RESEARCH ON A DETECTION ALGORITHM FOR DRY-DIRECT SEEDED RICE BASED ON YOLOv11N-DF

/

基于 YOLOv11n-DF 的旱直播水稻检测算法研究

Mingyang LI, Bin ZHAO*), **Song WANG, Di WANG**
College of engineering, Heilongjiang Bayi Agricultural University, Daqing/ China
Tel: +86-13589180106; E-mail: limingyangwf@126.com
DOI: https://doi.org/10.35633/inmateh-76-45

Keywords: *YOLOv11n, Automatic detection, Dry direct-seeded rice, DSConv, FASFFHead*

ABSTRACT
Identifying dry-direct seeded rice seedlings provides valuable information for field management. To address the challenges of seedling detection in cold-region dry-direct seeded rice fields, this study proposes an enhanced YOLOv11n-DF model. Key innovations include: 1) integrating DSConv into the C3k2 module to optimize phenotypic feature extraction, and 2) employing the FASFF strategy to improve scale invariance in the convolutional head. Experimental results show that the improved model achieves an mAP50 of 96%, with high recall, precision, and a processing speed of 251.5 FPS, outperforming the original YOLOv11n by 5 percentage points in mAP50, and surpassing YOLOv7–YOLOv10 in detection accuracy. The proposed algorithm effectively addresses challenges such as seedling occlusion and non-uniform distribution, offering a robust solution for automated seedling monitoring in precision agriculture.

摘要
识别旱直播秧苗，可为田间管理提供一定苗情信息。本研究为应对寒地旱直播稻田的幼苗检测难度大的问题，提出改进 YOLOv11n-DF 方法。该方法引入了两个关键创新：1）DSConv 引入 C3k2 模块以优化水稻表型特征提取，2）利用 FASFF 策略优化卷积头增强模型特征尺度不变性。实验结果表明，改进后模型的 mAP50、召回率（R）、检测精度（P）分别为 87.12%、77.07%和 84.11%，处理速度为 49.5 FPS，mAP50 较改进前 YOLOv11n 提高 4.28 个百分点，在检测精度方面超过了同类目标检测网络（YOLOv7~10）。该算法有效地处理了旱直播秧苗相互遮挡和非均匀分布挑战，为精准农业应用中的自动秧苗监测提供了新的解决方案。

INTRODUCTION
Dry-direct seeding is a rice cultivation method that replaces traditional transplanting with direct mechanical seeding. This planting approach - often referred to as dry-direct seeding - offers advantages such as water conservation, labor savings, and greater planting efficiency (*Li et al., 2023*). As a major production area for high-quality rice, Heilongjiang Province adopted dry-direct seeding technology in response to labor shortages and the development of drought-tolerant rice varieties (*Liu et al., 2022*). However, the emergence rate of dry-direct seeded rice is highly sensitive to climatic conditions, particularly in the cold regions of Northeast China, often resulting in uneven or unstable seedling emergence, which impairs rational dense planting (*Li et al., 2024*). Accurate identification and counting of seedlings provides a scientific basis for evaluating planting density, which reflects both sowing quality and varietal performance. Notably, the accuracy of seedling identification directly impacts the reliability of seedling counting (*Liu et al., 2024*).

The traditional rice seedling emergence number counting is determined manually, which is time-consuming, costly and difficult to implement (*Zhu et al., 2024*). In the past decade of research, machine vision technology has been gradually applied to the field of agricultural engineering, which greatly improves the reliability of agricultural information acquisition and provides a new solution idea for the detection of dry-direct seeded rice seedlings. For example, *Lu et al. (2021)* proposed the use of the TasselNetv3 segmentation network to achieve accurate rice plant counting. Similarly, *Takaho et al. (2022)* employed the Super Green Index (SGI) algorithm to predict seedling growth.

Interference from occlusion and background noise can affect image segmentation techniques. In addition, target detection is an important method of acquiring agricultural information. *Yi et al.* (*2022*) used a peak detection-based algorithm to count sunflower and soybean seedlings. *Rojas et al.* (*2022*) used a principal component analysis combined with an SVM detection algorithm to detect weeds. This type of hand-designed algorithm has strong specificity for target detection and better performance on tidy crops with obvious characteristics. Along with the introduction of deep learning methods for target detection, the way of agricultural information acquisition has become more diversified. Velumani et al. proposed the density statistics of maize seedlings by Faster R-CNN. *Shen et al.* (*2024*) utilized the improved YOLOv8 algorithm to achieve the detection of pod pepper. *Qiu D.F.* (*2024*) realized the detection of locust targets by improving the YOLOv7. improvement to achieve locust target detection. The single-stage target detection algorithm in deep learning networks directly realizes the simultaneous regression of target localization and classification through the end-to-end network architecture, and compared with the two-stage target detection algorithm, it discards the candidate region generation, reduces the computational redundancy and maintains a better detection accuracy at the same time.

