C2f_SAConv module.

To further visually compare and evaluate segmentation accuracy and understand the model's decision basis, heatmaps of maize seedling row image detection were used, as shown in Fig. 12. It can be seen that the heatmap of the SA-YOLO model is closer to the actual maize seedling row regions. In Fig. 12a, the boundaries indicated by the two straight lines are the boundaries that the model should accurately segment. Fig. 12b shows the heatmap of the YOLOv12n-seg baseline model, where the soil regions between adjacent crop rows exhibited weak boundary discrimination ability, leading to blurred boundaries between rows and incomplete segmentation at farther distances. Fig. 12c and Fig. 12d are heatmaps of the models with the A2C2f_SAConv convolution module and the SCSA attention mechanism module, respectively. The introduction of the A2C2f_SAConv module provided more complete activation of long maize seedling rows by expanding the receptive field, and the heatmap values at the far right were higher than those of the baseline model. The introduction of the SCSA attention mechanism module helped the model better distinguish the difference between maize seedling rows and the background, resulting in higher heatmap values at the maize seedling rows and clearer differentiation at the boundaries. Fig.12e shows the heatmap of the SA-YOLO model. Compared with the YOLOv12n-seg baseline model, the boundaries are clearer and the activation of maize seedling rows is more complete.

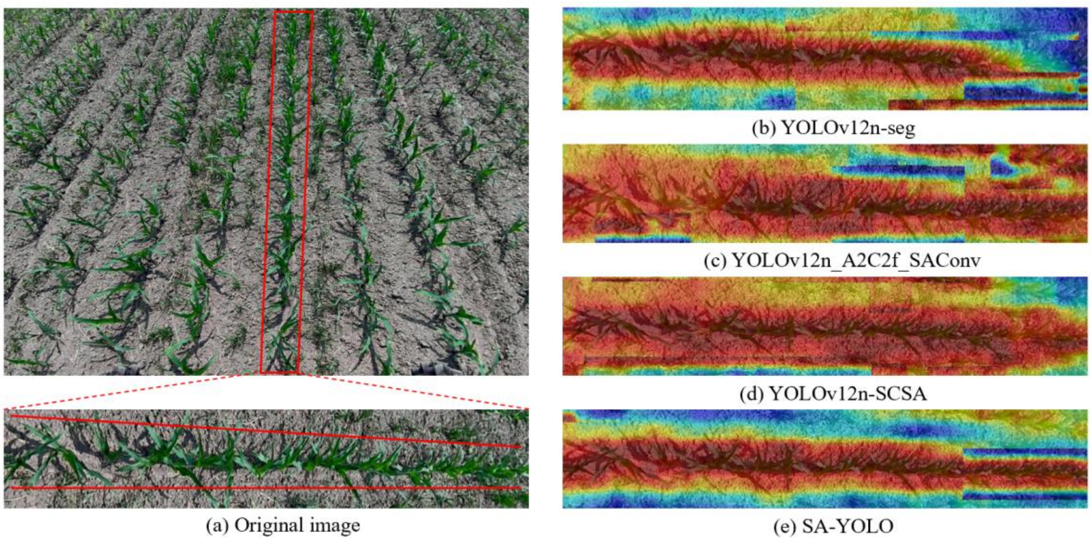


Fig. 12 - Heatmaps of maize seedling row image detection
The red regions represent the areas of greatest focus for the model, the yellow regions represent areas of secondary importance, and the blue regions represent areas with the least influence, typically representing redundant information.

