



Research Paper

YOLO-light-pruned: A lightweight model for monitoring maize seedling count and leaf age using near-ground and UAV RGB images

Tiantian Jiang^{a,b,c}, Liang Li^a, Zhen Zhang^b, Xun Yu^a, Yanqin Zhu^{a,b}, Liming Li^{a,b}, Yadong Liu^{a,c}, Yali Bai^{a,c,d}, Ziqian Tang^{a,c}, Shuaibing Liu^{a,c}, Yan Zhang^{a,b}, Zheng Duan^{a,e}, Dameng Yin^{a,c,*}, Xiuliang Jin^{a,c,*}^a State Key Laboratory of Crop Gene Resources and Breeding/Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, PR China^b School of Geomatics, Anhui University of Science and Technology, Huainan 232001, China^c National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya 572024, PR China^d Information Technology Group, Wageningen University & Research, Wageningen 6708 PB, the Netherlands^e School of Mechanical Engineering, Xinjiang University, Urumqi 830046, China

ARTICLE INFO

Article history:

Received 8 January 2025

Received in revised form 28 September 2025

Accepted 1 October 2025

Available online 02 October 2025

Keywords:

Maize seedling monitoring

Leaf age

Plant counting

RGB images

Deep learning

UAV

ABSTRACT

Maize seedling count and leaf age are critical indicators of early growth status, essential for effective field management and breeding variety selection. Traditional field monitoring methods are time-consuming, labor-intensive, and prone to subjective errors. Recently, deep learning-based object detection models have gained attention in crop seedling counting. However, many of these models exhibit high computational complexity and implementation costs, making field deployment challenging. Moreover, maize leaf age monitoring in field environments is barely investigated. Therefore, this study proposes two lightweight models, YOLOv8n-Light-Pruned (YOLOv8n-LP) and YOLOv11n-Light-Pruned (YOLOv11n-LP), for monitoring maize seedling count and leaf age in field RGB images. Our proposed models are improved from YOLOv8n and YOLOv11n by incorporating the DAttention mechanism, an improved BiFPN, an EfficientHead, and layer-adaptive magnitude-based pruning. The improvement in model complexity and model efficiency was significant, with the number of parameters reduced by over 73 % and model efficiency upgraded by up to 42.9 % depending on the device computation power. High accuracy was achieved in seedling counting (YOLOv8n-LP/ YOLOv11n-LP: AP = 0.968/0.969, $R^2 = 0.91/0.94$, rRMSE = 6.73 %/5.59 %), with significantly reduced model size (YOLOv8n-LP/ YOLOv11n-LP: parameters = 0.8 M/0.7 M, trained model size = 1.8 MB/1.7 MB). The robustness was validated across datasets with varying leaf ages (rRMSE = 4.07 % – 7.27 %), resolutions (rRMSE = 3.06 % – 6.28 %), seedling compositions (rRMSE = 1.09 % – 9.29 %), and planting densities (rRMSE = 3.38 % – 10.82 %). Finally, by integrating plant counting and leaf age estimation, the proposed models demonstrated high accuracy in leaf age detection using near-ground images (YOLOv8n-LP/ YOLOv11n-LP: rRMSE = 5.73 %/7.54 %) and UAV images (rRMSE = 9.24 %/14.44 %). The results demonstrate that the proposed models excel in detection accuracy, deployment efficiency, and adaptability to complex field environments, providing robust support for practical applications in precision agriculture.

© 2025 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Maize is among the most planted crops globally, the consumption of which has significantly increased in the past decade despite the decreasing cultivation area (Rehman et al., 2019). Enhancing maize yield is crucial for meeting the increasing needs of the ever-growing population (Ray et al., 2012). Seedling emergence and leaf development are key early growth indicators directly influencing subsequent yield

performance (Doebley et al., 1997; Liu et al., 2018; Xia et al., 2022). The emergence rate of maize, defined as the percentage of seeds that successfully germinate and grow into seedlings, is a critical factor in field management and breeding evaluation (Gao et al., 2023b; Liu et al., 2022). Leaf age, defined as the number of fully expanded leaves (Boyes et al., 2001), is also closely monitored by agronomists and farmers. Timely assessment of these two parameters can not only guide field management decisions such as seedling replenishment and irrigation (Bai et al., 2023), but also facilitate breeding material selection (Gao et al., 2023b).

In real practice, manual inspection in the field is still the primary way to monitor seedling emergence and development. However, manual monitoring through visual inspection is labor-intensive, time-

* Corresponding authors at: State Key Laboratory of Crop Gene Resources and Breeding/ Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, PR China.

E-mail addresses: yindameng@caas.cn (D. Yin), jinxuliang@caas.cn (X. Jin).

consuming, and subject to human error (Buters et al., 2019). Real-time or near real-time data feeds providing information on the seedling count and leaf age are necessary for smart farming and breeding (Forcella, 1998; Tan et al., 2022). Lately, we have been consulted by many researchers and decision-makers about portable devices with real-time visualization of seedling monitoring results. However, a big obstacle for developing such devices is the high requirements for battery and computing capacity, as well as the heavy weight that comes with them. Therefore, there is an urgent need for efficient and automated methods to monitor maize count and leaf age (Bai et al., 2023). We hope to develop a model that can be deployed on-board unmanned aerial vehicles (UAVs) or embedded devices with limited memory and computational resources without adding on too much weight.

Leveraging high-spatial-resolution RGB remote sensing as the data source, machine learning and deep learning methods are the main pathways for plant counting and leaf age estimation. For plant counting, traditional machine learning methods such as template matching, computer vision-based peak detection algorithm, and corner detection model have been applied for seedling counting from high-resolution UAV imagery (Bai et al., 2022; Koh et al., 2019; Liu et al., 2022; Yuan et al., 2024). Additionally, regression models such as multiple linear regression, support vector regression, regression trees, and gaussian process regression have been adopted to derive seedling counts from spectral and morphological vegetation features (Banerjee et al., 2021; Liu et al., 2022). Image-based leaf age estimation studies, however, are rare. Bai et al. (2023) input pre-defined features such as texture, shape features, and spectral indices from UAV RGB images in regression models to estimate maize leaf age. The Gradient Boosting Decision Tree (GBDT) model achieved the best performance with an R^2 of 0.88 and an RMSE of 0.33. A common problem in these traditional machine learning methods is that they heavily depend on image quality and manual feature extraction, frequently overlooking critical features and limiting their ability to generalize under complex environmental conditions. In contrast, deep learning, especially convolutional neural networks (CNNs), automatically learns multi-level features directly from raw images, capturing complex details like leaf shape, texture, and structure (Liu et al., 2022). This enables deep learning models to handle diverse conditions such as varying environments, lighting, and view angles, making them more suitable for real-world applications like maize seedling and leaf analysis.

Notable and influential deep learning frameworks in object detection include the Faster R-CNN (Ren et al., 2016), Single Shot MultiBox Detector (SSD) (Liu et al., 2016), and You Only Look Once (YOLO) models. The Faster R-CNN model is a two-stage model that first proposes potential regions and then identifies target objects from these regions. When applied in maize seedling detection, over 97.71 % precision has been reported (Quan et al., 2019). However, it still faces challenges such as large model sizes, limited real-time performance, and restricted generalization capabilities. The SSD combines feature fusion, optimized framework design, and improved loss functions to become a significant algorithm in object detection due to its speed and ease of use. However, it faces challenges in small object detection (Zhang et al., 2023). Recently, the YOLO family of models has made significant breakthroughs in object detection accuracy and efficiency (Jiang et al., 2022). Notably, YOLOv5, which integrates an attention mechanism, has been utilized for wheat ear counting. This model effectively alleviated environmental noise in the field, achieving a mean average precision (mAP) of 94.3 % (Li and Wu, 2022). In the past five years, updated versions YOLOv7, YOLOv8, and even YOLOv11 have been published, each time with enhanced accuracy and efficiency on object detection and segmentation tasks (Jegham et al., 2024; Yu et al., 2025).

The latest research endeavors in deep learning model development have focused on transforming large-parameter models into lightweight alternatives to fulfill real-time applications (Wang et al., 2022). These

studies typically reduce model parameters by leveraging lightweight feature extraction layers (Bao et al., 2021), channel attention mechanisms (Zhou et al., 2024), or depth-wise separable convolutions (DSC) (Gao et al., 2022a). However, the optimizations primarily target heavy-weight models, reducing the model size from 42.7 to 224.0 MB to 19.82–71.69 MB (Gao et al., 2022a; Zhu et al., 2024). As a result, these lightweight models still demand considerable computational resources. Recently, Jia et al. (2024) assessed the performance of multiple more lightweight models on maize seedling counting, including YOLOv8n (parameter size 3.20 M), YOLOv5 (parameter size 1.90 M), YOLOv3-tiny (parameter size 8.44 M), and more. They found that YOLOv8n achieved the highest accuracy with a mean average precision (mAP) of 0.976, demonstrating the potential of much smaller models in counting maize seedlings. However, the YOLOv8n model still requires 8.1 G floating-point operations (GFLOPs), i.e., 8.1×10^9 floating-point operations. The wide desire of instant inference and visualization devices by breeding scientists and decision makers require lightweight and robust models. Hence, to facilitate the deployment of cost-effective mobile devices for accurate and real-time crop monitoring, efforts in developing lightweight seedling counting models with fewer than a million parameters are critical.

Small as the model we want to be, robustness to diverse conditions is still essential for deployment in practical field environments. Real-field applications can be challenging due to complexities such as varying illumination conditions, varied planting density, different growth status, messy backgrounds (soil wetness or weeds), limited selections of cameras, etc. Most studies have shown strong performance in addressing one to two specific factors, including varying planting densities (Liu et al., 2023), growth stages (Quan et al., 2019), and flight altitudes (Gao et al., 2023a). However, field conditions are often more complex, posing significant challenges to model transferability and adaptability. The investigation of model robustness over multiple factors in more complicated field settings is highlighted (Jia et al., 2024).

Compared to the prosperity in seedling counting studies, research into leaf age detection is rare. Most were tested exclusively on single maize plants cultivated in controlled indoor environments (Ning et al., 2024; Xie et al., 2023). Despite the high precision, they lack the complexity of field environments and thus fail to capture the actual growth variations. We found two publications working on leaf counting in the field. Both adopted a two-stage deep learning approach, first segmenting the exact boundary of individual maize plants with a segmentation model and then identifying the plant leaves with an object detection model (Xu et al., 2023; Xu et al., 2022). While the leaf detection results are accurate (precision up to 93 %), segmentation models require annotations of the exact shape of targets, making the labeling process tedious. Additionally, both studies used the YOLOv5x model in the leaf detection step. YOLOv5x is the largest model (about 86.7 M parameters) in the YOLOv5 series, which brings us back to the problem of substantial computational costs and unfeasibility in real-time applications. Therefore, there remains a critical need for models that simultaneously monitor both maize seedling number and leaf age with high accuracy and low computational cost in field scenarios.

To address the abovementioned challenges, this study proposes a lightweight and robust deep-learning model to monitor maize seedling number and leaf age in field environments. The specific tasks are as follows: (1) proposing an improved lightweight YOLO model and validating its performance in maize seedling monitoring, (2) evaluating the robustness of the proposed model for maize seedling counting, and (3) monitoring the leaf age of maize at both the individual plant and plot levels. Given its lightweight and robustness, we anticipate that the proposed model will provide a significant tool for the instant monitoring of maize seedling development. Moreover, it could be highly applicable to similar crops such as sorghum and possibly transferable to other row-planting crops such as cotton and sugar beet.

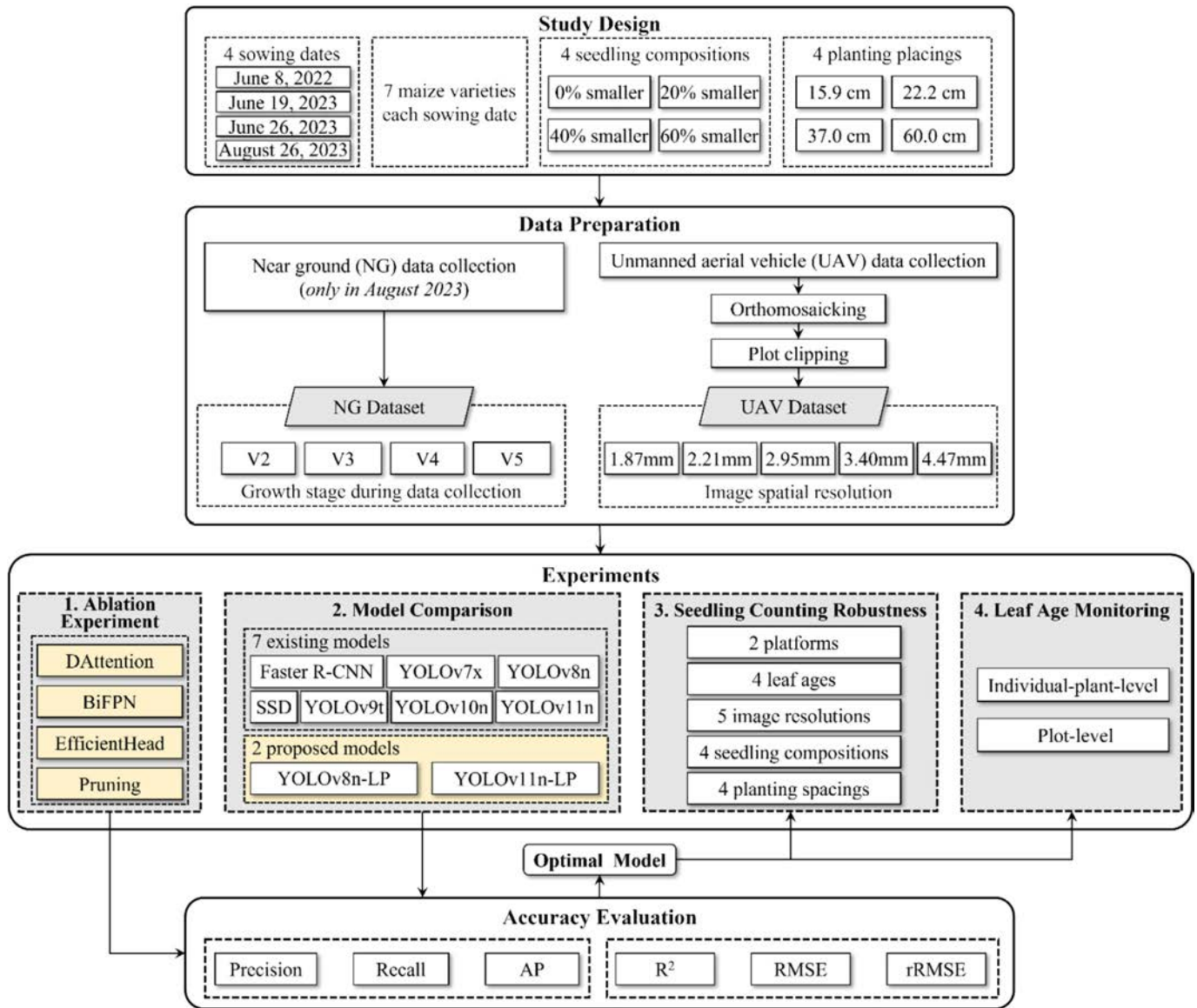


Fig. 1. Flowchart of the experimental design.

2. Materials and methods

The experimental design of this study is shown in Fig. 1. The following five subsections detail the study site and design (Section 2.1), the data preparation process (Section 2.2), principals of the proposed models (Section 2.3), four experiments to find and validate the optimal models for seedling counting and leaf age monitoring (Sections 2.4), and the accuracy evaluation criteria (Section 2.5).

2.1. Study site and design

The experimental site was at the Xinxiang Base of the Chinese Academy of Agricultural Sciences (113°47'E, 35°10'N) in Henan Province, China. The region is located in Huanghuaihai Plain with annual precipitation of 500–900 mm, 45–65 % of which is summer rainfall, providing sufficient moisture for the growth of summer maize (Liu et al., 2010). This plain is the largest concentration of maize production in China, accounting for more than 30 % of the Chinese country maize area (Bai et al., 2023).

Three rounds of experiments were conducted, planting maize seeds on June 8, 2022 (Fig. 2a), June 19, 2023 (Fig. 2b), June 26, 2023 (Fig. 2b), and August 24, 2023 (Fig. 2b). In each round, three treatment factors were considered, i.e., maize variety (Table 1), seedling composition, and planting density. In 2022, the first set of experiments had four maize varieties (ND108, ZD958, JNK728, and DH605) planted with four levels of seedling composition (0 %, 20 %, 40 %, and 60 % smaller seedlings). To create the different seedling composition level scenarios, seedlings in the plot were removed at the V2 stage (two leaves fully developed) and replaced with newly planted seeds. Each plot covers a $3.6 \text{ m} \times 5.4 \text{ m}$ area, with 6 rows at 0.6 m row spacing and 0.225 m plant spacing. The second experiment set had four maize varieties (DH 605 and KH699, YD9953, and ZYY432) planted at four levels of plant spacing (60.0 cm, 37.0 cm, 22.2 cm, and 15.9 cm plant spacing). Each plot covered a $3.6 \text{ m} \times 7.2 \text{ m}$ area. Each set was planted in two plots to avoid random errors due to local soil conditions. Both plots were sown on June 8, 2022 (Fig. 2b). In 2023, the same first experiments were carried out. The four varieties for different seedling composition levels were DH605, DK688, ZD958, and FDCY10. Each set was again

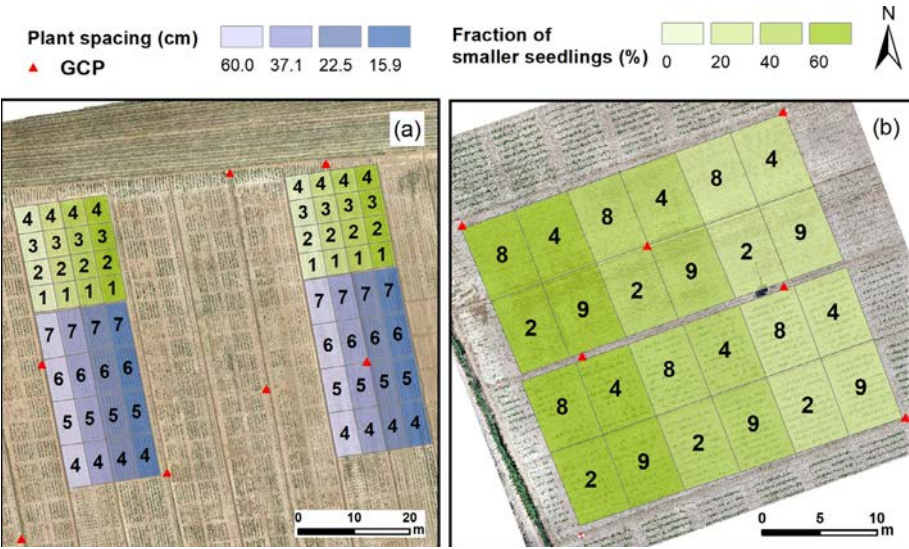


Fig. 2. Study site: (a) Experiment plots in 2022, (b) Plots in 2023.

planted in two plots to avoid random errors. One plot was sown on June 19, 2023, and the other one week later on June 26, 2023 (Fig. 2b), to avoid random errors due to the sowing date. All plots covered a 3.0 m × 4.0 m area. The 24 plots with 20 %, 40 %, and 60 % smaller seedlings were entirely removed and planted again on August 24, 2023, to increase the sample size.

