

Article

Not peer-reviewed version

---

# Comparison of drone-assisted maize seedling detection models based on deep learning

---

[Zhijie Jia](#) , Xinglong Zhang , Hongye Yang , Yuan Lu , Jiale Liu , Xun Yu , Dayun Feng , Kexing Gao , [Jianfu Xue](#)  
\*, [Bo Ming](#) \*, Chenwei Nie , [Shaokun Li](#)

Posted Date: 12 March 2024

doi: [10.20944/preprints202403.0629.v1](https://doi.org/10.20944/preprints202403.0629.v1)

Keywords: Maize Seedling Detection, Counting, Unmanned Aerial Vehicle



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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.

## Article

# Comparison of Drone-Assisted Maize Seedling Detection Models Based on Deep Learning

Zhijie Jia <sup>1</sup>, Xinglong Zhang <sup>2</sup>, Hongye Yang <sup>3</sup>, Yuan Lu <sup>4</sup>, Jiale Liu <sup>4</sup>, Xun Yu <sup>3</sup>, Dayun Feng <sup>3</sup>, Kexing Gao <sup>3</sup>, Jianfu Xue <sup>1,\*</sup>, Bo Ming <sup>3,\*</sup>, Chenwei Nie <sup>5</sup> and Shaokun Li <sup>3</sup>

<sup>1</sup> College of Agriculture, Shanxi Agricultural University, Taigu, Jingzhong 030801, China; 511998815@qq.com(Z.J.)

<sup>2</sup> Zhangjiakou Academy of Agricultural Sciences, Zhangjiakou, Hebei 075000, China; 274387042@qq.com

<sup>3</sup> Institute of Crop Sciences, Chinese Academy of Agricultural Sciences / State Key Laboratory of Crop Gene Resources and Breeding, Beijing100081, China; 82101231082@caas.cn (H.Y.); 82101211308@caas.cn(X.Y.); 18735269762@163.com(D.F.); 1582152032@qq.com(K.G.); lishaokun@caas.cn(S.L.)

<sup>4</sup> College of Agriculture, Qingdao Agricultural University, Qingdao, Shandong 266000, China; luyuan5251@163.com(Y.L.); 1327558758@qq.com(J.L.)

<sup>5</sup> Surveying and Municipal Engineering, Zhejiang Institute of Water Resources and Hydropower, Hangzhou, Zhejiang 310000, China; niechwu@zjweu.edu.cn

\* Correspondence: fudange95@163.com(J.X.) and 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 presents an in-depth analysis of multiple object detection models used by drones to count maize seedlings, while concurrently investigating the influence of planting density, flight altitude, and plant growth stage. The findings of this study demonstrate that one-stage models are able to more accurately detect maize seedlings than two-stage models. In particular, YOLOv8n exhibits outstanding performance, achieving an F1-score of over 0.92 under various density conditions and maintaining a stable performance, especially under densities below 105,000 plants/ha. Additionally, planting density and growth stage were observed to significantly affect detection accuracy, with performance declining as density and growth stage increased. Image resolution and detection were impacted by flight altitude, with lower flights producing higher-quality results. Ultimately, YOLOv8n displays optimal performance under various conditions, providing crucial data to support intelligent decision-making. This research offers valuable insights into the application of object detection technology and provides a basis for the future development of precision agriculture.

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

## 1. Introduction

Maize is among the world's most widely cultivated and traded crops, serving various purposes including human consumption, animal feed, and 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 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 in 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.

Unmanned Aerial Vehicles (UAV) are revered for their affordability, portability, and flexibility. They can be equipped with a diverse array of sensors such as RGB[4], multispectral[5], hyperspectral[6], and LiDAR[7] to ensure robust data capture. Initially, images captured by UAVs required traditional processing techniques such as skeletonization algorithms and multiple despeckling processes to extrapolate pertinent information, including seedling count[8]. While these approaches were effective, they necessitated intricate processing workflows and high-quality images. As the technology 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[9]. Moreover, corner detection techniques have enabled the effective counting of overlapping leaves which tend to complicate data, especially as plants mature[10].

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<sup>[11]</sup>. The ongoing development of deep learning is continuously broadening the scope of agricultural applications. 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<sup>[12]</sup>. Additionally, the model is able to automatically recognize and record different developmental stages of rice panicles, a previously labor-intensive manual process<sup>[13]</sup>. The multiple required complex processing stages 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<sup>[14]</sup>. 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<sup>[15]</sup>. 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<sup>[16]</sup>. 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<sup>[17]</sup>.

