A RICE GROWTH STAGE IDENTIFICATION MODEL BASED ON AN OPTIMIZED YOLOV12n ARCHITECTURE

/

基于改进 YOLOv12n 的水稻生育期识别模型

Sen LI, Sheng-xue ZHAO*), Heng ZHANG
College of Engineering, Heilongjiang Bayi Agricultural University, Daqing/P.R.China
Tel: +86-15945585971; E-mail: 15945585971@163.com
Corresponding author: Sheng-xue ZHAO

DOI: https://doi.org/10.35633/inmateh-78-41

Keywords: *Deep learning; Object detection; Developmental stages; YOLOv12n; Cold region agriculture*

ABSTRACT

Accurately identifying rice growth stages in cold regions is challenging due to subtle morphological differences between key stages (e.g., tillering, jointing-booting) and complex field conditions, often leading to inefficient water use. This paper proposes an improved YOLOv12n-based method for rice growth stage recognition. First, a Deformable Large Kernel Attention (D-LKA) mechanism is introduced to enhance the model's ability to capture multi-scale morphological features of rice plants through large receptive field convolution and adaptive sampling grids. Second, a weighted Bidirectional Feature Pyramid Network (BiFPN) and a Dynamic Upsampling module (DySample) are employed to construct an efficient multi-scale feature interaction pathway, improving the model's perception of details in key plant parts. Finally, the Normalized Wasserstein Distance (NWD) is adopted to optimize the small-target detection strategy, effectively mitigating the under-detection of small-scale features. A cold-region rice growth stage image dataset was constructed for training and evaluation. Results show that the proposed YOLOv12n-DBD model achieves a precision of 93.4%, a recall of 89.7%, a mean Average Precision (mAP@0.5) of 94.4%, and an inference speed of 121.9 FPS. The mAP@0.5 represents an improvement of 4.9 percentage points over the baseline model, outperforming current mainstream detection models while maintaining real-time performance. A mobile recognition system was also developed to provide a convenient solution. The proposed YOLOv12n-DBD model effectively balances recognition accuracy and computational efficiency in the complex environments of cold-region rice fields, offering reliable technical support for growth-stage-specific field management.

摘要

针对寒地水稻关键生育时期（如分蘖期、拔节孕穗期）形态特征存在细微差异、田间环境复杂导致识别难度大进而造成的水资源浪费问题。本文提出基于改进 YOLOv12n 模型的水稻生育时期识别方法。首先，引入可变形大核注意力机制（D-LKA），通过大感受野卷积与自适应采样网格增强模型对水稻植株多尺度形态特征的捕获能力；其次，采用加权双向特征金字塔网络（BiFPN）与动态上采样模块（DySample），构建高效的多尺度特征交互通路，提升模型对水稻关键部位的细节感知能力；最后，通过归一化 Wasserstein 距离（NWD）优化小目标检测策略，有效缓解小尺度特征的漏检问题。构建寒地水稻生育期图像数据集，对改进后的模型进行训练与验证。结果表明：YOLOv12n-DBD 模型的准确率（P）、召回率（R）、平均精度均值（mAP）和推理速度分别为 93.4%、89.7%、94.4%和 121.9fps，其中 mAP@0.5 较基线模型提升 4.9 个百分点。在保持实时性的同时优于当前主流检测模型。构建移动端识别系统，可提供便捷识别方案。本文提出的 YOLOv12n-DBD 模型有效平衡了寒地水稻复杂环境下的识别精度与计算效率，能够为寒地水稻的生育期田间管理提供技术支持。

INTRODUCTION

The cultivation of rice in cold regions currently faces numerous challenges, including water scarcity, complex climatic conditions, and a short growing cycle. Rice water demand varies significantly across different growth stages (such as tillering, heading, and flowering). Therefore, accurately identifying the growth stages of rice to guide irrigation strategies is crucial for improving yield and water use efficiency in these regions *(Tang., 2022; Bai et al., 2022; Xue et al., 2024; Li et al., 2024)*.

Sen LI, M.S. Stud.; Sheng-Xue ZHAO, Prof. Ph.D.; Heng ZHANG, M.S. Stud.

With the rapid advancement of technologies such as image recognition and deep learning, object detection has become an important tool in agriculture by enabling plant phenotyping *(Wang et al., 2024; Xu., 2024; Oraif., 2023)*.