2.2. Data preparation

2.2.1. Near-ground image acquisition and processing

Near-ground (NG) images were collected in August 2023, using a Daheng MER2–302-56U3M/C digital camera (Daheng Imaging, Beijing, China) with a 2448 × 2048 pixels resolution fixed on a 4.1 m tall pole (Fig. 3). The image spatial resolution was 1.18 mm. As the camera could only capture four rows of seedlings at a time, two images were taken at each plot to cover the entire plot. These two locations were marked and reused for each time of data collection. We took photos of the seedling composition plots 16 times, starting from the V2 stage (two fully unfolded leaves) to the V5 stage (five fully unfolded leaves) (Fig. 3d). In total, 1918 near-ground images were acquired. These images formed the NG dataset.

2.2.2. UAV image acquisition and processing

High-spatial-resolution RGB images were acquired using a Sony α 7 II digital camera (Sony Corporation, Tokyo, Japan) loaded on a DJI M600 Pro UAV (DJI, Shenzhen, China). Two different camera lenses were used, one with a focal length of 40 mm and the other with 50 mm. Flights

were taken at different flight altitudes and speeds to acquire images of different spatial resolutions (Table 2). By making these variations, we aimed to train the models to work well on UAV images with little restrictions on the data collection process. The latitude and longitude coordinates of a total of 13 ground control points (GCPs) were measured using a real-time kinematic (RTK) positioning system, GNSS RTK-G970II (UniStrong, Beijing, China). The data-acquisition devices are shown in Fig. 4a.

The pre-processing of the UAV data includes stitching the multiple photos into an orthomosaic, georeferencing orthomosaics captured on different dates, clipping the orthomosaics into plot-level images, and labeling all the seedlings (Fig. 4). The GCP coordinates were input into the Agisoft Metashape software (Agisoft LLC, Russia) to assist in mosaicking the UAV digital images (Fig. 4b). The result orthomosaic is a geometrically corrected, seamless composite image that provides full coverage of the entire study area, serving as a reliable foundation for subsequent target identification and analytical procedures. The spatial resolutions of the generated orthomosaic images were 1.87 mm, 2.21 mm, 2.95 mm, 3.40 mm, and 4.47 mm, respectively (Table 2). ArcGIS 10.8 (Environmental Systems Research Institute, Inc., Redlands, USA) was used to clip the UAV orthomosaic to individual-plot images (Fig. 4c). The per-plot image size varied depending on the image resolution, as detailed in Table 2. The labeling procedure and division of datasets for UAV images mirrored that of the near-ground images.

In this study, 744 UAV maize plot images were obtained, including 288 images from 2022 and 456 images from 2023. Due to the high demand for large sample sizes in deep learning models, we augmented the data to enhance the training effectiveness and generalizability of the model. Three data augmentation methods were applied: Contrast Limited Adaptive Histogram Equalization (CLAHE augment), histogram equalization (hiscolor augment), and adjustments to colour settings (brightness, contrast, and saturation) (i.e., illumination augment) (Shorten and Khoshgoftaar, 2019) (Fig. 5). These methods increased the diversity and volume of the training set, enabling the model to learn more robust features and enhancing its performance in various real-world scenarios.

2.2.3. Dataset configuration

2.2.3.1. Seedling count. The collected NG and UAV images were organized into multiple datasets to facilitate five different experiments, for the assessment of model robustness. Details of these datasets are summarized in Table 3.

Table 1
Experimental varieties design.

ID	Maize variety	Maize variety abbreviation	Sowing date			
			Jun 8, 2022	Jun 19, 2023	Jun 26, 2023	Aug 24, 2023
1	Nongda 108	ND 108	✓	–	–	–
2	Zhengdan958	ZD 958	✓	✓	✓	✓
3	Jingnongke728	JNK 728	✓	–	–	–
4	Denghai605	DH 605	✓	✓	✓	✓
5	Kehe699	KH 699	✓	–	–	–
6	Yudan9953	YD 9953	✓	–	–	–
7	Zhengyuanyu432	ZYY 432	✓	–	–	–
8	Dika688	DK 688	–	✓	✓	✓
9	Fondecunyu10	FDCY 10	–	✓	✓	✓

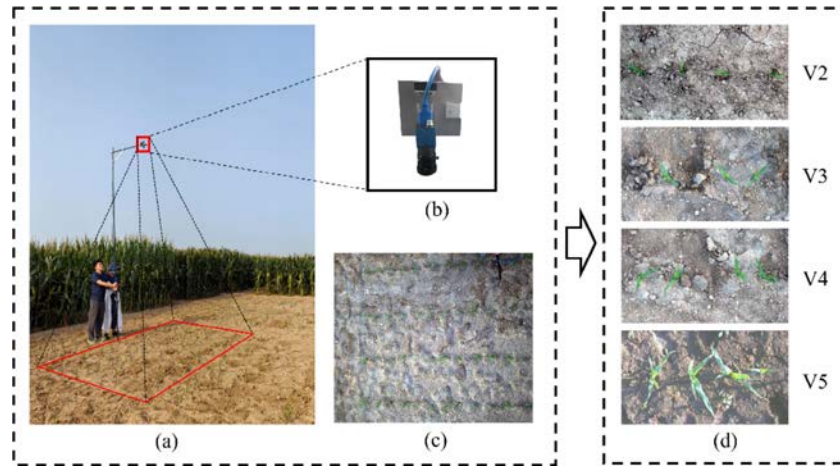


Fig. 3. Near-ground image acquisition. (a) Overview. (b) The camera. (c) An image acquired after processing. (d) Examples of maize seedlings at different growth stages.

Table 2

Flight parameters.

Year	Flight altitude (m)	Flight speed (m/s)	Side overlap (%)	Front overlap (%)	Camera focal length (mm)	Plot image size (pixels)	Spatial resolution (mm)	Repetitions
2022	30	2.0	75	80	40	1184 × 1325	3.40	7
	40	2.4	75	80	40	916 × 1025	4.47	2
2023	20	1.8	75	80	40	3040 × 3460	2.21	1
		1.8	75	80	50	3800 × 4200	1.87	12
	30	2.8	80	80	40	2024 × 2300	3.40	2
		2.8	80	80	50	2400 × 2667	2.95	11

Images collected using the NG and UAV approaches were organized into the NG and UAV datasets. The Labelling tool (Tzutalin, 2015) was used for manual annotation, drawing bounding boxes around each maize seedling. The NG dataset contained 1918 images (74,887 seedlings were labeled), with 90 % (1725 images) allocated to the NG-train dataset and the remaining 10 % (193 images) to the NG-test dataset. The NG-train dataset was used to train the seedling counting model for the NG platform, with 8/9 of the data allocated for training and 1/9 for validation. The model performance was evaluated using the NG-test dataset. The UAV dataset contained 744 images (90,176 seedlings were labeled), with 90 % forming the UAV-train dataset (2674 images after data augmentation) and the remaining 77 images designated for the UAV-test dataset. Similarly, 8/9 of the UAV-train dataset was used

for training and 1/9 for validation, while the UAV-test dataset was used to assess model accuracy.

2.2.3.2. Leaf age monitoring. We further estimated leaf age at both the individual plant level and plot level by training the proposed models to identify leaf tips, using both the NG and UAV datasets. Because maize leaves tend to overlap as the plants grow, we worked on plots with average leaf ages ranging from V2 and V3, where individual-plant-level leaf ages ranged from V1 to V6. Based on the seedling counting result, we generated cropped images of individual maize plants. The leaf tips in these individual maize images were labeled using labeling to create a new dataset related to leaf age. The NG platform dataset comprised 17,712 plant images, with the training set (NG-Leaf-train) containing

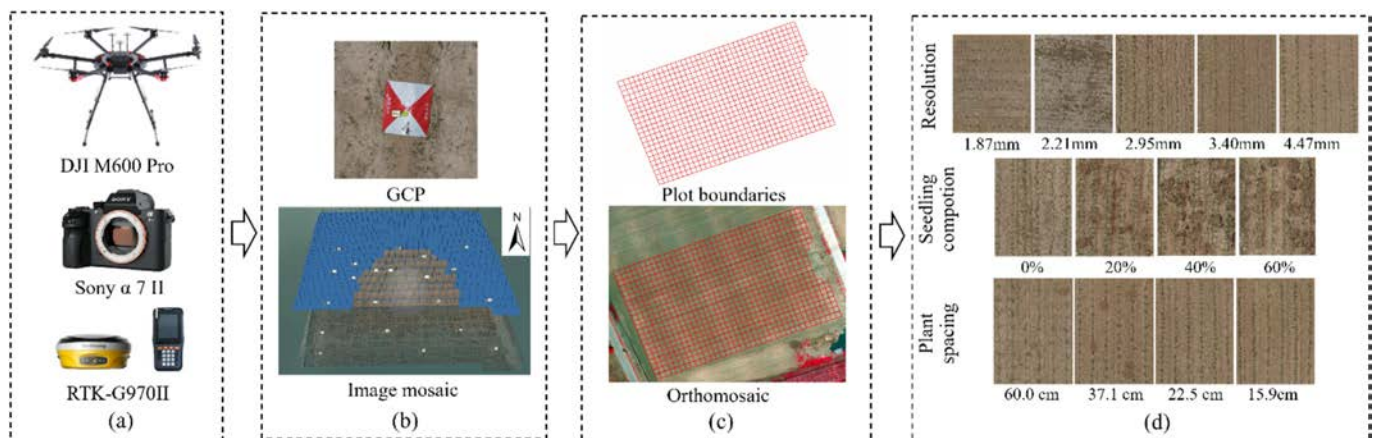


Fig. 4. UAV data preprocessing. (a) Data acquisition devices. (b) Image stitching. (c) Plot clipping. (d) Plot-level UAV images with different resolution, seedling composition and plant density.

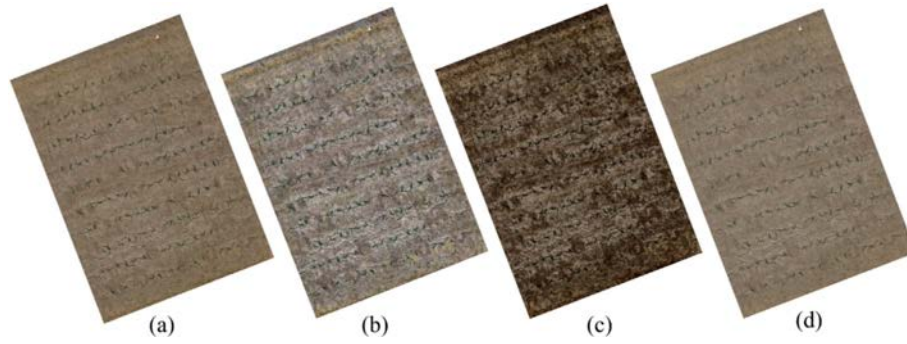


Fig. 5. Examples of augmented images. (a) Raw image. (b) CLAHE augment. (c) hiscolor augment. (d) illumination augment.

15,940 images and the testing set (NG-Leaf-test) containing 1772 images. The UAV platform dataset contained 14,244 plant images, with the training set (UAV-Leaf-train) comprising 12,819 images (25,639 images after data augmentation) and the testing set (UAV-Leaf-test) comprising 1425 images (Table 4).

2.3. Proposed models

The YOLO models are a series of deep learning models widely adopted for object detection (Jocher et al., 2023). In the YOLO family, YOLOv8 and YOLOv11 are the latest Ultralytics official versions to the date of submission, with “x” representing the most accurate but extra-large models and “n” (nano) representing the most efficient models. Therefore, aiming for real-time applications in portable devices, we proposed two novel lightweight models based on their nano versions (YOLOv8n and YOLOv11n), by redesigning the major components and adding a pruning step. The redesign targets the principal challenges of early-stage crop monitoring—tiny objects, heavy occlusion, complex illumination, and limited computational resources—through systematic modifications of the Backbone, Neck, and Head. The proposed models are YOLOv8n-Light-Pruned (YOLOv8n-LP) and YOLOv11n-Light-Pruned (YOLOv11n-LP), respectively.

Table 3
Datasets in the seedling count experiment.

Factor	Category	Data source	No. images	Dataset	
				training	testing
Platform	NG	all NG data	1918	NG-train	NG-test
	UAV	all UAV data	744	UAV-train	UAV-test
Growth stage	V2	NG data	534	NG-train	NG-V2-test
	V3		610		NG-V3-test
	V4		397		NG-V4-test
	V5		377		NG-V5-test
	1.87		204		UAV-1.87-test
Resolution (mm)	2.21	UAV data (see Table 2)	24	UAV-train	UAV-2.21-test
	2.95		204		UAV-2.95-test
	3.40		248		UAV-3.40-test
	4.47		64		UAV-4.47-test
Seedling composition (ratio of smaller seedlings)	0	UAV data	72	UAV-train	UAV-0-test
	20		224		UAV-20-test
	40		224		UAV-40-test
	60		224		UAV-60-test
Plant spacing (cm)	60.0	UAV data collected in 2022 using the 40 mm camera at 30 m flight height (see Table 2)	24	–	UAV-60.0
	37.0		24	–	UAV-37.0
	22.2		24	–	UAV-22.2
	15.9		24	–	UAV-15.9

2.3.1. YOLOv8n-LP

YOLOv8n-LP (Fig. 6) is an optimized variant derived from YOLOv8n, with modifications in the Backbone, Neck, and Head components, aiming to reduce the parameter size without compromising accuracy. These module optimizations and replacements make YOLOv8n-LP suitable for deployment in resource-constrained environments, such as mobile devices or real-time monitoring systems.

2.3.1.1. Backbone. We replaced the last C2f module with C2f_DAttention, which integrates deformable attention (DAttention) (Xia et al., 2022) mechanisms to improve the quality of feature representation. With DAttention, the weights of different channels and spatial locations were dynamically adjusted during feature extraction. This module amplifies fine textures (e.g., leaf tips and venation) while suppressing soil and weed noise with negligible parameter overhead.

2.3.1.2. Neck. The original Path Aggregation Feature Pyramid Network (PAFPN) (Liu et al., 2018) used in YOLOv8n struggles with effective cross-scale feature fusion, especially when dealing with occluded or small-scale objects. In our optimization, we replaced PAFPN with an enhanced Bidirectional Feature Pyramid Network (BiFPN). The BiFPN allows for both top-down and bottom-up feature propagation, thus improving multi-scale feature fusion (Tan et al., 2020). Additionally, we incorporated a lightweight 3×3 convolution layer before each feature fusion in BiFPN to compress feature channels, reducing computational costs while enhancing the representation of local context. Hereafter we refer to this enhanced BiFPN as Conv-BiFPN.

Table 4
Datasets for leaf age monitoring.

Platform	Leaf age	Individual-plant-level dataset			Plot-level dataset
		total	training	testing	testing
NG	V1	1054	954	100	0
	V2	8497	7616	881	267
	V3	6087	5498	589	267
	V4	1863	1680	183	0
	V5	195	178	17	0
	V6	16	14	2	0
	Total	17,712	NG-leaf-train (15940)	NG-leaf-test (1772)	NG-leaf-plot (534)
UAV	V1	1815	1628	187	4
	V2	8280	7468	812	99
	V3	2743	2471	272	22
	V4	1178	1051	127	6
	V5	220	195	25	0
	V6	8	6	2	0
	Total	14,244	UAV-leaf-train (12819)	UAV-leaf-test (1425)	UAV-leaf-plot (131)

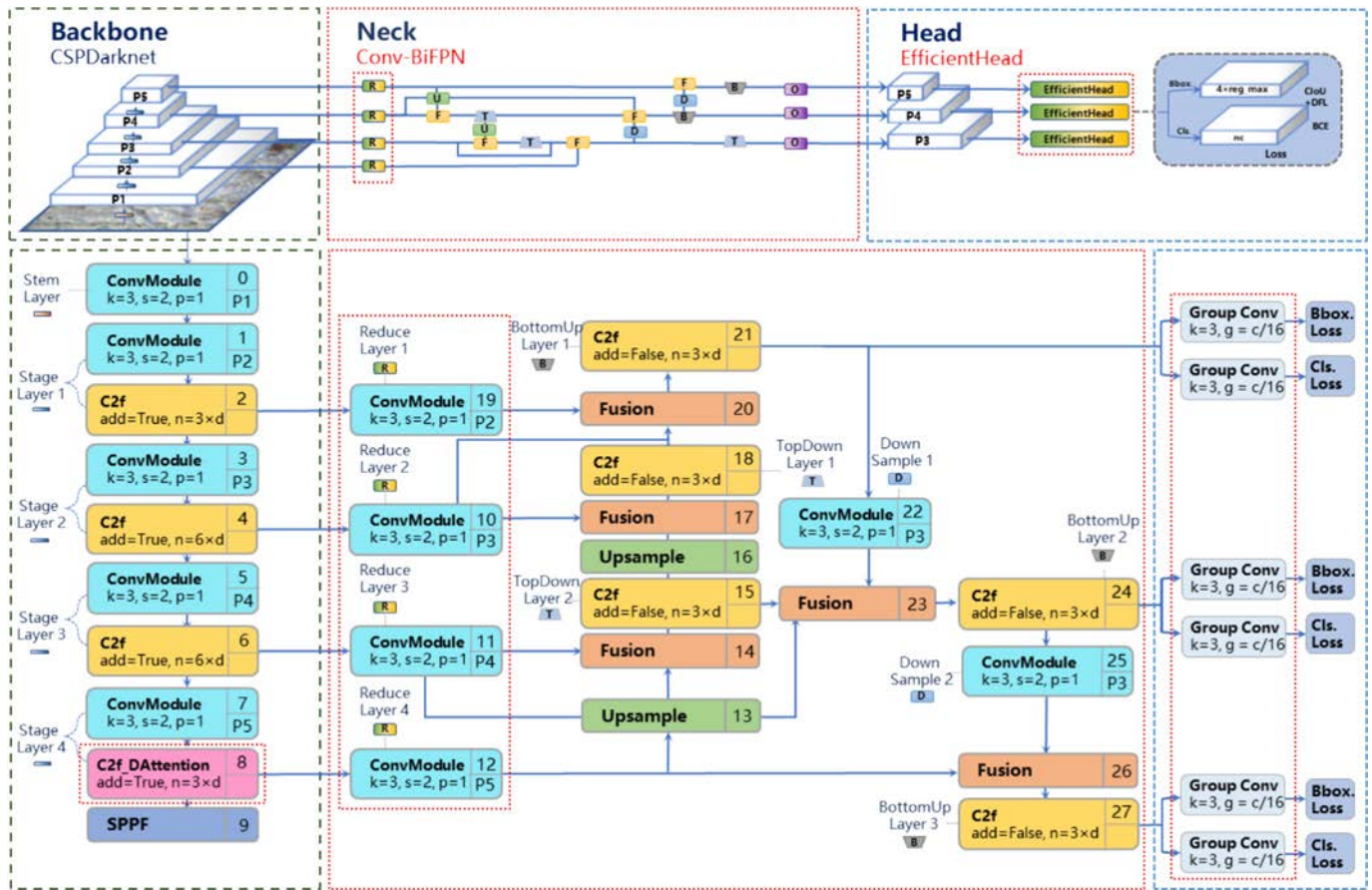


Fig. 6. Network architecture of YOLOv8n-LP: using C2f_DAttention in the Backbone, Conv-BiFPN in the Neck, and EfficientHead in the Head. Note: the red dotted rectangles mark modified modules. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3.1.3. Head. In the original YOLOv8n, the Decoupled Head (Varghese and Sambath, 2024) uses separate convolutional operations for classification, bounding box regression, and confidence score prediction. This increases computational and memory overhead. To reduce these costs, we replaced the Decoupled Head with an EfficientHead. This EfficientHead replaced the original 1×1 projection in YOLOv8n with two 3×3 group convolutions (Xie et al., 2017), with every 16 channels processed as one group. This grouping strategy reduces the memory and computational load while maintaining decoupled outputs for each task.

