

Article

Comparison and Optimal Method of Detecting the Number of Maize Seedlings Based on Deep Learning

Zhijie Jia ¹, Xinlong Zhang ², Hongye Yang ³, Yuan Lu ⁴, Jiale Liu ⁴, Xun Yu ³, Dayun Feng ³, Kexin Gao ³, Jianfu Xue ^{1,*}, Bo Ming ^{3,*}, Chenwei Nie ⁵ and Shaokun Li ³

¹ College of Agriculture, Shanxi Agricultural University, Taigu, Jinzhong 030801, China; z20213543@stu.sxau.edu.cn

² Zhangjiakou Academy of Agricultural Sciences, Zhangjiakou 075000, China; cropzxl@outlook.com

³ Institute of Crop Sciences, Chinese Academy of Agricultural Sciences/State Key Laboratory of Crop Gene Resources and Breeding, Beijing 100081, China; 82101231082@caas.cn (H.Y.); 82101211308@caas.cn (X.Y.); Feng82101222192@caas.cn (D.F.); Gao82101225018@caas.cn (K.G.); lishaokun@caas.cn (S.L.)

⁴ College of Agriculture, Qingdao Agricultural University, Qingdao 266000, China; ly20201000037@stu.qau.edu.cn (Y.L.)

⁵ Surveying and Municipal Engineering, Zhejiang Institute of Water Resources and Hydropower, Hangzhou 310000, China; niechw@zjweu.edu.cn

* Correspondence: Xuefudange95@sxau.edu.cn (J.X.); mingbo@caas.cn (B.M.)

Abstract: Effective agricultural management in maize production operations starts with the early quantification of seedlings. Accurately determining plant presence allows growers to optimize planting density, allocate resources, and detect potential growth issues early on. This study provides a comprehensive analysis of the performance of various object detection models in maize production, with a focus on the effects of planting density, growth stages, and flight altitudes. The findings of this study demonstrate that one-stage models, particularly YOLOv8n and YOLOv5n, demonstrated superior performance with AP₅₀ scores of 0.976 and 0.951, respectively, outperforming two-stage models in terms of resource efficiency and seedling quantification accuracy. YOLOv8n, along with Deformable DETR, Faster R-CNN, and YOLOv3-tiny, were identified for further examination based on their performance metrics and architectural features. The study also highlights the significant impact of plant density and growth stage on detection accuracy. Increased planting density and advanced growth stages (particularly V6) were associated with decreased model accuracy due to increased leaf overlap and image complexity. The V2–V3 growth stages were identified as the optimal periods for detection. Additionally, flight altitude negatively affected image resolution and detection accuracy, with higher altitudes leading to poorer performance. In field applications, YOLOv8n proved highly effective, maintaining robust performance across different agricultural settings and consistently achieving rRMSEs below 1.64% in high-yield fields. The model also demonstrated high reliability, with Recall, Precision, and F1 scores exceeding 99.00%, affirming its suitability for practical agricultural use. These findings suggest that UAV-based image collection systems employing models like YOLOv8n can significantly enhance the accuracy and efficiency of seedling detection in maize production. The research elucidates the critical factors that impact the accuracy of deep learning detection models in the context of corn seedling detection and selects a model suited for this specific task in practical agricultural production. These findings offer valuable insights into the application of object detection technology and lay a foundation for the future development of precision agriculture, particularly in optimizing deep learning models for varying environmental conditions that affect corn seedling detection.



Citation: Jia, Z.; Zhang, X.; Yang, H.; Lu, Y.; Liu, J.; Yu, X.; Feng, D.; Gao, K.; Xue, J.; Ming, B.; et al. Comparison and Optimal Method of Detecting the Number of Maize Seedlings Based on Deep Learning. *Drones* **2024**, *8*, 175. <https://doi.org/10.3390/drones8050175>

Academic Editors: Fei Liu and Yangquan Chen

Received: 13 March 2024

Revised: 22 April 2024

Accepted: 26 April 2024

Published: 28 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: unmanned aerial vehicle; maize seedling; object detection; counting

1. Introduction

Maize is among the world's most widely cultivated and traded crops, serving various purposes including human consumption, animal feed, and the production of adhesives [1].

Maize yield is significantly influenced by factors like emergence rate and planting density, necessitating growers to carefully observe their crops [2]. Inspections during the early phases of maize cultivation allow growers to identify and reseed areas with no germination. Therefore, the rapid detection and quantification of maize seedlings are key prerequisites for ensuring a maximal yield. Current traditional methods of seedling detection rely on manual visual assessments of selected plots. As global maize production shifts towards large-scale operations, manual surveys are becoming increasingly time-intensive. This method is also prone to human error, resulting in insufficient or inaccurate planting information [3]. Alternatively, advancements in drone technology have enabled the rapid and accurate collection of data on large-scale plantations. This information provides support for intelligent decision-making regarding field management strategies. Additionally, precision agriculture significantly increases efficiency while reducing time and labor costs [4].

Unmanned aerial vehicles (UAVs) are revered for their affordability, portability, and flexibility. They can be equipped with a diverse array of sensors, such as RGB, multispectral, hyperspectral, and LiDAR, to ensure robust data capture [5]. Initially, images captured by UAVs required image processing techniques such as skeletonization algorithms and multiple despeckling processes to extrapolate pertinent information, including seedling counts [6]. While these approaches were effective, they necessitated intricate processing workflows and high-quality images. As the technology behind image processing and analysis algorithms has evolved, the integration of computer vision algorithms into UAV image analysis has significantly improved the efficiency and precision of crop counting. Peak detection algorithms have especially improved the localization and enumeration of crop rows and seedlings in high-resolution images [7]. Moreover, corner detection techniques have enabled the effective counting of overlapping leaves, which tend to complicate data, especially as plants mature [8].

UAV data extrapolation still relies on traditional image processing techniques, which face many challenges such as complex target feature design, poor portability, slow operation speed, and cumbersome manual design [9]. The ongoing development of deep learning is continuously broadening the scope of agricultural applications [10]. Currently, object detection technology presents one of the most practical methods of identifying plants in various background environments. A variety of deep learning models have recently been developed to enhance the accuracy and efficiency of crop detection. For instance, the Faster R-CNN model has been incorporated into field robot platforms, enabling them to accurately identify corn seedlings at different growth stages and distinguish them from weeds [11]. Additionally, the model is able to automatically recognize and record different developmental stages of rice panicles, a previously labor-intensive manual process [12]. The multiple complex processing stages required render R-CNN models relatively slow, limiting their application potential in large-scale operations. Based on the improvements of YOLOv4, Gao et al. proposed a lightweight model for seedling detection with an enhanced feature extraction network, a novel attention mechanism, and a k-means clustering algorithm. Results show that the harmonic mean, recall rate, average precision, and accuracy rate of the model on all test sets are 0.95%, 94.02%, 97.03%, and 96.25%, respectively [13]. Zhang et al. further improved the efficacy and speed of maize male cob detection by optimizing the feature extraction network and introducing a multi-head attention mechanism, achieving an accuracy of 95.51% [14]. Later, Li et al. released an enhanced YOLOv5, which included downsampling to improve the detection of small targets and introduced a CBAM attention mechanism to eliminate gradient vanishing. The experimental results show that the improved algorithm has an mAP of 94.3%, an accuracy of 88.5%, and a recall of 98.1% [15]. Finally, a wheat head detection model was proposed by Zhang et al. based on a one-level network structure, which improves accuracy and generalizability by incorporating an attention module, a feature fusion module, and an optimized loss function. When compared to various other mainstream object detection networks, this model outperforms them, with a mAP of 0.688 [16].