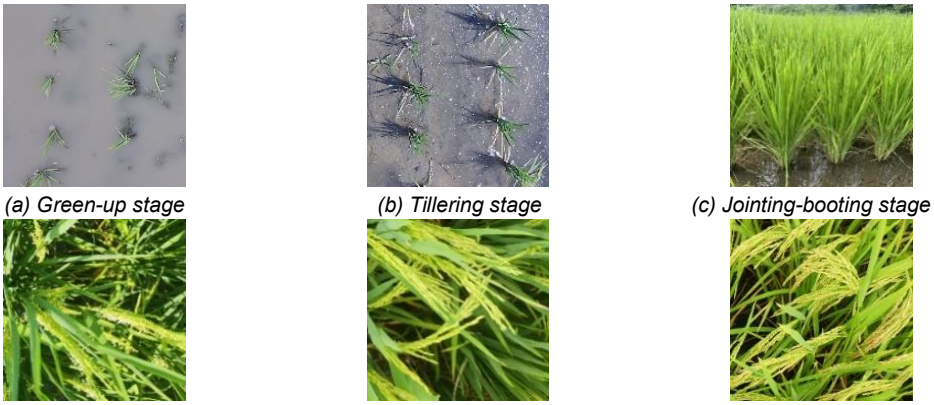
For instance, *Fan et al., (2020),* developed a deep learning model based on convolutional neural networks (CNNs) to classify defective apples. Implemented in custom software, their model achieved 92% accuracy with a processing time under 72 ms. *GEJILETU et al., (2025),* employed deep neural networks, to detect and quantify Mongolian cattle herds, overcoming the labor-intensive and time-consuming limitations of traditional methods with 97.3% accuracy, demonstrating the potential of image recognition for efficient detection in unique settings. Similarly, to improve the segmentation of apple leaves and lesion areas, *Wang et al., (2023),* proposed a U2Net semantic segmentation network. Comparative analysis with classical models like DeepLabV3+ and UNet showed that their U2Net achieved a Mean Pixel Accuracy (MPA) of 98.87% and a Mean Intersection over Union (MIoU) of 84.43%, providing a theoretical basis for precise apple disease control. *Jeyaraj et al., (2022),* created a cost-effective, non-contact rice grading system using an AlexNet-based deep learning model, which attained an average accuracy of 98.2% and a sensitivity of 97.6% on two public datasets. *Xu et al., (2025),* proposed an improved ResNet34 model for rice leaf disease recognition, incorporating a non-local attention mechanism and depth wise separable convolutions to enhance feature extraction. Their model achieved 96% accuracy, outperforming other mainstream algorithms. To address rice seedling counting in complex environments, *Cui et al., (2025),* introduced an enhanced YOLOv5 model that incorporates the Normalized Wasserstein Distance (NWD) metric for bounding box evaluation, effectively counting small seedlings from drone perspectives with 95.9% accuracy.

In summary, although object detection algorithms have demonstrated significant potential in agricultural applications, most research on rice identification has focused on rice diseases and seedling counting. Studies on the identification of rice throughout its entire growth stages remain insufficient, still facing challenges such as inadequate accuracy and high computational costs. To address this, this paper proposes a lightweight detection method based on an improved YOLOv12n model, tailored to the distinct features of different rice growth stages. By optimizing the feature extraction process in the neck network, the model enhances its ability to capture deep features from multi-scale, dense seedlings. This approach aims to improve recognition accuracy for key rice growth periods while controlling computational complexity, thereby meeting the requirements for real-time and reliable detection on standard local PC devices.

MATERIALS AND METHODS

Image Acquisition and Dataset Construction

Rice growth stage image data were collected at the Qinglongshan Farm in Jian Sanjiang, Jiamusi City, Heilongjiang Province (coordinates: 132°56'11.98"E, 47°34'30.32"N). The images were captured between June 10 and August 30, 2024, using a Xiaomi 13 Ultra smartphone. The shooting height was adjusted according to the growth stage: entire plants were photographed during the green-up, tillering, and jointing-booting stages, while the focus was on panicles during the flowering, milk-ripe, and yellow-ripe stages. All images have a resolution of 4000 × 3000 pixels. A total of 3500 in-situ field images were acquired. An additional 667 high-quality rice growth stage images were selected from existing public and laboratory datasets *(Qin et al., 2024)*. The final compiled dataset of 4167 images was saved in 24-bit RGB true-color JPG format. The six distinct growth stages of rice observed during the study period are illustrated in Fig. 1.