2.3.1.4. Pruning. The literature has indicated that removing unimportant weights from a network would improve the model generalization and learning speed (LeCun et al., 1990; Menghani, 2023). Therefore, we further applied layer-adaptive magnitude-based pruning (LAMP) (Lee et al., 2020) on the refined model. The implementation of LAMP begins by quantifying the importance of each layer in the neural network through the analysis of metrics such as weight magnitude, activation patterns, and gradient flows during training. Based on these importance scores, LAMP dynamically adjusts the pruning rate for each layer, aggressively pruning less important layers while retaining more parameters in crucial ones. Redundant parameters were removed by zeroing out or eliminating weights that contribute minimally to the model performance. Finally, the model undergoes retraining and fine-tuning to optimize the remaining weights, ensuring that high levels of maize leaf detection and counting accuracy are maintained.

2.3.2. YOLOv11n-LP

YOLOv11n represents an advanced model that benefits from additional refinements over YOLOv8n, particularly in multi-scale feature

fusion, model efficiency, and real-time performance (Khanam and Hussain, 2024). We applied similar structural modifications as introduced in Section 2.3.1 to YOLOv11n to propose the YOLOv11n-LP model.

The network structure of YOLOv11n-LP is illustrated in Fig. 7. **Backbone:** We replaced the original C3k2 module with C3k2_DAttention, which introduces deformable attention (Xia et al., 2022) to improve feature representation, particularly for small and occluded targets. **Neck:** The original PAFPN was replaced by Conv-BiFPN to efficiently fuse multi-scale features while reducing computational costs. **Head:** The original head was the Decoupled Head with DSC. The DSC-Decoupled Head was replaced by the abovementioned EfficientHead to reduce the model parameter size and computational complexity, allowing for real-time inference on edge devices. **Pruning:** After the modifications with these three core components, we integrated the LAMP module, which combines lightweight attention mechanisms with multi-scale feature fusion.

2.4. Experiments

2.4.1. Ablation experiment of the proposed model

We conducted ablation experiments to evaluate the contributions of individual modifications within the proposed model. Specifically for YOLOv8n-LP, each essential module (i.e., C2f_DAttention in the Backbone, Conv-BiFPN in the Neck, EfficientHead in the Head, and LAMP) was individually substituted with original configurations (i.e., C2f, PAFPN, Decoupled Head, and no LAMP) to observe changes in the detection accuracy and computational efficiency. For YOLOv11n-LP, the comparison was between each modification (i.e., C3k2_DAttention in the Backbone, Conv-BiFPN in the Neck, EfficientHead in the Head, and

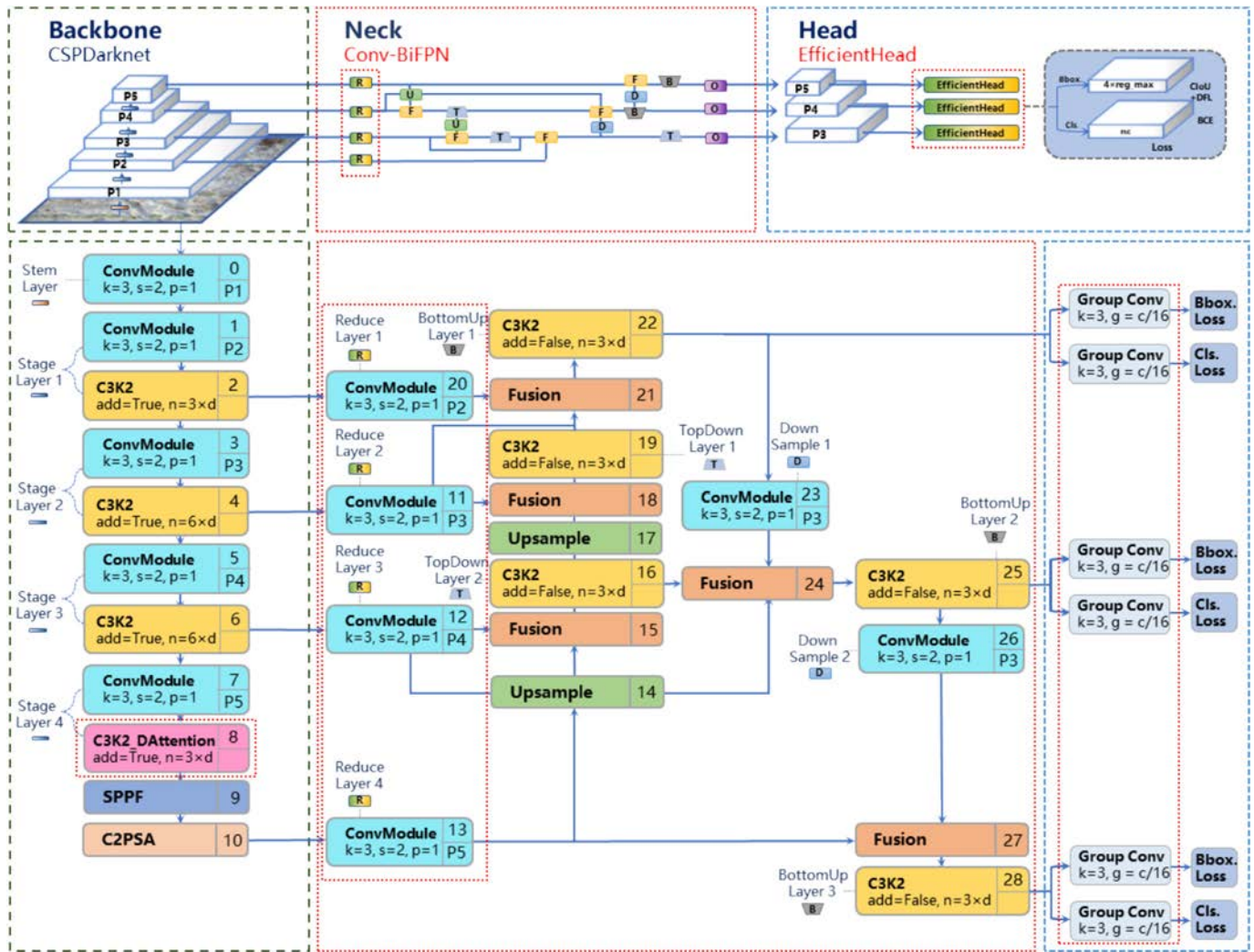


Fig. 7. Network architecture of YOLOv11n-LP: using C3K2_DAttention in the Backbone, Conv-BiFPN in the Neck, and EfficientHead in the Head. Note: the red dotted rectangles mark modified modules. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

LAMP) with their original configurations (i.e., C3k2, PAFPN, DSC-Decoupled Head, and no LAMP).

The ablation experiments were conducted consistently using the NG-train dataset for training and the NG-test for evaluation. All models used the same training parameters: a batch size of 16, a learning rate of 0.0001, and 100 epochs.

2.4.2. Comparison of the proposed model with existing ones

Two proposed lightweight models, YOLOv8n-LP and YOLOv11n-LP, were compared with seven state-of-the-art models to validate their performance on real-time maize seedling counting. The seven state-of-the-art models are Faster Region-Convolutional Neural Network (Faster R-CNN) (Ren et al., 2016), Single Shot MultiBox Detector (SSD) (Liu et al., 2016), YOLOv7x (Wang et al., 2023), YOLOv8n (Varghese and Sambath, 2024), YOLOv9n (Wang et al., 2022), YOLOv10n (Wang et al., 2024a), YOLOv11n (Khanam and Hussain, 2024). All models were trained on an NVIDIA GeForce RTX 4090 graphical processing unit (GPU), using pre-trained models from the Common Objects in Context (COCO) dataset. All models were trained using the NG-train dataset and tested using the NG-test dataset (Table 3).

2.4.3. Robustness of the proposed model for maize seedling counting

This experiment was designed to evaluate the robustness and generalizability of the proposed YOLOv8n-LP and YOLOv11n-LP models for

maize seedling counting in a series of meticulously structured scenarios. The specific data distribution is shown in Table 3. The experiment factors included: (1) Platform. Both the NG and UAV datasets were considered. While NG data collection is more suitable for real-time examination of each plot, the UAV platform allows for more efficient data collection for large-scale monitoring. Hence, we trained and tested models for the NG and UAV datasets separately to examine the model performance on different platforms. (2) Leaf age. The NG dataset was used to check the model performance in detecting maize seedlings at different growth stages. A single maize seedling detection model was trained using the NG-train dataset and tested using individual test datasets for each leaf age. (3) Image resolution. High-resolution images provide finer details, while lower-resolution images are more friendly in data storage and processing. By analyzing how the models perform with images of different resolutions, we aimed to determine their adaptability and effectiveness in diverse operational scenarios. (4) Seedling composition. As the seed vigor may vary, there could be smaller and larger seedlings simultaneously in the field. To ensure the model works in these situations, we applied the UAV-trained model on individual test datasets for different seedling composition scenarios (smaller seedling ratios: 0 %, 20 %, 40 %, and 60 %). (5) Planting density. Maize planting density differs depending on the maize variety and environmental conditions. We applied the UAV-trained model in the different planting density scenarios (plant spacing 60.0 cm, 37.0 cm, 22.2 cm, and

15.9 cm). To check the generalizability of the model, the high- and low-density samples were not included in model training.

By systematically varying these parameters, the study aimed to provide a comprehensive evaluation of the capabilities and limitations of the YOLOv8n-LP and YOLOv11n-LP models in counting maize seedlings under diverse real-world conditions, offering insights into their practical precision agriculture applications.

2.4.4. Application of the proposed model for leaf age monitoring

The proposed models, YOLOv8n-LP and YOLOv11n-LP, were further trained to monitor leaf age development. Models were built for the NG and UAV platform datasets separately (Table 4). Accuracy was evaluated at both the individual plant level and the plot level. For plot-level evaluation, the average leaf age for each plot was computed by averaging the leaf ages of all maize plants within that plot. Both the NG and UAV datasets were analyzed to check model adaptability to different data collection platforms.

2.5. Model performance evaluation

2.5.1. Object detection accuracy metrics

This study employs average precision (AP), recall (r), and precision (p) to assess model accuracy in object detection (Yu et al., 2025). Additionally, parameters and model size were recorded to evaluate the model efficiency and lightweight characteristics. Intersection over union (IoU) provides a measure of spatial overlap between the predicted and ground truth bounding boxes, which is crucial for determining the quality of object localization (Everingham et al., 2010). The formula is as follows:

$$IoU = \frac{I_{AO}}{I_{AU}} \quad (1)$$

I_{AO} represents the area of overlap between the predicted bounding boxes and the ground truth. I_{AU} denotes the area of union between the predicted bounding boxes and the ground truth. Following Yu et al. (2025), we utilized a threshold of 0.5. The predicted bounding box is classified as true positive (TP) if $IoU \geq 0.5$ and false positive (FP) if $IoU < 0.5$. In cases where the labeled bounding box was not detected, we have a false negative (FN) (Padilla et al., 2020). Subsequently, we computed r and p .

$$r = \frac{TP}{TP + FN} \quad (2)$$

$$p = \frac{TP}{TP + FP} \quad (3)$$

The precision-recall (PR) curve was constructed, with r values on the horizontal axis and p on the vertical. The area under this curve is AP.

Full-point interpolation was used to ensure monotonicity of the PR curve before the AP estimation (Padilla et al., 2020; Tassinari et al., 2021).

$$AP = \int_0^1 \max_{r' \geq r} (r') dr \quad (4)$$

2.5.2. Seedling counting and leaf age monitoring accuracy metrics

Three metrics were employed to measure the performance of seedling counting: the coefficient of determination (R^2), root mean square error (RMSE) (Chai and Draxler, 2014), and the relative RMSE (rRMSE) (Yu et al., 2025). The R^2 and RMSE were specifically utilized to evaluate the performance of maize seedlings during monitoring.

$$R^2 = 1 - \frac{\sum_{i=1}^n (t_i - c_i)^2}{\sum_{i=1}^n (t_i - \bar{t})^2} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (t_i - c_i)^2} \quad (6)$$

$$rRMSE = RMSE / \bar{t} \quad (7)$$

where n is the number of samples, t_i is the ground truth number of target objects, c_i is the number of objects detected by the model, and $\bar{t}$ is the average value of all t_i . In quantifying the seedling counting performance, each sample is one plot image and the target object is each seedling. In quantifying the leaf counting performance at the individual plant level, each sample is one seedling image and the target object is each leaf tip. In quantifying the leaf counting performance at the plot level, each sample is one plot image and the ground truth number of target objects is the average leaf age per plot, calculated as the ratio of the total number of leaves to the total number of plants in each plot.

2.5.3. Model complexity and efficiency

We evaluate the model complexity according to three core metrics: model size, number of parameters, and GFLOPs. Model size refers to the storage size required for the model, and is measured in megabytes (MB). Number of parameters is the number of all the weights and biases learned during training, and is measured in millions (M). GFLOPs is the number of floating-point operations required for implementing the model measured in giga (10^9). Smaller values of these metrics generally indicate less computing complexity (Molchanov et al., 2016).

Further, to demonstrate the improvement of the model complexity and efficiency, we ran the different models on multiple devices and recorded the processing time. The five different devices included two desktops, one laptop, and two tablet personal computers (PCs), as listed in Table 5. The most powerful device is a desktop (device 1), with 32.0 GB installed random access memory (RAM), 13th Gen Intel

Table 5
Devices used for model efficiency test.

	Device 1	Device 2	Device 3	Device 4	Device 5
Device type	Desktop	Desktop	Tablet PC	Laptop	Tablet PC
Processor	13th Gen Intel(R) Core(TM) i5-13490F 2.50 GHz	12th Gen Intel(R) Core(TM) i7-1255U 1.70 GHz	12th Gen Intel(R) Core(TM) i7-1255U 1.70 GHz	Intel(R) Core(TM) i7-10510U CPU @ 1.80GHz 2.30 GHz	Intel(R) Atom(TM) CPU E3845 @ 1.91GHz 1.91 GHz
Installed RAM	32.0 GB (31.8 GB usable)	16.0 GB (15.7 GB usable)	16.0 GB (15.7 GB usable)	8.00 GB (7.76 GB usable)	4.00 GB (3.85 GB usable)
Graphics card	NVIDIA GeForce RTX 4060 Ti (8 GB)	AMD Radeon(TM) Graphics	Intel(R) Iris(R) Xe Graphics	NVIDIA Quadro P520 (2GB)	Intel(R) HD Graphics (64 MB)
CUDA acceleration	Capable	No	No	Capable	No
System type	64-bit	64-bit	64-bit	64-bit	64-bit
Operating system	Windows 11 Pro	Windows 11 Pro	Windows 10 Pro	Windows 11 Home Basic	Windows 10 Home Basic
Python	3.9.21	3.8.0	3.10.15	3.11.4	3.8.0

(R) Core(TM) i5-13490F 2.50 GHz central computing unit (CPU), and a NVIDIA GeForce RTX 4060 Ti (8 GB) GPU. The device with the least computing power is a tablet PC (device 5), with 4.0 GB installed RAM, Intel(R) Atom(TM) CPU E3845 @ 1.91GHz 1.91 GHz and Intel (R) HD Graphics (64 MB). All devices had 64-bit Windows operating system. With each device, we used the seven state-of-the-art models and the two proposed models to process the same 200 images. The 200 images are real maize seedling images with 2448×2048 pixels. The average time used for seedling detection in each image was recorded. The average processing time, as well as the standard deviation, were compared. For device 1 and device 4 which had NVIDIA GPUs, we also conducted the comparison with GPU acceleration enabled by the compute unified device architecture (CUDA) parallel computing.

