RESEARCH ON CORN SEEDLING DETECTION AND COUNTING ALGORITHM BASED ON MEI-YOLOv11
/
基于 MEI-YOLOV11 的玉米幼苗检测计数算法研究

Yiting LIU, Xiuying XU*), Jinkai QIU, Kai MA, Yanxu JIAO, Ye KANG
College of engineering, Heilongjiang Bayi Agricultural University, Daqing/ China
Tel: +86-13945925608; E-mail: xuxiuying@byau.edu.cn
Corresponding author: Xiuying Xu
DOI: https://doi.org/10.35633/inmateh-77-36

Keywords: *Corn seedlings, Seedling emergence count, Deep learning, MEI-YOLOv11 model, Efficient upsampling*

ABSTRACT
Accurately counting the number of corn seedlings is the key to evaluating the growth status of corn. To address the problem of difficult detection and counting of corn seedlings in complex field environments, this study proposes an improved MEI-YOLOv11 model. By introducing MANet, EUCB module, and Inner-SIOU loss function, the ability to extract features and recognize small targets in complex environments is enhanced. The results showed that the mAP0.5, P, and R of the model reached 97.0%, 94.2%, and 95.7%, respectively, which were 2.8, 2.7, and 2.4 percentage points higher than YOLOv11, respectively. The parameter count and inference time only increased by 1.28 M and 0.4 ms, respectively, and the detection accuracy was better than other detection models. The accuracy of multi weather counting is above 90%, with the highest accuracy of 91.23% on sunny days (RMSE=4.5044, R ²=0.8508). This method can effectively identify corn seedlings in complex backgrounds, providing technical support for accurate detection and counting of corn seedlings in multiple weather conditions.

摘要
准确统计玉米苗数是评估玉米生长状况的关键，本研究为应对复杂田间环境玉米幼苗检测计数难的问题，提出改进 MEI-YOLOv11 模型。通过引入 MANet、EUCB 模块及 Inner-SIoU 损失函数，增强复杂环境下特征提取与小目标识别能力。结果表明，该模型的 mAP0.5、P 和 R 分别达到 97.0%、94.2%和 95.7%，分别比 YOLOv11 提高 2.8、2.7 和 2.4 个百分点。参数量和推理时间分别只增加了 1.28M 和 0.4ms，检测精度优于其他检测模型。多天气计数准确率均在 90%以上，晴天准确率最高，为 91.23% （RMSE=4.5044, R²=0.8508）。该方法能够有效识别复杂背景下玉米幼苗，为多天气玉米幼苗准确检测计数提供技术支持。

INTRODUCTION
The emergence of corn seedlings is one of the key factors affecting corn yield. Accurate detection of corn seedlings can accurately obtain their emergence status, providing a basis for subsequent evaluation of sowing quality, field water and fertilizer management, and replanting of missing seedlings (*Ma et al., 2014*). Traditional seedling counting and detection equipment mainly relies on ground mobile platforms, which have problems such as single device functions and low operating efficiency, making it difficult to meet the needs of modern crop information collection fields (*Jia et al., 2015*). In recent years, unmanned aerial vehicle remote sensing platforms have been in a rapid development stage. Compared with traditional satellite remote sensing, it has the advantages of strong timeliness, high spatial resolution, and low equipment operating costs (*Liu et al., 2024*).

Early studies mostly employed machine vision and image processing methods for seedling counting, mainly based on crop phenotypic characteristics (*Lu et al., 2025*), vegetation color information (*Bryson et al., 2010*), and morphological features (*Han et al., 2021*), etc., combined with morphological operations for image segmentation. This type of method is easily affected by factors such as terrain, climate, shooting tools and crop spacing, and requires manual input of features.

Yiting Liu, currently pursuing a master's degree; Xiuying Xu, Associate Professor, Doctor of Engineering; Jinkai Qiu, currently a doctoral student; Kai Ma, Master of Agriculture; Yanxu Jiao, currently pursuing a master's degree; Ye Kang, Teaching assistant, Master of Engineering.

In complex environments, the recognition accuracy is difficult to guarantee. *Gnadinger et al. (2017)* used the HSV model to detect ground coverage and combined it with a de-stretching contrast enhancement program to achieve the statistics of corn plant numbers, but it was easily affected by weeds and the growth period of crops. *Varela et al. (2018)* achieved an accuracy rate of 93.0% using super-green segmentation of decision trees, but the computational cost was very high.

With the rapid development of deep learning, the combination of UAVs and object detection algorithms can quickly, efficiently and accurately identify and count seedlings, gradually replacing traditional machine vision methods. Shahid et al. proposed three tobacco counting methods. Among them, the average F1 score of the YOLOv7 combined with the SORT tracking algorithm reached 96.7%, showing potential for real-time application, but its performance declined when there were too many weeds (*Shahid et al., 2024*). Vong et al. estimated the number of corn plants in the V2 period based on the U-Net network. The determination coefficient was the best in the minimum tillage corn-soybean rotation, which was 0.95 (*Vong et al., 2021*). Sun et al. used RC-Dino to detect early corn seedlings, with a recall rate of 0.779 and an average accuracy of 0.714 (*Sun et al., 2025*). Barreto et al. utilized a fully convolutional network (FCN) to achieve corn counting, with a prediction error of less than 4%, but it was susceptible to interference from crop spacing and growth stages (*Barreto et al., 2021*). Zhang et al. proposed a FE-YOLO model based on feature enhancement. The detection mAP of corn seedlings reached 87.22%, but the robustness improvement against weed and environmental noise interference was limited (*Zhang et al., 2021*).