Although previous research has focused on optimizing algorithms and achieving excellent accuracy, few have delved into the limitations of models and the factors affecting them. This study primarily investigates key factors such as varying planting densities, different growth stages, and multiple flight altitudes to comprehensively assess model performance. We screened nine object detection models and selected four—YOLOv8n, Deformable DETR, Faster R-CNN, and YOLOv3-tiny—based on their performance metrics and architectural differences between one-stage and two-stage detection frameworks. After conducting a thorough evaluation of these models under varied conditions, based on these validation results, field validations will be conducted in farmers' fields to further confirm the practical applicability of the model. The main contributions of this paper are:

- (1) Establishment of a comprehensive corn seedling dataset at different corn planting densities, growth stages, and flight altitudes.
- (2) Application of various object detection models to different datasets for validation and comparison, assessing their performance under diverse conditions.
- (3) Validation of the model's detection effectiveness at different planting densities, growth stages, and flight altitudes further identifies the most suitable growth stages for detecting corn seedlings while also revealing the limitations of model detection.
- (4) Field validation of the model in actual agricultural production environments confirms its effectiveness and feasibility in practical applications.

2. Materials and Methods

2.1. Field Experiments

Field experiments were conducted during 2021 and 2023 in Tongliao City (43°42' N, 122°25' E) and Liaohe Town (43°43' N, 122°10' E) in Inner Mongolia (Figure 1). This region features a semi-arid continental monsoon climate with 2500–2800 h of annual sunshine, an average daily temperature of 21.0 °C, a cumulative ≥ 10 °C temperature of 3000–3300 °C·d, a frost-free period of 150–169 d, and an average annual precipitation of 280–390 mm during the maize growing season (11 May). Both fields consisted of sandy loam soil and had previously been used for maize cultivation. A wide-narrow planting pattern was implemented, with alternating rows spaced at 80 cm and 40 cm. Irrigation was supplied through shallow, buried drip lines at a rate of 300 m³/ha. Base fertilizer with an N, P, and K ratio of 13:22:15 was applied at a rate of 525 kg/ha through water-fertilizer integration methods.

The maize variety Xianyu 335 was selected for the density experiment conducted in Qianxiaili Village. Trials were planted on 10 May 2021 at densities of 30,000, 45,000, 60,000, 75,000, 90,000, 105,000, 120,000, and 135,000 plants/ha. Data were collected on 29 May (2 leaves unfolded), 1 June (3 leaves unfolded), 5 June (4 leaves unfolded), and 16 June (6 leaves unfolded).

The maize variety Dika 159 was selected for the flight altitude experiment conducted in Dongsheng Village. Trials were planted on 8 May 2023 at a density of 90,000 plants/ha. Data were collected on 27 May (2 leaves unfolded), 31 May (3 leaves unfolded), 4 June (4 leaves unfolded), 7 June (5 leaves unfolded), and 11 June (6 leaves unfolded).

In 2021 and 2023, we carried out validation trials on the agricultural lands of local farmers. These trials included three distinct types of cultivation areas: high-yielding fields in 2021, agricultural cooperative plots in 2023, and peasant household farmland in 2021. Maize varieties Jingke 968, Tianyu 108, and Dika 159 were planted in each plot at densities of 100,000, 80,000, and 65,000 plants/ha. UAV visible light images were collected at noon during the 3-leaf stage from 8 sample areas (5 m × 2.4 m) within each field. Additional images were collected from 20 randomly selected sample areas (11.66 m²) in each field, which were monitored at noon on 1 June 2021 (3 leaves unfolded) and 31 May 2023 (3 leaves unfolded).

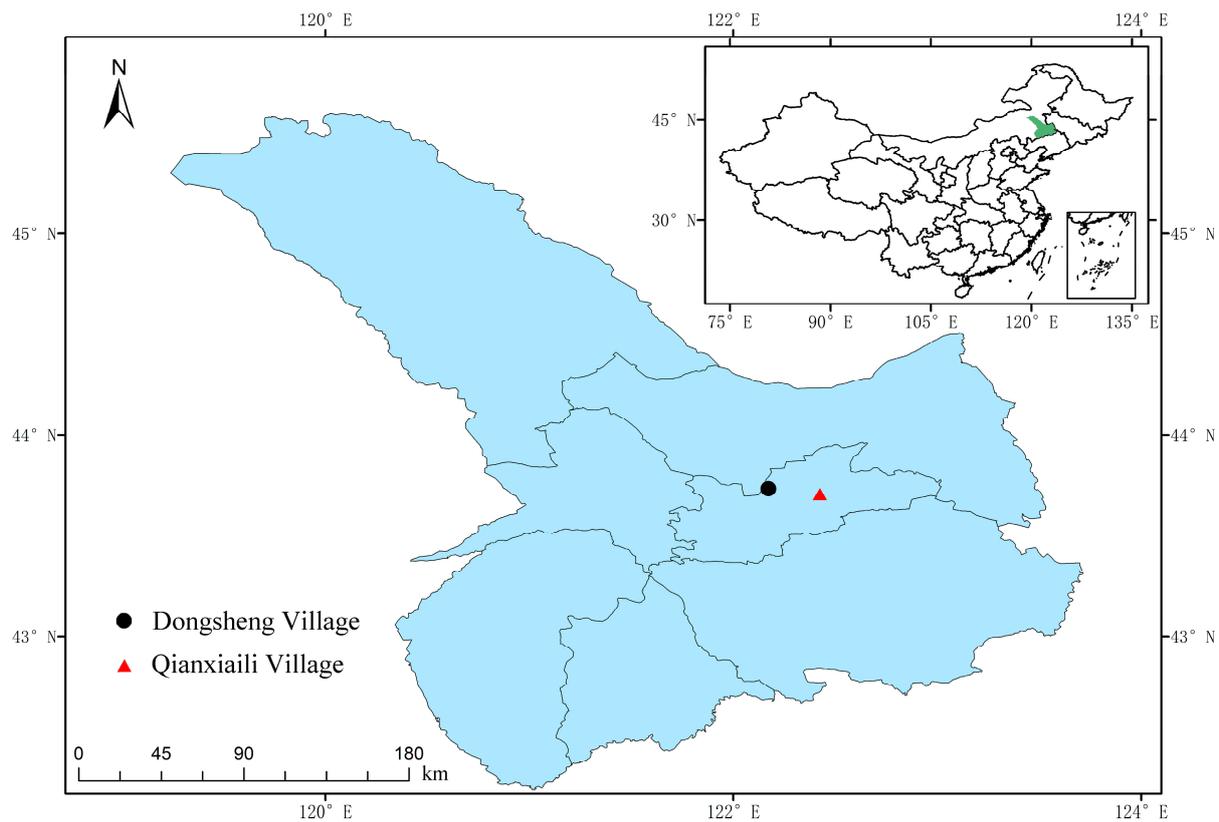


Figure 1. Experimental area.

2.2. UAV Image Collection

High-resolution images of maize seedlings were captured with a UAV-based RGB camera mounted perpendicular to the ground onto a DJI M600 drone with a Ronin-MX gimbal. GPS and barometers were used to control horizontal position and altitude within approximately 2 m and 0.5 m, respectively. Drone images were collected every 3 days between 10 a.m. and 2 p.m. for the duration of the experiment. Detailed image collection information is listed in Table 1.

Table 1. UAV image collection parameters.