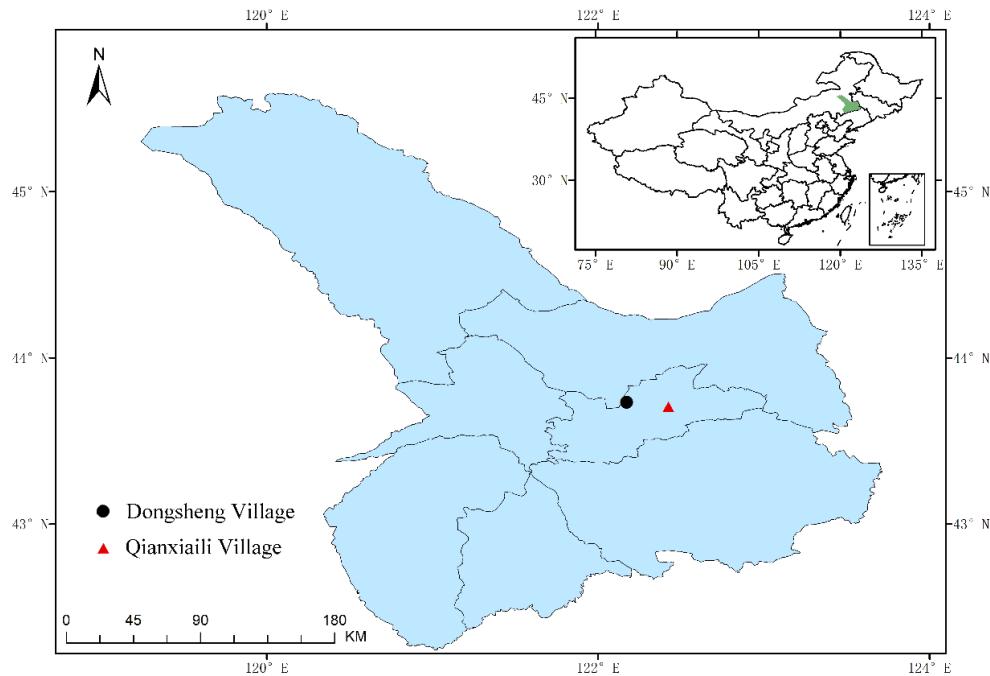
While previous studies have achieved impressive accuracy by focusing on algorithm optimization, there has been little exploration of how external considerations may impact performance. In this study, we investigate the influence of key factors such as planting density, growth stage, and flight altitude to comprehensively evaluate the performance of deep learning models used for maize seedling target detection. Future field trials will be conducted to further validate these results and confirm practicality in an active farming operation.

## 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. This region features a semi-arid continental monsoon climate with 2,500-2,800 h of annual sunshine, an average daily temperature of 21.0°C, a cumulative  $\geq 10^{\circ}\text{C}$  temperature of 3,000-3,300°C·d, a frost-free period of 150-169 d, and an average annual precipitation of 280-390 mm during the maize growing season (May-Eleventh). 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, 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 Qianxialili Village. Trials were planted on May, 10, 2021 at densities of 30,000, 45,000, 60,000, 75,000, 90,000, 105,000, 120,000, and 135,000 plants/ha. Data was collected on May 29<sup>th</sup> (2 leaves unfolded), June 1<sup>st</sup> (3 leaves unfolded), June 5<sup>th</sup> (4 leaves unfolded), and June 16<sup>th</sup> (6 leaves unfolded).



**Figure 1.** Experimental area.

The maize variety Dika 159 was selected for the flight altitude experiment conducted in Dongsheng Village. Trials were planted on May, 8, 2023 at a density of 90,000 plants/ha. Data was collected on May 27<sup>th</sup> (2 leaves unfolded), May 31<sup>st</sup> (3 leaves unfolded), June 4<sup>th</sup> (4 leaves unfolded), June 7<sup>th</sup> (5 leaves unfolded), and June 11<sup>th</sup> (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 m<sup>2</sup>) in each field, which were monitored at noon on June 1<sup>st</sup>, 2021 (3 leaves unfolded) and May 31<sup>st</sup>, 2023 (3 leaves unfolded).

## 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 barometer 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	Qianxialili Village	High Yielding Field	Peasant Household	Dongsheng Village	Agricultural Cooperative

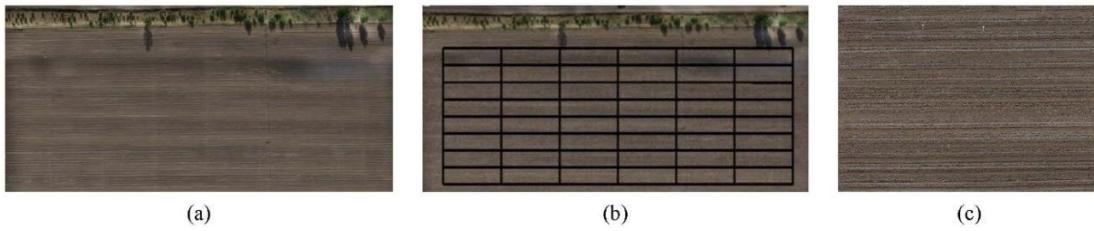
Image acquisition	2, 3, 4, 6	3	3	2-6	3
Stage (leaves)					
Flight speed (m/s)	2.1	2.3	2.0	2.0	2.5
photo interval (s)	1	2	2	2	2
Height above ground (m)	20	20	20	20, 40, 60	20
Overlap rate along tracks(%)	75	73	75	80	75
Overlap rate across tracks(%)	85	75	80	80	75

Images were collected with a Sony α7 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 1Hz with an intervalometer-controlled camera.