In summary, although corn seedling recognition based on drones and deep learning has the advantages of high accuracy and good real-time performance, corn seedling detection still faces challenges such as small target size, susceptibility to weed and straw background interference, and difficulty in feature extraction. In addition, there is a lack of seedling counting under multiple weather conditions. In response to the above issues, this article proposes the MEI-YOLOv11 corn seedling detection and counting model. The YOLOv11 architecture is enhanced in three aspects: the backbone, the neck, and the loss function. Specifically, the MANet module is integrated into the backbone network, the sampling strategy in the neck network is replaced with the EUCB module, and Inner-IoU is adopted as the loss function. Images of corn seedlings at the V3 growth stage were collected using an unmanned aerial vehicle, and an all-weather dataset was constructed through data augmentation. The result is a model capable of accurate detection and counting of corn seedlings under diverse weather conditions.

MATERIALS AND METHODS

Data Acquisition and Pre-processing

The experimental site of this study was located in the experimental field of Jianshan Farm, Heilongjiang Province, and the previous crop was soybeans. On June 4, 2024, the DJI Mavic 3M Multispectral Edition drone was used to obtain field image information of corn seedlings during the V3 stage. During the shooting, the wind speed did not exceed level 3, the row overlap was set at 80%, the side overlap was set at 70%, the flight altitude was set at 12 meters, and the ground sampling distance was 0.56 pixels/cm. Seedling data were collected along the "S" -shaped flight path. The shooting mode was equidistant fixed-point hovering, and 884 images were collected. The data collection of corn in the experimental area is shown in Fig. 1.

Fig. 1 - Corn data collection diagram in the experimental area

Considering the relatively large size of the collected single image and the dense distribution of seedlings, the pixel area occupied by each seedling in the image is relatively small. Therefore, the original images (5280 × 3956 pixels) were cropped into patches of 640 × 640 pixels. Images that were blurred, overexposed, or distorted were removed. A total of 1,000 images were randomly selected to construct the dataset. The images were annotated using LabelImg in YOLO format, with all corn seedlings labeled as the category “seedling. In order to further enhance the robustness and generalization ability of the model under complex weather conditions, four weather conditions - cloudy, foggy, rainy, and sandstorm - were simulated before training data augmentation. The dataset has been expanded to 5000 images. Randomly divide 5000 image data into training set, validation set, and test set in a ratio of 7:2:1, with corresponding quantities of 3500, 1000, and 500, respectively. The Expansion of Corn Seeding Dataset is shown in Fig. 2.

(a) Sunny (b) Cloudy (c) Foggy (d) Rainy (e) Sandstorm

Fig. 2 - Expansion of Corn Seedling Dataset

Improved YOLOv11 model

Aiming at the problem of poor detection effect of YOLOv11 due to the small volume of corn seedlings, numerous weed interferences and complex straw distribution, this study selects the lightweight YOLOv11n as the benchmark model for improvement. The improved MEI-YOLOv11 model is mainly optimized from three aspects: the backbone network, the neck network and the loss function. The network structure of MEI-YOLOv11 is shown in Fig. 3.

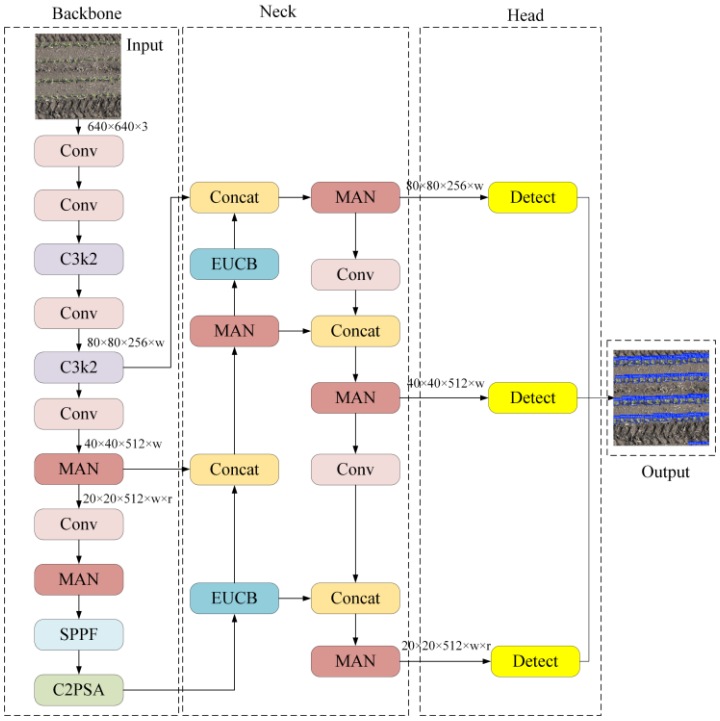

Fig. 3 - MEI-YOLOv11 Network Structure

In the backbone and neck networks, several C3k2 modules were replaced with MANet modules to enhance feature extraction capability. In addition, the sampling strategy in the neck was upgraded to the EUCB module, improving the representation of fine-grained features associated with small seedlings. The Inner-SIoU loss function was employed in place of CIoU, enabling more effective discrimination between seedlings and visually similar weeds by leveraging subtle morphological differences. Together, these modifications substantially improve the detection accuracy of corn seedlings in complex field environments.

MANet module

In the detection of corn seedlings, when the seedlings are in a complex environment formed by the random distribution of straw, the YOLOv11 network may fail to accurately identify the position of the seedlings, thereby resulting in a decrease in detection accuracy. Thus, a hybrid aggregation network (MANet) was proposed (*Feng et al., 2024*).