Year	2021			2023	
Station	Qianxiaili Village	High-Yielding Field	Peasant Household	Dongsheng Village	Agricultural Cooperative Plots
Image acquisition stage (leaves)	2, 3, 4, 6	3	3	2–6	3
Flight speed (m/s)	2.1	2.3	2.0	2.0	2.5
Photo interval (s)	1	2	2	2	2
Height above ground (m)	20	20	20	20, 40, 60	20
Overlap rate along tracks (%)	75	73	75	80	75
Overlap rate across tracks (%)	85	75	80	80	75

Images were collected with a Sony $\alpha 7$ II camera with a 35 mm sensor and a maximum resolution of 6000×4000 pixels. Shutter speed was prioritized, and the ISO was set to automatically adjust (1600 maximum value). RGB images were captured at a frequency of 1 Hz with an intervalometer-controlled camera.

2.3. Data Construction and Preprocessing

2.3.1. Image Preprocessing

After RGB images were exported from the UAV, Agisoft Metashape Professional software was used for image stitching. Feature points in each image were initially automatically calculated and then matched in the image sequence through multiple iterations. Next, dense point clouds were generated before the final images were produced (Figure 2a).

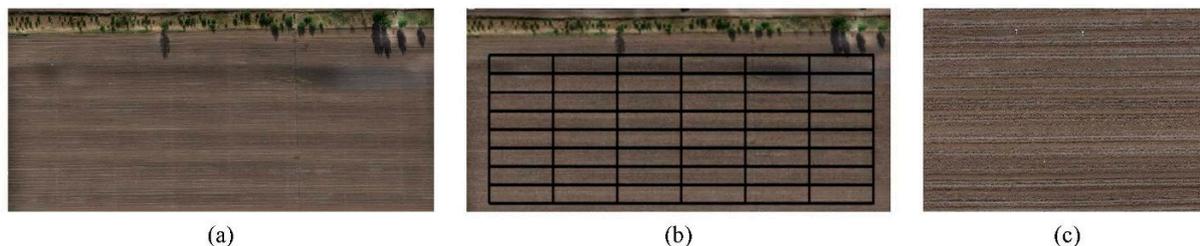


Figure 2. Orthophotos of (a) the test area, (b) experimental plot layout, and (c) plot cropping.

Experimental fields were cropped and divided into multiple plots (Figure 2b). Original high-quality maize seedling images were cropped to 1000×1000 pixels by a sliding step. Poor-quality images, including those with large shooting angles, obvious occlusions, and uneven illumination, were removed. Final images (1000 total) were categorized by quantity of spreading leaves and quantity of straw (Figure 3). The location and size of each seedling are labeled using Labellmg in each image. The location and size of each seedling are labeled using Labellmg in each image. The dataset was then divided, with 90% of images used for model training and 10% for validating model performance.

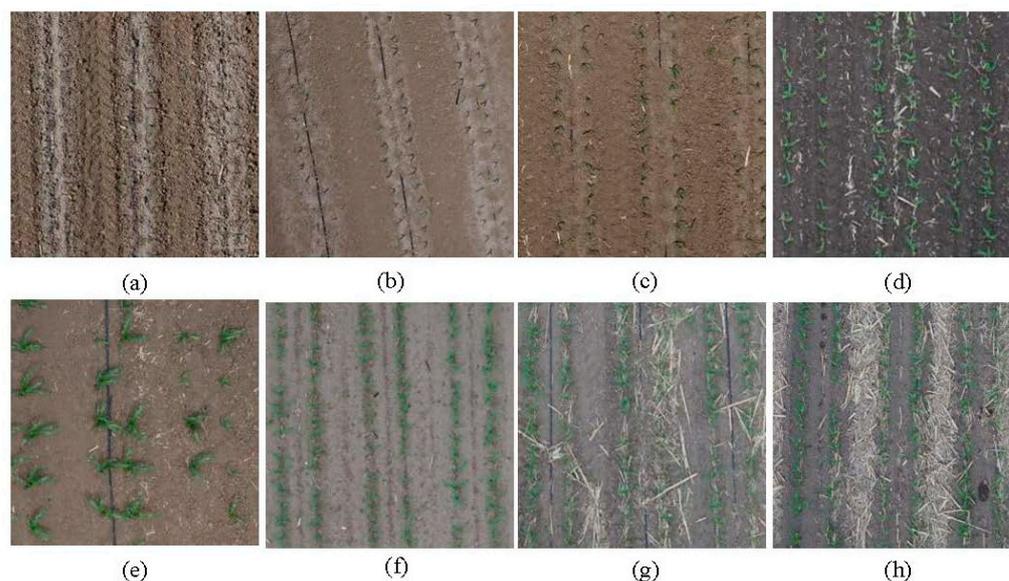


Figure 3. Maize seedlings at (a) V2, (b) V3, (c) V4, (d) V5, (e) and V6 stages. (f) Low, (g) moderate, and (h) high quantities of stover. Note: V2–V6 correspond to the 2-leaf stage (unfurled leaves), 3-leaf stage (unfurled leaves), 4-leaf stage (unfurled leaves), 5-leaf stage (unfurled leaves), and 6-leaf stage (unfurled leaves), respectively.

2.3.2. Data augmentation

Data augmentation is crucial for expanding the quantity and diversity of samples and enhancing the training model's ability to generalize across different conditions. This process not only increases the robustness of the model by introducing a wider range of scenarios but also helps prevent overfitting by simulating real-world variations. Images were adjusted during the data augmentation process, including horizontal and vertical

flipping, random contrast and hue adjustments, and resolution alterations (Figure 4). These modifications were used to simulate the effects of varying lighting and environmental factors during different times of day and flight altitudes. The final training dataset was expanded to a total of 16,200 images.

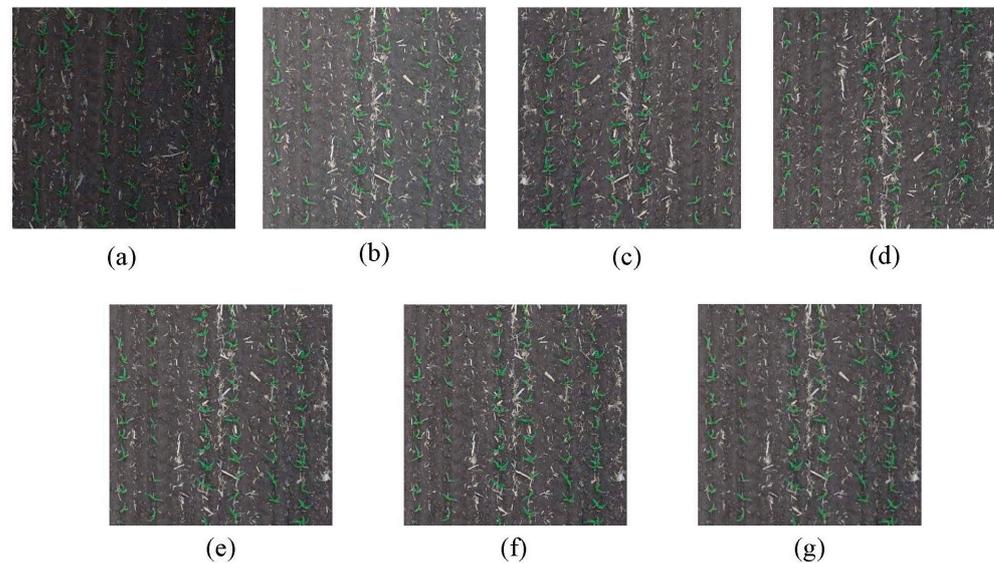


Figure 4. (a) Random contrast, (b) random hue adjustment, (c) horizontal flip, and (d) vertical flip. Images were modified to resolutions of (e) 800×800 , (f) 500×500 , and (g) 300×300 .