(a) Green-up stage *(b) Tillering stage* *(c) Jointing-booting stage*

(d) Flowering stage *(e) Milk-ripe stage* *(f) Yellow-ripe stage*

Fig. 1 - Growth period of rice

After excluding unqualified images, a total of 3,018 samples were retained. Data augmentation techniques—including random scaling, cropping, rotation, noise addition, and color transformation—were applied to expand the dataset, thereby enhancing the model's generalization capability and effectively mitigating potential overfitting. All images were uniformly resized to 640 × 640 pixels, resulting in a final dataset of 15,090 image samples. The entire dataset was randomly partitioned into training, validation, and test sets in a ratio of 8:1:1. These images were annotated using LabelImg, and labels were generated in the YOLO format.

Model Construction Based on Improved YOLOv12n

The newly introduced YOLOv12n model incorporates a progressive feature pyramid and an adaptive attention mechanism in its architecture *(Tian et al., 2025)*, which enhances its sensitivity to subtle feature variations across rice growth stages. To improve the overall performance of the model for identifying rice in cold regions, this study proposes an enhanced model based on the YOLOv12n architecture, named YOLOv12-DBD.

First, a Deformable Large Kernel Attention (D-LKA) mechanism is introduced to enhance the feature extraction capability of the neck network, utilizing adaptive sampling grids and large receptive field convolutions to improve the modeling of multi-scale morphological features in rice. Second, a Bidirectional Weighted Feature Pyramid Network (BiFPN) is adopted to reconstruct the multi-scale feature fusion pathway. Its fast normalized fusion mechanism strengthens the model's focus on features critical to specific growth stages. Subsequently, a Dynamic Upsampling Module (DySample) replaces traditional interpolation methods, improving the quality of feature map resolution reconstruction and enhancing the preservation of fine details in panicles and leaves. Finally, the Normalized Wasserstein Distance (NWD) is incorporated to optimize the small-object detection strategy, significantly improving the detection of densely distributed small targets and subtle transitional features. The structure of the optimized model is illustrated in Fig. 2.

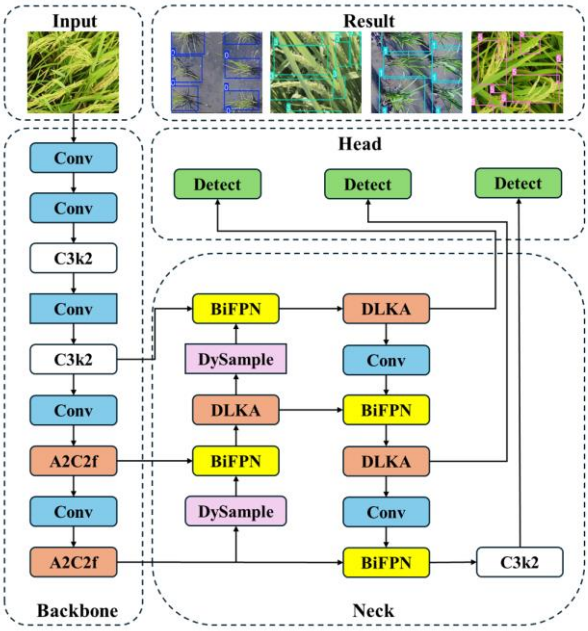


Fig. 2 - Structure diagram of the improved YOLOv12n model

Dynamic Upsampling Module

The nearest-neighbor interpolation method used in YOLOv12n lacks content-aware capability when processing images from stages such as jointing-booting. It struggles to preserve fine details like panicle edges and leaf vein textures, leading to significant information loss during feature map reconstruction. To address this limitation, a Dynamic Upsampling module (DySample) is introduced in the Neck section to replace traditional interpolation methods *(Liu et al., 2023)*. Compared to other dynamic upsampling methods, DySample employs a point-based resampling approach. It adaptively scales the input feature maps, enabling precise capture of detailed features in rice panicles and leaf veins, thereby improving detection accuracy while incurring lower inference latency and requiring fewer parameters. The structure of DySample is illustrated in Fig 3.

DySample generates sampling points using a static range factor and a dynamic range factor. The feature map X is processed through linear layer sampling to produce a feature map of corresponding dimensions. The pixel shuffle technique is then combined with the range factors to generate the offset O, which is finally added to the original grid positions G to obtain the sampling set S, as illustrated in Fig. 4.