The network first extracted fundamental features using the YOLO backbone, and then recalibrated the channels through a 1×1 convolution to emphasize key representations. Depthwise separable convolution was introduced to enable efficient multi-scale spatial feature encoding. In addition, the C2f module was integrated to facilitate cross-stage connections and feature feedback, thereby fusing low-level detail with high-level semantic information. These enhancements collectively improved feature richness and the detection accuracy for corn seedlings. The architecture of the MANet module is shown in Fig. 4.

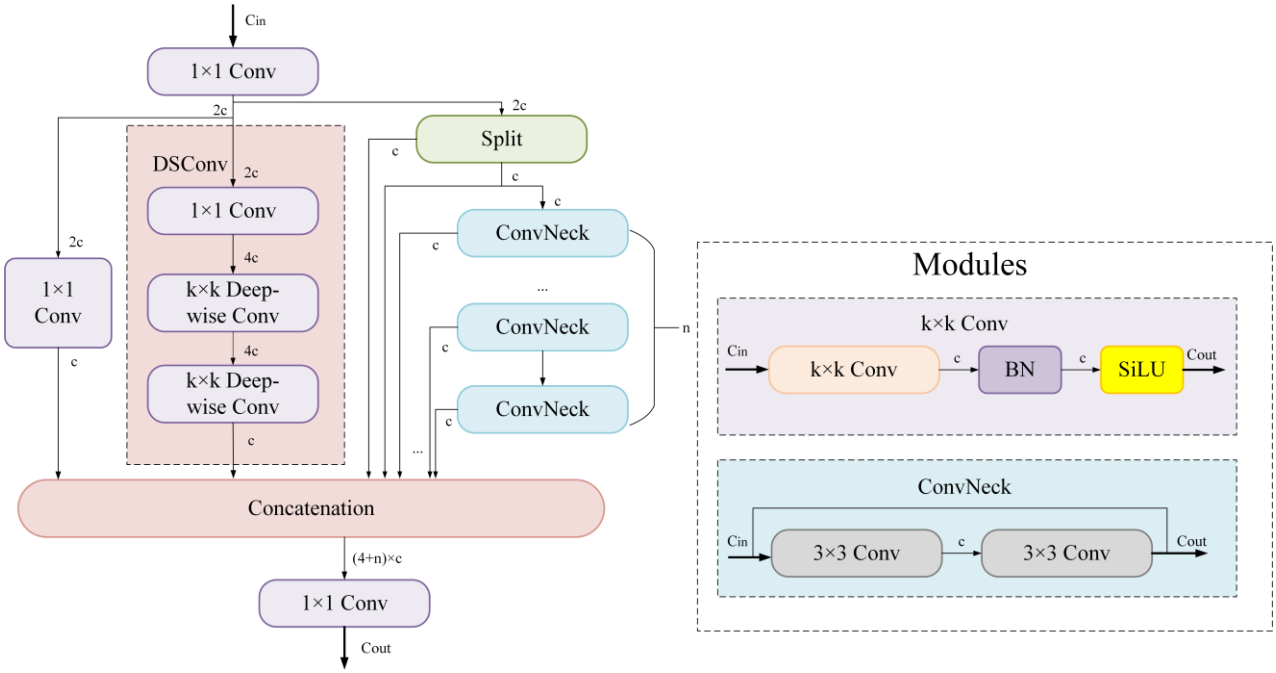

Fig. 4 - MANet module

EUCB module

To preserve the edge detail information of small-volume corn seedlings during upsampling, an efficient upsampling convolutional block (EUCB) was introduced (Rahman et al., 2024). A 2× upsampling operation was performed on the input seedling feature map to enlarge the feature scale, after which the features were processed using depthwise separable convolution (DWC) with batch normalization and ReLU activation. Finally, a 1×1 convolution was applied to adjust the channel dimension so that the upsampled feature map was compatible with the number of channels in the subsequent stages. This approach further enhanced the feature representation of small-sized seedlings while maintaining a relatively low computational cost. The network structure of the EUCB module is shown in Fig. 5.

Fig. 5 - EUCB module

Inner-SIoU loss function

When corn seedlings were located within weed-dense environments, the model exhibited limited self-adaptive regulation and generalization ability. To address this issue, the original CIoU loss was replaced with the Inner-SIoU loss. By incorporating the angular difference between the predicted bounding box and the target box, the bounding box localization was further optimized, thereby improving the detection accuracy of seedlings. The structure of the Inner-SIoU loss function is shown in Fig. 6.

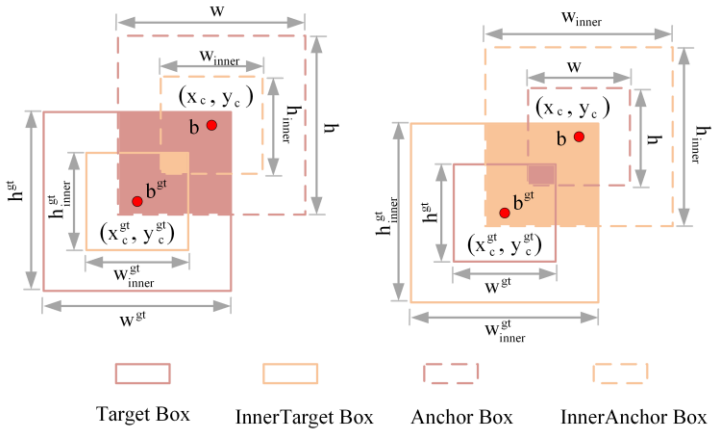

Fig. 6 - Inner-SIOU loss function

The Inner-SIoU loss function introduced a dynamic auxiliary bounding box mechanism based on SIoU *(Zhang et al., 2023)*. For samples with high IoU, smaller auxiliary boxes were employed to achieve fine-grained adjustment, whereas for samples with low IoU, larger auxiliary boxes were used to expand the correction range. This scale-adaptive strategy significantly improved both the regression accuracy and the generalization capability of the detection bounding boxes. The calculation formula is as follows:

$$b_l^{gt} = x_c^{gt} - \frac{w^{gt} \times ratio}{2}, b_r^{gt} = x_c^{gt} + \frac{w^{gt} \times ratio}{2} \quad (1)$$

$$b_t^{gt} = y_c^{gt} - \frac{h^{gt} \times ratio}{2}, b_b^{gt} = y_c^{gt} + \frac{h^{gt} \times ratio}{2} \quad (2)$$

$$b_l = x_c - \frac{w \times ratio}{2}, b_r = x_c + \frac{w \times ratio}{2} \quad (3)$$

$$b_t = y_c - \frac{h \times ratio}{2}, b_b = y_c + \frac{h \times ratio}{2} \quad (4)$$

$$inter = \left(\min\left(b_r^{gt}, b_r\right) - \max\left(b_l^{gt}, b_l\right)\right) \times \left(\min\left(b_b^{gt}, b_b\right) - \max\left(b_t^{gt} - b_t\right)\right) \quad (5)$$

$$union = \left(w^{gt} \times h^{gt}\right) \times \left(ratio\right)^2 + \left(w \times h\right) \times \left(ratio\right)^2 - inter \quad (6)$$

$$IoU^{inner} = \frac{inter}{union} \quad (7)$$

$$L_{Inner-SIoU} = L_{SIoU} + IoU - IoU^{inner} \quad (8)$$

where: b^{gt} represents the target box and b represents the anchor box. The target box and the center point inside the target box are represented by (x_c^{gt}, y_c^{gt}). (x_c, y_c) represents the anchor box and the center point inside the anchor box. The width and height of the target box are represented as w^{gt} and h^{gt}, respectively, while the width and height of the anchor box are represented as w and h, respectively.

Experimental setup

The software configuration for experimental training and testing is based on the deep learning framework PyTorch 2.0.1, developed in Window11 (64 bit), with Python version 3.9 and CUDA version 11.8, respectively. The hardware configuration is Intel (R) Core (TM) i7-14650HX CPU, NVIDIA GeForce RTX 4060 GPU, 16GB of memory. The iteration count is set to 100, the batch size is 16, the learning rate is 0.01, and the optimizer uses stochastic gradient descent algorithm (SGD).

Evaluation Metrics

To evaluate the detection performance of corn seedlings, the following evaluation criteria were selected: precision (P), recall rate (R), F1 score and mean precision (mAP0.5) were used to assess the accuracy of the detection model. The speed and efficiency of the model are evaluated by inference time (Reasoning time), floating-point operation volume (FLOPs), and the number of parameters (Params).

The corn seedling counting model was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and corn seedling counting accuracy (P_t). These indicators were selected based on the research on crop seedling counting by *Wang et al., (2025)*, *Xue et al., (2024)*, *Liu et al., (2021)* etc. R^2 represents the explanatory power of model detection on observed data variation, with a range of 0 to 1.

The closer the R^2 value is to 1, the better the model fitting performance. *RMSE* is a key indicator for evaluating the prediction accuracy of a model, where a lower RMSE indicates higher accuracy and more stable model performance. P_t represents the comparison between the model's seedling count and the actual number of seedlings in the image, and is used to calculate the counting accuracy. The calculation formulas are as follows:

$$R^2 = 1 - \frac{\sum_1^n \left(m_i - p_i\right)^2}{\sum_1^n \left(m_i - \overline{m_i}\right)^2} \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_1^n \left(m_i - p_i\right)^2}{n}} \quad (10)$$

$$P_t = \left(1 - \frac{\left|m_i - p_i\right|}{m_i}\right) \times 100\% \quad (11)$$

where:

m_i is the number of artificially labeled corn seedlings in the i-th image,

$\overline{m_i}$ is the average number of artificially labeled corn seedlings in the i-th image,

p_i is the predicted number of corn seedlings by the model, and n is the number of tested images.

RESULTS

Comparative experiments of different models

To further verify the effectiveness of the improved MEI-YOLOv11 model in corn seedling detection, comparative experiments were conducted against several advanced mainstream object detection models using the same training dataset. The comparison included YOLOv5, YOLOv8, YOLOv9, YOLOv10, YOLOv11, YOLOv12, RT-DETR, and the proposed MEI-YOLOv11 model. The experimental results for the different models are presented in Table 1. A comparison of the comprehensive performance indicators is shown in Fig. 7, where each curve corresponds to one model. The closer the vertex of a curve lies to the outer boundary of the coordinate system, the better the model's performance on that metric. Additionally, a larger enclosed area indicates stronger overall performance.

It can be seen from Table 1 that the mAP0.5, P, R and F1 of the MEI-YOLOv11 model are all superior to other models in detection accuracy, reaching 97.0%, 94.2%, 95.7% and 95.0% respectively. In terms of inference time, it outperforms the YOLOv5 and RTDETR models. In terms of computational load, although it is 2 to 3G and 1 to 2M higher than other YOLO models, it is 91.8 G and 24.57 M lower than the RTDETR model.