2.4. Overview of Object Detection Models

This study encompasses a range of advanced object detection models, categorized into one-stage and two-stage models, to evaluate their performance across various types of complex image datasets. The selection of these models was based on their processing speed, accuracy, and capability to handle different types and complexities of objects and scenes. One-stage object detection models such as YOLOv8n, YOLOv5n, YOLOv3-tiny, SSD (Single-Shot MultiBox Detector), FCOS (Fully Convolutional One-Stage Object Detection), and RetinaNet predict bounding boxes and class probabilities directly at the network output layer, offering rapid detection speeds that are particularly suitable for real-time processing requirements. In contrast, two-stage object detection models, including Faster R-CNN, Cascade R-CNN, and Deformable DETR (Deformable End-to-End Transformer-based Object Detection), first generate potential object candidates, then refine these through more precise classification and bounding box regression to enhance detection accuracy.

2.4.1. One-Stage Object Detection Models

The YOLO series is renowned for its rapid detection speed and effective performance balance. From the lightweight design of YOLOv3-tiny [17], optimized for resource-constrained devices, to YOLOv5n, which features various sizes and training optimizations, and the latest YOLOv8n, each version strives to find a better balance between detection speed and accuracy. YOLOv8n further improves detection precision and speed through an enhanced network architecture and more efficient training strategies. SSD performs object prediction at multiple scales simultaneously, effectively capturing objects of varying sizes [18]. FCOS, as an innovative anchor-free detection model, eliminates the complexity of anchor choice, simplifies the training process, and enhances the model's versatility and flexibility [19]. RetinaNet addresses the issue of class imbalance with focal loss, significantly improving performance in complex environments [20].

2.4.2. Two-Stage Object Detection Models

Faster R-CNN, through its innovative Region Proposal Network (RPN), efficiently generates high-quality candidate areas, significantly enhancing subsequent detection accuracy [21]. Cascade R-CNN, with its unique multi-stage refinement strategy, effectively improves recognition of occluded and small targets [22]. Deformable DETR utilizes a deformable attention mechanism to optimize the handling of dynamic scenes and complex backgrounds, resulting in superior performance on large-scale image datasets [23].

2.5. Assessment of Indicators

The maize seedling detection and quantification abilities of each model were evaluated by calculating the Precision (Equation (1)), Recall (Equation (2)), F1-Score (Equation (3)), AP (Equation (4)), and rRMSE (root mean square error) (Equation (5)) values according to the following formulas:

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

$$\text{F1-Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (3)$$

$$\text{AP} = \int_0^1 P(R) dR \quad (4)$$

$$\text{rRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sum_{i=1}^n y_i} \times 100\% \quad (5)$$

True positive (TP) and false positive (FP) represent the number of correctly and incorrectly detected maize seedlings, respectively, while false negative (FN) indicates the number of those missed. F1 represents the harmonic mean of Precision and Recall. Average Precision (AP) measures the average precision of the classification model at all Recall levels and n denotes the number of samples. y_i represents the actual value of each sample point, \hat{y}_i denotes the predicted value of each sample point according to the regression model, and \bar{y} is the mean of the actual observed values of the dependent variable.

2.6. Test Parameter Setting and Training Process Analysis

The computer specifications and software environments used are described in Table 2. The training parameters were tailored to the characteristics of the task dataset. The training settings were tested with a batch size of 8, an image size of 640, a confidence threshold (conf_thres) of 0.3, and an intersection over union threshold (iou_thres) of 0.2.

Table 2. Model training specifications.

Experimental Environment			
Processor	12th Gen Intel(R) Core(TM) i5-12600KF3.69 GHz		
Operating system	Windows 10		
Ram	64 GB		
Graphics card	NVIDIA GeForce RTX 3060		
Programming language	Python 3.8		
Model	YOLOv8n	YOLOv5n	Other
Deep learning libraries	CUDA11.7	CUDA11.1	CUDA 10.2
Software	Ultralytics = 8.0.105 Opencv = 4.7.0.72	Opencv = 4.1.2 Numpy = 1.18.5	Mmcv = 2.0.0 Mmdet = 3.0.0 Mmengine = 0.9.1

3. Results

3.1. Model Comparison

To further validate the performance of YOLOv8n, multiple one-stage and two-stage object detection models were trained and evaluated based on metrics such as AP_{50} , $AP_{50:95}$, params, and FLOPs (Table 3, Figure 5).

Table 3. Comparison of object detection model performances.

Category	Model	Backbone	Image Size	AP_{50}	$AP_{50:95}$	Params	FLOPs
One-stage	YOLOv8n	New CSP-Darknet53	640×640	0.976	0.643	3.20 M	8.7 G
	YOLOv5n	CSP-Darknet53	640×640	0.951	0.510	1.90 M	4.5 G
	SSD	VGG16	416×416	0.942	0.526	23.75 M	137.1 G
	FCOS	Resnet50	640×640	0.922	0.493	31.84 M	78.6 G
	YOLOv3-tiny	Tiny-Darknet	640×640	0.919	0.452	8.44 M	13.3 G
	RetinaNet	Resnet50	300×300	0.861	0.411	36.10 M	81.7 G
Two-stage	Deformable DETR	Resnet50	640×640	0.939	0.471	36.10 M	27.4 G
	Cascade R-CNN	Resnet50	640×640	0.884	0.565	68.94 M	80.1 G
	Faster R-CNN	Resnet50	640×640	0.882	0.529	41.12 M	78.1 G

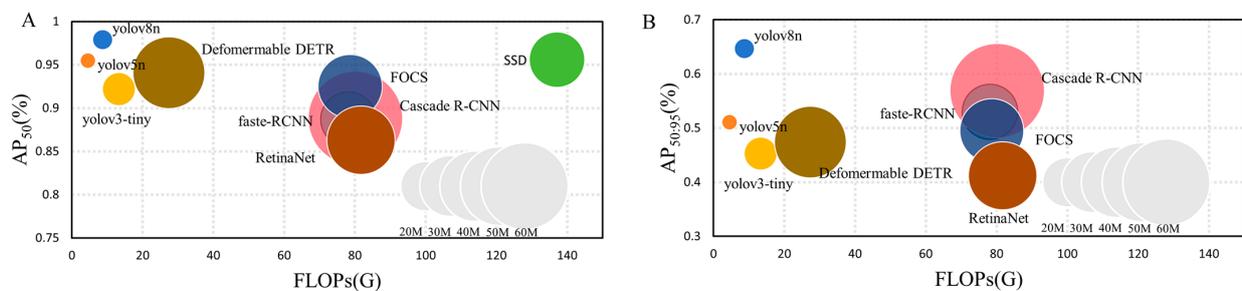


Figure 5. Performance comparison of different models. (A) Relationship between AP_{50} , Params, and FLOPs. (B) Relationship between $AP_{50:95}$, Params, and FLOPs.

The performances of YOLOv8n and YOLOv5n stand out among the single-stage models, achieving AP_{50} values of 0.976 and 0.951, respectively, at an input image size of 640×640 . Although SSD and FCOS also performed highly, their increased number of parameters and computational requirements under the same conditions render them less suitable for resource-constrained scenarios. YOLOv3-tiny and RetinaNet demonstrated slightly reduced performances and are better suited to environments with limited resources.

Deformable DETR showed the highest performance out of the two-stage models, achieving an AP_{50} value of 0.939 at an input image size of 640×640 . Moreover, the model has a reduced number of parameters and computational requirements, exhibiting an optimal balance between performance and efficiency. Comparatively, Faster R-CNN and Cascade R-CNN perform similarly at the same size but have more requirements, making them less ideal for resource-limited situations.