Currently, there is limited research on target detection in complex scenarios such as dry-direct seeded rice seedlings, which are characterized by high planting density, complex backgrounds, and severe mutual occlusion. For instance, *Huang et al. (2024)* and *Li et al. (2023)* optimized YOLOv5 for the plant counting of oilseed rape seedlings, achieving improved detection accuracy. Although studies targeting this specific type of context-aware object detection remain scarce, such research is crucial for advancing precision agriculture.

In summary, while deep learning-based target detection algorithms demonstrate strong adaptability in agricultural applications, their detection accuracy remains limited when applied to dry-direct seeded seedlings, which often exhibit multi-scale morphological features and densely distributed, non-uniform row patterns. To address this issue, this paper proposes an improved YOLOv11n-based method tailored for the dry strip sowing environment. The approach enhances the feature extraction capabilities of the backbone network and improves the model's ability to capture deep, multi-scale features of densely distributed seedlings, thereby achieving more accurate detection of dry-direct seeded rice plants.

MATERIALS AND METHODS

Data Acquisition and Pre-processing

Image data was acquired on 18 June 2024 at the Jian Sanjiang Qixing Farm in Jiamusi City, Heilongjiang Province, China. Acquisition took place between 08:00 and 11:00 under clear weather conditions. Figure 1 shows the image acquisition device and the acquisition route diagram. The DJI Mavic 3T UAV was used at a height of 3 m and a speed of 1 m/s, in accordance with the planned route, to acquire images at an orthographic angle. The original image was filtered to remove some of the edges of the paddy field images and 640×640 pixel images were randomly extracted from different areas of the image, totaling 1489.

Fig. 1- Image acquisition equipment and range

The rice variety under study is "Magic Rice 119", which was planted according to the pattern of alternating wide and narrow row spacing (300 mm/200 mm). Field management followed the "dry sowing and water pipe" cultivation technique, in which the paddy field is irrigated after the four-leaf stage of rice growth. To accurately capture the morphological characteristics of dry-direct seeded rice seedlings and avoid the influence of water reflections on plant feature extraction, images were collected during the pre-irrigation stage. This standardized image acquisition process provides a reliable data foundation for subsequent rice phenotyping.

(a) Rice field environment (b) Background Straw (c) Post-irrigation environment

Fig. 2- Live streaming of dry rice paddy environment

Data set construction

Adobe Lightroom was used to preprocess 1489 of the images: adjusting contrast, exposure and saturation to enhance the distinction between straw and seedlings, and combining sharpening and noise reduction to highlight the details of the plants and reduce background noise, resulting in a darker background and more distinctive morphological features of the seedlings, which facilitated the algorithm's extraction of plant features. Mixed preprocessing and original images are used to enhance the adaptability of the model, and 90504 "rice" target frames are labeled by Labelimg. Finally, the independent training set, validation set and test set are divided according to the ratio of 8:1:1, which ensures the reliability of model training and testing while guaranteeing the richness of data.

Model selection and construction

The heterogeneity of plant morphology, the narrow, linear structure of rice leaves, and the low contrast between dark green moss and light green leaves on wet soil surfaces make classical target detection algorithms highly susceptible to lighting and background interference, hindering the effective capture of seedling features. Furthermore, due to the nature of strip sowing and variable soil moisture, seedling distribution is often non-uniform, exhibiting both aggregated bands and random sparsity. This complexity requires the detection model to possess both robust feature extraction capabilities and the ability to adapt to irregular spatial distributions.

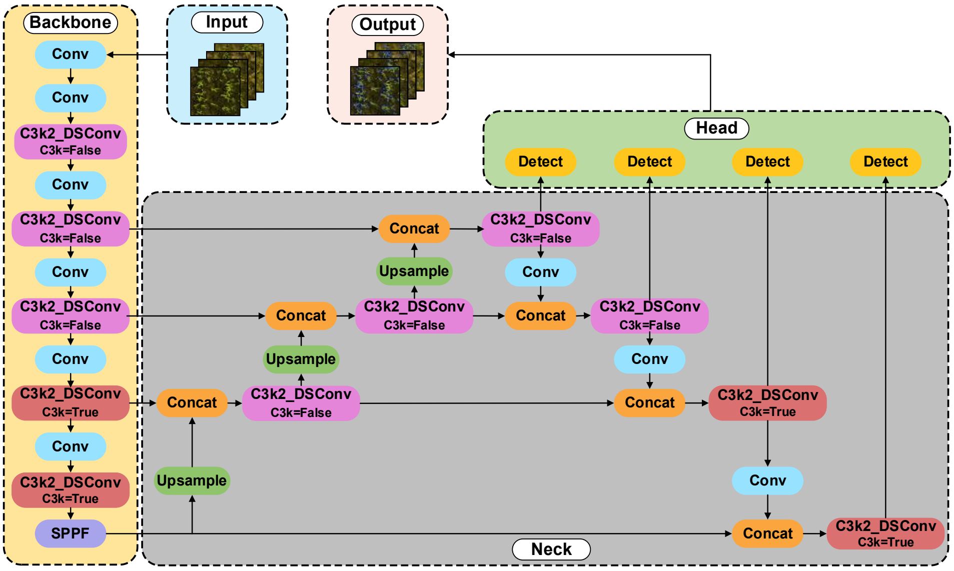


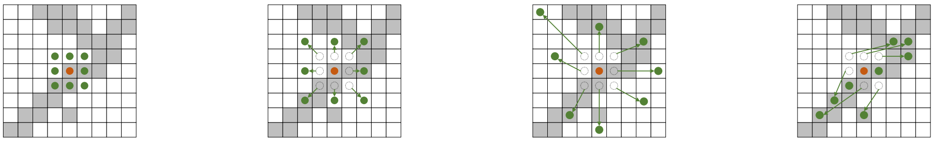
Fig. 3- YOLOv11n-DF model structure diagram