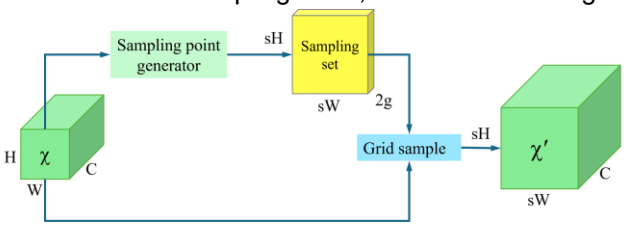
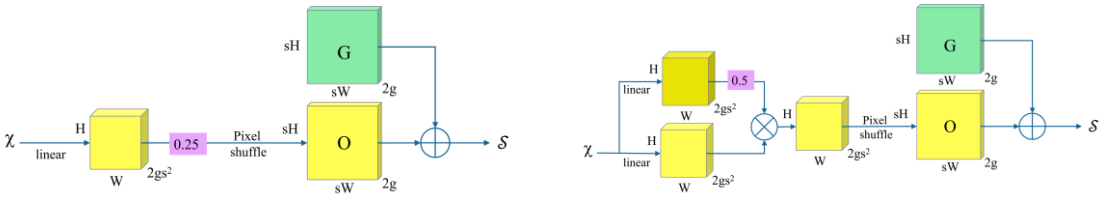


Fig. 3 - DySample structure diagram

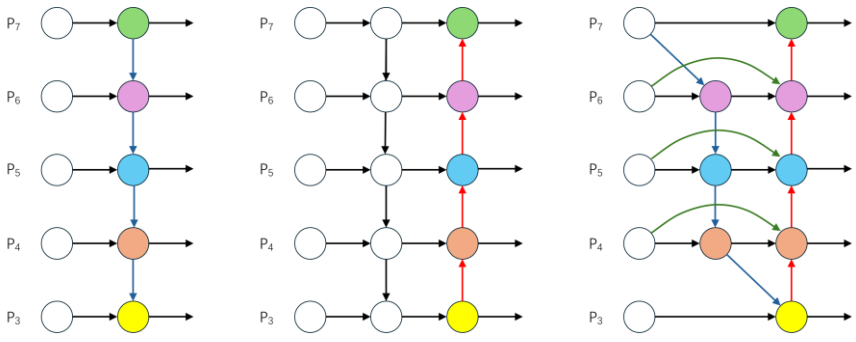


(a) Static Scope Factor *(b) Dynamic Scope Factor*

Fig. 4 - Sampling point generator structure

BiFPN Bidirectional Weighted Feature Pyramid Module

The neck network of YOLOv12n employs an FPN+PAN structure. However, the Feature Pyramid Network (FPN) fuses different input features through a simple top-down pathway and feature concatenation, without distinguishing the importance of different features *(Lin et al., 2017)*. Although the Path Aggregation Network (PAN) adds an extra bottom-up pathway for feature aggregation, it increases the number of parameters and computational cost *(Liu et al., 2018)*. To address these issues, this paper introduces a Bidirectional Weighted Feature Pyramid Network (BiFPN) *(Tan et al., 2020)*. Utilizing a multi-scale feature fusion mechanism, BiFPN enables sufficient interaction between low-level details and high-level semantics through bidirectional cross-scale connections, achieving more efficient and accurate multi-scale feature fusion and thereby enhancing feature representational capacity. Its architecture is shown in Fig. 5.



(a) FPN *(b) PAN* *(c) BiFPN*

Fig. 5 - Comparison of pyramid Networks with different features

D-LKA Deformable Large Kernel Attention Module

The Deformable Large Kernel Attention (D-LKA) mechanism integrates large-kernel convolution with deformable convolutions, building upon developments in areas like medical image segmentation *(Azad et al., 2024)*. The traditional YOLOv12 model uses standard convolution operations, whose kernels have a fixed geometric structure, limiting the model's adaptability to the morphological features of rice plants. When applied to cold-region rice growth stage identification, standard convolutions struggle to effectively capture the spatial distribution of densely packed small targets and have a limited ability to perceive the geometric deformations of panicle details, especially from the flowering stage onward. To address these limitations, a D-LKA module is added to the neck network. This enables the network to integrate multi-dimensional information and better emphasize crucial features, thereby enhancing its ability to extract features from targets of varying scales and irregular shapes. The 2D structure of the D-LKA module is shown in Fig. 6.