### 2.3. Data Construction and Preprocessing

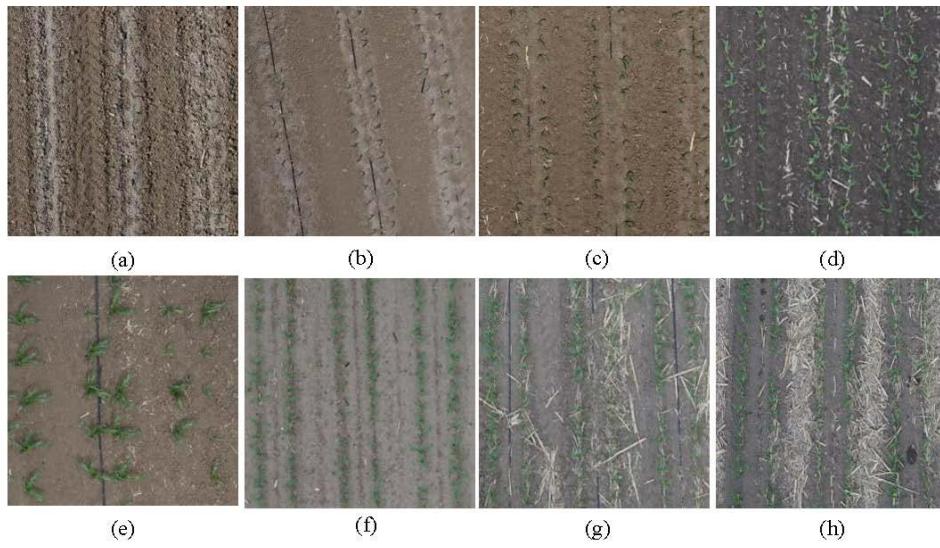
#### 2.3.1. Image Preprocessing

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).



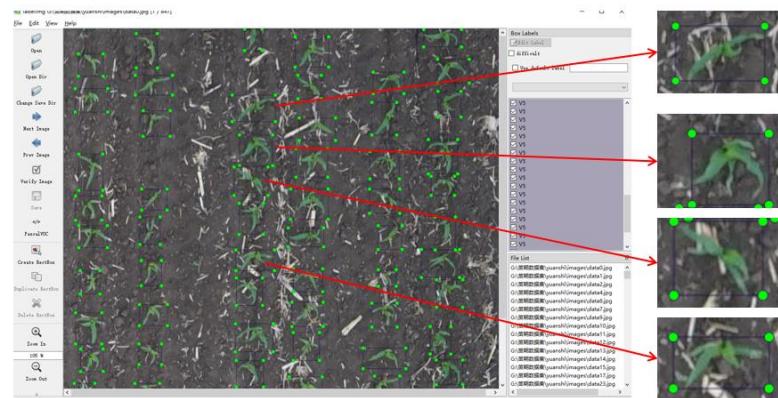
**Figure 2.** (a) Orthophoto of test area, (b) experimental plot layout, and (c) plot cropping.

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[18]. Poor-quality images including those with large shooting angles, obvious occlusions, and uneven illumination were removed. Final images (900 total) were categorized by quantity of spreading leaves and quantity of straw.



**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.

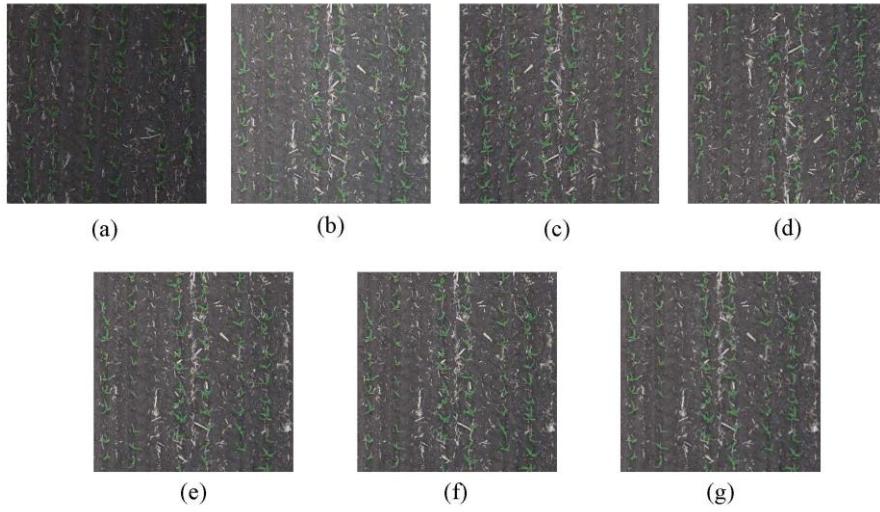
The location and size of each seedling are labeled in each image[18] (Figure 4).