3. Results

3.1. Ablation experiment on model improvement

As shown in Tables 6 and 7, the ablation study results indicate that the base models YOLOv8n and YOLOv11n achieved test set precision scores of 0.945 and 0.955, respectively, with corresponding AP values of 0.970 and 0.976 (Test 1). When all three modules (DAttention, BiFPN, and EfficientHead) were integrated (marked with “√”), the models achieved the highest detection precision of 0.956 and 0.959 with much reduced model size and parameters (Test 8), indicating positive contribution of the modifications. Further applying LAMP reduced the model size to below 2 MB with fewer than a million parameters, which significantly enhances the applicability of the model in devices with limited computing power.

Checking the contributions of the individual modifications, similar patterns were found for both the YOLOv8n-LP and YOLOv11n-LP models. **DAttention** module in the backbone enhanced model accuracy but expanded the model size (Test 1 vs 2, 3 vs 5, 4 vs 6, and 7 vs 8). The accuracy improvement is possibly due to improved feature representation for fine-grained details. **Conv-BiFPN** module in the neck reduced model size and computational costs but also degraded the model accuracy (Test 1 vs 3, 2 vs 5, 4 vs 7, and 6 vs 8). The improvement in model size is possibly due to improved multi-scale feature fusion. The incorporation of **EfficientHead** module in the head generally reduced model size but had mixed effects in accuracy. After applying LAMP **pruning**, the models showed a significant reduction in model size and computational complexity, though with a slight decrease in accuracy. Nevertheless, despite the slight decrease in AP and precision, good recall score was maintained (Test 8 vs. 9). In YOLOv8n-LP (Test 9 in Table 6), the pruned model achieved $p = 0.953$, $r = 0.946$, and $AP = 0.968$, with only 0.8 M parameters. Similarly in YOLOv11n-LP (Test 9 in Table 7), the pruned model achieved $p = 0.948$, $r = 0.949$, and $AP = 0.969$, with only 0.7 M parameters.

3.2. Comparison of different models for maize seedling counting

Among the seven state-of-the-art models tested, the YOLO models demonstrated similar accuracy, with YOLOv11n achieving the highest accuracy ($p = 0.955$, $r = 0.951$, $AP = 0.976$, $R^2 = 0.96$, and $rRMSE = 4.96\%$) (Fig. 8). The two relatively older models (i.e., Faster R-CNN and SSD) had much lower accuracy compared to the YOLO models, which was examined by checking the result images (Fig. 9) and scatterplots (Fig. 10). Faster R-CNN exhibited the highest frequency of repeated detections (Fig. 9a), leading to an overestimation of the seedling count (Fig. 10a). In contrast, SSD tended to miss detections (Fig. 9b), resulting in an underestimation of the seedling count (Fig. 10b). Among the YOLO models, a general increase of accuracy with a three-level staircase trend was observed, which was particularly obvious if we look at the R^2 and $rRMSE$ lines (Fig. 8). In the first level was YOLOv7x, with $R^2 = 0.84$ and $rRMSE = 9.94\%$. In the second level were YOLOv8n, YOLOv9t, and YOLOv10n, with $R^2 = 0.90$ – 0.91 and $rRMSE = 7.46\% \sim 8.53\%$. Finally, the third level hosted YOLOv11n, YOLOv8n-LP, and YOLOv11n-LP, with $R^2 = 0.91$ – 0.96 and $rRMSE = 4.96\% \sim 6.73\%$.

The two proposed models YOLOv8n-LP and YOLOv11n-LP maintained high accuracy for maize seedling counting, with $rRMSE = 6.73\%$ and 5.59% respectively, which were only below YOLOv11n ($rRMSE = 4.96\%$) and well above the other state-of-the-art models ($rRMSE = 7.46\% \sim 22.70\%$) (Fig. 8). The model precision and recall, although not as high as YOLOv11n, were still with good balance (YOLOv8n-LP: $p = 0.953$, $r = 0.946$; YOLOv11n-LP: $p = 0.948$, $r = 0.949$).

The proposed YOLOv8n-LP and YOLOv11n-LP not only maintained the high seedling counting accuracy of YOLOv11n, but also obtained significantly reduced computing complexity. Similar with the accuracy trend, the model complexity also exhibited a three-level staircase distribution (Fig. 11). In the first level were Faster R-CNN, SSD, and YOLOv7x, with over 90 MB model size, over 26.3 million parameters, and more than 60 GFLOPs. The second level accommodated four models, i.e., YOLOv8n, YOLOv9t, YOLOv10n, and YOLOv11n. The model size was between 3.7 and 6.1 MB, the parameter size was between 1.8 and 3.0 million, while the computation amount was between 6.3 and 10.7 GFLOPs. The two proposed models comprised the third group, with the least computation complexity. Among the nine models, they were the only two with less than a million parameters and required less than 2 GFLOPs.

Running the models in real seedling detection applications, YOLOv8n-LP was the fastest regardless of devices (Table 8). With CUDA acceleration activated, all the lightweight YOLO models achieved very high computing efficiency (less than 0.1 s/image) which makes the improvement less important. However, CUDA acceleration is only available with certain NVIDIA GPUs and requires over 3 GB installation itself. With the very powerful device 1, the efficiency was improved by about 17.9% (from 0.056 ± 0.004 s/image by YOLOv8n to 0.046 ± 0.003 s/

Table 6
Ablation experiment based on YOLOv8n-LP.

Test	DAttention	Conv-BiFPN	Efficient Head	LAMP	Parameters (M)	Model size (MB)	Dataset: NG-train			Dataset: NG-test		
							<i>p</i>	<i>r</i>	AP	<i>p</i>	<i>r</i>	AP
1	–	–	–	–	3.0	6.0	0.958	0.950	0.980	0.945	0.938	0.970
2	√	–	–	–	3.8	6.1	0.960	0.952	0.982	0.950	0.936	0.971
3	–	√	–	–	2.0	4.0	0.954	0.945	0.975	0.952	0.933	0.967
4	–	–	√	–	2.4	4.8	0.960	0.948	0.975	0.958	0.942	0.971
5	√	√	–	–	2.1	4.2	0.955	0.944	0.975	0.952	0.929	0.968
6	√	–	√	–	3.9	7.7	0.962	0.952	0.976	0.959	0.952	0.973
7	–	√	√	–	1.6	3.3	0.955	0.941	0.972	0.949	0.942	0.969
8	√	√	√	–	1.8	3.7	0.959	0.950	0.974	0.956	0.948	0.972
9	√	√	√	√	0.8	1.8	0.956	0.942	0.970	0.953	0.946	0.968

Note: “√” means module used and “–” means module removed.

Table 7
Ablation experiment based on YOLOv11n-LP.

Test	DAttention	Conv-BiFPN	Efficient Head	LAMP	Parameters (M)	Model size (MB)	Dataset: NG-train			Dataset: NG-test		
							p	r	AP	p	r	AP
1	–	–	–	–	2.6	5.2	0.962	0.958	0.983	0.955	0.951	0.976
2	✓	–	–	–	2.6	5.3	0.962	0.956	0.982	0.957	0.948	0.977
3	–	✓	–	–	1.9	4.0	0.957	0.942	0.974	0.958	0.938	0.971
4	–	–	✓	–	2.3	4.7	0.962	0.952	0.979	0.955	0.949	0.974
5	✓	✓	–	–	2.0	4.0	0.958	0.944	0.974	0.951	0.945	0.972
6	✓	–	✓	–	2.3	4.8	0.961	0.952	0.979	0.957	0.95	0.976
7	–	✓	✓	–	1.7	3.6	0.956	0.943	0.973	0.956	0.938	0.971
8	✓	✓	✓	–	1.8	3.7	0.957	0.946	0.974	0.959	0.941	0.972
9	✓	✓	✓	✓	0.7	1.7	0.956	0.946	0.974	0.948	0.949	0.969

Note: “✓” means module used and “–” means module removed.

image by YOLOv8n-LP). As the computing power downgraded, the improvement was increasingly higher and YOLOv11n-LP started to exceed all state-of-the-art models and rank 2nd. Using the device with the least computing power (device 5), YOLOv8n-LP was the only model that could finish seedling detection under 1 s/image, immediately followed by YOLOv11n-LP (Table 8).

3.3. Robustness of the YOLOv8n-LP and YOLOv11n-LP models in seedling counting

This study evaluated the performance of the YOLOv8n-LP and YOLOv11n-LP models for maize seedling counting under varying observation platforms, leaf ages, image resolutions, seedling distributions, and planting densities. The results are summarized in Table 9, presenting the R^2 , RMSE, and rRMSE metrics for each experiment. The YOLOv8n-LP and YOLOv11n-LP models exhibited consistent performance trends across the various factors examined.

3.3.1. Seedling detection using NG and UAV images

The proposed models, YOLOv8n-LP and YOLOv11n-LP, demonstrated high accuracy across both NG and UAV datasets, as shown in Table 9. The NG dataset contained between 10 and 70 seedlings per image, with an RMSE of less than three seedlings. In contrast, the UAV dataset contained 60 to 180 seedlings per image, with RMSE values of 6.74 and 6.96 seedlings, respectively. However, the normalized accuracy

metric, rRMSE, remained below 7 % for both models with both NG and UAV platforms.

The strong correlation between the true and estimated number of seedlings was evident, with YOLOv8n-LP and YOLOv11n-LP achieving R^2 values of 0.91 and 0.94 on the NG dataset, and 0.96 and 0.95 on the UAV dataset. This correlation is further illustrated by the data points aligning closely with the 1:1 line in Fig. 12. Some instances of severe underestimation were observed, as indicated by the yellow circles in Fig. 12. Despite these underestimations, the testing data showed a close alignment with the 1:1 line, demonstrating the models' strong generalization capability on new data, even with slight biases in the training predictions.

3.3.2. Seedling detection at different leaf ages

Examples of the seedling detection results across different leaf age stages (V2, V3, V4, V5) are illustrated in Fig. 13. The model detected bounding boxes aligned well with the annotated ground truths, which reflected the high accuracy (Table 9) and strong robustness of both models. This was also embodied by the scatterplots, where all samples were located close to the 1:1 line (Fig. 14). At the V2 stage, both YOLOv8n-LP ($R^2 = 0.96$, rRMSE = 4.50 %) and YOLOv11n-LP ($R^2 = 0.96$, rRMSE = 4.07 %) demonstrated optimal accuracy, indicating that early-stage seedling morphology is highly conducive to detection. As the leaves continued to develop, the performance of both models downgraded but remained high (RMSE ≤ 2.95 and rRMSE ≤ 7.27 %).

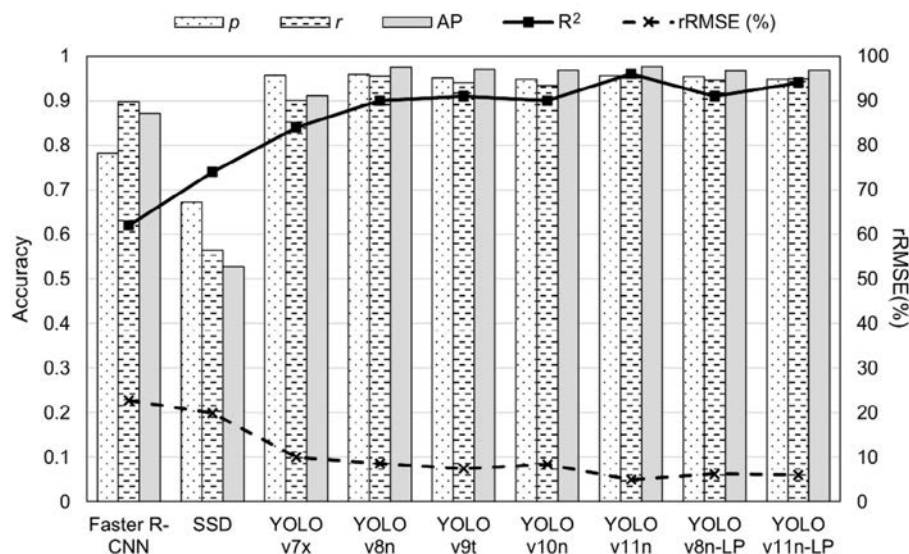


Fig. 8. Maize seedlings detection accuracy of different models.

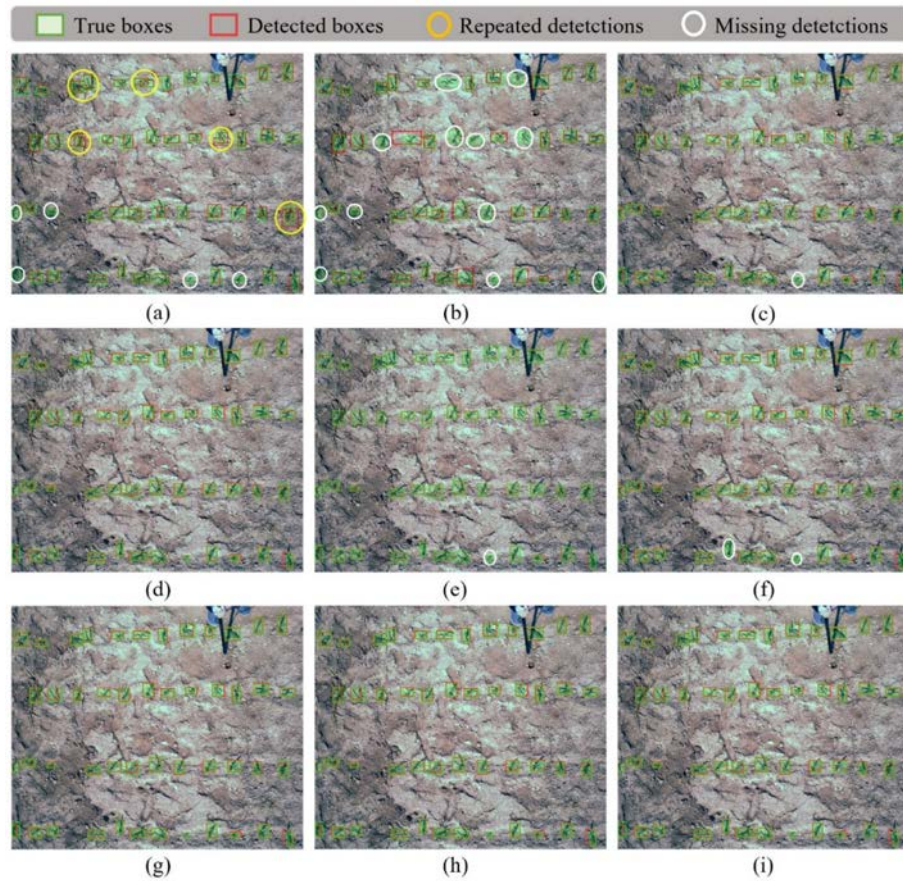


Fig. 9. Maize seedlings detection in different models: (a) Faster R-CNN, (b) SSD, (c) Yolo V7x, (d) YOLOv8n, (e) YOLOv9t, (f) YOLOv10n, (g) YOLOv11n, (h) YOLOv8n-LP, (i) YOLOv11n-LP.

3.3.3. Seedling detection at different spatial-resolution images in UAV

Across five different image resolutions (i.e., 1.87 mm, 2.21 mm, 2.95 mm, 3.40 mm, and 4.47 mm), both YOLOv8n-LP and YOLOv11n-LP exhibited high consistency with the ground truth (Fig. 15). The performance of the two models showed a similar decreasing pattern as the image resolution became finer (Table 9, Fig. 16). Both models achieved low rRMSEs at 3.40 mm (YOLOv8n-LP: 3.30 %; YOLOv11n-LP: 3.34 %) and 4.47 mm (YOLOv8n-LP: 3.06 %; YOLOv11n-LP: 4.11 %). The dataset at the 2.21 mm resolution (UAV-2.21-test) contained only two images, so accuracy is not considered when observing the trend. This trend suggests that higher-resolution images might introduce unnecessary complexity, which does not necessarily improve detection accuracy. At the finest image resolution (1.87 mm), the number of seedlings was underestimated (Fig. 14a and Fig. 14f) due to the omission of small seedlings (Fig. 15a and Fig. 15f).

3.3.4. Seedling detection at different seedling compositions in UAV

Examples of the seedling detection results of different seedling compositions are illustrated in Fig. 17. The distribution of the big and small bounding boxes clearly reflected the proportion of small seedlings, which indicated that the detected results correctly reflect the seedling compositions. As the proportion of small seedlings increased, more detection errors occurred (Fig. 18), although the accuracy remained high (Table 9). The highest detection accuracy was achieved when all seedlings had similar sizes (the 0 % scenario), with an rRMSE of 1.09 % for YOLOv8n-LP and 2.28 % for YOLOv11n-LP. As the small seedling proportion increased to 20 %, 40 %, and 60 %, the detection accuracy of YOLOv8n-LP decreased to rRMSE values of 5.55 %, 4.36 %, and 9.29 %,

while YOLOv11n-LP demonstrated rRMSE values of 4.26 %, 5.64 %, and 9.20 %, respectively. Despite the decrease in accuracy, both the YOLOv8n-LP (Fig. 18 a-d) and YOLOv11n-LP (Fig. 18 e-h) models maintained strong correlations between the true number of seedlings and the estimated values, demonstrating their robustness for seedlings with uneven growths.

3.3.5. Seedling detection at different planting densities in UAV

The model performance was further evaluated under varying seedling densities (Fig. 19), although only trained with images at the 22.2 cm spacing. YOLOv8n-LP achieved its best performance at a spacing of 22.2 cm, with rRMSE = 3.38 %, while YOLOv11n-LP performed the best at the densest spacing of 15.9 cm, with rRMSE = 3.96 %. At wider plant spacings of 37.0 cm and 60.0 cm, both models experienced a decline in performance. For instance, YOLOv8n-LP's rRMSE increased to 5.28 % and 7.93 %, while YOLOv11n-LP recorded rRMSE values of 8.87 % and 8.71 % (Table 9). These results suggest that YOLOv8n-LP performs better under moderate seedling densities (22.2 cm and 37.0 cm), while YOLOv11n-LP is more trustworthy at the extreme density levels (i.e., 15.9 cm) (Fig. 20).