The YOLO model, with the advantage of single-phase detection architecture, exhibits excellent detection performance in the complex environment. Among them, YOLOv11 embeds multi-attention through the C2PSA mechanism, which strengthens the feature fusion ability and can cope with the scale changes and background interference in the large field environment (*He et al., 2025*). The YOLOv11 algorithm provides five different depth models, YOLOv11n, YOLOv11s, YOLOv11m, YOLOv11l, and YOLOv11x, in order to cope with different detection scenarios. It is important to note that the number of floating-point computations and parameters increases with the increase of the model depth. Among them, the model depth of YOLOv11n is the shallowest, which requires the least number of floating-point computational parameters to maintain better accuracy, while the training cycle is shorter than other models (*Wang et al.,2025*). This is consistent with the application scenario of high-efficiency real-time dry-direct seeded rice detection. Therefore, this paper selects the YOLOv11n algorithm for optimizing the detection algorithm for dry-direct seeded rice.

In this study, the C3k2 module in the YOLOv11n network was optimized by integrating DSConv convolution, and the original Detect head was replaced with the FASFFHead to improve the model's consistency in handling seedling phenotypic feature scales. The structure of the improved YOLOv11n-DF algorithm is shown in Fig. 3.

DSConv convolution

The leaves of the dry-direct seeded rice seedlings sown in strips, exhibit distinct narrow and slender morphological characteristics at the four-leaf stage, forming interlocking bands that provide mutual shading. The convolutional layer, a critical component in shallow phenotypic feature extraction, has a direct impact on the model's ability to assess target features. As illustrated in Fig. 4, which compares different convolution strategies, the standard convolutional kernel extracts local features with spatial invariance using a fixed receptive field, based on local connections, weight sharing, and hierarchical abstraction. However, the stacked structure with globally shared weights struggles to adapt to the irregular morphology of seedlings. Therefore, employing structure-adaptive convolution is essential for achieving accurate detection.



(a) Conv (b) Dilated Conv (c) Deformable Conv (d) DSConv

Fig. 4 - Comparison of different convolution strategies

Dynamic serpentine convolution focuses on slender and meandering local structures by adaptively adjusting the offsets of sampling paths, which enhances the model's ability to capture features of elongated structures (*Qi et al., 2023*). These characteristics make DSConv significantly more effective than the rigid grid sampling used in traditional convolutional kernels for extracting features from non-rigid and non-linear structures in complex scenes (*Zhang et al., 2024*). To address the challenge of capturing the narrow, elongated structures of dry-direct seeded rice seedlings, this study introduces DSConv into the C3k2 module. This integration takes into account the distributional characteristics of dry-direct seeded seedlings and enhances the geometric perception of convolutions by dynamically adjusting the sampling paths of convolution kernels, allowing them to continuously extend along thin, curved leaf structures. A lightweight subnetwork is used to predict incremental offsets at each sampling point, with each point's position determined by the cumulative offset from the previous point. This design ensures a smooth and continuous sampling path and prevents perceptual drift - a common issue in traditional deformable convolution caused by independent offset predictions.

The incremental offset of each sampling point is calculated as follows:

$$\delta \mathrm{p}_k = f_{\mathrm{offset}}(\mathrm{X};\mathrm{p}_0,\mathrm{p}_k) \tag{1}$$

where f_{offset} is the offset prediction network; X is the input feature map; p_0 denotes the coordinates of the center position of the current convolution kernel; p_k represents the coordinates of the kth sampling point; along the x-axis.

Since the offset values may be floating-point numbers, sampling at non-integer coordinates becomes necessary. To address this, bilinear interpolation is employed to compute the pixel values at these non-integer positions, ensuring smooth and continuous feature extraction. The interpolation is performed using the following formula:

$$\mathrm{X}(\mathrm{p}_0+\mathrm{p}_k+\Delta\mathrm{p}_k)=\sum B(q,\mathrm{p}_0+\mathrm{p}_k+\Delta\mathrm{p}_k)\cdot\mathrm{X}(q) \tag{2}$$

where B is a bilinear interpolation kernel decomposed into one-dimensional interpolations along the x and y directions; $\mathrm{p}_k+\Delta\mathrm{p}_k$ is the dynamically adjusted coordinate of the kth sampling point; q refers to the four surrounding integer-coordinate neighbors.