**Figure 4.** Example of annotated image. Green boxes represent labeled maize seedlings.

### 2.3.2. Data Enhancements

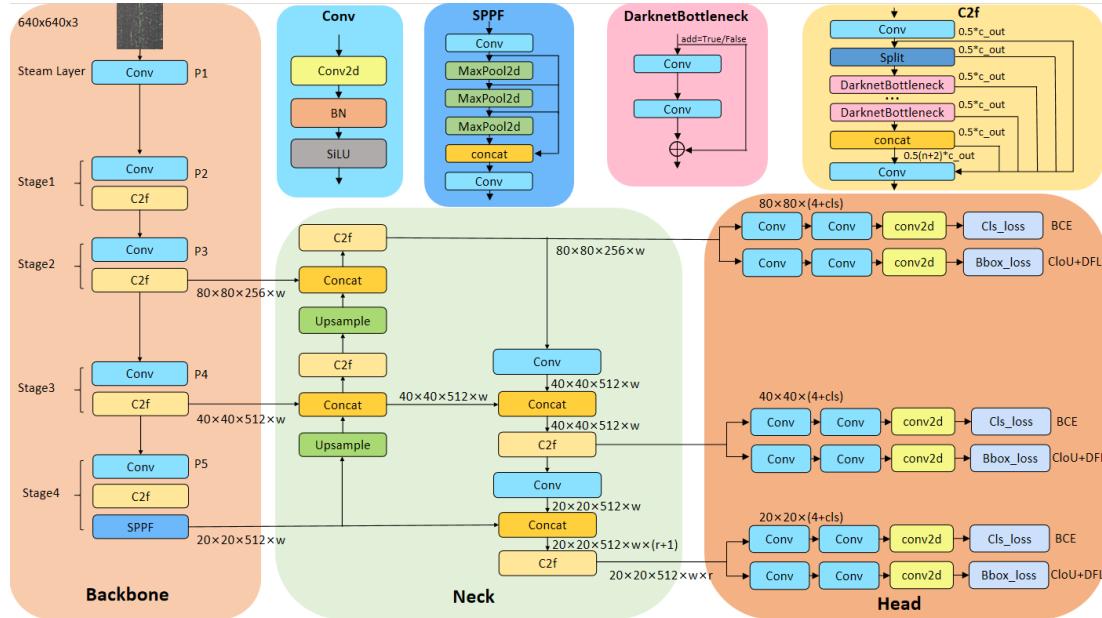
Images were adjusted during the data augmentation process, including horizontal and vertical flipping, random contrast and hue adjustments, and resolution alterations. These modifications were used to simulate the effects of varying lighting and environmental factors during different times of day and flight altitudes. Training and testing datasets were created from the resulting 14,815 images. The dataset was then divided, with 90% of images used for model training and 10% for validating model performance.



**Figure 5.** (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. YOLOv8 Model Framework

YOLO (You Only Look Once) is a highly popular single-stage object detection framework, which exhibits enhanced processing speeds, performance, efficiency, and flexibility when compared with two-stage algorithms. YOLOv8 is comprised of a Backbone, Neck, and Head (Figure 6).



**Figure 6.** YOLOv8 framework.

##### 2.4.1. Backbone and Neck

The YOLOv8 Backbone employs the Darknet-53 network structure. Through effective feature learning and residual connections, the model avoids the gradient vanishing typical of deep neural networks, enabling it to capture high-level semantic features[19]. This module primarily functions to extract and fuse these higher-level features through multiple iterations of maximum pooling[20].

The C3 module in YOLOv8n is replaced with the more gradient-rich C2f module, which adjusts the number of channels based on the intended model scale to minimize weight[21]. This change enhances efficiency and efficacy when processing images of various sizes and complexities.

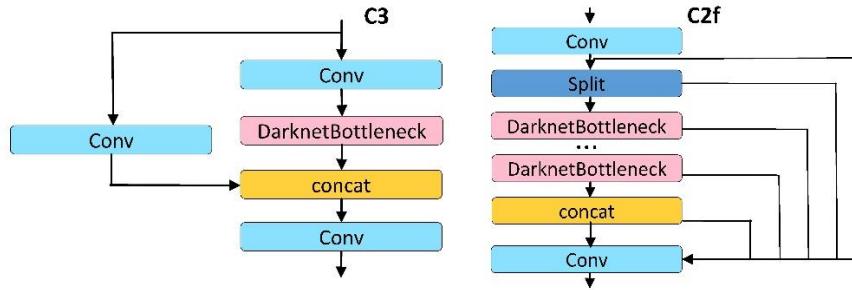


Figure 7. C3 and C2f modules.

#### 2.4.2. Head

The YOLOv8n Head structure exhibits two major improvements when compared to YOLOv5. Firstly, the newer model has separate classification and detection heads (Decoupled-Head), a structure type that has been increasing in popularity. Secondly, the model transitions away from the Anchor-Based approach to an Anchor-Free approach.

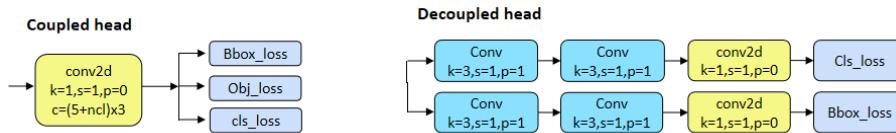


Figure 8. Coupled-Head and Decoupled-Head modules.

#### 2.4.3. Loss Calculation

Loss calculation in the YOLOv8n model is simplified by removing the objectness branch and only considering the classification and regression branches. The classification branch continues to utilize Binary Cross-Entropy (BCE) Loss. YOLOv8n abandons the traditional IOU matching and unilateral proportion allocation methods in favor of the Task-Aligned Assigner for positive and negative sample matching. Additionally, the model introduces Distribution Focal Loss (DFL)[22] to improve the precision and reliability of object detection tasks. Loss calculations are conducted according to the following formulas:

$$DFL(S_i, S_{i+1}) = -((y_{i+1} - y) \log(S_i) + (y - y_i) \log(S_{i+1})) \quad (1)$$

In two consecutive distributions,  $S_i$  and  $S_{i+1}$ , the values  $y_i$  and  $y_{i+1}$  represent the corresponding probability density values at specific points or for the distributions as a whole. The weighting of the loss function is adjusted based on  $y$ , to reflect changes or differences in these probability densities.