3.4. Leaf age monitoring using YOLOv8n-LP and YOLOv11n-LP

3.4.1. Individual-plant-level leaf counting

The proposed models also demonstrated outstanding performance in leaf age detection, whether on NG devices or UAVs. There is a high correlation between actual and predicted leaf ages at the individual plant scale (Fig. 21). The predicted data was concentrated around the

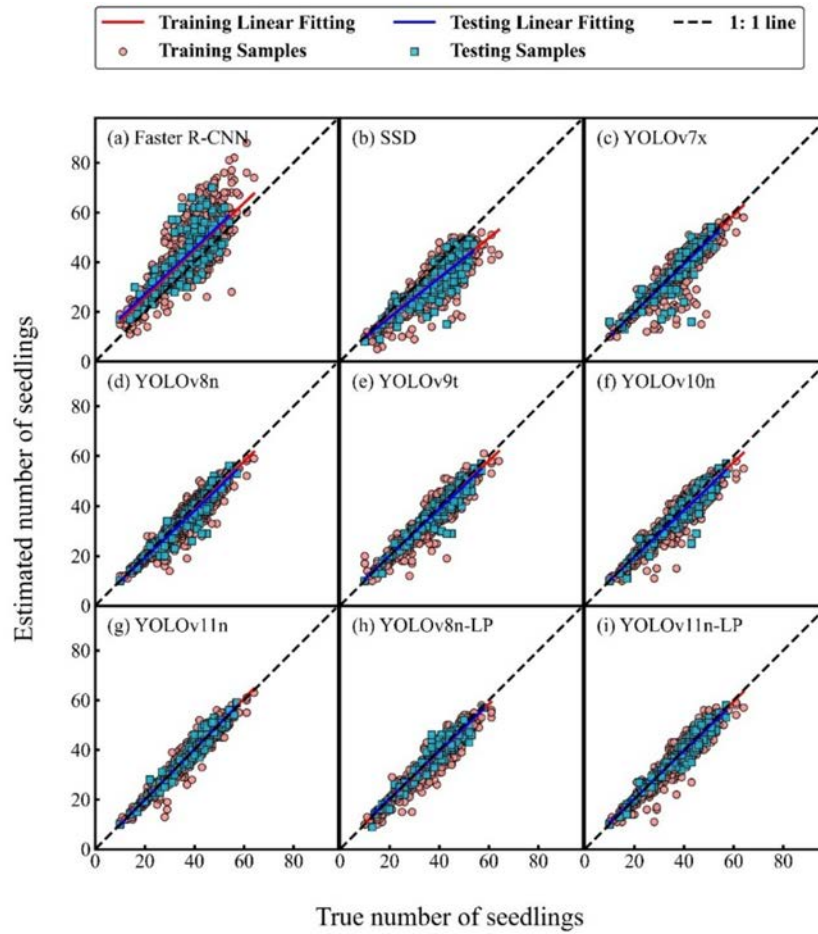


Fig. 10. Comparison of ground truth and estimated number of seedlings based on NG-train and NG-test datasets.

1:1 line, indicating that the model predictions are very close to the actual results. The statistical distribution of the individual plant leaf ages was well reflected in the model detected results. The RMSE of both models on the NG dataset was half a leaf. On the UAV dataset, the YOLOv8n-LP model still maintained RMSE = 0.503. However, the YOLOv11n-LP had much higher RMSE = 1.024, with severe underestimation errors.

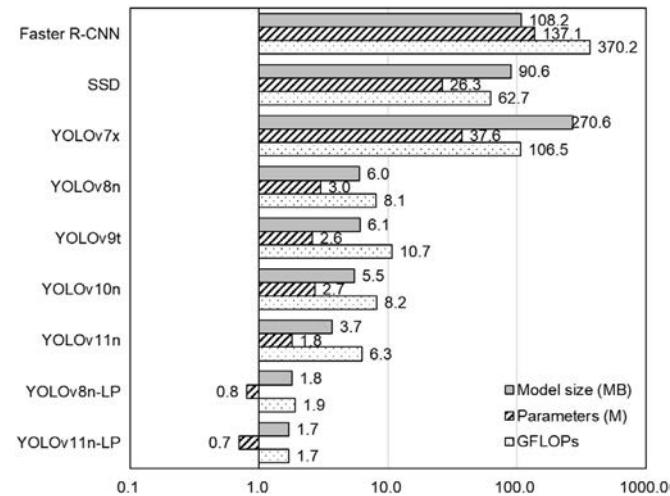


Fig. 11. Model complexity.

3.4.2. Plot-level leaf age monitoring

Table 10 presents the performance evaluation of YOLOv8n-LP and YOLOv11n-LP across two platforms (NG and UAV). YOLOv8n-LP outperformed YOLOv11n-LP on both platforms. On the NG platform, YOLOv8n-LP achieved an R^2 of 0.88, RMSE of 0.143, and rRMSE of 5.73 %, while YOLOv11n-LP achieved an R^2 of 0.75, RMSE of 0.188, and rRMSE of 7.54 %. On the UAV platform, YOLOv8n-LP reaches an R^2 of 0.93, RMSE of 0.204, and rRMSE of 9.24 %, clearly outperforming YOLOv11n-LP, which had an R^2 of 0.89, RMSE of 0.319, and rRMSE of 14.44 %.

On the NG platform, YOLOv8n-LP tended to underestimate the leaf age, with more samples labeled as V2 (Fig. 22a), while YOLOv11n-LP tended to overestimate the leaf age, with more samples labeled as V3 (Fig. 22b). The two models had similar accuracy. For the UAV platform, the trend was opposite with YOLOv8n-LP tending to give higher estimates and YOLOv11n-LP favoring lower estimates. As a result, YOLOv11n-LP had higher classification accuracy for V1 and YOLOv8n-LP had high accuracy for V2–V4.

4. Discussion

In this study, we proposed two lightweight models, YOLOv8n-LP and YOLOv11n-LP, for maize seedling counting and leaf age monitoring in the field. Despite the compact size (1.8 MB and 1.7 MB), the models demonstrated exceptional performance in detecting and tracking maize seedling growth using multi-temporal RGB images. They achieved satisfactory results across images captured from both NG and UAV platforms, demonstrating high robustness and reliability, thereby

Table 8
Image processing time (s/image).

Model	Device 1 - CUDA	Device 4 - CUDA	Device 1	Device 2	Device 3	Device 4	Device 5
Faster R-CNN	0.175 ± 0.028	8.647 ± 1.104	23.931 ± 0.221	4.906 ± 0.080	120.253 ± 22.038	95.606 ± 5.088	3936.663 ± 332.272
SSD	0.245 ± 0.013	0.980 ± 0.207	0.238 ± 0.009	0.634 ± 0.021	0.390 ± 0.026	1.013 ± 0.243	6.364 ± 0.384
YOLOv7x	0.171 ± 0.033	1.189 ± 0.651	0.700 ± 0.012	2.068 ± 0.037	2.046 ± 0.623	3.865 ± 0.590	23.608 ± 0.518
YOLOv8n	0.026 ± 0.015	0.053 ± 0.022	0.056 ± 0.004	0.145 ± 0.006	0.123 ± 0.006	0.201 ± 0.033	1.603 ± 0.775
YOLOv9t	0.037 ± 0.020	0.077 ± 0.022	0.148 ± 0.006	0.280 ± 0.010	0.151 ± 0.012	0.326 ± 0.039	2.548 ± 0.244
YOLOv10n	0.037 ± 0.023	0.061 ± 0.045	0.111 ± 0.006	0.178 ± 0.007	0.177 ± 0.043	0.250 ± 0.031	1.855 ± 0.300
YOLOv11n	0.040 ± 0.026	0.058 ± 0.051	0.094 ± 0.004	0.157 ± 0.013	0.134 ± 0.013	0.203 ± 0.024	1.513 ± 0.266
YOLOv8n-LP	0.021 ± 0.016	0.053 ± 0.025	0.046 ± 0.003	0.103 ± 0.010	0.087 ± 0.004	0.127 ± 0.022	0.916 ± 0.105
YOLOv11n-LP	0.037 ± 0.023	0.055 ± 0.029	0.066 ± 0.003	0.130 ± 0.012	0.105 ± 0.015	0.156 ± 0.019	1.046 ± 0.129

Note: in each column, the top two models are bolded and the top one model is italic.

making them highly efficient for field applications with minimal computational overhead.

4.1. Significance of the proposed models

The lightweight YOLOv8n-LP and YOLOv11n-LP models proposed in this study make significant contributions to the field of maize seedling and leaf age monitoring, addressing three key challenges in current research. First, they provide a balanced solution by offering high accuracy and reduced model complexity. Second, increased environmental adaptability was achieved. Third, the new models were capable of simultaneously monitoring maize seedling count and leaf age in complex field environments. Hence, they represent a promising tool for precision agriculture and can significantly enhance decision-making in maize growth management. It is also fair to infer that the model has high potential to be applied to similar crops such as sorghum. With proper training, it may also be transferred to other row-planting crops.

4.1.1. Comparison with existing models

First, the proposed models demonstrate significant improvement in accuracy, computational efficiency, and model size. Earlier lightweight models targeting large YOLOv4 or YOLOv5 models were over 40 MB in model size (Gao et al., 2022b; Li and Wu, 2022), while the maize seedling detection precision varied from 88.5 % to 96.25 %. The high demand for storage and computational resources made them less practical in resource-constrained environments. A recent study shrank the model size to 5.04 MB by modifying the YOLOX-tiny model, which was

successfully applied in a maize tassel detection task (Song et al., 2023). The YOLOv8n-LP model proposed in this study is as small as 1.8 MB with only 0.8 M parameters, while YOLOv11n-LP is only 1.7 MB with 0.7 M parameters. This reduction in model size significantly lowers the computational burden. The time required for processing individual images was reduced by at least 20 %, which got more significant when the device had less computing power. We have installed the YOLOv8n-LP model in a Raspberry Pi4B device which was only 8 cm in size with a 64-bit 1.5 GHz quad-core CPU, and accurately counted maize seedlings within 0.4 s for a single image. Meanwhile, a high level of accuracy was maintained. The accuracy is even higher than the heavyweight models Faster R-CNN, SSD, and YOLOv7x, while comparable to YOLOv8n, YOLOv9t, and YOLOv10n (Fig. 8). The optimal balance between accuracy and model size enables the model to be deployed effectively in real-time monitoring systems and on mobile devices. For future applications, these lightweight modifications can be implemented to create small models to be installed on more affordable devices.

Second, the proposed models exhibit strong robustness under various field environments. Applied to multiple complex scenarios with different platforms, leaf ages, resolutions, and seedling compositions, the maize seedling counting error ranged from $rRMSE = 1.09\%$ to $rRMSE = 9.29\%$ (Table 9). Most existing studies have considered only one to two factors (e.g., Bai et al., 2022; Liu et al., 2023). Jia et al. (2024) assessed five deep learning models (i.e., YOLOv8n, YOLOv5n, Deformable DETR, Faster R-CNN and YOLOv3-tiny) in counting maize seedling with varying planting densities, growth stages, and flight altitudes, tests of five

Table 9
Evaluation of YOLOv8n-LP and YOLOv11n-LP across five different factors.

Experiment	Category	Dataset	YOLOv8n-LP			YOLOv11n-LP		
			R ²	RMSE	rRMSE	R ²	RMSE	rRMSE
Platform	NG	NG-test	0.91	2.64	6.73 %	0.94	2.19	5.59 %
	UAV	UAV-test	0.96	6.74	5.66 %	0.95	6.96	5.85 %
Leaf age	V2	NG-V2-test	0.96	1.84	4.50 %	0.96	1.67	4.07 %
	V3	NG-V3-test	0.94	2.39	7.27 %	0.95	2.18	6.23 %
	V4	NG-V4-test	0.68	2.43	5.63 %	0.77	2.04	4.73 %
	V5	NG-V5-test	0.56	2.56	6.93 %	0.60	2.95	6.94 %
	1.87	UAV-1.87-test	0.96	8.64	6.74 %	0.95	8.74	6.82 %
Resolution (mm)	2.21*	UAV-2.21-test*	N/A*	5.00*	3.82 %*	N/A*	2.00*	1.53 %*
	2.95	UAV-2.95-test	0.95	8.57	6.31 %	0.94	8.95	6.59 %
	3.40	UAV-3.40-test	0.90	3.46	3.30 %	0.95	3.48	3.34 %
	4.47	UAV-4.47-test	0.93	3.02	3.06 %	0.81	4.05	4.11 %
	0	UAV-0-test	0.96	1.06	1.09 %	0.84	2.20	2.28 %
Seedling composition (%)	20	UAV-20-test	0.97	7.43	5.55 %	0.96	5.71	4.26 %
	40	UAV-40-test	0.93	5.22	4.36 %	0.94	6.75	5.64 %
	60	UAV-60-test	0.94	9.87	9.29 %	0.95	9.77	9.20 %
	60.0	UAV-60.0	0.78	4.77	7.93 %	0.69	5.24	8.71 %
Density (cm) (not included in the training process)	37.0	UAV-37.0	0.73	5.66	5.28 %	0.53	9.52	8.87 %
	22.2	UAV-22.2	0.79	5.51	3.38 %	0.60	9.79	6.00 %
	15.9	UAV-15.9	0.33	21.65	10.82 %	0.48	7.92	3.96 %

Note: UAV-2.21-test* only contained one image.

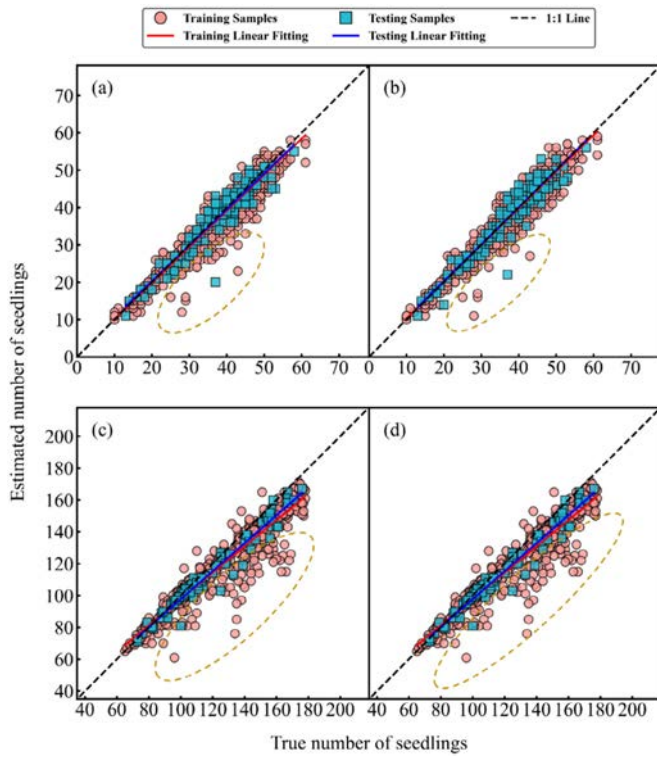


Fig. 12. Comparison of ground truth and estimated number of maize seedlings. (a) YOLOv8n-LP model applied to the NG data. (b) YOLOv11n-LP model applied to the NG data. (c) YOLOv8n-LP model on UAV data. (d) YOLOv11n-LP model on UAV data.

models. While working only on UAV images, they had similar findings to our study, that accuracy declined as planting density, canopy overlap, and altitude increased. Our study went one-step further, considering not only UAVs but also near-ground or handheld devices. The results demonstrate that the YOLOv8n-LP and YOLOv11n-LP models maintain robust performance under complex field conditions, thereby providing a reliable and efficient solution for real-time field monitoring.

Additionally, the YOLOv8n-LP and YOLOv11n-LP, with a plant-to-leaf two-step application, can monitor the maize leaf age together with seedling count from images captured in the field environment. Compared to indoor single-plant maize leaf detection (Ning et al.,

2024; Xie et al., 2023), this method more accurately captures the variations in actual growing conditions, overcoming the complexities of field environments that indoor experiments cannot replicate. Compared to existing in-field leaf counting studies that segment plants with one deep learning model and then identify leaves with another (Xu et al., 2023; Xu et al., 2022), our proposed approach grants three advantages. First, the annotation process for plant identification is markedly simplified. Instead of digitizing the exact maize plant boundary, pulling simple bounding boxes is enough. Second, because a single model structure is used in both steps, the configuration of the coding environment becomes much more streamlined. Third, computational cost is significantly reduced. Switching to YOLOv8n-LP and YOLOv11n-LP from the larger models, the two-step process can be carried out with a minimal 1.4–1.6 million parameters.

4.1.2. Visualization of model attention

We have conducted visual interpretability analysis using gradient-weighted class activation mapping (Grad-CAM) (Selvaraju et al., 2017) to highlight the attention regions of the proposed model during inference. Four examples are provided, with two showing the seedling detection task (Fig. 23a and Fig. 23b) and two for the leaf-tip detection task (Fig. 23c and Fig. 23d). The top row displays the original RGB images and annotated bounding boxes, the middle row shows the attention heatmaps by YOLOv8n-LP overlaid on the input images, while the bottom row shows the attention heatmaps by YOLOv11n-LP overlaid on the input images. The example in Fig. 23a has maize seedlings of different orientations and sizes, with small weeds and cracked soil as the background. The example in Fig. 23b has maize seedlings with small weeds and flat terrain as the background. Fig. 23c shows a maize seedling at V3 stage exposed to intense illumination, while Fig. 23d shows a seedling having four leaf-tips with moderate illumination. In all the examples, both models focus on the central region of each target. These visualizations demonstrate that the model is not only learning spatially meaningful patterns but also aligning well with human-interpretable plant structures.

4.1.3. Ablation interpretation

The mechanism behind the significance of the proposed models is tested through the ablation experiment and discussed as follows. The ablation study results (Table 6, Table 7) indicate that the combination of the DAttention mechanism, Conv-BiFPN module, EfficientHead, and LAMP significantly reduced computational cost while maintaining good model accuracy.

The DAttention mechanism, for both YOLOv8n-LP and YOLOv11n-LP, improved the model accuracy. We suppose the enhanced accuracy

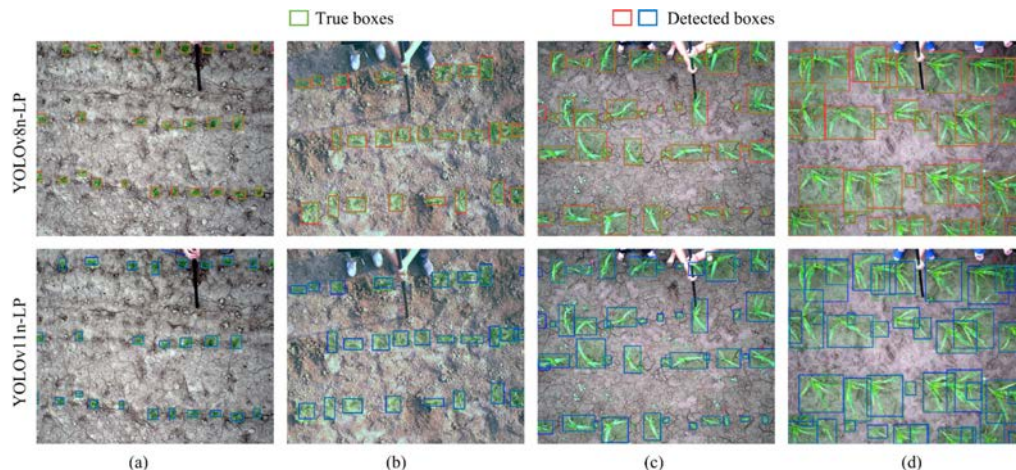


Fig. 13. Seedling detection results of different leaf age stages (V2, V3, V4, and V5 respectively from left to right). (a–d) YOLOv8n-LP results. (e–h) YOLOv11n-LP results.