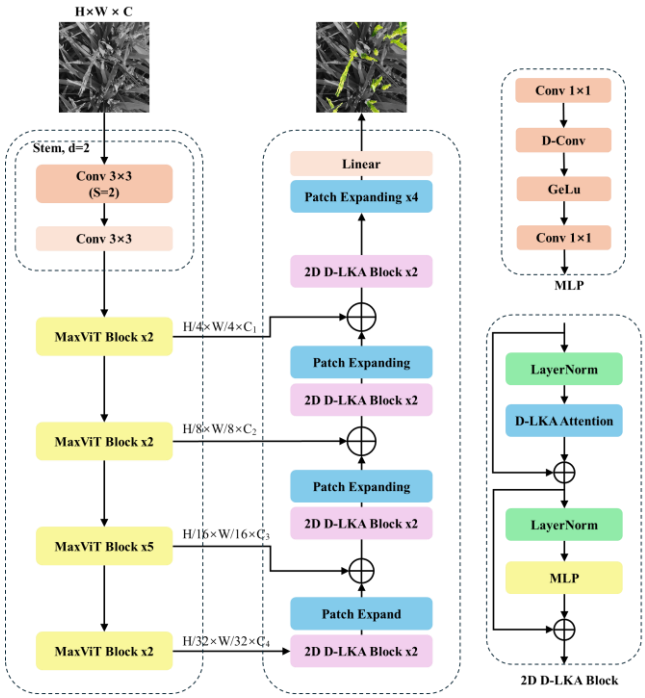


Fig. 6 - D-LKA module network structure

NWD Loss Function

In the original YOLOv12n network, the bounding box regression loss (Box Loss) is employed to ensure that the model can accurately localize targets. It typically uses Intersection over Union (IoU) to measure the degree of overlap between the predicted box and the ground truth box. However, as a primary similarity metric, IoU is overly sensitive to positional deviations in small targets. To address this limitation, this study introduces the Normalized Wasserstein Distance (NWD) to measure the similarity between predicted and ground truth boxes *(Wang et al., 2021)*, thereby mitigating the inherent sensitivity of IoU to the position of small targets. This method is robust to minor positional variations and can effectively compute similarity even for non-overlapping bounding boxes.

The bounding box is first modeled as a 2D Gaussian distribution, where the central pixel has the highest weight, and the weight decreases from the center towards the boundaries. For a horizontal bounding box $R=(cx,cy,\omega,h)$, its corresponding Gaussian distribution can be represented as:

$$f(x/\mu,\Sigma)=\frac{exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{2\pi/\sum^{\frac{1}{2}}} \quad (1)$$

where x represents the pixel coordinates, μ is the mean vector, and Σ is the covariance matrix. By adjusting the parameters of the covariance matrix, the shape and size of the Gaussian distribution can be controlled.

The Wasserstein distance is then employed to calculate the distance between two Gaussian distributions. For two Gaussian distributions, μ_1 and μ_2, their second-order Wasserstein distance is defined as:

$$W_2^2(\mu_1,\mu_2)=\|m_1-m_2\|_2^2+Tr\left(\sum_1+\sum_2-2\left(\sum_2^{1/2}\sum_1\sum_2^{1/2}\right)^{1/2}\right) \quad (2)$$

For the Gaussian distributions μ_a and μ_b modeled from bounding boxes $A=(cx_a,cy_a,\omega_a,h_a)$ and $B=(cx_b,cy_b,\omega_b,h_b)$, the distance can be further simplified to:

$$W_2^2(N_a,N_b)=\left\|\left(\left[cx_a,cy_a,\frac{\omega_a}{2},\frac{h_a}{2}\right]^T,\left[cx_b,cy_b,\frac{w\omega_b}{2},\frac{h_b}{2}\right]^T\right)\right\|_2^2 \quad (3)$$

where W_2^2 is a distance metric and cannot be directly used as a similarity measure (i.e., a value between 0 and 1 like IoU).

Therefore, it is normalized using its exponential form to derive a new metric, the Normalized Wasserstein Distance (NWD), expressed as:

$$NWD(N_a,N_b)=exp\left(-\frac{\sqrt{W_2^2(N_a,N_b)}}{C}\right) \quad (4)$$

where C is a constant closely related to the dataset, controlling the range of NWD values.

RESULTS AND DISCUSSIONS

EXPERIMENTAL DESIGN AND ANALYSIS

Experimental environment and evaluation indicators

The experimental hardware configuration for this study included an AMD Ryzen 5 7500F 6-Core Processor, an NVIDIA GeForce RTX 4070 GPU (with 12 GB of video memory), 32 GB of RAM, a 2 TB solid-state drive, and the 64-bit Windows 10 operating system. The software environment consisted of PyTorch 2.7.1, CUDA 13.0, and Python 3.11. The relevant hyperparameters were set as follows: the number of training epochs was 200, the batch size was 16, and the input image resolution was 640 × 640 pixels. To enhance the convergence rate, the initial learning rate was set to 0.01.