During the image acquisition process, the morphology of field-grown seedlings exhibits significant heterogeneity. To enhance the model's generalization capability for seedling detection, a multi-feature fusion approach is proposed. This involves extracting feature maps $f^l(K_x)$, $f^l(K_y)$ along the x-axis and y-axis, respectively, at the lth layer, using multi-morphology convolutional kernel templates generated by DSConv. The goal is to capture the shallow morphological features of the seedlings and to efficiently cluster key features, thereby improving detection robustness across varying seedling shapes.

During the target detection process, the model often fails to capture the shallow structural features of seedlings due to mutual occlusion among the leaves of dry-direct seeded rice. To reduce detection inaccuracies caused by disrupted seedling targets and to preserve the overall linear structure of the seedlings as much as possible, a topological continuity constraint loss is introduced into DSConv. This is achieved by calculating the bi-directional Hausdorff distance between the persistence maps of the predicted and ground truth labels, as defined by the following equation:

$$d_H^*(P_O,P_L)=\max(\max_{u\in P_O}\min_{v\in P_L}\|u-v\|,\max_{v\in P_L}\min_{u\in P_O}\|v-u\|) \quad (3)$$

where P_O is the persistence map of the prediction, recording the birth and death times of topological features; P_L is the persistence map of the ground truth labels used for comparison with P_O; d^*_H represents the bi-directional Hausdorff distance, which quantitatively measures the topological discrepancy between P_O and P_L.

Serpentine convolution has been proven effective in extracting fine and narrow structural features of seedling leaves. Additionally, the C3k2 module has been optimized to enhance the model's perception of seedling features through four key mechanisms: chained accumulation of dynamic offsets, bilinear interpolation of fractional coordinates, a multi-view feature fusion strategy and a topological continuity constraint loss.

Multi-scale detection head

Seedling scale diversity in field environments leads to inconsistent seedling features in feature maps, affecting single-stage target detectors. The paper proposes a FASFFHead for the head network as a four-head detection mechanism based on an adaptive spatial feature fusion design to maintain feature scale consistency in the seedling detection model (*Liu et al., 2019*).

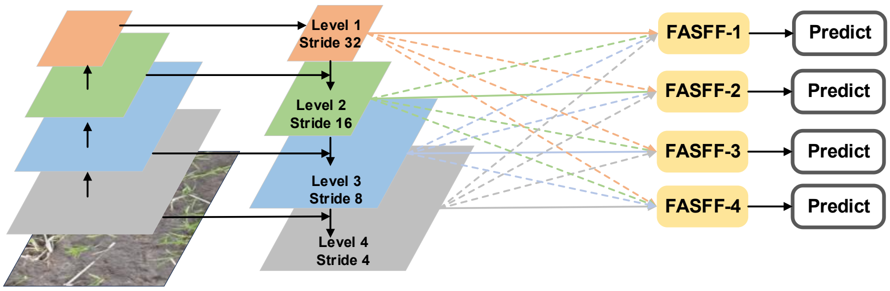


Fig. 5 - Schematic diagram of FASFF

The FASFFHead model adopts an improved feature fusion mechanism based on ASFF and introduces an additional small-scale target detection layer to the original 3-layer ASFF architecture, resulting in a 4-layer FASFF feature fusion framework (*Li et al., 2023*). This enhancement improves the algorithm's ability to capture and integrate features of small seedlings, enabling cross-scale information fusion through dynamic optimization of multi-layer feature strategies. As shown in Fig. 5, the working principle diagram of the FASFF mechanism illustrates how Level 1 through Level 4 (represented by different colors) correspond to varying spatial resolutions in the feature pyramid, each encoding seedling features at different scales. The labels FASFF-1, FASFF-2, FASFF-3, and FASFF-4 denote the fusion operations applied at each respective level, enabling layer-wise feature fusion across different resolutions.

To balance seedling feature representation, information retention, and computational efficiency, the FASFF method introduces a unified hierarchical structure for feature fusion. The core mechanism, Feature Aggregation and Selection (FAS), aims to standardize the resolution of feature maps from different hierarchical levels. For low-resolution seedling features, the number of channels is first compressed to match the target hierarchy using 1×1 convolution, followed by bilinear interpolation for up-sampling, thereby enhancing spatial resolution. In contrast, for high-resolution features, the resolution is first reduced via max pooling, and then 3×3 convolution is applied to achieve effective down-sampling.

Building on this foundation, a learnable spatial attention weight matrix is constructed. Through a parameterized weight learning module, the model dynamically learns the spatial contribution of features at each level, enabling adaptive aggregation of shallow features across multiple layers in the spatial dimension. This mechanism also preserves the local responses of features sensitive to different target scales, while effectively suppressing cross-level semantic conflicts and filtering out interfering information during feature fusion.

To address the scale inconsistency inherent in single-stage detection algorithms, FASFF enhances scale invariance by introducing an adaptive spatial feature fusion strategy. This approach improves the consistency across different feature scales by learning to suppress conflicting features. As a result, the model can flexibly determine which feature layers contribute most significantly to final predictions, depending on the contextual information at each specific location and scale.