Classification loss is typically calculated using cross-entropy loss, which measures the difference between the predicted class probabilities and the ground truth class labels:

$$L_{cls} = -\frac{1}{N} \sum_{i \in posUNeg} \sum_{c=1}^C y_c^{(i)} \log(\hat{y}_c^{(i)}) \quad (2)$$

In these formulas,  $i$  represents the sample index and  $i \in posUNeg$  denotes the set of positive and negative samples.  $c$  represents the category index and  $C$  denotes the total number of categories. The real label of the  $i^{\text{th}}$  sample for category  $c$  is denoted by  $y_c^{(i)}$ , which indicates whether the sample actually belongs to category  $c$  (typically 1 for yes, 0 for no). The model's predicted probability that the  $i^{\text{th}}$  sample belongs to category  $c$  is represented by  $\hat{y}_c^{(i)}$ .

$$L = \lambda_{loc} L_{loc} + \lambda_{cls} L_{cls} \quad (3)$$

$L$  represents the total loss function.  $\lambda_{cls}$  and  $\lambda_{loc}$  are hyperparameters controlling the relative importance of localization and classification loss.

### 2.5. Assessment of Indicators

The maize seedling detection and quantification abilities of each model were evaluated by calculating the Precision (Eq.4), Recall (Eq.5), F1-Score (Eq.6), AP(Eq.7) and rRMSE (Root Mean Square Error) (Eq.8) values according to the following formulas:

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (5)$$

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (6)$$

$$AP = \int_0^1 P(R) d^R \quad (7)$$

$$rRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2}}{\sum_{i=1}^n y_i} \times 100\% \quad (8)$$

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 3. 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) i5-12600KF 3.69 GHz		
Operating system	Windows 10		
Ram	64GB		
Graphics card	NVIDIA GeForce RTX 3060		
Programming language	Python 3.8		
Model	YOLOv8n	YOLOv5, 3	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

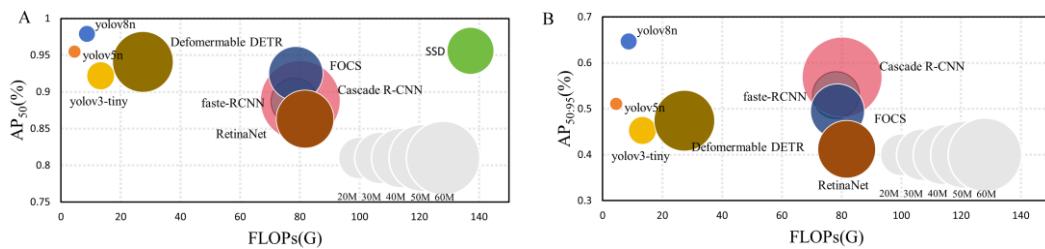
## 3. Results

### 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<sub>50</sub>, AP<sub>50:95</sub>, params, and FLOPs (Table 4, Figure 11).

**Table 3.** Comparison of object detection model performances.

Category	Model	Backbone	Image size	AP <sub>50</sub>	AP <sub>50:95</sub>	Params	FLOPs
One-stage	YOLOv8n	New CSP-Darknet53	640×640	<b>0.979</b>	<b>0.647</b>	3.20M	8.7G
	YOLOv5n	CSP-Darknet53	640×640	0.955	0.511	1.90M	4.5G
	SSD	VGG16	416×416	0.946	0.530	23.75M	137.1G
	FCOS	Resnet50	640×640	0.925	0.495	31.84M	78.6G
	YOLOv3-tiny	Tiny-Darknet	640×640	0.922	0.453	8.44M	13.3G
	RetinaNet	Resnet50	300×300	0.863	0.412	36.10M	81.7G
Two-stage	Deformable DETR	Resnet50	640×640	0.941	0.474	36.10M	27.4G
	Cascade R-CNN	Resnet50	640×640	0.888	0.569	68.94M	80.1G
	Faster R-CNN	Resnet50	640×640	0.887	0.530	41.12M	78.1G

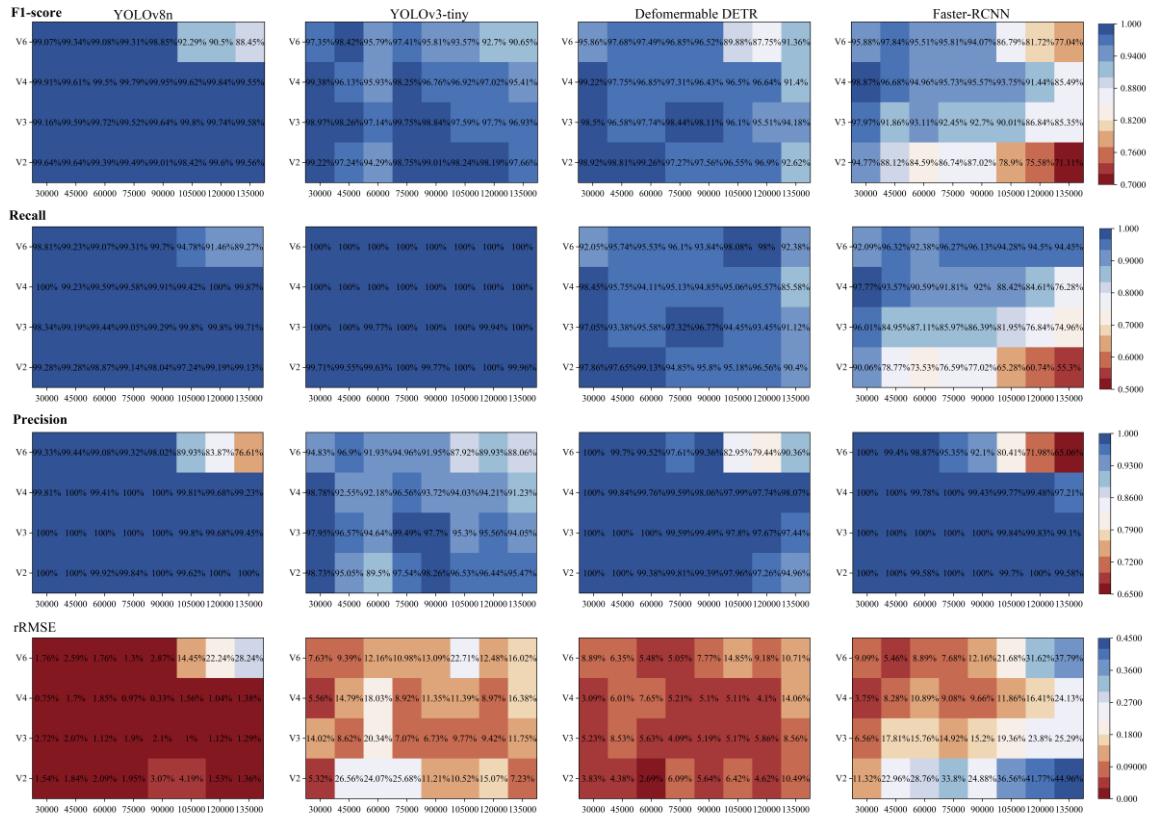
**Figure 9.** Performance comparison of different models. (a) Relationship between AP<sub>50</sub>, Params, and FLOPs. (b) Relationship between AP<sub>50:95</sub>, Params, and FLOPs.