Taking into account both performance and architectural differences (between one-stage and two-stage detection frameworks), we have selected the top-performing YOLOv8n and Deformable DETR, as well as the lower-performing Faster R-CNN and YOLOv3-tiny, for further in-depth study.

3.2. Impact of Planting Density and Growth Stage on Seedling Detection

Planting density and growth stage were found to significantly affect the estimation accuracy of maize seedling detection models. In this study, YOLOv8n, YOLOv3-tiny, Deformable DETR, and Faster R-CNN were analyzed for key metrics such as accuracy rate, miss rate, false detection rate, and rRMSE. Experimental validations were conducted across four growth stages (V2, V3, V4, and V6) at eight different planting densities (30,000, 45,000,

60,000, 75,000, 90,000, 105,000, 120,000, and 135,000 plants/ha). For each density, 20 images were selected, resulting in a total of 640 images for inference.

Our findings demonstrate that as density increases, overall detection accuracy measured by the F1-score, rRMSE, Recall, and Precision declines (Figure 6). Moreover, our analysis of V2–V6 growth stages revealed a trend of increasing and then decreasing detection performance. Performance seemed to improve up until the V3 stage but declined as time progressed to the V6 stage.

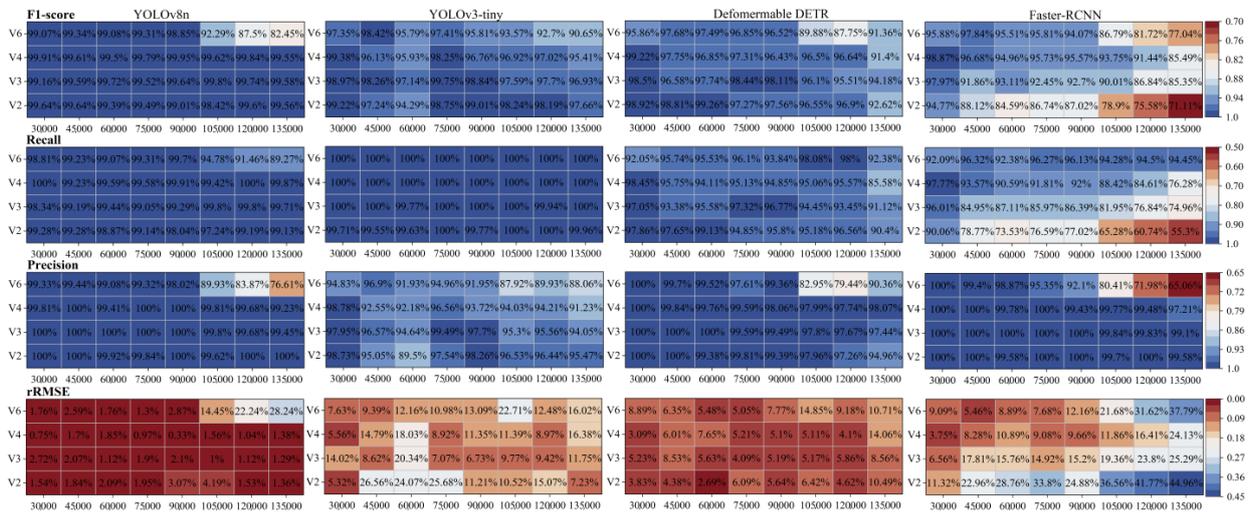


Figure 6. Performances of YOLOv8n, YOLOv3-tiny, Deformable-DETR, and Faster R-CNN in terms of F1-score, Recall, Precision, and rRMSE across various planting densities and growth stages.

Our study showed that as planting density increases, YOLOv8n exhibits a relatively stable rRMSE performance compared to YOLOv3-tiny, especially at higher densities (from 105,000 to 135,000). The Deformable-DETR model exhibited relatively steady performance across different densities, with only minor fluctuations. In contrast, Faster R-CNN performed poorly at high densities, with a significantly increased rRMSE. Taken together, these results demonstrate the superior performance and stability of YOLOv8n across all densities. Additionally, model performance was significantly impacted by the plant growth stage (V2–V6). YOLOv8n and Faster R-CNN achieved their highest performances at the V4 stage, while YOLOv3-tiny and Deformable-DETR peaked at the V3 stage. While the optimal growth stage differed between the models, all displayed a similar trend of declining performance with increasing density in terms of Recall and Precision. These findings highlight the performance variations between different models across various planting densities and growth stages, providing a foundation for model selection based on growth operation requirements.

3.3. Impact of Flight Altitude and Growth Stage on Detection

In this study, plant detection was conducted through UAV flights at various growth stages (V2, V3, V4, V5, and V6) and altitudes (20 m, 40 m, and 60 m). Metrics such as accuracy rate, miss rate, false detection rate, and rRMSE were calculated to explore potential impacts. Twenty images per altitude across five growth stages were collected, resulting in a total of 300 images for inference. Overall, changes in altitude were found to affect image resolution and coverage area.

The performance metrics of all four models decreased across all growth stages (V2–V6) as flight altitude increased. From Figure 7, it is evident that for YOLOv8n, at 20 m, the performance across various stages remains strong, with F1-scores ranging between 97.82 and 99.62% and rRMSE staying below 5.36%. At 40 m, V3 to V5 still maintain high detection performance, but V2 and V6 see a slight decline, with rRMSEs of 7.52% and 15.85%, respectively. For 60 m, V6 significantly drops, with the F1-score falling to

54.55% and rRMSE rising to 62.73%. YOLOv3-tiny shows a gradual decrease in detection effectiveness at 20 m as leaf age increases, with its F1-score dropping from 99.00% in V2 to 89.48% in V6, and rRMSE increasing from 2.47% in V2 to 17.33% in V6. At 40 m, the F1-score remains above 90% from V2 to V5, but drops to 85.96% in V6, with rRMSE also rising to approximately 22.23%. At 60 m, the F1-score significantly decreases to 35.19%, with the rRMSE increasing to 78.38%. Deformable DETR demonstrates good performance at a flight altitude of 20 m. At 40 m, the F1-score significantly decreases to between 55.60 and 67.35%, with similar trends observed in Recall and Precision and a notable increase in rRMSE, reaching 56.68% in V2. At 60 m, there is a further decline, with the F1-score dropping from 40.31% to 19.17%, especially notable in Recall, which falls from 25.72% to 10.37%. rRMSE further increases from 74.42% to 89.36%. For Faster R-CNN at 20 m, the F1-score and rRMSE show a trend of initial increase followed by a decrease, peaking in V3 at 95.59% and 8.6%, respectively. At 40 m, the model’s F1-score, Recall, and rRMSE significantly decrease, with V2 being the worst at 11.37%, 6.07%, and 94.15%. At 60 m, the decline in F1-score and Recall is even more pronounced, with V2 performing the worst at 0.96% and 0.48%, respectively. rRMSE remains exceedingly high, exceeding 91.3% across all leaf stages.