RESULTS

Experimental Environment

The hardware equipment platform in this experiment includes: CPU Intel Core I7-13700KF, RAM 32GB, GPU Nvidia GeForce RTX 4070TI Super 16G; the software platform includes: 64-bit Windows 10 operating system and deployed Cuda11.8, Cudnn11.x with Python 3.11, using Pytorch 2.0.1 depth learning framework, PyCharm2023.3.5 editor for network model training. The relevant hyper-parameters were set: the number of epochs was 400, the batch size was 16, the optimizer was selected as SGD, and the initial learning rate was set to 0.02 in order to improve the convergence rate.

Evaluation index

In order to verify the applicability of the improved model for dry-direct seeded rice seedling detection, this study selects key performance indicators for multi-dimensional evaluation: in terms of detection accuracy, Precision - *P,* Recall - *R* and mAP50 are used as the core indicators; for the model complexity and model volume, the floating point operations (GFLOPs) and model Size are selected for quantitative analysis, and the detection speed of the model is evaluated using the detection speed (FPS). By comparing the experiments with mainstream target detection algorithms, the proposed method is comprehensively verified to see if the enhancement effect of the proposed method in model accuracy meets the expectation.

Model Performance Analysis

The training set of dry-direct seeded rice seedlings was used to train the improved model. The results showed that the YOLOv11n-DF model achieved an mAP50 of 87.12%, precision (*P*) of 84.11%, and recall (*R*) of 77.07%. Compared with the pre-improvement baseline, the mean average precision (mAP50) increased by 4.28 percentage points. Additionally, the model size remained compact at 12.7 MB, supporting efficient deployment.

The loss function curve, shown in Fig. 6, illustrates that both training loss and validation loss decrease rapidly and in synchronization during the first 50 epochs. After approximately 300 epochs, both curves converge in parallel, indicating that the model has reached a Pareto optimal state. Throughout the training process, the two models exhibit similar trends, and no significant fluctuations are observed in the validation loss. The final convergence difference between training and validation loss remains below 2%. The YOLOv11n-DF model, designed specifically for dry-direct seeding detection, does not exhibit overfitting or underfitting. Furthermore, it demonstrates strong memory retention and excellent generalization performance.

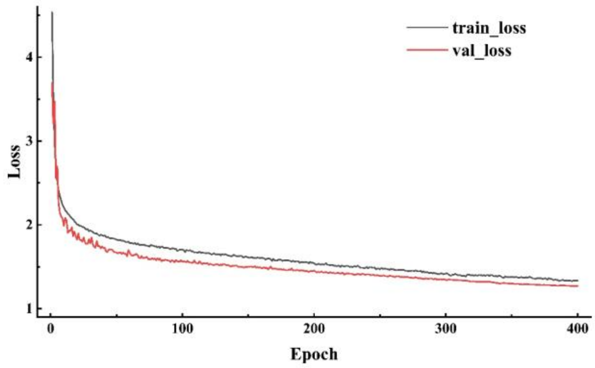


Fig. 6 - Loss function curve

As shown in Fig. 7, the detection results for both pre-processed and unpreprocessed images across different scenes are compared before and after the model improvement. The improved YOLOv11n-DF algorithm demonstrates a significantly enhanced ability to accurately detect dry-direct seeded rice seedlings in both types of images. Compared to the pre-improvement version, the improved model can more effectively identify individual seedlings, reduce redundant detections, and accurately focus on small targets - particularly in sparsely populated regions. These results highlight the enhanced robustness and adaptability of the improved algorithm under various image conditions.