The performances of YOLOv8n and YOLOv5n stand out among the single-stage models, achieving AP<sub>50</sub> values of 0.979 and 0.955, 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<sub>50</sub> value of 0.941 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.

### 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 in 8 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.



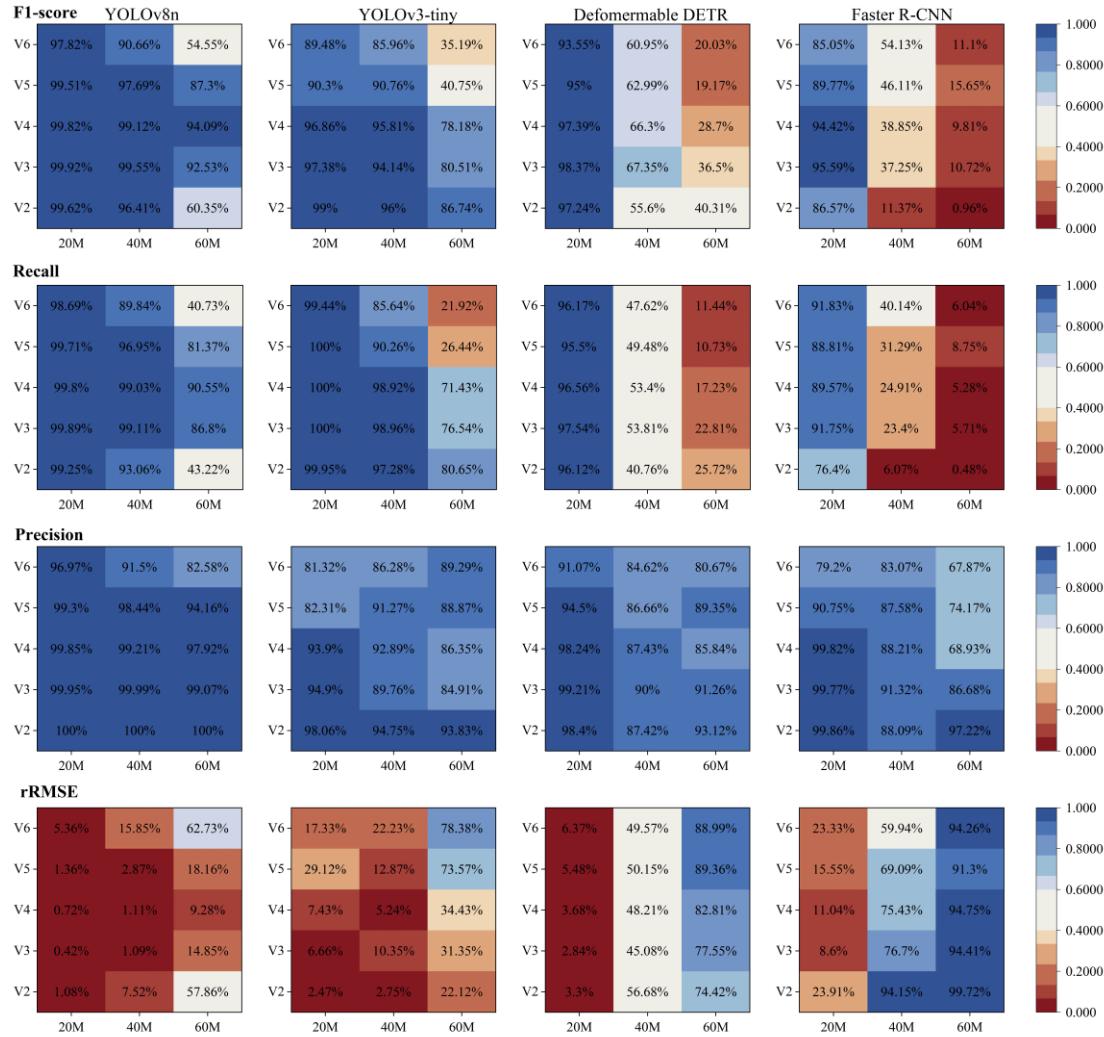
**Figure 10.** Performances of YOLOv8n, YOLOv3-tiny, Deformable-DETR, and Faster R-CNN in terms of F1-score, Recall, Precision, and rRMSE across various planting densities and growth stages.