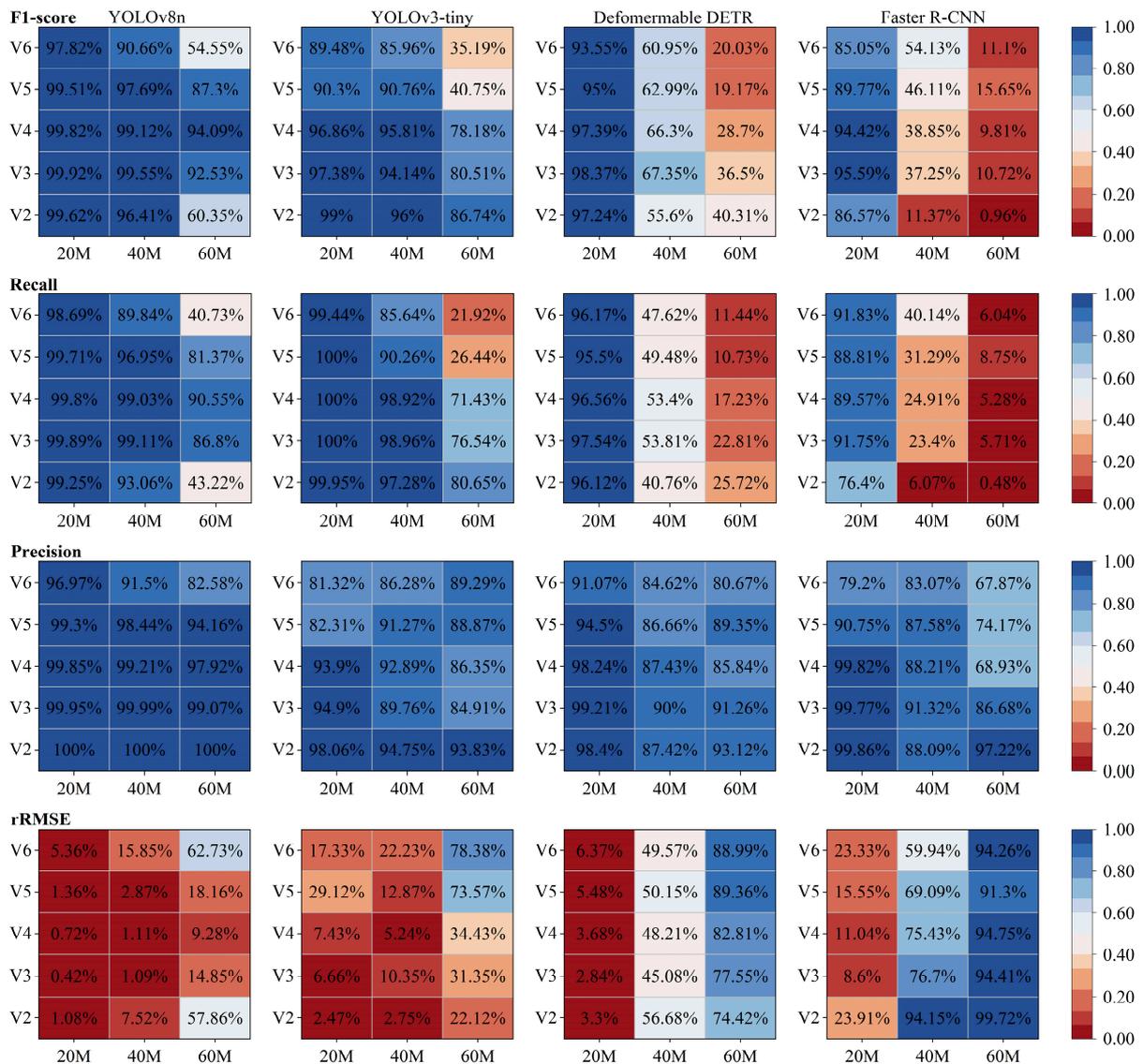


Figure 7. Performance of YOLOv8n, YOLOv3-tiny, Deformable-DETR, and Faster R-CNN in terms of F1-score, Recall, Precision, and rRMSE at different flight altitudes and growth stages.

3.4. Validation of the YOLOv8n Seedling Counting Algorithm

To validate the applicability and accuracy of YOLOv8n, the model's performance was comprehensively evaluated under various planting conditions in active growing operations across different years and locations. This evaluation process involved six total repetitions, each consisting of the acquisition of 20 images from one location at a flight altitude of 20 m, cumulatively amounting to a total of 120 images. The model exhibited outstanding performance across all planting methods, with notably significant results in high-yield fields characterized by high-density planting conditions.

Our analysis illustrates that predicted values are closely aligned with the actual values, maintaining a highly consistent, near 1:1 line (Figure 8). Exceptional performance was observed in high-yield fields 1 and 2, with rRMSE values of roughly 1.64% and 1.33%, respectively. The Recall, Precision, and F1 scores all exceeded 99%. Stellar performance was also observed under cooperative and individual farmer planting conditions; rRMSE values are all below 1.23%. Together, these results affirm the model's stability and accuracy. The consistency between the algorithm's predicted planting density and the actual planting density further validates the model's reliability and precision.

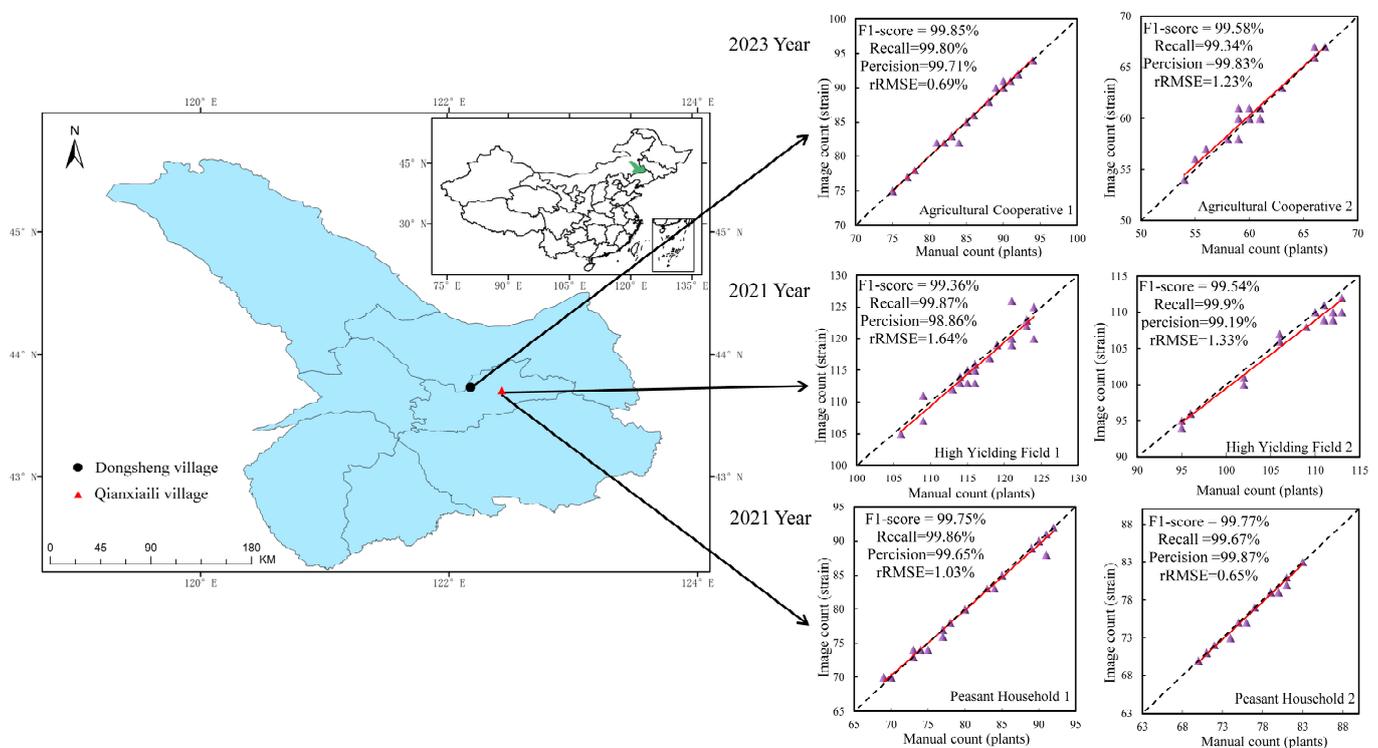


Figure 8. YOLOv8n model validation across different farms presented by linear fitting curves of actual values versus predicted values.