3.4 Crop row centreline detection experiments

3.4.1 Crop row centreline detection using different methods

From a test dataset of 189 maize seedling row images, crop row centrelines were extracted using PCA *(Zhang et al.,2024a)*, LSM *(Ju et al.,2024)*, and the proposed PCA-LSM method, and were also compared with the traditional HT method *(Chen et al.,2025)* and the RANSAC method *(Zhang et al.,2024b)*. The quantitative results are shown in Table 3. As shown in the table, the PCA method achieved a A_{mean} of 0.613°, a L_{mean}^{norm} of 0.359%, and a mean pixel lateral error of 6.90 pixels. The LSM method achieved a A_{mean} of 0.590°, a L_{mean}^{norm} of 0.357%, and a mean pixel lateral error of 6.85 pixels. The results of the PCA-LSM method were exactly the same as those of the LSM method, with all metrics identical. The reason for this is that the maize rows in the test images were approximately vertical, causing PCA-LSM to always select the vertical fitting form, which is mathematically equivalent to LSM. Although PCA was slightly better in stability, LSM had a slight advantage in mean values. Overall, all three methods could effectively extract crop row centrelines, with angle errors all below 0.7° and pixel errors below 7 pixels, meeting the basic accuracy requirements for visual navigation of agricultural robots. The PCA-LSM method automatically degenerated to LSM on the vertical row dataset, verifying the correctness of its adaptive selection.

Table 3 Comparison of crop row centreline detection results using different methods

Fitting method	Scene	Mean angle error (°)	Mean normalized lateral error (%)	Mean pixel lateral error (px)
HT *(Chen et al.,2025)*	Maize seedling	2.600	-	10.78
RANSAC *(Zhang et al.,2024b)*	Rice transplanting	0.776	0.730	4.69
PCA	Maize seedling	0.613	0.359	6.90
LSM	Maize seedling	0.590	0.357	6.85
This study	Maize seedling	0.590	0.357	6.85

Compared with the RANSAC method, its A_{mean} of 0.776° and L_{mean}^{norm} of 0.730% were both higher than those of the proposed algorithm, while its mean pixel lateral error of 4.69 pixels was better than that of our method because of the different image resolutions. The image size in this study was 1920×1536 pixels, while the image size in that study was 1280×1024 pixels. Under the same maize seedling scene, the HT method achieved a A_{mean} of 2.600°, far higher than that of the PCA-LSM method used in this study. The reason for this is discussed as follows. Most fitting algorithms are based on object detection methods, using the center point or bottom midpoint of the detection box as feature points, which may introduce fitting errors due to discrete or missing plants. Alternatively, instance segmentation or semantic segmentation methods obtain feature points by calculating the centroids of connected regions or directly extract the skeleton of inter-row regions, both of which can cause overall deviation of the navigation line due to segmentation breaks or mis-segmentation. Centreline extraction methods that directly act on the crop itself usually use corner detection algorithms such as FAST to extract feature points. The feature points extracted by such methods are mainly from the leaf regions of the crop rather than the main stem regions of the crop rows, ultimately resulting in fitted centrelines that deviate from the crop main stem and are unstable. Compared with existing mainstream methods, the row-unit instance segmentation annotation adopted in this study provides a more structurally consistent learning target for the model. Meanwhile, the skeleton feature point extraction method proposed in this study directly acts on the crop itself, using an image thinning algorithm to extract the topological midline of the crop. This method not only preserves the complete geometric orientation information of the crop rows but also has a natural ability to suppress noise at segmentation edges. Traditional methods require high segmentation accuracy, whereas this study only uses segmentation to extract the ROI region to reduce noise interference, avoiding overall deviation of the navigation line due to segmentation breaks or mis-segmentation, and providing a feature point set with more uniform spatial distribution and higher geometric fidelity for subsequent navigation line fitting.