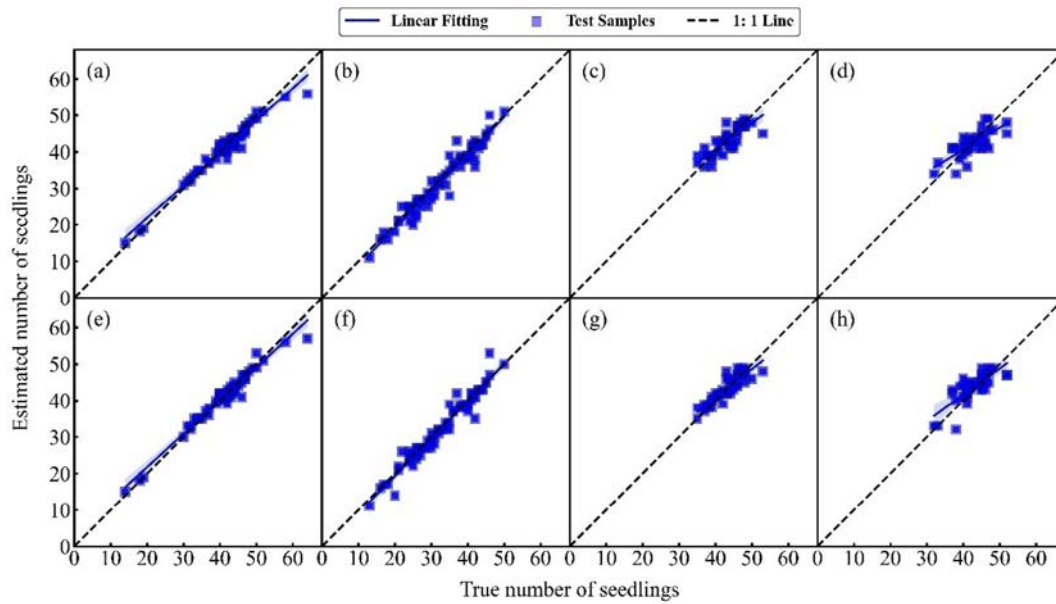


Fig. 14. Comparison of ground truth and estimated seedling numbers at different leaf stages (V2, V3, V4, and V5 respectively from left to right). (a-d) YOLOv8n-LP results. (e-h) YOLOv11n-LP results.

can be attributed to the improved efficacy in capturing fine-grained details. This is supported by Wang et al. (2024b), which similarly highlighted that the DAttention module significantly enhanced model performance, primarily due to its ability to capture key features and improve overall model accuracy. The conclusion aligns with the findings of this study, further supporting the potential application of this mechanism in complex scenarios.

The Conv-BiFPN module slightly degraded model accuracy while significantly reducing the model size, by integrating multi-scale feature fusion. These results demonstrate the superior ability of BiFPN to efficiently capture key information across different scales. In comparison, Mo and Wei (2024) introduced the traditional BiFPN into the segmentation model YOLOv8n-seg, which significantly enhanced multi-scale feature integration and improved segmentation accuracy. However, the

computational complexity increased, challenging the model efficiency. The Conv-BiFPN module proposed in this study reduces computational load through architectural optimization. This improvement achieves a better balance between efficiency and performance, providing enhanced practicality and adaptability for object detection tasks in complex scenarios.

The EfficientHead module further reduced computational complexity. By introducing the group convolution strategy, the computation load was significantly reduced. While the DSC in YOLOv11n is already a quite effective lightweight structure compared to the regular dense convolution in YOLOv8n, the EfficientHead adopted in this study successfully further reduced the model size by about 10 %. Compared to original decoupled head, the EfficientHead achieved a good balance between efficiency and performance, providing an effective and practical

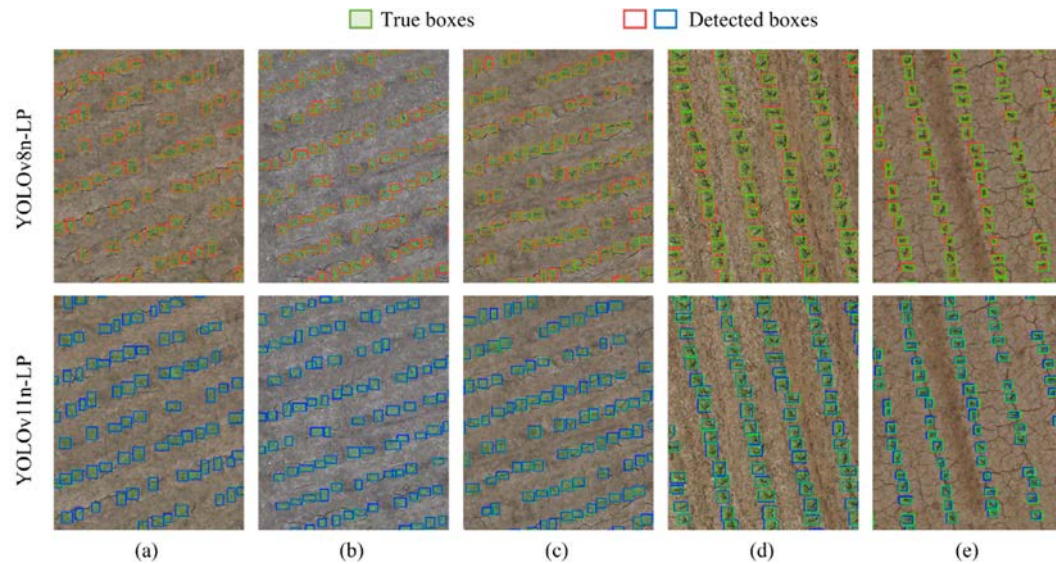


Fig. 15. Seedling detection results of different spatial resolutions (1.87 mm, 2.21 mm, 2.95 mm, 3.40 mm, and 4.47 mm, respectively from left to right). (a-e) YOLOv8n-LP results. (f-j) YOLOv11n-LP results.

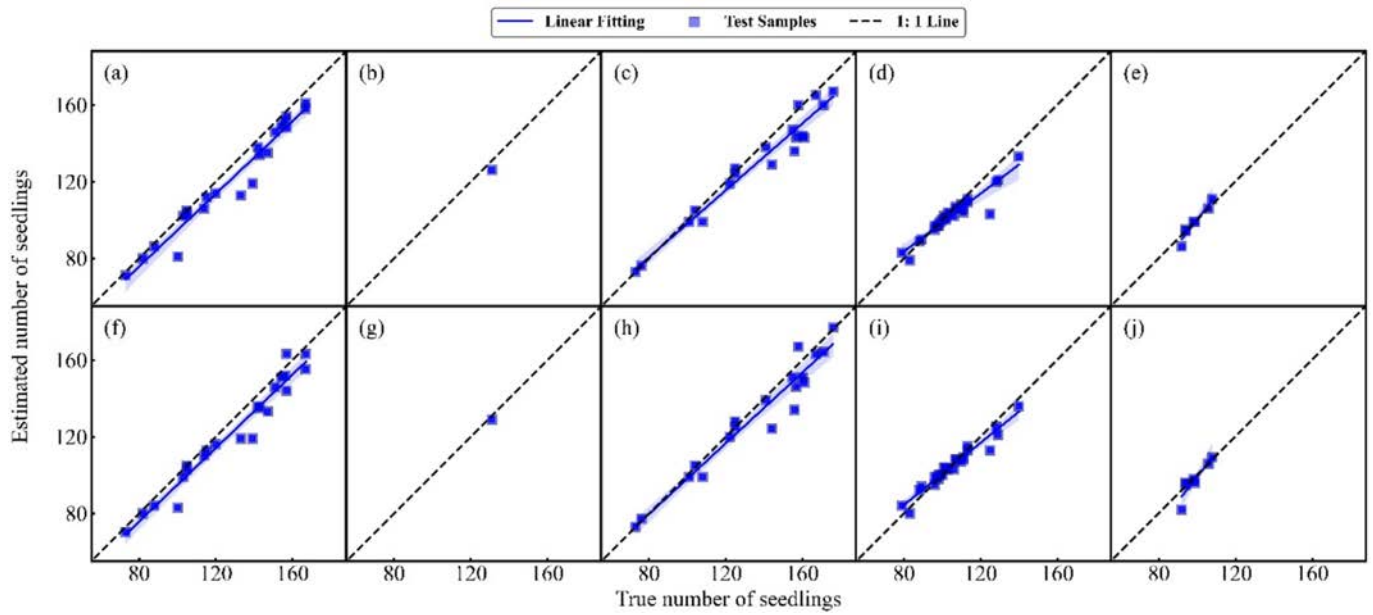


Fig. 16. Comparison of ground truth and estimated seedling numbers based on different spatial-resolution images using YOLOv8n-LP (a-e) and YOLOv11n-LP (f-j). The resolutions from left to right are 1.87 mm, 2.21 mm, 2.95 mm, 3.40 mm, and 4.47 mm.

solution for object detection tasks in complex scenarios. This echoes conclusions in previous studies that lightweight design for detection head structures can significantly enhance inference speed and resource efficiency while maintaining detection accuracy (Tan et al., 2020).

Finally, by utilizing LAMP pruning techniques, the model parameter and overall model size were reduced by over 50 %. Although rarely used in crop seedling detection, over three thousand papers on pruning techniques have been published in the past five years (Cheng et al., 2024). The LAMP algorithm used in this study belongs to the unstructured pruning category, which is the finest-grained pruning (Cheng et al., 2024; Su et al., 2020). It is applied after training, following a pretrain-prune-retrain process. In the future, it is worthwhile to test other types of pruning techniques for seedling detection tasks.

4.2. Robustness of proposed models in seedling counting

4.2.1. Robustness across trained scenarios

The proposed models demonstrated excellent and stable detection performance across different platforms, leaf ages, resolutions, and seedling compositions.

4.2.1.1. Different platforms. When applied to UAV imagery, the model exhibited robust performance, with R^2 and rRMSE values comparable to those obtained in NG scenarios. The minimal differences underscore the model's adaptability in diverse environments (Yu et al., 2024). We notice that the training dataset exhibited some underestimation (Fig. 12), primarily from images containing the highest proportion of

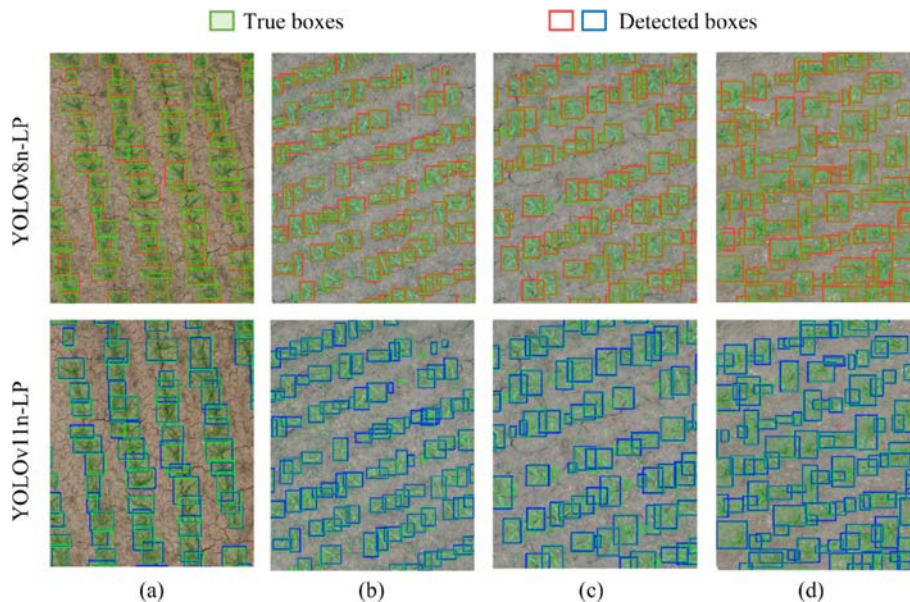


Fig. 17. Seedling detection results of different seedling compositions, with the proportion of small seedling being 0 %, 20 %, 40 %, and 60 % from left to right. (a-d) YOLOv8n-LP results. (e-h) YOLOv11n-LP results.

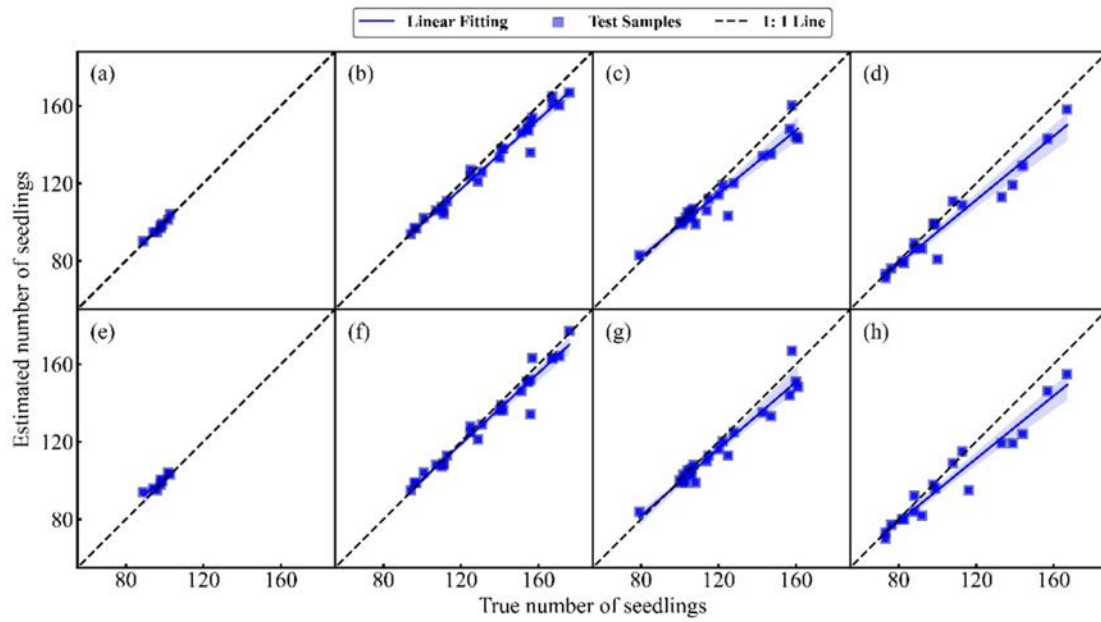


Fig. 18. Comparison of ground truth and estimated number of seedlings based on different seedling composition images by YOLOv8n-LP (a-d) and YOLOv11n-LP (e-h). The proportion of small seedlings from left to right is 0 %, 20 %, 40 %, and 60 %, respectively.

small seedlings and their augmented dataset (Fig. 24). Two key factors likely contributed to this underestimation: the significant morphological differences between small and large seedlings during the early growth stage and the favorable conditions for weed growth at the early seedling stage, which resulted in dense weed coverage that interfered with detection. Future research could prioritize improving the ability of the model to detect small seedlings. Despite these challenges, the YOLOv8n-LP and YOLOv11-LP models maintained strong and

consistent performance in maize seedling detection across both NG and UAV scenarios, supporting efficient crop monitoring under varying conditions and scales.

The UAV dataset was augmented for model training. To check the impact of data augmentation on the model performance, we examined the model accuracy with and without augmentation. We trained two YOLOv8n-LP models, one trained using only the original data, and the other trained using the augmented dataset. The accuracy of both models

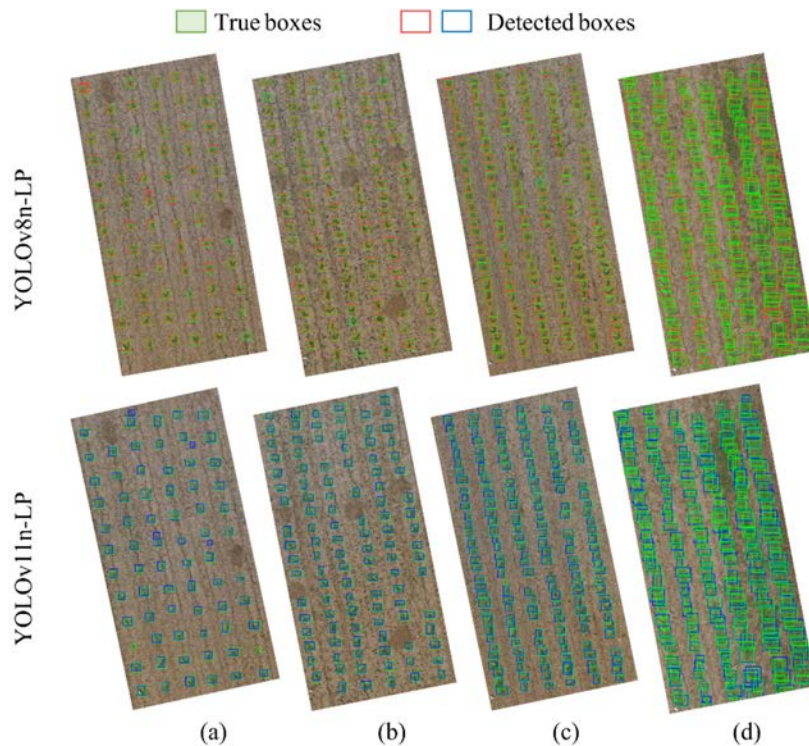


Fig. 19. Seedling detection results of different plant spacings (60.0 cm, 37.0 cm, 22.2 cm, and 15.9 cm from left to right). (a-d) YOLOv8n-LP results. (e-h) YOLOv11n-LP results.