4. Discussion

Our model comparison experiments indicate that one-stage models generally outperform two-stage ones. This is potentially due to the direct collection of target location and category information in an end-to-end manner, eliminating the need for candidate box generation. This direct transmission of position, scale, and category information between targets through the supervision signal also allows for a more simple and rapid way of determining relationships between targets, thereby achieving better detection results [24]. YOLOv8n showed the highest detection performance, followed closely by YOLOv5n, then YOLOv3-tiny. The two-stage Deformable DETR model also exhibits high performance, which may be attributed to the introduction of a modified transformer with local and sparsely efficient attention mechanisms [23]. The YOLO deep learning series was found

to be highly accurate, with fast detection speeds and small sizes. YOLOv8n contains a decoupled head instead of the coupled one employed by YOLOv5, potentially contributing to its increased accuracy. Attempting to perform classification and localization on the same feature map may lead to a “misalignment” problem and poor results [25]. Instead, the decoupled head uses distinct branches for computation, thus improving performance [26]. Contrary to expectations, Faster R-CNN was less accurate in detecting small objects. This is likely due to the low resolution of the feature maps generated by the backbone network, which causes the minute features to blur or lose their clarity during processing. Additionally, the RoI generation method may not be accurate enough for small object localization. In addition to small object sizes, background noise may also affect detection accuracy. Moreover, Faster R-CNN may lack the ability to adapt to large-scale target changes when processing small objects, making it difficult for the model to capture and recognize object size variations or changes [24].

Plant density was found to have the most significant influence on the accuracy of maize seedling quantification. Our results indicate that increased overlapping between leaves is responsible for much of the declining accuracy [27]. However, YOLOv8n was less affected by planting density compared to the other models, and its detection capability only destabilizes when density surpasses 105,000 plants/ha. Increased density is a persistent challenge to the efficacy of various plant detection methods [28], representing an important direction for future research. Dense planting techniques have been increasingly favored due to their higher crop yields, especially regarding maize cultivation. For instance, the average maize planting density in regions like Xinjiang, China, has already passed 105,000 plants/ha. This represents a major hurdle in the application of this technology in an agricultural setting. To tackle the challenge of detecting corn in high-density fields where plant occlusion is significant, it is crucial to adopt strategies that enhance the ability to differentiate individual plants despite heavy overlap. Employing three-dimensional imaging technologies, such as LiDAR or structured light scanning, can provide depth information, enabling more accurate differentiation between overlapping plants [29]. Enhancing machine learning models with algorithms designed to process 3D data [30] or recognize patterns in occluded environments can improve detection accuracy [31]. This study explores the limits of model detection at high densities, providing a basis for future research and development. Our results reveal variations in detection performance characteristics between different models, highlighting the need to match the model to the growing operation.

Detection accuracy varies greatly between maize growth stages, highlighting the importance of timing drone image-capturing operations. If images are captured too early, the seedlings may be too small for detection, but if captured too late, there is increased leaf presence and overlapping, which can lead to a decline in detection performance [32]. Significant overlapping was documented during the V6 stage in this study, causing notable difficulties in plant detection [7]. Plants often fail to germinate or grow in a production setting, necessitating additional plantings to fill in gaps and maximize crop yield. The optimal replanting period is during the 2–3 leaf stage—missing this short window may negatively impact crop growth and yield. YOLOv8n and Deformable-DETR were found to be more effective than other models in detecting small targets, with rRMSE and F1-scores at the V2 stage of 2.19% and 5.52%, and 99.34% and 97.32%, respectively.

Image ground resolution is mainly determined by the UAV sensor and the flight altitude. We tested flight altitudes of 20, 40, and 60 m to comprehensively evaluate their effects on detection accuracy. Increases in flight altitude were correlated with decreases in seedling detection. This phenomenon is not only caused by the decrease in image ground resolution but also by the reduction of details in the acquired images [33,34]. Such a loss of detail may blur the features of the seedlings, directly affecting the model’s ability to recognize them. Changes in flight altitude can affect the visibility of maize features in collected images, putting a higher demand on dataset construction.

During this study, we worked directly with farmers to explore the practical applications of this technology in agriculture. Previous studies have validated the field use of this

technology by exploring its efficacy in various soil types, meteorological conditions, and growing operations. The existing models could be utilized in future studies to construct a maize emergence quality assessment model, which currently lacks an assessment index. UAVs have been increasingly used for precisely assessing maize seedling emergence and quality [35]. These assessments can provide growers with information crucial for making intelligent management decisions. These decisions can have dramatic impacts on crop growth, yield, and quality [28].

5. Conclusions

This study analyzed the performance of various object detection models used in maize production. Additionally, we explored the impacts of planting density, growth stages, and flight altitudes on model accuracy. Our results show that one-stage models such as YOLOv8n and YOLOv5n, with AP₅₀ scores of 0.976 and 0.951, respectively, generally outperformed two-stage models in quantifying maize seedlings while maintaining lower resource demands. SSD and FCOS, though effective, required higher computational resources with FLOPs of 137.1G and 78.6G, respectively, which may limit their practical use. YOLOv3-tiny and RetinaNet, while more resource-efficient, achieved lower performance efficiencies. Among two-stage models, Deformable DETR achieved an AP₅₀ of 0.939, indicating strong performance, whereas models like Faster R-CNN and Cascade R-CNN, though less resource-efficient, provided useful data. Based on their performance and architectural features, YOLOv8n, Deformable DETR, Faster R-CNN, and YOLOv3-tiny were selected for further detailed exploration.

Plant density and growth stage significantly impacted the seedling detection accuracy of all models. An increase in either factor complicated the obtained image and decreased accuracy. The V6 growth stage was especially difficult to quantify, as the increase in leaf overlap leads to detection difficulties. The optimal detection period was identified as the V2–V3 stages. YOLOv8n was the most stable model, only losing detection abilities at planting densities of more than 105,000 plants/ha. Additionally, flight altitude was negatively correlated with image resolution and detection results, causing decreased detection at higher altitudes.

In practical field applications across diverse agricultural settings, YOLOv8n demonstrated high accuracy and robustness. Specifically, in high-yield fields 1 and 2, the model achieved rRMSEs of only 1.64% and 1.33%, respectively. Furthermore, it consistently exceeded expectations, with Recall, Precision, and F1 scores all surpassing 99%. The model also performed exceptionally well in both cooperative and individual farmer scenarios, with all rRMSE values remaining below 1.23%.

Taken together, these results provide the framework for the application of UAV image collection models in an agricultural setting and highlight potential areas of future research. Lower flight altitude was favorable to maintaining good detection results, and the performance of the model gradually decreased with increasing altitude.

Author Contributions: Conceptualization, Z.J. and B.M.; methodology, B.M. and H.Y.; software, Z.J. and X.Y.; validation, Z.J., Y.L. and J.L.; formal analysis, Z.J. and K.G.; investigation, D.F.; resources, X.Z.; data curation, Z.J. and C.N.; writing—original draft preparation, Z.J.; writing—review and editing, Z.J. and J.X.; visualization, Z.J.; supervision, J.X.; project administration, B.M.; funding acquisition, S.L. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for this project was provided by the Inner Mongolia Science and Technology Major Project (2021ZD0003), the earmarked fund for China Agriculture Research System (CARS-02), and the Agricultural Science and Technology Innovation Program (CAAS-ZDRW202004).