To evaluate the performance of the improved model in detecting multiple rice growth stages, precision, recall, and mean average precision (mAP) were employed as the core metrics. Model complexity was quantified by analyzing the floating-point operations (FLOPs). Real-time performance was assessed using the frames per second (FPS). Comparative experiments with mainstream object detection algorithms were conducted to comprehensively validate the enhanced balance between accuracy and efficiency achieved by the proposed model.

Conclusions on rice growth stage identification

Upon the successful completion of the model training phase, the optimal weight file, best.pt, was obtained. The performance metrics for each rice growth stage category are summarized in Table 1. Subsequently, images from the test set were utilized to comprehensively assess the model's performance in identifying growth stages. To provide a visual representation of the results of the growth stage identification experiment, the outcomes are depicted in the form of a confusion matrix, as presented in Figure 7.

From the normalized confusion matrix, it can be observed that the improved YOLO model exhibits superior performance in identifying rice during the early growth stages, but its performance is relatively insufficient in the later stages. This decline in accuracy after the milky stage is attributed to the high visual similarity between the ripening rice panicles, which transition from green to yellow, and the surrounding rice leaves.

Table 1

Per-class performance metrics for rice growth stages

Class	P/%	R/%	mAP50/%
Green-up stage	99.5%	99.9%	99.5%
Tillering stage	99.4%	100%	99.5%
Jointing-booting stage	93.4%	93.1%	97.3%
Flowering stage	91.5%	82.7%	92.2%
Milk-ripe stage	85.3%	76.1%	84.2%
Yellow-ripe stage	91.4%	86.1%	93.5%

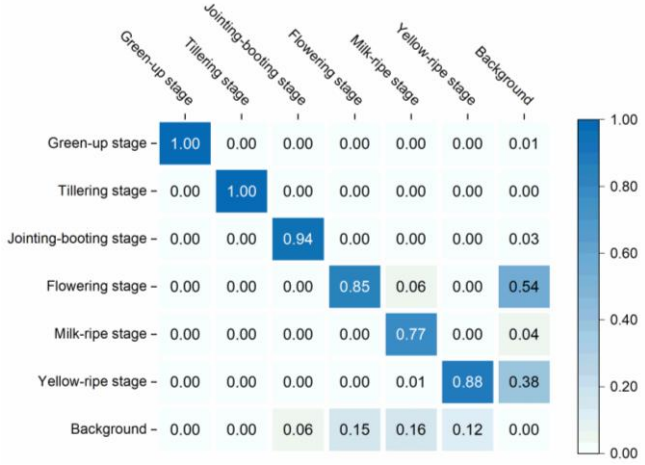


Fig. 7 – Confusion Matrix Normalized

Loss function curve analysis

Fig. 8 shows the loss function curves over the 500-iteration training process. In the first 10 iterations, both the training and validation losses decreased rapidly. After approximately 70 iterations, the two loss curves converged and ran nearly parallel, indicating the model was approaching its optimal state. Notably, during the final 10 iterations, the training loss experienced a sharp decline due to the deactivation of the Mosaic data augmentation, eventually fluctuating and stabilizing around a value of 2.0. Throughout the training, the trends of the two loss values remained consistent without significant divergence. These results indicate that, during training, the model for cold-region rice growth stage identification exhibited no signs of overfitting or underfitting, demonstrating good generalization capability.

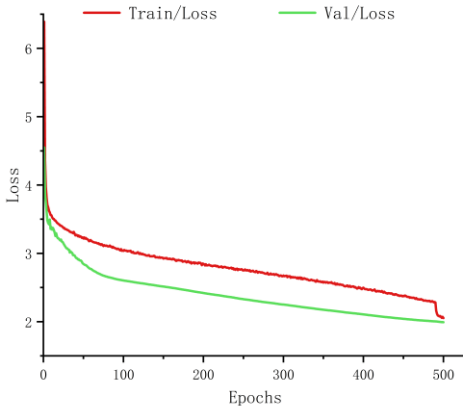


Fig. 8 - Loss Function Curve

Ablation Study

To further verify the effectiveness of the proposed improvement measures for the rice growth stage recognition model, an ablation study was conducted using YOLOv12n as the baseline model. As shown in Table 2, the results indicate that the D-LKA module improves the detection accuracy while maintaining a reasonable computational cost; BiFPN significantly enhances the efficiency of multi-scale feature utilization, achieving this goal through weighted feature fusion. In the comprehensive experiments, these modules demonstrated good compatibility and synergy. The final model achieved an accuracy of 93.4%, a recall rate of 89.7%, and an average precision (mAP@0.5) of 94.4%, which was 4.9 percentage points higher than the baseline model and reduced the recognition speed by 27.3 FPS.