Table 1

Comparison test results of different models

Model	P[%]	R[%]	F1[%]	mAP0.5[%]	FLOPs[G]	Params [M]	Reasoning time[ms]
YOLOv5	91.2	93.2	92.0	94.0	**5.8**	2.18	3.2
YOLOv8	91.2	93.5	92.0	94.1	6.8	2.68	2.4
YOLOv9	91.1	93.5	92.0	93.9	6.4	**1.73**	2.7
YOLOv10	88.4	89.1	89.0	92.2	6.5	2.27	**1.7**
YOLOv11	91.5	93.3	92.0	94.2	6.3	2.58	2.5
YOLOv12	91.1	93.3	92.0	94.0	6.3	2.54	2.9
RTDETR	86.4	91.1	89.0	90.3	100.6	28.46	4.9
MEI-YOLOv11	**94.2**	**95.7**	**95.0**	**97.0**	8.8	3.86	2.9

The comprehensive performance comparison between the MEI-YOLOv11 model and other models is shown in Fig. 7. At the cost of a slight increase in computational load, the detection accuracy is significantly improved while maintaining real-time performance. Comprehensive comparison shows that this model has both the best detection performance and real-time processing capability, and thus has been selected as the final model for corn seedling detection and counting.

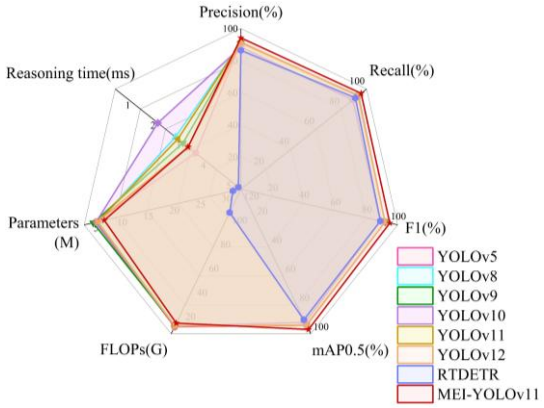

Fig. 7 - Performance Comparison Results of Different Models

Ablation experiments

To verify the effectiveness of each improved module in the MEI-YOLOv11 model in the detection of corn seedlings, ablation experiments were conducted for different improved modules, and the results are shown in Table 2. A module model was randomly added. Through multi-scale feature fusion, the MANet module increased P by 1.2%, R by 0.9%, mAP0.5 by 1.1%, and the number of parameters and reasoning time increased by 1.17 M and 0.9 ms respectively, effectively alleviating the interference of complex backgrounds. The EUCB module enhanced the extraction of detailed features. The mAP0.5 was improved by 0.9%, the number of parameters was increased by 0.09 M, and the inference time was shortened by 0.1 ms simultaneously. The Inner-SIoU loss function increased mAP0.5 by 0.8% without increasing the number of parameters and effectively solved the problem of category confusion between weeds and seedlings. Then, two modules were randomly added, and their model P, recall rate R, and mAP0.5 were all higher than those of adding one module. Finally, three modules were added. The MEI-YOLOv11 model demonstrated a stronger detection effect. Compared with the original YOLOv11, P increased by 2.7 percentage points to 94.2%, R increased by 2.4 percentage points to 95.7%, and mAP0.5 increased by 2.8 percentage points to 97.0%. Only 1.28 M of parameters and 0.4 ms of reasoning time were added, effectively overcoming the interference of weeds and straw and the difficulty in detecting small-sized seedlings, providing technical support for the precise counting of corn seedlings.

Table 2

Results of ablation experiments

Model	MANet	EUCB	Inner-SIoU	P[%]	R[%]	mAP0.5[%]	Params [M]	Reasoning time[ms]
YOLOv11	-	-	-	91.5	93.3	94.2	**2.58**	**2.5**
M-YOLOv11	✓	-	-	92.7	94.2	95.3	3.77	3.4
E-YOLOv11	-	✓	-	92.2	94.0	95.1	2.67	2.4
I-YOLOv11	-	-	✓	92.3	93.9	95.0	2.58	2.4
ME-YOLOv11	✓	✓	-	93.4	94.6	96.1	3.86	2.9
MI-YOLOv11	✓	-	✓	93.2	94.4	95.7	3.77	3.2
EI-YOLOv11	-	✓	✓	92.6	94.6	95.4	2.67	3.1
MEI-YOLOv11	✓	✓	✓	**94.2**	**95.7**	**97.0**	3.86	2.9

Note: ✓ indicates that the module is used; - indicates that the module is not used

Visualization analysis of heat map

Grad-CAM is a method used to interpret the decision-making process of convolutional neural networks (CNNS), which utilizes gradient information to locate the key regions in the image and thereby generate visualized images (Selvaraju et al., 2017). Through the visualization analysis of Grad-CAM, it was found that the attention distribution of the original YOLOv11 model to corn seedlings was relatively scattered under different weather conditions, and there was obvious background interference of straw and weeds. The improved MEI-YOLOv11 model shows a more concentrated and extensive deep red heatmap area, indicating a high level of attention to the corn seedling area in the image. The results are shown in Fig. 8. The model can accurately identify seedlings and has high robustness and adaptability for the detection of corn seedlings in different weather conditions.