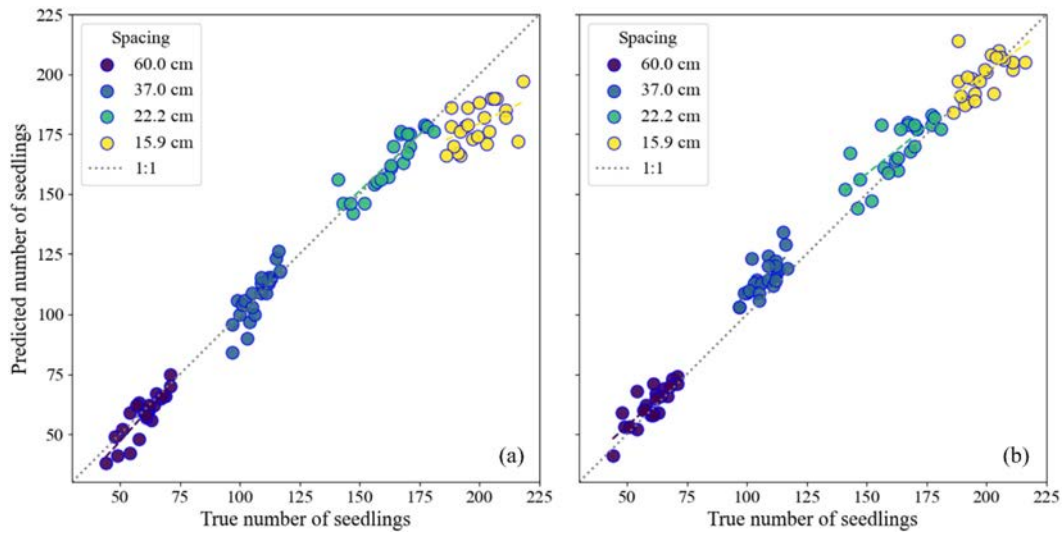


Fig. 20. Comparison of ground truth and estimated number of seedlings in plots with different planting densities by YOLOv8n-LP (a) and YOLOv11n-LP (b).

was tested using an independent test dataset comprised with only original data. The model precision increased from 0.948 to 0.951, recall increased from 0.870 to 0.934, and AP increased from 0.912 to 0.964. The same pattern was found for YOLOv11n-LP, with precision, recall, and AP increasing from 0.948, 0.862, and 0.908 to 0.948, 0.918, and 0.956, respectively. These results demonstrate that the data augmentation step did not add noise to the model but rather brought positive impacts.

The NG platform, although a bit slow to operate in the field, holds several advantages as a maize seedling monitoring device. First, the data collection is straightforward compared to UAV data collection which requires professional software, training, and pre-processing. Second, the NG data has higher spatial resolution than UAV images, which contributed to the higher accuracy of the models. Third, the NG

platform is useful toward the development of more convenient equipment in the future. Four, it is much more affordable than UAV. We propose several pathways to streamline the NG data collection: (1) to mount the pole on the self-propelled crane with precise GPS or plot tracking, and (2) to integrate the model into other NG devices such as smartphones. Such modifications would further enhance portability, lower hardware costs, and make the system more accessible for broader agricultural applications.

4.2.1.2. Different leaf ages in NG images. YOLOv8n-LP and YOLOv11n-LP exhibited a consistent performance trend across datasets with different leaf-age stages. The model performed exceptionally well during the early stages of development (V2), achieving high detection accuracy due to the relatively simple structure of the seedlings, where leaves are more distinct and less likely to overlap. The low accuracy during the V3 stage (Table 9) was due to the interference caused by experiments with seedlings of varying sizes (Bai et al., 2022). In the same plot, seedlings with late emergence were photographed at the VE stage, appearing very similar to weeds (Fig. 25a), which caused a significant error. The accuracy of the identification results improved substantially after two days, when the maize seedlings had grown (Fig. 25b). When detecting maize seedlings beyond the VE stage, weed interference has minimal impact on model accuracy (Fig. 26). However, the model accuracy declined as the leaf age progressed to the V5 stage. This decline can be attributed to increased leaf overlap during seedling development, which complicates identifying individual leaves (Bai et al., 2022). Overlapping leaves obscure key features within the image, making it difficult for the model to detect and classify the seedlings accurately.

In future experiments, spectral information can be utilized to further differentiate between weeds and seedlings at the VE stage. Spectral information provides the reflectance characteristics of plants at different wavelengths, which is very effective for distinguishing plants that are morphologically similar but have different physiological characteristics.

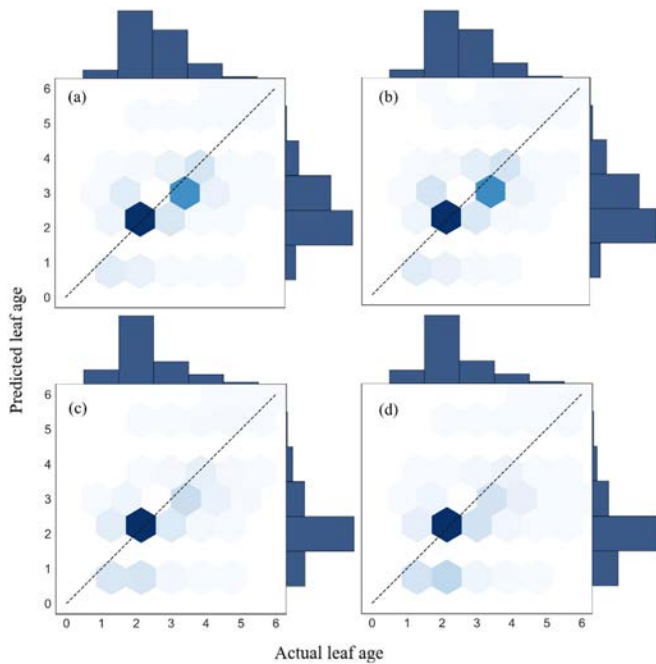


Fig. 21. Evaluation of individual-plant-level leaf counting. (a) YOLOv8n-LP on the NG-leaf-test dataset. (b) YOLOv11n-LP on the NG-leaf-test dataset. (c) YOLOv8n-LP on the UAV-leaf-test dataset. (d) YOLOv11n-LP on the UAV-leaf-test dataset.

Table 10
Evaluation of the proposed models across two platforms in plot-level leaf age estimation.

Platform	Model	R^2	RMSE	rRMSE
NG	YOLOv8n-LP	0.88	0.143	5.73 %
	YOLOv11n-LP	0.75	0.188	7.54 %
UAV	YOLOv8n-LP	0.93	0.204	9.24 %
	YOLOv11n-LP	0.89	0.319	14.44 %

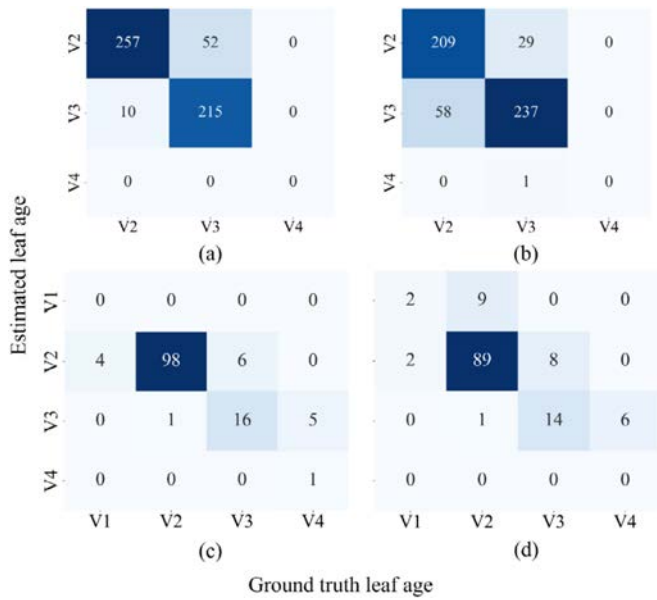


Fig. 22. Confusion Matrix of plot-level leaf counting. (a) YOLOv8n-LP on the NG dataset. (b) YOLOv11n-LP on the NG dataset. (c) YOLOv8n-LP on the UAV dataset. (d) YOLOv11n-LP on the UAV dataset.

By integrating spectral information with existing detection models, the accuracy of maize seedling detection can be significantly improved, reducing the occurrence of error points and thereby providing more reliable data support for maize field management.

4.2.1.3. Different spatial resolution in UAV images. The YOLOv8n-LP and YOLOv11n-LP models demonstrated strong robustness identifying maize seedlings from UAV images of varying resolutions. Typically, higher image resolution provides more detailed features, contributing to improved detection accuracy. However, in this study, higher-resolution images performed less effectively in detection tasks. This discrepancy is attributed to information loss during image preprocessing, as the model resizes all input images to a uniform size of 640×640 pixels. Images with resolutions of 1.87 mm and 2.95 mm, initially sized at 3985×4472 and 2486×2781 pixels, were resized by factors

of 6 and 4, respectively. We suppose that the down-sampling obscured critical texture and small object details, which ultimately resulted in poor detection performance. In contrast, images with lower resolution (such as 3.40 mm and 4.47 mm) were able to retain relatively more useful features during compression, which contributed to improved detection performance. These issues demonstrate that in detection tasks involving high-resolution images, it is not only the original resolution that matters but also the balance between image compression, information retention, and model adaptation, which is crucial for achieving optimal detection performance (Yu et al., 2024).

4.2.1.4. Different seedling compositions in UAV images. Based on Fig. 17, it is evident that the YOLOv8n-LP and YOLOv11n-LP models maintain a strong correlation between the actual and predicted number of maize seedlings across varying proportions of small seedlings. As the proportion of small seedlings increases (from 0 % up to 60 %), there is a slight increase in the deviation from the 1:1 line, indicating a gradual rise in the rRMSE. This trend suggests that the complexity of the seedling composition deteriorate the model accuracy, which is consistent with the findings of Bai et al. (2022). However, both models still demonstrate a high level of accuracy, showing robustness even in more challenging scenarios with mixed seedling sizes. These findings underscore the model's adaptability in diverse field conditions and suggest that YOLOv8n-LP and YOLOv11n-LP are well-suited for practical applications in UAV-based maize seedling monitoring.

4.2.2. Generalizability to untrained scenarios

Although the training dataset only contained images of maize plots with 22.2 cm spacing, both YOLOv8n-LP and YOLOv11n-LP were able to accurately identify maize seedlings at lower and higher planting densities. YOLOv8n-LP showed superior generalizability to moderate plant densities (22.2 cm and 37.0 cm spacings), which are typical of conventional field conditions. Under such conditions, the canopy overlap was moderate and the detection mainly depended on clear object boundaries, which YOLOv8n-LP could capture effectively. In contrast, YOLOv11n-LP demonstrated greater robustness under extreme densities at 15.9 cm. At such planting density, the canopy overlap increased substantially, resulting in more occlusion and reduced visible gaps between plants. These challenging conditions placed higher demands on the model's ability to extract discriminative features from crowded individuals. YOLOv11n-LP demonstrated greater robustness in this scenario than YOLOv8n-LP, probably thanks to its added module,

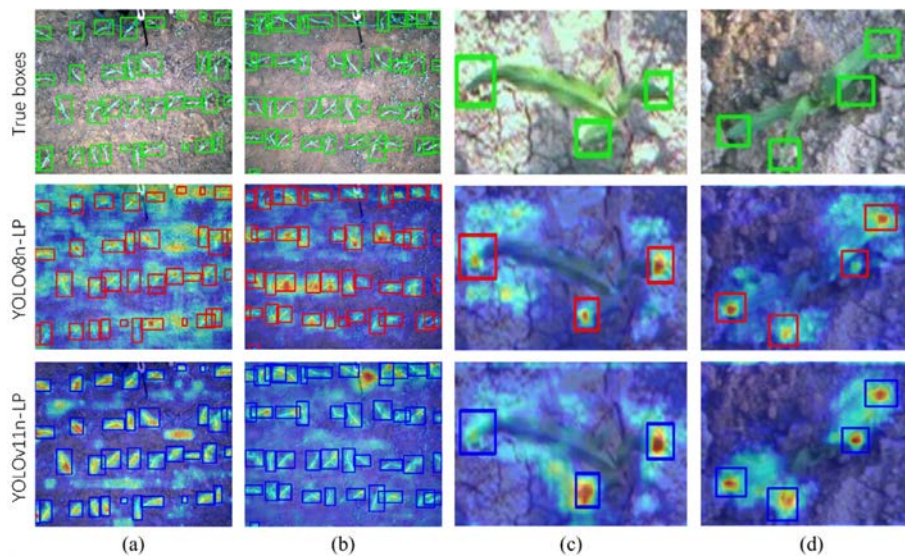


Fig. 23. Visualizations for the seedling counting task based on Grad-CAM: (a, b) Maize plots for seedling detection. (c, d) Maize seedlings for leaf-tip detection.

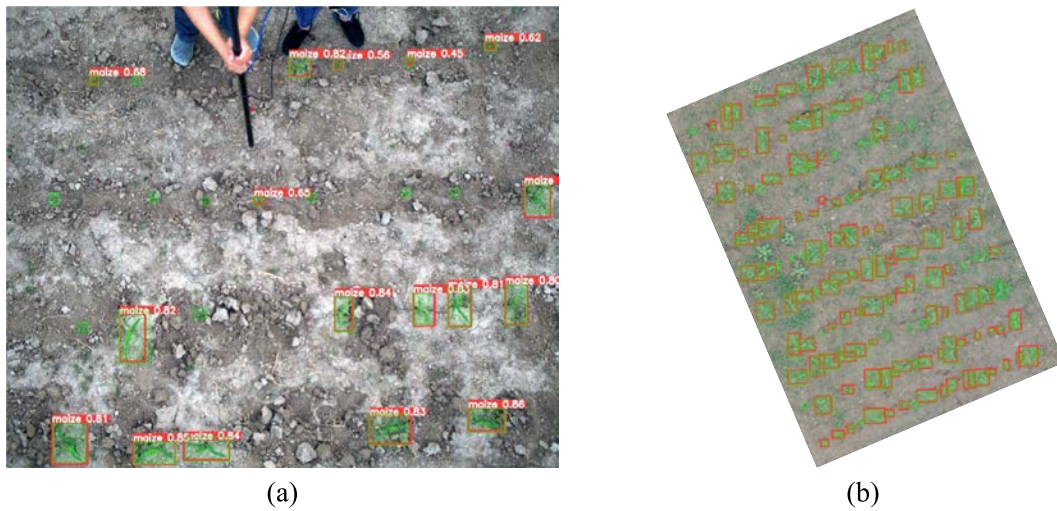


Fig. 24. Maize seedling detection results. (a) Image with 60 % small seedlings at NG dataset (Sep. 7, 2023). (b) Image with 60 % small seedlings at UAV dataset (July 13, 2023).

i.e., Cross-Channel Position-aware Spatial Attention (C2PSA). C2PSA incorporates a parallel spatial attention mechanism that strengthens the model's ability to distinguish fine-grained spatial features under conditions of severe occlusion (Khanam and Hussain, 2024). This possibly has led to the improved performance of YOLOv11n-LP in discriminating highly overlapped plants, thereby outperforming YOLOv8n-LP at extreme planting densities. This characteristic is of practical importance: as modern maize cultivation increasingly shifts toward high-density planting to maximize land use efficiency (Jafari et al., 2024), the stronger performance of YOLOv11n-LP in dense planting conditions suggests it may offer greater potential for deployment in precision agriculture applications.

4.3. Leaf age monitoring using proposed models

The YOLOv8n-LP and YOLOv11n-LP models exhibited strong performance in leaf age detection on both near-ground and UAV-acquired datasets, showing high precision and accuracy, especially during the early V2 stage of maize seedling growth. Results indicate that the model can predict leaf age with minimal error, as demonstrated by

the strong correlation between predicted and actual values. However, several limitations were identified. The detection accuracy declined in the later stages (V5), likely due to increased leaf overlap, which complicates feature extraction (Xu et al., 2022). This observation is consistent with the results discussed in Section 4.2.1, where plant detection accuracy also decreased during the V5 stage.

In the context of plot-level leaf age estimation, YOLOv8n-LP exhibited better generalization and adaptability, outperforming YOLOv11n-LP on both NG and UAV platforms. Both models had lower accuracy processing the UAV data (rRMSE = 9.24 %–14.44 %) than the NG data (rRMSE = 5.73 %–7.54 %). This is possibly due to lost fine-grained features essential for seedling detection and growth stage assessment (Zhang et al., 2024). The higher RMSE values observed on the UAV dataset call for more effort on leaf age identification from low-resolution or aerial imagery.