Due to the need for higher precision in identifying the rice growth stages in the complex environment of paddy fields to make irrigation decisions, and since the YOLOv12n-DBD model has greater advantages in terms of accuracy and less reduction in recognition speed, the improved model is more superior compared to the original model.

Table 2

Ablation test results

Yolo v12n	DySample	BiFPN	D-LKA	NWD	P/%	R/%	mAP50/%	GFLOPs	FPS
✓	-	-	-	-	89.1	84.2	89.7	5.8×10^9	149.2
✓	✓	-	-	-	89.1	85.2	90.2	6.3×10^9	105.2
✓	-	✓	-	-	88.9	84.8	89.9	6.3×10^9	147.1
✓	-	-	✓	-	88.6	84.9	89.9	6.1×10^9	121.9
✓			-	✓	89.7	86.2	91	6.3×10^9	142.9
✓	✓	✓	-	-	91.9	88.1	93.3	6.3×10^9	111.1
✓	✓	✓	✓	-	93	87.8	93.4	6.4×10^9	91.7
✓	✓	✓	✓	✓	93.4	89.7	94.4	6.4×10^9	121.9

Note: "-" indicates not using this module;
"√" indicates the adoption of this module;

Comparative Experiment on Different Models

Comparative experiments were conducted to evaluate the performance differences between the improved model and other leading models in detecting multiple rice growth stages. The proposed YOLOv12n-

DBD model was compared with Faster R-CNN and mainstream YOLO variants. The results are summarized in Table 3, and the performance differences are visually presented in Fig. 9.

In these comparative tests, the proposed model demonstrated significant advantages in key metrics such as precision, recall, and mean average precision, while maintaining a relatively low computational cost (GFLOPs), reflecting its excellent overall detection capability. Although its inference speed on the test platform was slightly lower than that of some comparative models, it still meets the real-time requirements for rice growth stage image recognition. This indicates a favorable balance between accuracy and efficiency for practical application scenarios.

Table 3

Comparative experimental results of different models

Model	P/%	R/%	mAP50/%	GFLOPs	FPS
SSD	82.6	66.5	72.5	3.1×10^{10}	77.1
Faster R-CNN	68.3	82.2	81.4	1.3×10^{11}	20.1
YOLO v5n	82.3	77.7	82.5	7.1×10^{9}	86.9
YOLO v8n	84.8	80.1	85.0	8.1×10^{9}	200
YOLO v10n	85.4	82.2	86.6	8.2×10^{9}	129.8
YOLO v11n	86.8	81.7	87	6.3×10^{9}	82.6
YOLO v12n	89.1	84.2	89.7	5.8×10^{9}	149.2
YOLO v12n-DBD	93.4	89.7	94.4	6.4×10^{9}	121.9

Note: P is the precision rate; R is the recall rate;
mAP50 is the average precision mean;
GFLOPs are floating-point operations

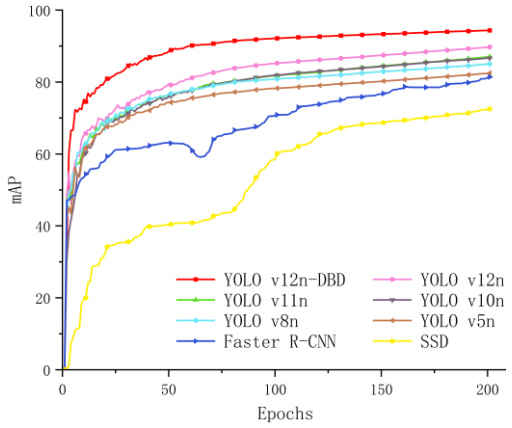


Fig. 9 - mAP50 curves for each comparison model

Rice growth period identification system

Based on the improved YOLOv12n algorithm for cold-region rice growth stage identification, a dedicated recognition system was developed. The system adopts a modular architecture and features a comprehensive visual interface built on the PyQt5 framework. The overall interface layout is shown in Fig. 10.

The system's core functionalities comprise four main modules:

- Operation Module: Users import single or batch rice growth stage images via the "Upload Single Image" or "Batch Upload Images" buttons.
- Identification Module: After clicking the "Start Detection" button, the improved YOLOv12n model performs feature extraction and growth stage identification, accurately distinguishing key stages such as tillering and jointing-booting.
- Result Output Module: Detection results are displayed in real-time on the right panel of the interface, visualizing the distribution of rice growth stages using bounding boxes and category labels. Simultaneously, detailed identification statistics are output in a dedicated results window below.
- Export Module: This module generates a structured data file containing plant locations, growth stage categories, and confidence scores, providing data support for subsequent agricultural research such as growth progression analysis and population vigor assessment.