Data Availability Statement: The data presented in this study are available on request from the corresponding author (accurately indicate status).

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Erenstein, O.; Jaleta, M.; Sonder, K.; Mottaleb, K.; Prasanna, B.M. Global maize production, consumption and trade: Trends and R&D implications. *Food Secur.* **2022**, *14*, 1295–1319. [[CrossRef](#)]
2. Tollenaar, M.; Lee, E. Yield potential, yield stability and stress tolerance in maize. *Field Crops Res.* **2002**, *75*, 161–169. [[CrossRef](#)]
3. Kimmelshue, C.L.; Goggi, S.; Moore, K.J. Seed size, planting depth, and a perennial groundcover system effect on corn emergence and grain yield. *Agronomy* **2022**, *12*, 437. [[CrossRef](#)]
4. Bongiovanni, R.; Lowenberg-DeBoer, J. Precision agriculture and sustainability. *Precis. Agric.* **2004**, *5*, 359–387. [[CrossRef](#)]
5. Kayad, A.; Paraforos, D.S.; Marinello, F.; Fountas, S. Latest advances in sensor applications in agriculture. *Agriculture* **2020**, *10*, 362. [[CrossRef](#)]
6. Liu, S.; Yin, D.; Feng, H.; Li, Z.; Xu, X.; Shi, L.; Jin, X. Estimating maize seedling number with UAV RGB images and advanced image processing methods. *Precis. Agric.* **2022**, *23*, 1604–1632. [[CrossRef](#)]
7. Bai, Y.; Nie, C.; Wang, H.; Cheng, M.; Liu, S.; Yu, X.; Shao, M.; Wang, Z.; Wang, S.; Tuohuti, N.; et al. A fast and robust method for plant count in sunflower and maize at different seedling stages using high-resolution UAV RGB imagery. *Precis. Agric.* **2022**, *23*, 1720–1742. [[CrossRef](#)]
8. Liu, T.; Yang, T.; Li, C.; Li, R.; Wu, W.; Zhong, X.; Sun, C.; Guo, W. A method to calculate the number of wheat seedlings in the 1st to the 3rd leaf growth stages. *Plant Methods* **2018**, *14*, 101. [[CrossRef](#)]
9. Xu, B.; Chen, L.; Xu, M.; Tan, Y. Path Planning Algorithm for Plant Protection UAVs in Multiple Operation Areas. *Trans. Chin. Soc. Agric. Mach.* **2017**, *48*, 75–81.
10. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
11. Quan, L.; Feng, H.; Lv, Y.; Wang, Q.; Zhang, C.; Liu, J.; Yuan, Z. Maize seedling detection under different growth stages and complex field environments based on an improved Faster R-CNN. *Biosyst. Eng.* **2019**, *184*, 1–23. [[CrossRef](#)]
12. Zhang, Y.; Xiao, D.; Liu, Y.; Wu, H. An algorithm for automatic identification of multiple developmental stages of rice spikes based on improved Faster R-CNN. *Crop J.* **2022**, *10*, 1323–1333. [[CrossRef](#)]
13. Gao, J.X.; Tan, F.; Cui, J.P.; Ma, B. A Method for Obtaining the Number of Maize Seedlings Based on the Improved YOLOv4 Lightweight Neural Network. *Agric. Basel* **2022**, *12*, 1679. [[CrossRef](#)]
14. Zhang, X.; Zhu, D.; Wen, R. SwinT-YOLO: Detection of densely distributed maize tassels in remote sensing images. *Comput. Electron. Agric.* **2023**, *210*, 107905. [[CrossRef](#)]
15. Li, R.; Wu, Y. Improved YOLO v5 wheat ear detection algorithm based on attention mechanism. *Electronics* **2022**, *11*, 1673. [[CrossRef](#)]
16. Zhang, Y.; Li, M.Z.; Ma, X.X.; Wu, X.T.; Wang, Y.J. High-Precision Wheat Head Detection Model Based on One-Stage Network and GAN Model. *Front. Plant Sci.* **2022**, *13*, 787852. [[CrossRef](#)] [[PubMed](#)]
17. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
18. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Part I 14. pp. 21–37.
19. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: A simple and strong anchor-free object detector. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 1922–1933. [[CrossRef](#)] [[PubMed](#)]
20. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
21. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
22. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
23. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv* **2020**, arXiv:2010.04159.
24. Li, X.; Chen, J.; He, Y.; Yang, G.; Li, Z.; Tao, Y.; Li, Y.; Li, Y.; Huang, L.; Feng, X. High-through counting of Chinese cabbage trichomes based on deep learning and trinocular stereo microscope. *Comput. Electron. Agric.* **2023**, *212*, 108134. [[CrossRef](#)]
25. Song, G.; Liu, Y.; Wang, X. Revisiting the sibling head in object detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11563–11572.
26. Wu, Y.; Chen, Y.; Yuan, L.; Liu, Z.; Wang, L.; Li, H.; Fu, Y. Rethinking classification and localization for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10186–10195.
27. Liu, S.; Baret, F.; Andrieu, B.; Burger, P.; Hemmerlé, M. Estimation of wheat plant density at early stages using high resolution imagery. *Front. Plant Sci.* **2017**, *8*, 232042. [[CrossRef](#)] [[PubMed](#)]
28. Liu, M.; Su, W.-H.; Wang, X.-Q. Quantitative Evaluation of Maize Emergence Using UAV Imagery and Deep Learning. *Remote Sens.* **2023**, *15*, 1979. [[CrossRef](#)]
29. Debnath, S.; Paul, M.; Debnath, T. Applications of LiDAR in agriculture and future research directions. *J. Imaging* **2023**, *9*, 57. [[CrossRef](#)] [[PubMed](#)]

30. Anifantis, A.S.; Camposeo, S.; Vivaldi, G.A.; Santoro, F.; Pascuzzi, S. Comparison of UAV photogrammetry and 3D modeling techniques with other currently used methods for estimation of the tree row volume of a super-high-density olive orchard. *Agriculture* **2019**, *9*, 233. [[CrossRef](#)]
31. Sun, T.; Zhang, W.; Miao, Z.; Zhang, Z.; Li, N. Object localization methodology in occluded agricultural environments through deep learning and active sensing. *Comput. Electron. Agric.* **2023**, *212*, 108141. [[CrossRef](#)]
32. Feng, Y.; Chen, W.; Ma, Y.; Zhang, Z.; Gao, P.; Lv, X. Cotton Seedling Detection and Counting Based on UAV Multispectral Images and Deep Learning Methods. *Remote Sens.* **2023**, *15*, 2680. [[CrossRef](#)]
33. Gong, Y.; Yu, X.; Ding, Y.; Peng, X.; Zhao, J.; Han, Z. Effective fusion factor in FPN for tiny object detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual Conference, 5–9 January 2021; pp. 1160–1168.
34. Sankaran, S.; Khot, L.R.; Espinoza, C.Z.; Jarolmasjed, S.; Sathuvalli, V.R.; Vandemark, G.J.; Miklas, P.N.; Carter, A.H.; Pumphrey, M.O.; Knowles, N.R. Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *Eur. J. Agron.* **2015**, *70*, 112–123. [[CrossRef](#)]
35. Liu, T.; Li, R.; Jin, X.; Ding, J.; Zhu, X.; Sun, C.; Guo, W. Evaluation of seed emergence uniformity of mechanically sown wheat with UAV RGB imagery. *Remote Sens.* **2017**, *9*, 1241. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.