Our findings demonstrate that as density increases, overall detection accuracy measured by the F1-score, rRMSE, Recall, and Precision declines (Figure 10). 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.

Our study showed that as planting density increases, YOLOv8n exhibits a relatively stable rRMSE performance compared to YOLOv3-tiny, especially at higher densities (105,000 to 135,000). The Deformable-DETR model exhibited a 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 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.

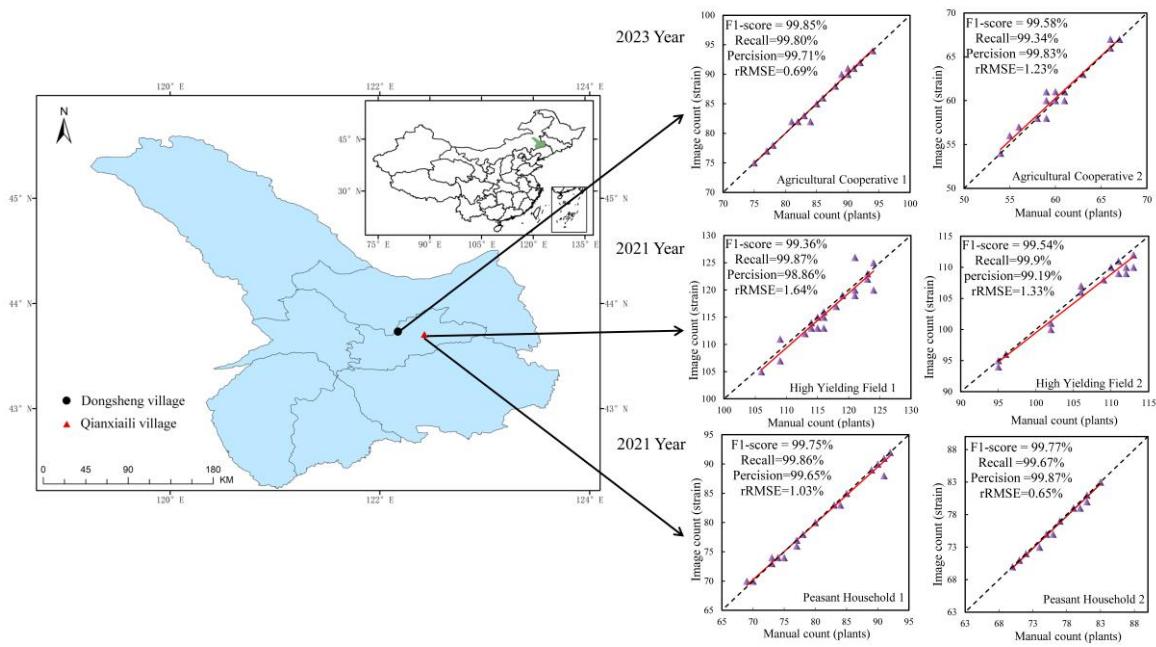


**Figure 11.** Performance of YOLOv8n, YOLOv3-tiny, Deformable-DETR, and Faster R-CNN in terms of F1-score, Recall, Precision, and rRMSE at different flight altitudes and growth stages.

The F1-scores of all four models decreased across all growth stages (V2-V6) as flight altitude increased. Performances were relatively high at 20 m but declined significantly at 60 m. The Recall and Precision metrics display a similar but less severe trend. Overall, these models demonstrated greater efficacy at lower altitudes.

### 3.4. Validation of 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 (six total repetitions). The model exhibited outstanding performance across all planting methods, with notably significant results in high-yield fields characterized by high-density planting conditions.



**Figure 12.** YOLOv8n model validation across different farms presented by linear fitting curves of actual values versus predicted values.

Our analysis illustrates that predicted values are closely aligned with the actual values, maintaining a highly consistent, near 1:1 line (Figure 12). 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. 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.

#### 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[23]. YOLOv8n showed the highest detection performance, followed closely by YOLOv5n, then YOLOv3-tiny. The two-stage Deformable DETR model also exhibits a high performance, which may be attributed to the introduction of a modified transformer with local and sparse efficient attention mechanisms[24]. 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 causing the minute features to blur or lose 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[27].

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[28]. 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[29], 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. 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, spotlighting 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[30]. Significant overlapping was documented during the V6 stage in this study, causing notable difficulties in plant detection[9]. 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 the reduction of details in the acquired images[31]. Such 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 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. 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.

## 5. Conclusions

This study analyzed the performances of various target detection models used in maize production. Additionally, we explored the impacts of planting density, flight height, and plant growth stages on model accuracy. Our results indicate that the one-stage models generally outperform two-stage models in maize seedling quantification. The one-stage models YOLOv8n and YOLOv5n demonstrated stable and excellent performances, especially at lower planting densities. The two-stage model, Deformable DETR, was relatively stable and outperformed Faster R-CNN, which showed significant performance degradation under highly dense planting conditions.