3.4.2 Crop row centreline detection of different models in complex scenes

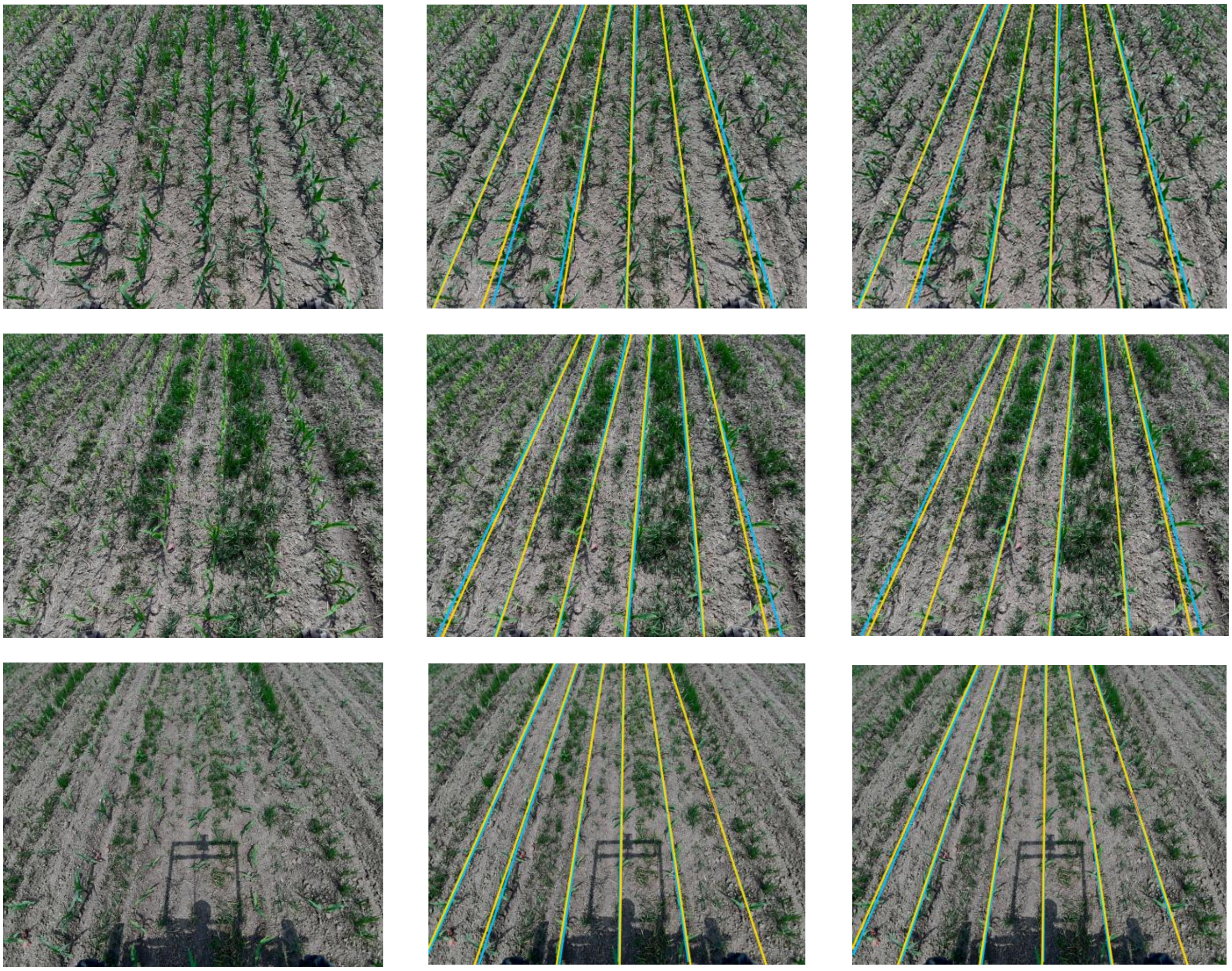
To validate the effectiveness of the proposed SA-YOLO model in crop row centreline extraction, comparative experiments were conducted against the baseline YOLOv12n-seg model. The error statistics of the two models are shown in Table 4.