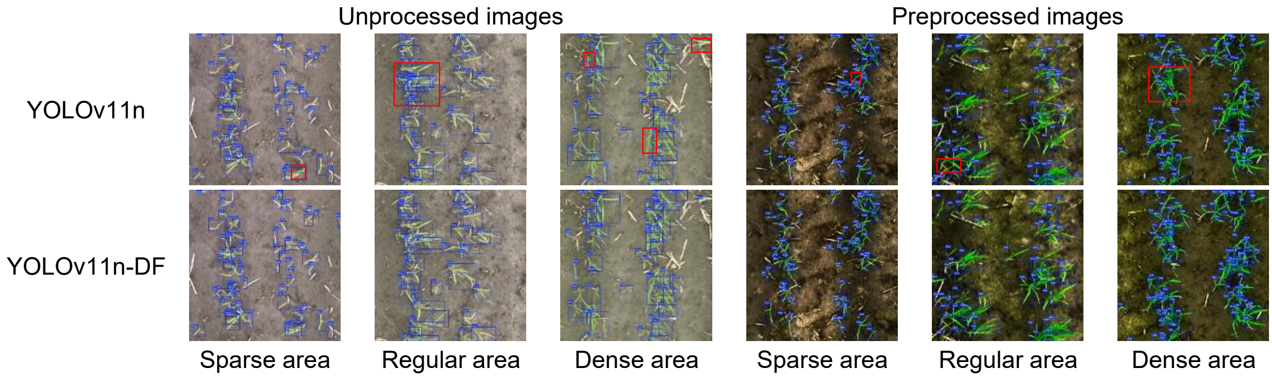


Fig. 7 - Sample images for image detection in different scenarios

Ablation experiments

To evaluate the effectiveness of the model improvements, ablation experiments were conducted based on the YOLOv11n baseline. The results demonstrated that both incorporating the DSConv mechanism into the C3k2 module and replacing the original detection head with the FASFFHead independently enhanced the model's ability to detect dry-direct seeded rice seedlings. In particular, the FASFFHead yielded a notable improvement in mAP50, although the addition of multiple detection heads also led to a corresponding increase in GFLOPs. Further module combination experiments confirmed the good compatibility of the improved components. Integrating both DSConv and FASFFHead significantly increased GFLOPs due to the multi-head structure, but this trade-off was justified by the sustained accuracy gains. The experimental findings indicate that the proposed improvements form a two-tier optimization architecture: DSConv enhances the backbone's ability to extract linear structural and distributional features of seedlings, while the four-head FASFFHead improves multi-scale detection performance. The final YOLOv11n-DF model achieves an mAP50 of 87.12%, representing a 4.28% increase over the baseline, with only a 13% increase in computational cost, thus meeting the expected performance improvements.

Table 1

Results of ablation experiment

YOLOv11n	DSConv	FASFFHead	mAP	Precious	Recall	FPS	GFLOPs	Size (M)
√	-	-	82.84%	80.62%	74.35%	**273.4**	**6.3**	**5.3**
√	√	-	84.62%	82.55%	74.63%	168.7	6.5	5.6
√	-	√	85.13%	83.20%	75.10%	154.7	7.1	11.8
√	√	√	**87.12%**	**84.11%**	**77.07%**	49.5	7.3	12.7

Note: The bolded data in the table represents the maximum value of the same group

Comparison of detection results across models

To more clearly highlight the performance differences in dry-direct seeded rice seedling detection, a comparative experiment was conducted between the improved YOLOv11n-DF model and several mainstream YOLO-based architectures. These comparisons provide a comprehensive evaluation of the model's effectiveness in handling the challenges of dry-direct seeded rice seedling detection. The detailed performance metrics and detection results are presented in Table 2.

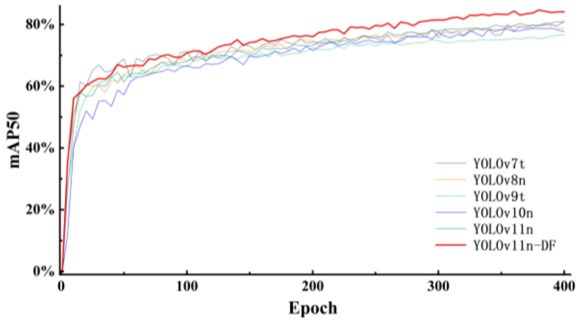
Table 2

Comparison of detection results by model

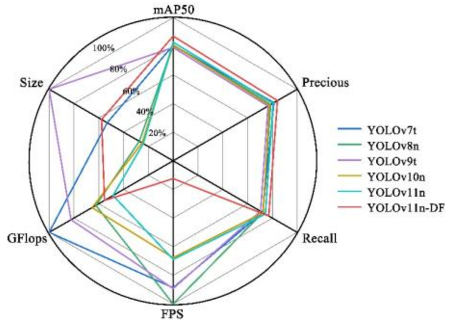
Models	mAP	Precious	Recall	FPS	GFLOPs	Size (M)
YOLOv7t	80.70%	81.25%	72.13%	351.6	13	11.7
YOLOv8n	80.91%	78.61%	72.22%	**396.8**	8.1	6
YOLOv9t	79.17%	76.49%	70.42%	354.3	10.7	22
YOLOv10n	80.83%	77.68%	72.53%	268.6	8.5	5.5
YOLOv11n	82.84%	80.62%	74.35%	273.4	**6.3**	**5.3**
YOLOv11n-DF	**87.12%**	**84.11%**	**77.07%**	49.5	7.2	12.7

Note: The bolded data in the table represents the maximum value of the same group

Fig. 8(a) shows the change curves of the mAP50 of different models with the number of training rounds. As the number of training rounds increases, the mAP50 of each model shows a general growth trend, and all models achieve basic convergence after about 300 epochs, with only small fluctuations. Among them, the pre-improved YOLOv11n model has a slower curve growth after 200 epochs, but still has some growth, and shows better learning ability in the detection of dry-direct seeding of rice seedlings. The improved YOLOv11n-DF grows second after the YOLOv7t model in the first 100 epochs, and the mAP does not fluctuate significantly in the training epoch, and still maintains a steady growth after 100 epochs, compared with the YOLOv11t model. After 300 epochs, the mAP50 stabilizes without significant growth, reaching a basic convergence.



(a) mAP50 comparison curve chart



(b) Model performance radar chart

Fig. 8 - Comparative experimental results

To visually compare the performance differences among the models, radar charts based on normalized results were generated, as shown in Fig. 8(b). The improved YOLOv11n-DS network demonstrates clear advantages in terms of mAP50, *P*, and *R*, aligning with the intended improvement objectives. These results confirm that the YOLOv11n-DS network, which integrates enhanced feature extraction in the backbone and a multi-head detection strategy, outperforms similar object detection models in the task of dry-direct seeded rice seedling detection.