One challenge in leaf age monitoring is that some leaves get occluded by other leaves above them because the NG and UAV images are taken from nadir view. Our strategy of targeting leaf tips instead of the entire leaves enabled us to significantly alleviate this problem. Nonetheless, looking ahead, to obtain oblique photos from multiple

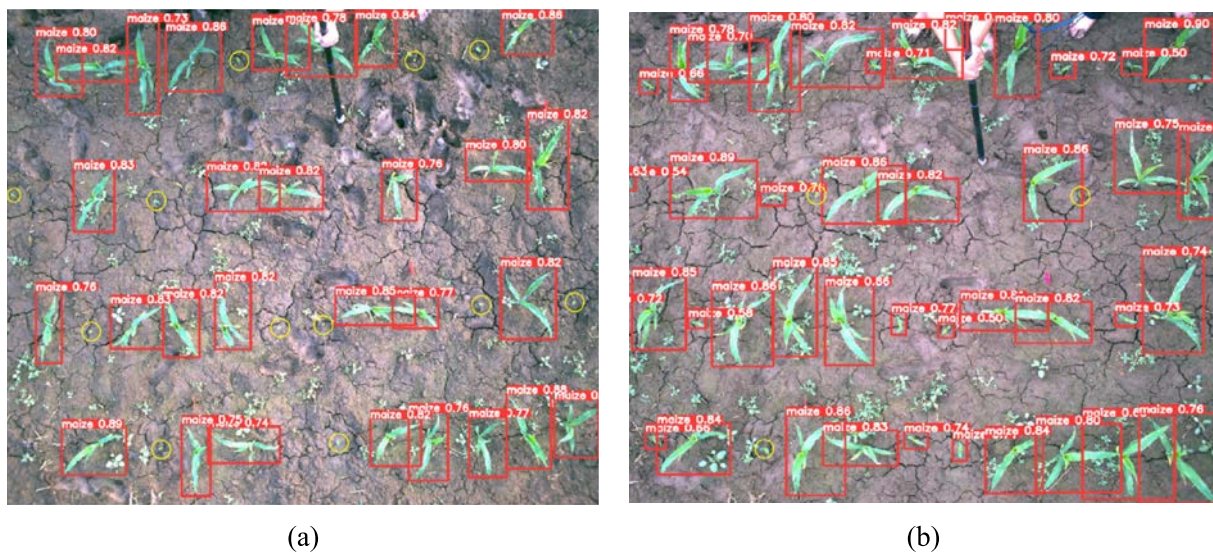


Fig. 25. Maize seedling detection results. Yellow circles mark maize seedlings that are difficult to distinguish from weeds. (a) Small seedlings at the VE stage (Sep. 7, 2023). (b) The same plot two days later (Sep. 9, 2023). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

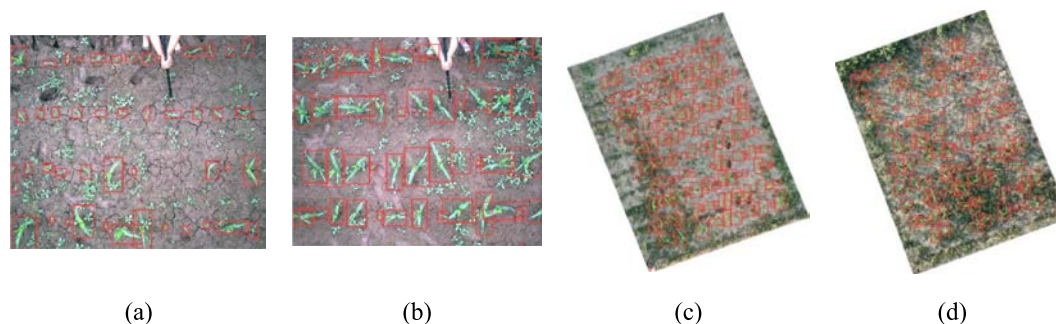


Fig. 26. Seedling detection results with different levels of weed interference: (a-b) NG dataset, (c-d) UAV dataset.

angles is worth investigating for a more complete characterization of the seedlings.

5. Conclusions

This study introduces two lightweight deep learning models, YOLOv8n-LP and YOLOv11n-LP, developed to track maize seedling count and leaf age from high-resolution RGB images captured by NG and UAV platforms. Compared to the state-of-the-art YOLOv8n and YOLOv11n models, YOLOv8n-LP (parameters = 0.8 M, model size = 1.8 MB) and YOLOv11n (parameters = 0.7 M, model size = 1.7 MB) reduce parameters by over 73 %. The image processing efficiency was significantly improved, especially on devices with limited computation power. Meanwhile, the proposed models maintain performance comparable to YOLOv11n while outperforming classic models such as Faster R-CNN, SSD, and YOLOv7x, YOLOv8n, YOLOv9t, and YOLOv10n.

The two proposed models demonstrate strong performance across various datasets, effectively managing differences in observation platforms, leaf age, and planting density. The model achieves its highest accuracy during the V2 stage of seedling growth and maintains reliable performance across all seedling proportion scenarios, with rRMSE values below 10 %. While detection accuracy decreases at lower image resolutions, they perform optimally at 3.40 mm and remain accurate between 1.87 and 4.47 mm. It also adapts well to varying seedling compositions, performing reliably even as the proportion of small seedlings increases. Although the models were trained exclusively at a plant spacing of 22.2 cm, they maintained satisfactory accuracy in scenarios with a spacing of 37.0 cm and 60.0 cm. While YOLOv8n-LP had slightly lower accuracy (rRMSE = 10.82 %) at higher density (15.9 cm spacing), likely due to interference from weeds and plant overlap, the YOLOv11n-LP model demonstrates greater robustness (rRMSE = 3.96 % at 15.9 cm spacing). Despite these challenges, the models effectively monitor seedling counts and leaf age development, making them well-suited for real-time in-field crop monitoring applications.

In future research, the YOLOv8n-LP and YOLOv11n-LP models could be deployed on portable devices for real-time maize monitoring, leveraging their lightweight architecture. Their applicability to monitoring other crops or other maize organs is worth investigating. Moreover, future efforts can focus on improving the model scalability for large-scale agricultural applications, which involves automated target region boundary identification, enhanced weed resistance, and validation across broader geographic areas.

CRedit authorship contribution statement

Tiantian Jiang: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Liang Li:** Supervision, Project administration, Funding acquisition. **Zhen Zhang:** Supervision, Project administration, Conceptualization. **Xun Yu:** Writing – review & editing, Visualization,

Validation, Methodology, Investigation, Data curation, Conceptualization. **Yanqin Zhu:** Investigation, Data curation. **Liming Li:** Validation, Investigation. **Yadong Liu:** Investigation, Data curation. **Yali Bai:** Investigation, Data curation. **Ziqian Tang:** Investigation, Data curation. **Shuaibing Liu:** Investigation, Data curation. **Yan Zhang:** Investigation, Data curation. **Zheng Duan:** Supervision, Project administration, Conceptualization. **Dameng Yin:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Funding acquisition, Conceptualization. **Xiuliang Jin:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

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

Acknowledgements

This research was supported by National Science and Technology Major Project (2022ZD0115701), National Natural Science Foundation of China (42471361, 42301427), Nanfan special project, CAAS (YBXM2401, YBXM2402, PTXM2402, PTXM2501), and The Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences.

References

- Bai, Y., Nie, C., Wang, H., Cheng, M., Liu, S., Yu, X., Shao, M., Wang, Z., Wang, S., Tuohuti, N., Shi, L., Ming, B., Jin, X., 2022. A fast and robust method for plant count in sunflower and maize at different seedling stages using high-resolution UAV RGB imagery. *Precis. Agric.* 23, 1720–1742.
- Bai, Y., Shi, L.S., Zha, Y.Y., Liu, S.B., Nie, C.W., Xu, H.G., Yang, H.Y., Shao, M.C., Yu, X., Cheng, M.H., Liu, Y.D., Lin, T., Cui, N.B., Wu, W.B., Jin, X.L., 2023. Estimating leaf age of maize seedlings using UAV-based RGB and multispectral images. *Comput. Electron. Agric.* 215, 108349.
- Banerjee, B.P., Sharma, V., Spangenberg, G., Kant, S., 2021. Machine learning regression analysis for estimation of crop emergence using multispectral UAV imagery. *Remote Sens.* 13 (15), 2918.
- Bao, W., Yang, X., Liang, D., Hu, G., Yang, X., 2021. Lightweight convolutional neural network model for field wheat ear disease identification. *Comput. Electron. Agric.* 189, 106367.
- Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R., Görlach, J., 2001. Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell* 13, 1499–1510.
- Buters, T., Belton, D., Cross, A., 2019. Seed and seedling detection using unmanned aerial vehicles and automated image classification in the monitoring of ecological recovery. *Drones* 3 (3), 53.
- Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* 7, 1247–1250.
- Cheng, H., Zhang, M., Shi, J.Q., 2024. A survey on deep neural network pruning: taxonomy, comparison, analysis, and recommendations. *IEEE Trans. Pattern Anal. Mach. Intell.* 46, 10558–10578.
- Doebley, J., Stec, A., Hubbard, L., 1997. The evolution of apical dominance in maize. *Nature* 386, 485–488.

- Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2010. The PASCAL visual object classes (VOC) challenge. *Int. J. Comput. Vis.* 88 (2), 303–338.
- Forcella, F., 1998. Real-time assessment of seed dormancy and seedling growth for weed management. *Seed Sci. Res.* 8, 201–210.
- Gao, A., Geng, A., Zhang, Z., Zhang, J., Hu, X., Li, K., 2022a. Dynamic detection method for falling ears of maize harvester based on improved YOLO-V4. *Int. J. Agric. Biol. Eng.* 15, 22–32.
- Gao, J., Tan, F., Cui, J., Ma, B., 2022b. A method for obtaining the number of maize seedlings based on the improved YOLOv4 lightweight neural network. *Agriculture* 12, 1679.
- Gao, M., Yang, F., Wei, H., Liu, X., 2023a. Automatic monitoring of maize seedling growth using unmanned aerial vehicle-based RGB imagery. *Remote Sens.* 15, 3671.
- Gao, X., Zan, X., Yang, S., Zhang, R., Chen, S., Zhang, X., Liu, Z., Ma, Y., Zhao, Y., Li, S., 2023b. Maize seedling information extraction from UAV images based on semi-automatic sample generation and mask R-CNN model. *Eur. J. Agron.* 147, 126845.
- Jafari, F., Wang, B., Wang, H., Zou, J., 2024. Breeding maize of ideal plant architecture for high-density planting tolerance through modulating shade avoidance response and beyond. *J. Integr. Plant Biol.* 66, 849–864.
- Jegham, N., Koh, C.Y., Abdelatti, M., Hendawi, A., 2024. Yolo evolution: a comprehensive benchmark and architectural review of yolov12, yolo11, and their previous versions. *arXiv preprint arXiv:2411.00201*.
- Jia, Z., Zhang, X., Yang, H., Lu, Y., Liu, J., Yu, X., Feng, D., Gao, K., Xue, J., Ming, B., 2024. Comparison and optimal method of detecting the number of maize seedlings based on deep learning. *Drones* 8, 175.
- Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B., 2022. A review of YOLO algorithm developments. *Procedia Comput. Sci.* 199, 1066–1073.
- Jocher, G., Qiu, J., Chaurasia, A., 2023. Ultralytics YOLO (Version 8.0.0). Available at: <https://github.com/ultralytics/ultralytics> accessed 25 September 2025.
- Khanam, R., Hussain, M., 2024. YOLOv11: an overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*.
- Koh, J.C., Hayden, M., Daetwyler, H., Kant, S., 2019. Estimation of crop plant density at early mixed growth stages using UAV imagery. *Plant Methods* 15, 1–9.
- LeCun, Y., Denker, J.S., Solla, S.A., 1990. Optimal brain damage. In: Touretzky, D. (Ed.), *Advances in Neural Information Processing Systems*. Morgan Kaufmann, Denver, CO, pp. 598–605.
- Lee, J., Park, S., Mo, S., Ahn, S., Shin, J., 2020. Layer-adaptive sparsity for the magnitude-based pruning. *arXiv preprint arXiv:2010.07611*.
- Li, R., Wu, Y., 2022. Improved YOLO v5 wheat ear detection algorithm based on attention mechanism. *Electronics* 11, 1673.
- Liu, S., Mo, X., Lin, Z., Xu, Y., Ji, J., Wen, G., Richey, J., 2010. Crop yield responses to climate change in the Huang-Huai-Hai plain of China. *Agric. Water Manag.* 97, 1195–1209.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016. SSD: Single shot multibox detector. *Proc. Eur. Conf. Comput. Vis.* Springer, Cham, pp. 21–37.
- Liu, S., Qi, L., Qin, H., Shi, J., Jia, J., 2018. Path aggregation network for instance segmentation. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* pp. 8759–8768.
- Liu, S., Yin, D., Feng, H., Li, Z., Xu, X., Shi, L., Jin, X., 2022. Estimating maize seedling number with UAV RGB images and advanced image processing methods. *Precis. Agric.* 23, 1604–1632.
- Liu, M.G., Su, W.H., Wang, X.Q., 2023. Quantitative evaluation of maize emergence using UAV imagery and deep learning. *Remote Sens.* 15 (8), 1979.
- Menghani, G., 2023. Efficient deep learning: a survey on making deep learning models smaller, faster, and better. *ACM Comput. Surv.* 55 (12), 1–37.
- Mo, H.H., Wei, L.J., 2024. Fine segmentation of Chinese character strokes based on coordinate awareness and enhanced BiFPN. *Sensors* 24 (11), 3480.
- Molchanov, P., Tyree, S., Karras, T., Aila, T., Kautz, J., 2016. Pruning convolutional neural networks for resource efficient inference. *arXiv preprint arXiv:1611.06440*.
- Ning, S., Tan, F., Chen, X., Li, X., Shi, H., Qiu, J., 2024. Lightweight corn leaf detection and counting using improved YOLOv8. *Sensors* 24 (16), 5279.
- Padilla, R., Netto, S.L., Da Silva, E.A., 2020. A survey on performance metrics for object-detection algorithms. *Proc. Int. Conf. Syst. Signals Image Process. (IWSSIP)*. IEEE, Niteroi, Brazil, pp. 237–242.
- Quan, L., Feng, H., Lv, Y., Wang, Q., Zhang, C., Liu, J., Yuan, Z., 2019. Maize seedling detection under different growth stages and complex field environments based on an improved Faster R-CNN. *Biosyst. Eng.* 184, 1–23.
- Ray, D.K., Ramankutty, N., Mueller, N.D., West, P.C., Foley, J.A., 2012. Recent patterns of crop yield growth and stagnation. *Nat. Commun.* 3, 1293.
- Rehman, T.U., Mahmud, M.S., Chang, Y.K., Jin, J., Shin, J., 2019. Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. *Comput. Electron. Agric.* 156, 585–605.
- Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149.
- Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proc. IEEE Int. Conf. Comput. Vis. IEEE, Venice, Italy*, pp. 618–626.
- Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. *J. Big Data* 6, 1–48.
- Song, C.Y., Zhang, F., Li, J.S., Xie, J.Y., Chen, Y., Hang, Z., Zhang, J.X., 2023. Detection of maize tassels for UAV remote sensing image with an improved YOLOX model. *J. Integr. Agric.* 22, 1671–1683.
- Tan, M., Pang, R., Le, Q.V., 2020. EfficientDet: scalable and efficient object detection. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. IEEE, Seattle, WA*, pp. 10781–10790.
- Su, Z., Fang, L., Kang, W., Hu, D., Pietikäinen, M., Liu, L., 2020. August). Dynamic group convolution for accelerating convolutional neural networks. *European conference on computer vision*. Springer International Publishing, Cham, pp. 138–155.
- Tan, C., Li, C., He, D., Song, H., 2022. Towards real-time tracking and counting of seedlings with a one-stage detector and optical flow. *Comput. Electron. Agric.* 193, 106683.
- Tassinari, P., Bovo, M., Benni, S., Franzoni, S., Poggi, M., Mammi, L.M.E., Mattoccia, S., Di Stefano, L., Bonora, F., Barbaresi, A., 2021. A computer vision approach based on deep learning for the detection of dairy cows in free stall barn. *Comput. Electron. Agric.* 182, 106030.
- Tzutalin, 2015. LabelImg. GitHub Repository. <https://github.com/tzutalin/labelimg> accessed 25 September 2025.
- Varghese, R., Sambath, M., 2024. YOLOv8: A novel object detection algorithm with enhanced performance and robustness. *Proc. Int. Conf. Adv. Data Eng. Intell. Comput. Syst. (ADICS)*. IEEE, Chennai, India, pp. 1–6.
- Wang, C.H., Huang, K.Y., Yao, Y., Chen, J.C., Shuai, H.H., Cheng, W.H., 2022. Lightweight deep learning: an overview. *IEEE Consum. Electron. Mag.* 13 (4), 51–64.
- Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M., 2023. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. IEEE, New Orleans, LA*, pp. 7464–7475.
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., Ding, G., 2024a. YOLOv10: real-time end-to-end object detection. *Adv. Neural Inf. Proces. Syst.* 37, 107984–108011.
- Wang, C.Y., Yeh, I.H., Liao, H.Y.M., 2024b. YOLOv9: Learning what you want to learn using programmable gradient information. *Proc. Eur. Conf. Comput. Vis.* Springer, Cham, pp. 1–21.
- Xia, Z., Pan, X., Song, S., Li, L.E., Huang, G., 2022. Vision transformer with deformable attention. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. IEEE, New Orleans, LA*, pp. 4794–4803.
- Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. IEEE, Honolulu, HI*, pp. 1492–1500.
- Xie, X., Ge, Y., Walia, H., Yang, J., Yu, H., 2023. Leaf-counting in monocot plants using deep regression models. *Sensors* 23 (4), 1890.
- Xu, X., Wang, L., Shu, M., Liang, X., Ghafoor, A.Z., Liu, Y., Ma, Y., Zhu, J., 2022. Detection and counting of maize leaves based on two-stage deep learning with UAV-based RGB image. *Remote Sens.* 14 (21), 5388.
- Xu, X., Wang, L., Liang, X., Zhou, L., Chen, Y., Feng, P., Yu, H., Ma, Y., 2023. Maize seedling leaf counting based on semi-supervised learning and UAV RGB images. *Sustainability* 15 (12), 9583.
- Yu, X., Yin, D., Xu, H., Pinto Espinosa, F., Schmidhalter, U., Nie, C., Bai, Y., Sankaran, S., Ming, B., Cui, N., Wu, W., Jin, X., 2024. Maize tassel number and tasseling stage monitoring based on near-ground and UAV RGB images by improved YOLOv8. *Precis. Agric.* 25 (4), 1800–1838.
- Yu, X., Jiang, T., Zhu, Y., Li, L., Fan, F., Jin, X., 2025. FEL-YOLOv8: a new algorithm for accurate monitoring soybean seedling emergence rates and growth uniformity. *IEEE Trans. Geosci. Remote Sens.* 63, 1–17.
- Yuan, J., Li, X., Zhou, M., Zheng, H., Liu, Z., Liu, Y., Wen, M., Cheng, T., Cao, W., Zhu, Y., 2024. Rapidly count crop seedling emergence based on waveform method (WM) using drone imagery at the early stage. *Comput. Electron. Agric.* 220, 108867.
- Zhang, X., Zhang, Y.A., Gao, T., Fang, Y., Chen, T., 2023. A novel SSD-based detection algorithm suitable for small object. *IEICE Trans. Inf. Syst.* E106D, 625–634.
- Zhang, Y., Jiang, Y., Xu, B., Yang, G., Feng, H., Yang, X., Yang, H., Liu, C., Cheng, Z., Feng, Z., 2024. Study on the estimation of leaf area index in rice based on UAV RGB and multispectral data. *Remote Sens.* 16, 3049.
- Zhou, Y., Yang, Y., Wang, D., Zhai, Y., Li, H., Xu, Y., 2024. Innovative ghost channel spatial attention network with adaptive activation for efficient rice disease identification. *Agronomy* 14, 2869.
- Zhu, X., Chen, F., Zheng, Y., Chen, C., Peng, X., 2024. Detection of Camellia oleifera fruit maturity in orchards based on modified lightweight YOLO. *Comput. Electron. Agric.* 226, 109471.