Table 4

Comparison of crop row centreline detection results of different models

Model	Angle error (°)		Normalized lateral error (%)	
	Mean	Std	Mean	Std
YOLOv12n-seg	0.622	0.530	0.379	0.433
SA-YOLO	0.590	0.498	0.357	0.428

The proposed SA-YOLO model outperformed the baseline YOLOv12n-seg model in all metrics. In terms of angle error, the mean of SA-YOLO decreased from 0.622° to 0.590°, a relative reduction of 5.1%; the standard deviation decreased from 0.530° to 0.498°, a relative reduction of 6.0%. This indicates that the row lines extracted by SA-YOLO are closer to the ground truth rows, with better stability across different scenes. In terms of normalized lateral error, the mean of SA-YOLO decreased from 0.379% to 0.357%, a relative reduction of 5.8%. The standard deviation remained essentially unchanged, indicating that SA-YOLO improved accuracy while maintaining stability. The introduction and design of the SCSA attention module and the A2C2f_SAConv module enriched contextual information and preserved shallow details, enabling the model to more accurately locate crop row centres. Fig. 13 shows the crop row centreline detection performance of YOLOv12n-seg and SA-YOLO in mild weed interference, severe weed interference, and sparse planting scenes, respectively.



(a) (b) (c)

Fig. 13 - Crop row centreline detection results under different scenes

(a) Detection scenes; (b) Detection results of YOLOv12n-seg model; (c) Detection results of SA-YOLO model.

3.4.3 Comparison with related studies

To further evaluate the superiority of the proposed navigation method, our method was compared with other related studies. As shown in Table 5, under field crop environments, our method achieved lower mean angle error and mean normalized lateral error than other state-of-the-art deep learning methods. In addition, it is worth noting that the mean pixel errors of the navigation line generation methods using YOLOX-Nano and Faster R-CNN architectures in jujube orchard and kiwifruit orchard environments, respectively *(Zheng et al., 2023; Ma et al., 2021)*, were slightly lower than our mean lateral pixel error in the maize seedling field. This phenomenon can be explained by the fact that the field scene is more complex than orchards, with weaker illumination, more occlusion, and unstructured environments. The main reason is also the larger image width in this study compared with those studies. Compared with other advanced methods *(Lai et al., 2026)*, the proposed method has stronger robustness and generalization ability and can meet the navigation requirements of maize seedling fields.

Table 5

Performance comparison of different crop row line extraction methods

Reference	Scene	Network	Mean angle error (°)	Mean normalized lateral error (%)	Mean pixel lateral error (px)
(Kong et al., 2025)	Heading rice	SEU-Enet	1.910	-	-
(Lai et al., 2026)	Multiple seedling crops	YOLO-SSR	-	0.720	-
(Zheng et al., 2023)	Jujube orchard	YOLOX-Nano	2.550	0.700	4.49
(Ma et al., 2021)	Kiwifruit orchard	Faster R-CNN	3.200	0.340	5.10
This study	Maize seedling	SA-YOLO	0.590	0.357	6.85

Although the proposed method has clear advantages in accuracy, there are still some limitations. The parameter count of the model is 3.72 M, which is higher than that of the SEU-Enet model of *Kong et al. (2025)* (approximately 0.094 M), leaving room for optimization in extreme lightweight scenarios. This study mainly focused on the accuracy analysis of navigation line extraction and has not yet performed end-to-end real-time testing on actual agricultural machinery platforms.

4. CONCLUSIONS

In this study, to address the problem of crop row centreline extraction for visual navigation of agricultural robots, the SA-YOLO model was constructed, and a crop-based skeleton feature point extraction method and a corresponding crop row centreline extraction method were designed. The model enhanced feature extraction capability by introducing the SCSA attention module and replacing the original A2C2f module with the designed A2C2f_SAConv module, achieving accurate segmentation of crop row regions. In the crop row centreline extraction stage, the constructed skeleton feature point extraction and PCA-LSM joint optimization method effectively improved the accuracy of crop row centrelines. The experimental results showed that the IoU_{mask}, AP_{50}, P, and R scores of the SA-YOLO model were 83.8%, 99.1%, 96.5%, and 98.0%, respectively. Compared with the other four models, it exhibited higher recognition accuracy while maintaining faster segmentation speed. Second, regarding the proposed algorithm, the A_{mean} is reduced to 0.590° and its A_{std} to 0.498°, while the L_{mean}^{norm} is reduced to 0.357% and its L_{std}^{norm} to 0.428%. All the above metrics satisfy the precision requirements of autonomous agricultural machinery. However, several aspects of the proposed model require further improvement. The validation of the model was mainly focused on the crop seedling stage, and its applicability in other growth stage scenes needs further verification. Therefore, future work will include more crop species, more growth stages, and more samples from complex environments to promote the practical application of the algorithm.