Visual Analysis

The visual analysis of the model can break through the limitations of the “black-box” model in the deep network. Heat map localization analysis can quantify the intensity with which the model focuses on the target region, verify whether the improved model suppresses noise and strengthens the target region of seedling plants, and assess the interpretability of the improved YOLOv11-DF model for seedling detection.

As shown in Fig. 9, Grad-CAM was used to visualize and analyze six detection scenarios. After integrating DSConv into the C3k2 module of the backbone network, the model exhibited a stronger focus on the banded distribution of dry-direct seeded rice seedlings, particularly in denser regions, compared to more sparse or scattered areas. The visualization shows that the model can effectively identify target regions in both preprocessed and original images, demonstrating enhanced generalization ability. Furthermore, the model’s attention is more concentrated on the seedling distribution in preprocessed images, suggesting that image preprocessing aids in the extraction of phenotypic features. Overall, the improved C3k2 module with DSConv significantly enhances the network’s ability to capture the structural and morphological characteristics of dry-direct seeded rice seedlings, especially those arranged in regular banded patterns.

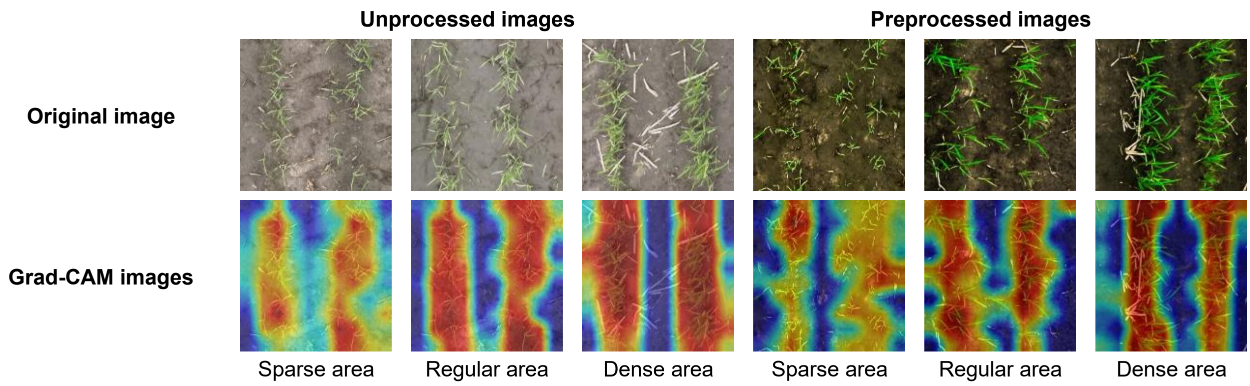


Fig. 9 - Grad-CAM visualization of detection scenarios

CONCLUSIONS

This study proposes an individual seedling detection algorithm based on an enhanced YOLOv11n network, aimed at addressing the challenges posed by the heterogeneous morphology and uneven distribution of dry-direct seeded rice. The following section outlines the results of this method:

(1) Use YOLOv11n to create a target algorithm that captures the visual traits of dry-direct seeded rice. The DSConv depthwise separable convolutional optimization C3k2 module strengthens the linear extraction of seedlings' traits, while a FASFF-based adaptive feature fusion mechanism improves the detection head's structure to better capture the multi-scale traits of seedlings.

(2) Use comparison and ablation experiments to verify the effectiveness of model improvement. The YOLOv11n-DF model's accuracy and recall reached 84.11% and 77.07%, respectively. Its mAP50 was 87.12%, showing better detection performance than the original model and mainstream algorithms. This model accurately detected multiscale seedlings in complex field environments.

(3) The interpretability of the model improvement is analyzed using visualization tools. DSConv improves the detection of linear phenotypic features of seedlings by the C3k2 module and focuses on seedlings with banded distributions. The four-detector-head architecture of FASFFHead mitigates the inconsistency in the scale of the fused multiscale features through dynamic weight allocation. The model improvement meets the design expectation, and YOLOv11n-DF has better generalization and interpretation.

(4) In summary, the improved model can accurately detect dry-direct seeded rice seedlings. Subsequent research could reduce the model's complexity using a pruning technique to meet video detection requirements in field management.

ACKNOWLEDGEMENT

This research, entitled "Research on a detection algorithm for dry-direct seeded rice based on YOLOv11n-DF", was funded by the National Natural Science Foundation of China Youth Fund Program (32201655).

REFERENCES

[1] He, L. H., Zhou, Y. Z., Liu, L., Cao, W., Ma, J. H. (2025). Research on object detection and recognition in remote sensing images based on YOLOv11. *Scientific Reports*, Vol. 15, pp. 14032, London/UK

[2] Huang, X., Zhang, W., Qiu, T., Zhu, Y., Xu, S., & Li, W. (2024). Rapeseed seedling detection and counting based on UAV videos (基于无人机视频影像的油菜苗检测与计数). *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 40, pp. 147-156, Beijing/China.