Plant density and growth stage significantly impacted the seedling detection accuracy of all models. Increased in either factor complicated obtained imaged and decreased accuracy. The V6 growth stage was especially difficult to quantify, as the increase of 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. 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 maintain 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.; 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:** Not applicable.

**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 a potential conflict 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 Security* **2022**, *14*, 1295-1319, doi:10.1007/s12571-022-01288-7.
2. LI shaoKun, Z.J., DONG ShuTing, ZHAO Ming, LI ChaoHai, CUI YanHong, LIU YongHong, GAO JuLin, XUE JiQuan, WANG LiChun, WANG Pu, LU WeiPing, WANG JunHe, YANG QiFeng, WANG ZiMing. . Advances and Prospects of Maize Cultivation in China. *Scientia Agricultura Sinica* **2017**, *50*, 1941-1959.
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. **2022**, *12*, 437.
4. Onishi, M.; Ise, T. Explainable identification and mapping of trees using UAV RGB image and deep learning. *Scientific reports* **2021**, *11*, 903.
5. Candiago, S.; Remondino, F.; De Giglio, M.; Dubbini, M.; Gattelli, M. Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. *Remote sensing* **2015**, *7*, 4026-4047.
6. Adão, T.; Hruška, J.; Pádua, L.; Bessa, J.; Peres, E.; Morais, R.; Sousa, J.J. Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote sensing* **2017**, *9*, 1110.
7. Wallace, L.; Lucieer, A.; Watson, C.; Turner, D. Development of a UAV-LiDAR system with application to forest inventory. *Remote sensing* **2012**, *4*, 1519-1543.
8. Liu Shuaibing, Y.G., Zhou Chengquan, Jing Haitao, Feng Haikuan, Xu Bo, Yang Hao. . Extraction of maize seedlingnumber information based on UA imagery. *Transactions of the Chinese Society of Agricultural Engineering* **2018**, *34*, 69-77.
9. 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. *Precision Agriculture* **2022**, *23*, 1720-1742, doi:10.1007/s11119-022-09907-1.
10. 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, doi:10.1186/s13007-018-0369-5.
11. Xu, B.; Chen, L.; Xu, M.; Tan, Y. Path Planning Algorithm for Plant Protection UAVs in Multiple Operation Areas. *Transactions of the Chinese Society for Agricultural Machinery* **2017**, *48*, 75-81.
12. 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. *Biosystems Engineering* **2019**, *184*, 1-23, doi:10.1016/j.biosystemseng.2019.05.002.
13. 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. *The Crop Journal* **2022**, *10*, 1323-1333.
14. 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. *Agriculture-Basel* **2022**, *12*, doi:10.3390/agriculture12101679.
15. Zhang, X.; Zhu, D.; Wen, R. SwinT-YOLO: Detection of densely distributed maize tassels in remote sensing images. *Computers and Electronics in Agriculture* **2023**, *210*, doi:10.1016/j.compag.2023.107905.
16. Li, R.; Wu, Y.P. Improved YOLO v5 Wheat Ear Detection Algorithm Based on Attention Mechanism. *Electronics* **2022**, *11*, doi:10.3390/electronics11111673.
17. 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. *Frontiers in Plant Science* **2022**, *13*, doi:10.3389/fpls.2022.787852.

18. Tzutalin, D. Labelimg. git code. **2015**.
19. Pathak, D.; Raju, U.J.O. Content-based image retrieval using feature-fusion of GroupNormalized-Inception-Darknet-53 features and handcraft features. **2021**, *246*, 167754.
20. He, K.; Zhang, X.; Ren, S.; Sun, J.J.I.t.o.p.a.; intelligence, m. Spatial pyramid pooling in deep convolutional networks for visual recognition. **2015**, *37*, 1904-1916.
21. Xiao, B.; Nguyen, M.; Yan, W.Q. Fruit ripeness identification using YOLOv8 model. *Multimedia Tools and Applications* **2023**, doi:10.1007/s11042-023-16570-9.
22. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J.J.A.i.N.I.P.S. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. **2020**, *33*, 21002-21012.
23. 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. *Computers and Electronics in Agriculture* **2023**, *212*, 108134.
24. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J.J.a.p.a. Deformable detr: Deformable transformers for end-to-end object detection. **2020**.
25. Song, G.; Liu, Y.; Wang, X. Revisiting the sibling head in object detector. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 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 Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020; pp. 10186-10195.
27. Li, X.; Chen, J.; He, Y.; Yang, G.; Li, Z.; Tao, Y.; Li, Y.; Li, Y.; Huang, L.; Feng, X.J.C.; et al. High-through counting of Chinese cabbage trichomes based on deep learning and trinocular stereo microscope. **2023**, *212*, 108134.
28. Liu, S.; Baret, F.; Andrieu, B.; Burger, P.; Hemmerlé, M.J.F.i.P.S. Estimation of wheat plant density at early stages using high resolution imagery. **2017**, *8*, 739.
29. Liu, M.; Su, W.-H.; Wang, X.-Q.J.R.S. Quantitative Evaluation of Maize Emergence Using UAV Imagery and Deep Learning. **2023**, *15*, 1979.
30. Feng, Y.; Chen, W.; Ma, Y.; Zhang, Z.; Gao, P.; Lv, X.J.R.S. Cotton Seedling Detection and Counting Based on UAV Multispectral Images and Deep Learning Methods. **2023**, *15*, 2680.
31. 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 Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2021; pp. 1160-1168.

**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.