5. ACKNOWLEDGEMENT

This work was supported by Opening Fund of National Key Laboratory of Agricultural Equipment Technology (SKL-2025-005), Scientific Research Start-up Foundation of Anhui Agricultural University (rc412405).

6. REFERENCES

[1] Arthur L., Mahnan S., He L., Hussain M., Heinemann P., Brunharo C., (2025). YOLOv7-CBAM and DeepSORT with pixel grid analysis for real-time weed localization and intra-row density estimation in apple orchards, *Computers and Electronics in Agriculture*, vol.239, article no.111071, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2025.111071

[2] Ban C., Wang L., Chi R., Su T., Ma Y., (2024). A camera-LiDAR-IMU fusion method for real-time extraction of navigation line between maize field rows, *Computers and Electronics in Agriculture*, vol.223, article no.109114, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2024.109114

[3] Bayareh Mancilla R., Tấn B., Daul C., Gutiérrez Martínez J., Leija Salas L., Wolf D., Vera Hernández A., (2021). Anatomical 3D modeling using IR sensors and radiometric processing based on structure from motion: Towards a tool for the diabetic foot diagnosis, *Sensors*, vol.21, no.11, article no.3918, Basel/Switzerland. DOI: https://doi.org/10.3390/s21113918

[4] Bronson K.F., French A.N., Conley M.M., Barnes E.M., (2021). Use of an ultrasonic sensor for plant height estimation in irrigated cotton, *Agronomy Journal*, vol.113, no.2, pp.2175-2183, Madison/USA. DOI: https://doi.org/10.1002/agj2.20552

[5] Cai C., Deng Y., Zhou Z., Zhang Y., Yu J., (2026). An end-to-end automated framework for pavement crack detection and dimensional quantification using foundation models and learned skeletonization, *Engineering Applications of Artificial Intelligence*, vol.171, article no.114239, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.engappai.2026.114239

[6] Chamara N., Bai G., Ge Y., (2023). AICropCAM: Deploying classification, segmentation, detection, and counting deep-learning models for crop monitoring on the edge, *Computers and Electronics in Agriculture*, vol.215, article no.108420, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2023.108420

[7] Chen J., Qiang H., Wu J., Xu G., Wang Z., Liu X., (2020). Extracting the navigation path of a tomato-cucumber greenhouse robot based on a median point Hough transform, *Computers and Electronics in Agriculture*, vol.174, article no.105472, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2020.105472

[8] Chen Q., Chen J., Jiang Z., Huang L., Gai J., Li P., Zhang M., (2025). Research on efficient and fast extraction of vineyard navigation path based on key point detection, *Engineering Applications of Artificial Intelligence*, vol.159, part A, article no.111549, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.engappai.2025.111549

[9] Chen Y., Geng C., Wang Y., Zhu G., Shen R., (2021). Extraction method for centerlines of rice seedings based on Fast-SCNN semantic segmentation (基于 Fast-SCNN 语义分割的秧苗列中心线提取方法), *INMATEH Agricultural Engineering*, vol.64, no.2, pp.335-344, Bucharest/Romania. DOI: https://doi.org/10.35633/inmateh-64-33

[10] Diao Z., Guo P., Zhang B., Zhang D., Yan J., He Z., Zhao S., Zhao C., Zhang J., (2023). Navigation line extraction algorithm for corn spraying robot based on improved YOLOv8s network, *Computers and Electronics in Agriculture*, vol.212, article no.108049, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2023.108049