[3] Li, C.P. (2025). *Research on Seedling Rape Plant Counting Algorithm Based on Improved Neural Network (基于改进卷积神经网络的苗期油菜植株计数算法研究)*, MSc dissertation, Anhui Agricultural University, Hefei/China.

[4] Li, M., Xue, M., Sha, Y., Li, H., Chen, Y., (2024). Synergistic and Optimal Allocation of Water, Land and Fertilizer Resources of Rice in Heilongjiang Province Based on Meta-analysis (基于 Meta 分析的黑龙江省水稻水土肥资源协同优化调配). *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 55, pp. 302-311, Beijing/China.

[5] Li, Y., Xue, J., Zhang, M., Yin, J., Liu, Y., Qiao, X., Li, Z. (2023). YOLOv5-ASFF: a multistage strawberry detection algorithm based on improved YOLOv5. *Agronomy*, Vol. 13, pp. 1901. Basel/Switzerland

[6] Liang, C., Li, Y., Zhang, K., Wu, Z., Liu, J., Liu, J., Zhou, C. Wang, S., Li, F., Sui, G. (2023). Selection and yield formation characteristics of dry direct seeding rice in Northeast China. *Plants*, Vol. 12, pp.3496, Basel/Switzerland.

[7] Liu, S., Huang, D., Wang, Y. (2019). Learning spatial fusion for single-shot object detection. *arxiv preprint* arxiv:1911.09516, USA.

[8] Liu, Y., He, C., Gai, D., Geng, Y., Guo, L., Shao, X. (2022). Morphological and physiological traits of roots and their relationships with shoot growth and grain yield in direct-seeded rice in northeastern China. *Crop and Pasture Science*, Vol. 73, pp. 1229-1244. Australian.

[9] Lu, H., Liu, L., Li, Y. N., Zhao, X. M., Wang, X. Q., & Cao, Z. G. (2021). TasselNetV3: Explainable plant counting with guided upsampling and background suppression. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, pp.1-15, USA.

[10] Qiu, D. (2024). Research on locust target detection algorithm based on YOLO v7-Mobilenetv3-CA. *INMATEH-Agricultural Engineering*, Vol. 74, pp.283-292, Bucharest/Romania.

[11] Lu, H., Liu, L., Li, Y. N., Zhao, X. M., Wang, X. Q., & Cao, Z. G. (2021). TasselNetV3: Explainable plant counting with guided upsampling and background suppression. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, pp. 1-15, USA.

[12] Qi, Y., He, Y., Qi, X., Zhang, Y., & Yang, G. (2023). Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. *In Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6070-6079, USA.

[13] Rojas, C. P., Guzmán, L. S., Toledo, N. V. (2017). Weed recognition by SVM texture feature classification in outdoor vegetable crops images. *Ingeniería e Investigación*, Vol. 37, pp. 68-74, Bogota/Spain.

[14] Shen, J., Jiang, Q., Liu, Y., Ma, N., (2024). Pod pepper target detection based on improved YOLOv8. *INMATEH-Agricultural Engineering*, Vol. 74, pp. 273-282, Bucharest/Romania.

[15] Ueno, T., Nagano, S., Moriyuki, S., Hayashi, T., Fukuda, H. (2022). Optimized Excess-Green Image Binarization for Accurate Estimation of Lettuce Seedling Leaf-Area in a Plant Factory. *Environmental Control in Biology*, Vol. 60, pp. 153-159, Japan.

[16] Velumani, K., Lopez-Lozano, R., Madec, S., Guo, W., Gillet, J., Comar, A., Baret, F. (2021). Estimates of maize plant density from UAV RGB images using faster-RCNN detection model: Impact of the spatial resolution. *Plant Phenomics*, pp. 9824843, USA.

[17] Wang, C., Song, X., Wang, J., Yan, X. (2025). An improved YOLOv11 algorithm for small object detection in UAV images. *Signal, Image and Video Processing*, Vol. 19, pp. 1-12, Germany.

[18] Bai, Y., Nie, C., Wang, H., Cheng, M., Liu, S., Yu, X., Shao, M., Wang, Z., Wang, S., Tuohuti, N., Shi, L., Ming, B., Jin, X. (2022). A fast and robust method for plant count in sunflower and maize at different seedling stages using high-resolution UAV RGB imagery. *Precision Agriculture*, Vol. 23, pp. 1720-1742, Germany.

[19] Zhang, J., Zhang, D., Liu, J., Zhou, Y., Cui, X., Fan, X. (2024). DSCONV-GAN: a UAV-BASED model for Verticillium Wilt disease detection in Chinese cabbage in complex growing environments. *Plant Methods*, Vol. 20, pp. 186, UK.

[20] Zhu, W., Dai, S., Feng Z., Duan, K., Shao, C. Wei, X. (2024). Estimation of Rice Basic Seedling Number Based on Mixed Pixel (基于混合像元分解的分蘖期水稻基本苗数量估测方法研究) *Transactions of the Chinese Society of Agricultural Machinery*, Vol. 55, pp. 202-209, China.