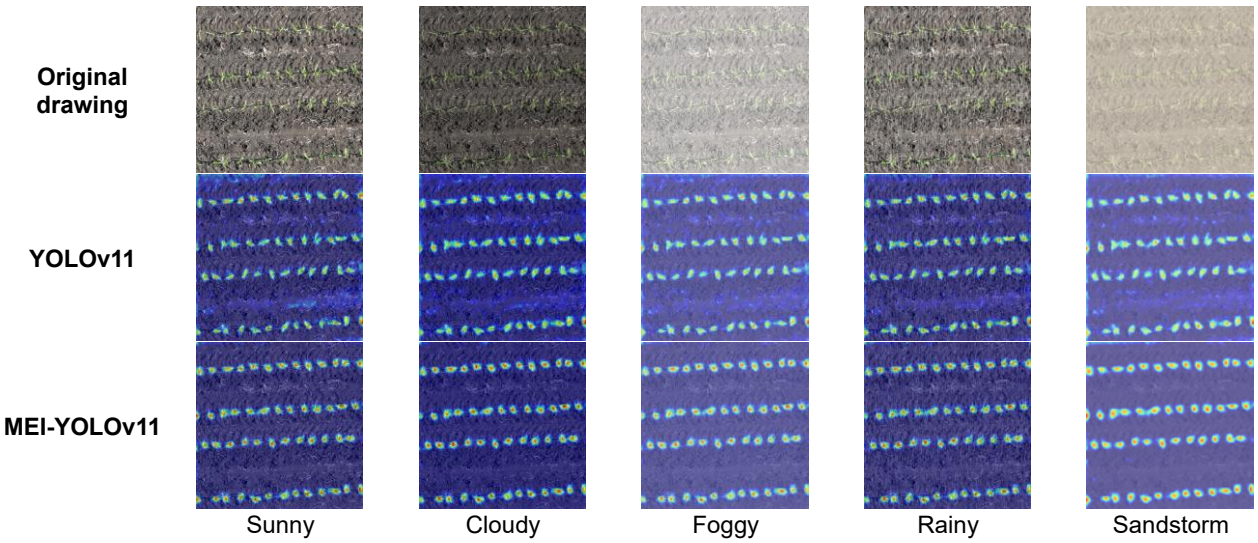

Fig. 8 - Visualization results of YOLOv11 model heatmap before and after improvement

Corn seedling counting experiment

To test the counting accuracy of the MEI-YOLOv11 model for corn seedlings under different weather conditions, the corn seedlings in five weather images were detected. The comparison results of the counting accuracy of the model before and after improvement are shown in Table 3. The multi weather linear fitting results of the model before and after improvement are shown in Fig. 9. The multi weather detection results of the model before and after improvement are shown in Fig. 10.

According to Table 3, the MEI-YOLOv11 model exhibits significant improvement in counting performance under different weather conditions. The highest accuracy of 91.23% was achieved in a sunny environment, which is 2.73 percentage points higher than the model. In cloudy, foggy, rainy, and sandstorm environments, the increase was 2.57, 3.02, 2.57, and 2.59 percentage points, respectively.

Comparison of Counting Accuracy of Models Before and After Improvement **Table 3**

Weather	Accuracy of counting corn seedlings (%)	
	YOLOv11 model	MEI-YOLOv11 model
Sunny	88.50	91.23
Cloudy	88.28	90.85
Foggy	87.95	90.97
Rainy	88.21	90.78
Sandstorm	88.16	90.75

From Figure 9, it can be seen that the MEI-YOLOv11 model outperforms the original model in predicting seedling accuracy and fitting performance under all weather conditions. The root mean square error is the lowest in sunny environments, and the accuracy of predicting corn seedlings is the highest. However, its fitting performance is only better than sandy and dusty environments, lower than the other three environments, but the difference is small, indicating that the MEI-YOLOv11 model is suitable for various weather environments and can meet the precise detection and counting of corn seedlings.

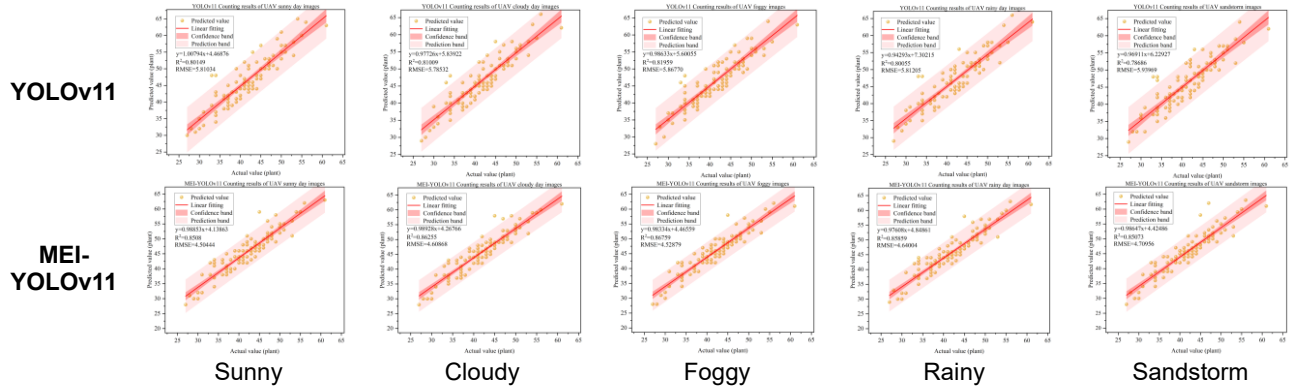

Fig. 9 - Multi weather linear fitting of corn seedling counting before and after improvement model

Fig. 10 visually shows that the MEI-YOLOv11 model performs the best in corn seedling detection and counting tasks, accurately detecting corn seedlings and counting their plant numbers. The detection box is closest to the true value, and the probability of missed or false positives is relatively low. This indicates that the model not only has the best detection accuracy, but also exhibits stronger environmental adaptability and generalization ability, fully verifying the effectiveness of the improved scheme.