[11] Fan J., Li Y., Yu S., Gou W., Guo X., Zhao C., (2023). Application of Internet of Things to agriculture—The LQ-FieldPheno platform: A high-throughput platform for obtaining crop phenotypes in field, *Research*, vol.6, article no.0059, Washington/USA. DOI: https://doi.org/10.34133/research.0059

[12] Gao X., Zan X., Yang S., Zhang R., Chen S., Zhang X., Liu Z., Ma Y., Zhao Y., Li S., (2023). Maize seedling information extraction from UAV images based on semi-automatic sample generation and Mask R-CNN model, *European Journal of Agronomy*, vol.147, article no.126845, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.eja.2023.126845

[13] Gao X., Wang G., Qi J., Wang Q., Xiang M., Song K., Zhou Z., (2024). Improved YOLOv7 for sustainable agriculture significantly improves precision rate for Chinese cabbage (Brassica pekinensis Rupr.) seedling belt detection, *Sustainability*, vol.16, no.11, article no.4759, Basel/Switzerland. DOI: https://doi.org/10.3390/su16114759

[14] Guo Y., Aggrey S.E., Yang X., Oladeinde A., Qiao Y., Chai L., (2023). Detecting broiler chickens on litter floor with the YOLOv5-CBAM deep learning model, *Artificial Intelligence in Agriculture*, vol.9, pp.36-45, Beijing/China. DOI: https://doi.org/10.1016/j.aiia.2023.08.002

[15] Han W., Xiang Y., Wu M., Zeng Y., Zou R., Jiang X., Quan W., (2026). Ridge navigation line extraction for rapeseed transplanters based on R-YOLOv11-Prune, *Smart Agricultural Technology*, vol.13, article no.101797, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.atech.2026.101797

[16] Hu B., Qin Y., Chen R., Ren T., Zhang Z., Wu R., Yuan X., Chen Z., Wang Q., (2025). YOLOv8n-COSD: A lightweight multi-mechanism network for straw plastic impurity detection, *Smart Agricultural Technology*, vol.12, article no.101204, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.atech.2025.101204

[17] Huang S., Wu S., Sun C., Ma X., Jiang Y., Qi L., (2020). Deep localization model for intra-row crop detection in paddy field, *Computers and Electronics in Agriculture*, vol.169, article no.105203, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2019.105203

[18] Jiang Y., Li W., Liu J., ur Rehman M.M., Xie B., Wang J., (2025). Generalized perception of tree-row with distribution-peak feature in 3D point cloud for various orchards navigation, *Smart Agricultural Technology*, vol.12, article no.101137, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.atech.2025.101137

[19] Jin L., Duan A., Shi Y., Fang H., (2025). EFNetPX: An improved panoptic perception network for inter-row navigation in orchard, *Computers and Electronics in Agriculture*, vol.239, part C, article no.111133, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2025.111133

[20] Ju J., Chen G., Lv Z., Zhao M., Sun L., Wang Z., Wang J., (2024). Design and experiment of an adaptive cruise weeding robot for paddy fields based on improved YOLOv5, *Computers and Electronics in Agriculture*, vol.219, article no.108824, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2024.108824

[21] Kim Y.H., Lee S.J., Yun C., Im S.J., Park K.R., (2023). LCW-Net: Low-light-image-based crop and weed segmentation network using attention module in two decoders, *Engineering Applications of Artificial Intelligence*, vol.126, article no.106890, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.engappai.2023.106890

[22] Kong X., Guo Y., Liang Z., Zhang R., Hong Z., Xue W., (2025). A method for recognizing inter-row navigation lines of rice heading stage based on improved ENet network, *Measurement*, vol.241, article no.115677, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.measurement.2024.115677

[23] Lai X., Huang J., Mo Y., Li H., Dong T., Wu T., He D., (2026). A lightweight deep learning model for synchronized crop stem detection and row segmentation at the seedling stage: Exploring their contribution to agricultural navigation line extraction, *Computers and Electronics in Agriculture*, vol.243, article no.111385, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2025.111385