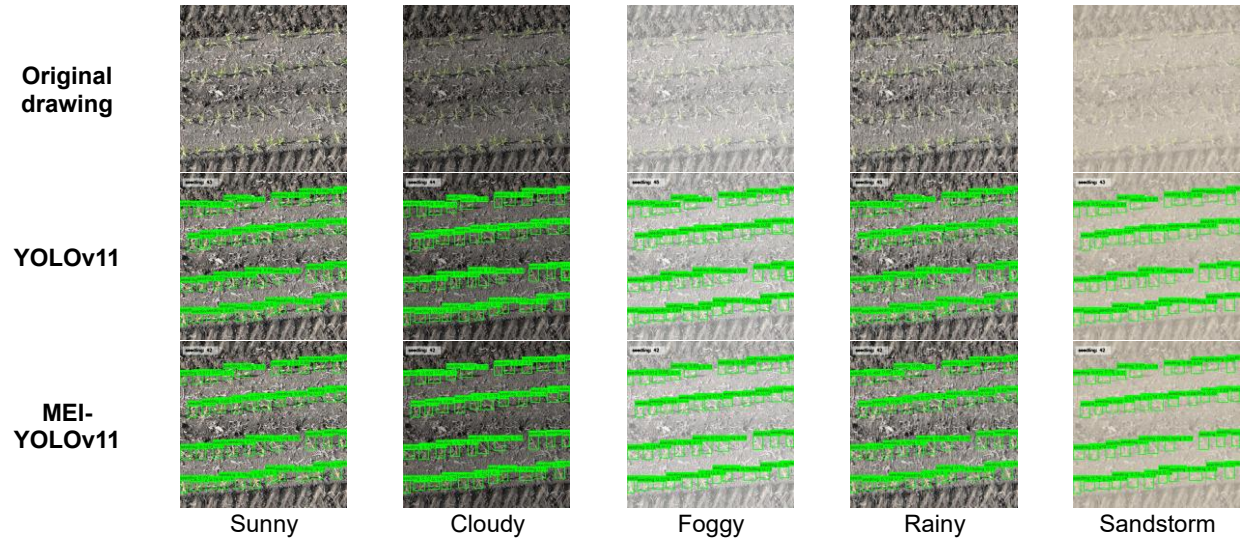

Fig. 10 - Multiple weather detection results before and after improvement of the model

CONCLUSIONS

Aiming at the problems that the detection and counting of traditional corn seedlings are easily affected by the complex field environment and rely on artificial features, a detection and counting method for corn seedlings based on the MEI-YOLOv11 model is proposed.

(1) Images of corn seedlings in sunny conditions were collected by using unmanned aerial vehicles (UAVs), and an all-weather dataset was constructed through data augmentation.

(2) Different detection models were compared, and YOLOv11 with better performance was selected as the benchmark model. To further improve the accuracy of the model's detection count, the MANet module was introduced into the backbone and neck networks. Sampling on the neck network was replaced by the EUCB module, and the Inner-SIoU loss function was replaced by the original CIoU loss.

(3) Compared with the original model, the mAP50, P and R of the MEI-YOLOv11 model have increased by 2.8, 2.7 and 2.4 percentage points respectively. Although the computational load has slightly increased, the difference is not significant, and the shorter reasoning time meets the real-time requirements of the detection. The detection accuracy is superior to other detection models, and the overall performance is better.

(4) Under multiple weather conditions, the counting accuracy of corn seedlings is above 90%. Among them, the counting accuracy on sunny days is the highest, reaching 91.23%, and the root mean square error is the lowest, only 4.5044. The coefficient of determination is 0.8508, which is 0.01679 less than that in foggy environments. The differences are very small, and it can achieve the detection and counting of corn seedlings in all-weather environments, which has broad application prospects and significant practical value.

Overall, the MEI-YOLOv11 model can reduce the missed detection and false detection of corn seedlings in complex field environments such as straw and weeds, and has high accuracy and robustness. It can simultaneously achieve detection and counting in various weather conditions, and has strong practicality. Furthermore, the limitation of the model lies in the fact that the computational load has increased by 1.28 M compared to the original model, and the training data comes from specific regions and corn varieties. In the future, data on corn seedlings of different varieties selected from various regions will be collected for experiments. Lightweight technologies such as knowledge distillation, pruning or quantization will be adopted to reduce computational complexity while ensuring accuracy, making it more suitable for deployment on edge devices with limited computing power.

ACKNOWLEDGEMENT

This study is supported by the doctoral funding project (XDB202501) initiated by Heilongjiang Bayi Agricultural University and the Graduate Innovation Research Project (YJSCX2024-Y38) of Heilongjiang Bayi Agricultural University.

REFERENCES

[1] Barreto, A., Lottes, P., Ispizua Yamati, F.R., Baumgarten, S., Wolf, N.A., Stachniss, C., Mahlein, A.K., & Paulus, S. (2021). Automatic UAV-based counting of seedlings in sugar-beet field and extension to maize and strawberry. *Computers and Electronics in Agriculture*, Vol. 191, pp 106493, Netherlands.

[2] Bryson, M., Reid, A., Ramos, F., & Sukkarieh, S. (2010). Airborne vision-based mapping and classification of large farmland environments. *Journal of Field Robotics*, Vol. 27, pp. 632–655, USA.

[3] Feng, Y., Huang, J., Du, S., Ying, S., Yong, J.H., Li, Y., Ding, G., Ji, R., & Gao, Y. (2024). Hyper-YOLO: When visual object detection meets hypergraph computation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, USA.

[4] Gnädinger, F., & Schmidhalter, U. (2017). Digital counts of maize plants by unmanned aerial vehicles (UAVs). *Remote Sensing*, Vol. 9, pp. 544, Switzerland.

[5] Han, Y.H., Su, H., Yu, Z.Y., Liu, H.J., Guan, H.X., & Kong, F.C. (2021). Estimation of rice leaf area index combining UAV spectrum, texture features and vegetation coverage (结合无人机光谱与纹理特征覆盖度的水稻叶面积指数估算). *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 37, pp. 64–71, Beijing/China.

[6] Jia, H., Wei, G., Zhuo, M., Ming, X., & Le, J. (2015). Methods and experiments of obtaining corn population based on machine vision (基于机器视觉的玉米植株数量获取方法与试验). *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 31, pp. 215–220, Beijing/China.

[7] Liu, H., Liu, T., Li, S.J., Li, L.H., Lv, C.Y., & Liu, S.P. (2021). Research on wheat ear regression counting based on deep residual network (基于深度残差网络的麦穗回归计数方法). *Journal of China Agricultural University*, Vol.26, pp. 170–179, Beijing/China.

[8] Liu, Y.F., She, J.Y., Yuan, Q.F., Zhou, R., & Qi, N.M. (2024). Real-time small target detection networks for UAV remote sensing (无人机遥感图像实时小目标检测方法). *Acta Aeronautica et Astronautica Sinica*, Vol. 45, pp. 59–78, Beijing/China.

[9] Lu, S.L., Li, Y.Y., Li, G., Jia, X.Z., Ju, Q.Q., & Qian, T.T. (2025). Lettuce Phenotype Estimation Using Integrat ed RGB Depth Image Synergy (基于 RGB 与深度图像融合的生菜表型特征估算方法). *Transactions of the Chinese Society of Agricultural Machinery*, Vol. 56, pp. 84–91, Beijing/China.

[10] Ma, S.Q., Wang, Q., Zhang, T.L., Yu, H., Xu, L.P., Ji, L.L. (2014). Response of maize emergence rate and yield to soil water stress in period of seeding emergence and its meteorological assessment in central area of Jilin Province (吉林省中部玉米出苗率和产量对播种-出苗期水分胁迫的反应及其气象评估). *Chinese Journal of Applied Ecology*, Vol. 25, pp. 451–457, Shenyang/China.

[11] Rahman, M.M., Munir, M., & Marculescu, R. (2024). EMCAD: Efficient multi-scale convolutional attention decoding for medical image segmentation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11769–11779, USA.

[12] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626, USA.

[13] Shahid, R., Qureshi, W.S., Khan, U.S., Munir, A., Zeb, A., & Moazzam, S.I. (2024). Aerial imagery-based tobacco plant counting framework for efficient crop emergence estimation. *Computers and Electronics in Agriculture*, Vol. 217, pp. 108557, Netherlands.

[14] Sun, Z., Yang, Z., Ding, Y., Sun, B., Li, S., Guo, Z., & Zhu, L. (2025). Adaptive spatial-channel feature fusion and self-calibrated convolution for early maize seedlings counting in UAV images. *Frontiers in Plant Science*, Vol. 15, pp. 1496801, Switzerland.

[15] Varela, S., Dhodda, P.R., Hsu, W.H., Prasad, P.V., Assefa, Y., Peralta, N.R., Griffin, T., Sharda, A., Ferguson, A., & Ciampitti, I.A. (2018). Early-season stand count determination in corn via integration of imagery from unmanned aerial systems (UAS) and supervised learning techniques. *Remote Sensing*, Vol. 10, pp. 343. Switzerland.

[16] Vong, C.N., Conway, L.S., Zhou, J., Kitchen, N.R., & Sudduth, K.A. (2021). Early corn stand count of different cropping systems using UAV-imagery and deep learning. *Computers and Electronics in Agriculture*, Vol. 186, pp. 106214, Netherlands.

[17] Wang, F., Jiang, H., Wu, J., Li, F., Zhao, B., Mao, W., Cheng, L., Li, M.Z., & Xu, Q. (2025). Efficient detection and counting method for maize seedling plots. *Smart Agricultural Technology*, Vol. 11, pp. 100914, Netherlands.

[18] Xue, X., Niu, W., Huang, J., Kang, Z., Hu, F., Zheng, D., Wu, Z., & Song, H. (2024). TasselNetV2++: A dual-branch network incorporating branch-level transfer learning and multilayer fusion for plant counting. *Computers and Electronics in Agriculture*, Vol. 223, pp. 109103, Netherlands.

[19] Zhang, H., Xu, C., & Zhang, S. (2023). Inner-IoU: More effective intersection over union loss with auxiliary bounding box. *arXiv preprint arXiv*: 2311.02877, USA.

[20] Zhang, H.M., Fu, Z.Y., Han, W.T., Yang, G., Niu, D.D., & Zhou, X.Y. (2021). Detection Method of Maize Seedlings Number Based on Improved YOLO (基于改进 YOLO 的玉米幼苗株数获取方法). *Transactions of the Chinese Society of Agricultural Machinery*, Vol.52, pp. 221–229, Beijing/China.