[24] Li D., Li B., Long S., Feng H., Xi T., Kang S., Wang J., (2023). Rice seedling row detection based on morphological anchor points of rice stems, *Biosystems Engineering*, vol.226, pp.71-85, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.biosystemseng.2022.12.012

[25] Liu S., Jin Y., Ruan Z., Ma Z., Gao R., Su Z., (2022). Real-time detection of seedling maize weeds in sustainable agriculture, *Sustainability*, vol.14, no.22, article no.15088, Basel/Switzerland. DOI: https://doi.org/10.3390/su142215088

[26] Liu T., Zhu S., Yang T., Zhang W., Xu Y., Zhou K., Wu W., Zhao Y., Yao Z., Yang G., Wang Y., Sun C., Sun J., (2024a). Maize height estimation using combined unmanned aerial vehicle oblique photography and LiDAR canopy dynamic characteristics, *Computers and Electronics in Agriculture*, vol.218, article no.108685, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2024.108685

[27] Liu W., Zhou Z., Xu X., Gu Q., Zou S., He W., Luo X., Huang J., Lin J., Jiang R., (2023). Evaluation method of rowing performance and its optimization for UAV-based shot seeding device on rice sowing, *Computers and Electronics in Agriculture*, vol.207, article no.107718, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2023.107718

[28] Liu Y., Guo Y., Wang X., Yang Y., Zhang J., An D., Han H., Zhang S., Bai T., (2024b). Crop root rows detection based on crop canopy image, *Agriculture*, vol.14, no.7, article no.969, Basel/Switzerland. DOI: https://doi.org/10.3390/agriculture14070969

[29] Lopes R.A., Carvalho M.V.L., Kitani E., Zampirolli F.A., Yoshioka L., Celiberto Junior L.A., Ibusuki U., (2025). Deep learning-based instance segmentation for enhanced navigation of agricultural vehicles, *Revista de Informática Teórica e Aplicada*, vol.32, no.1, pp.136-142, Porto Alegre/Brazil. DOI: https://doi.org/10.22456/2175-2745.143329

[30] Luo S., Liu W., Zhang Y., Wang C., Xi X., Nie S., Ma D., Lin Y., Zhou G., (2021). Maize and soybean heights estimation from unmanned aerial vehicle (UAV) LiDAR data, Computers and Electronics in Agriculture, vol.182, article no.106005, *Amsterdam/Netherlands*. DOI: https://doi.org/10.1016/j.compag.2021.106005

[31] Luo Y., Gao G., Jiang P., Sun C., Xu M., Zhou Z., Hu W., Tan Y., (2025). A navigation line extraction method based on semantic segmentation and centerline fitting, *Smart Agricultural Technology*, vol.12, article no.101492, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.atech.2025.101492

[32] Luz G.P.C.P. da, Sato G.M., Gomez Gonzalez L.F., Borin J.F., (2024). Smart parking with pixel-wise ROI selection for vehicle detection using YOLOv8, YOLOv9, YOLOv10, and YOLOv11, *arXiv preprint*, article no.2412.01983, Ithaca/USA. DOI: https://doi.org/10.48550/arXiv.2412.01983

[33] Lv P., Xue J., Wei W., Liu S., Gao W., Sun H., Miao H., Wang W., (2026). An advanced detector and dual-shortest distance intersection algorithm for navigation path extraction in complex orchards, *Engineering Applications of Artificial Intelligence*, vol.165, part A, article no.113246, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.engappai.2025.113246

[34] Ma C., Dong Z., Chen Z., Zhu Y., Shi F., (2021). Research on navigation line generation of kiwi orchard between rows based on root point substitution (基于根点替代的猕猴桃果园行间导航线生成方法研究), *Agricultural Research in the Arid Areas*, vol.39, no.5, pp.222-230, Yangling/China. DOI: https://doi.org/10.7606/j.issn.1000-7601.2021.05.29

[35] Meng F., Yan C., Yang Y., Xing R., Zhu D., Zhang A., Tang Q., Liao J., (2024). A lightweight improved YOLOv5s model-based rice blast detection method and mobile deployment (基于轻量化改进 YOLOv5s 模型的稻瘟病检测方法及移动部署), *INMATEH Agricultural Engineering*, vol.74, no.3, pp.771-786, Bucharest/Romania. DOI: https://doi.org/10.35633/inmateh-74-68

[36] Qiao S., Chen L.-C., Yuille A., (2021). DetectoRS: Detecting objects with recursive feature pyramid and switchable atrous convolution, *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.10208-10219, Nashville/USA. DOI: https://doi.org/10.1109/CVPR46437.2021.01008

[37] Rodríguez Rodríguez L.A., Castañeda-Miranda C.L., Moreno Lució M., Solís-Sánchez L.O., Castañeda-Miranda R., (2021). Quarternion color image processing as an alternative to classical grayscale conversion approaches for pest detection using yellow sticky traps, *Mathematics and Computers in Simulation*, vol.182, pp.646-660, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.matcom.2020.11.022

[38] Shahid R., Qureshi W.S., Khan U.S., Munir A., Zeb A., Moazzam S.I., (2024). Aerial imagery-based tobacco plant counting framework for efficient crop emergence estimation, *Computers and Electronics in Agriculture*, vol.217, article no.108557, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2023.108557

[39] Shao J., Zhao Q., Gong Z., Guo X., Geng S., Li Z., Li D., (2026). ENT-YOLO: An improved lightweight YOLO for cotton organ detection in mulched drip irrigation systems in southern Xinjiang, *Agricultural Water Management*, vol.323, article no.110054, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.agwat.2025.110054

[40] Si Y., Xu H., Zhu X., Zhang W., Dong Y., Chen Y., Li H., (2025). SCSA: Exploring the synergistic effects between spatial and channel attention, *Neurocomputing*, vol.634, article no.129866, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.neucom.2025.129866

[41] Song Y., Xu F., Yao Q., Liu J., Yang S., (2023). Navigation algorithm based on semantic segmentation in wheat fields using an RGB-D camera, *Information Processing in Agriculture*, vol.10, no.4, pp.475-490, Beijing/China. DOI: https://doi.org/10.1016/j.inpa.2022.05.002

[42] Surulivelu M., Hemanand D., Babu R.G., Subash T.D., (2026). AI-based tuberculosis detection and localization in chest X-rays using a binarized simplicial convolutional neural network for efficient and

accurate diagnosis, *Biomedical Signal Processing and Control*, vol.113, article no.108928, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.bspc.2025.108928

[43] Wang S., Yu S., Zhang W., Wang X., Li J., (2023). The seedling line extraction of automatic weeding machinery in paddy field, *Computers and Electronics in Agriculture*, vol.205, article no.107648, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2023.107648

[44] Yang Y., Zhou Y., Yue X., Zhang G., Wen X., Ma B., Xu L., Chen L., (2023). Real-time detection of crop rows in maize fields based on autonomous extraction of ROI, *Expert Systems with Applications*, vol.213, part A, article no.118826, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.eswa.2022.118826

[45] Zhang H., Ge Y., Xia H., Sun C., (2024a). Safflower picking points localization method during the full harvest period based on SBP-YOLOv8s-seg network, *Computers and Electronics in Agriculture*, vol.227, part 2, article no.109646, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2024.109646

[46] Zhang X., Lu X., Zhang Z., Yang G., He Y., Fang H., (2024b). Simultaneous detection of reference lines in paddy fields using a machine vision-based framework, *Computers and Electronics in Agriculture*, vol.221, article no.108923, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2024.108923

[47] Zheng Z., Hu Y., Li X., Huang Y., (2023). Autonomous navigation method of jujube catch-and-shake harvesting robot based on convolutional neural networks, *Computers and Electronics in Agriculture*, vol.215, article no.108469, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2023.108469