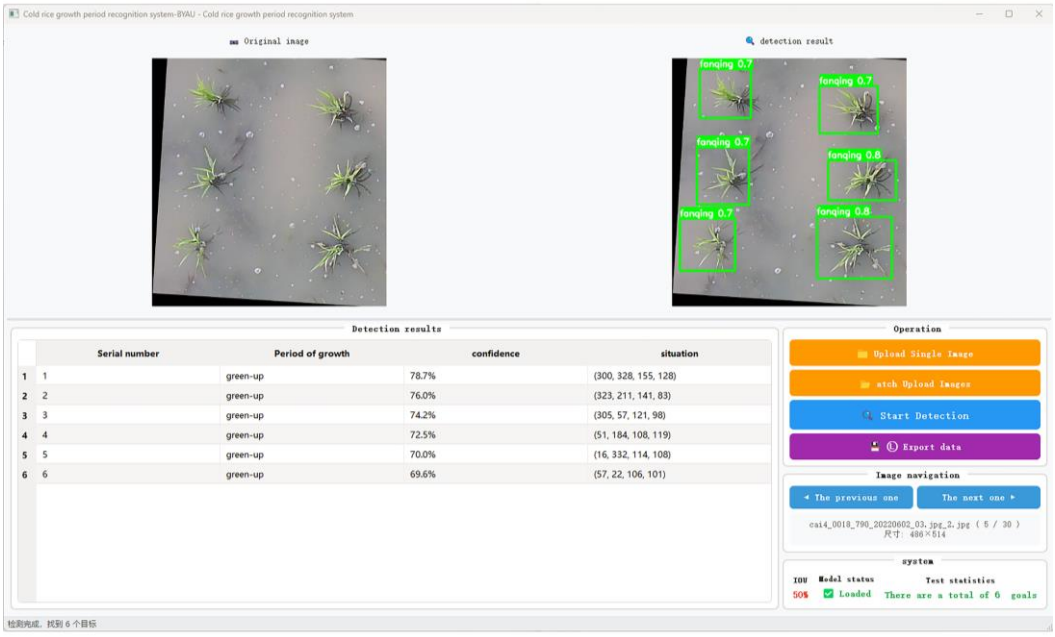


Fig. 10 - Visual interface of the cold-region rice growth period identification system

CONCLUSIONS

This study introduces YOLOv12n-DBD, an improved YOLOv12n-based model, addressing low recognition accuracy and poor generalization in cold-region rice growth stage identification caused by subtle morphological variations, complex field environments, and occlusions. The model enhances recognition precision and inference efficiency through synergistic optimization.

1. Enhancements to YOLOv12n include a deformable large kernel attention (D-LKA) module in the neck to improve multi-scale feature extraction and geometric modeling via adaptive sampling. A bidirectional weighted feature pyramid network (BiFPN) optimizes multi-scale feature fusion. The integration of a dynamic upsampling (DySample) and normalized Wasserstein distance (NWD) improves feature map reconstruction and small object detection, significantly increasing accuracy in identifying rice phenological stages.

2. The YOLOv12n-DBD architecture achieves a mAP@0.5 of 94.4%, with 93.4% precision and 89.7% recall on a cold-region rice phenology dataset. It requires 6.4 GFLOPs and 5.5 million parameters, combining high detection accuracy with computational efficiency. The model outperforms YOLOv5n, YOLOv8n, and Faster R-CNN, demonstrating robust performance and practical applicability in harsh cold environments.

3. Based on the improved model, a rice growth stage identification system for cold regions was constructed. A visual interactive interface was developed using PyQt5, enabling high-precision real-time recognition on local devices and automatic output of structured data including target locations and phenological stages. Future work will focus on model lightweighting and edge deployment to adapt to mobile terminals and field IoT nodes.

ACKNOWLEDGEMENT

This study was supported by the Collaborative Innovation Achievement Project for "Double First-Class" Disciplines of Heilongjiang Province (LJGXCG2024-F18).

REFERENCES

[1] Azad, R., Niggemeier, L., Hüttemann, M., Kazerouni, A., Aghdam, E. K., Velichko, Y., Bagci U., & Merhof, D., (2024). Beyond self-attention: Deformable large kernel attention for medical image segmentation, In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp.1287-1297.

[2] 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, no.5, pp. 20-42.

[3] Cui J., Ye W., Zheng H., Liu T., Qi L., & Xu Y., (2025). Rice Seedling Counting in Complex Environments Based on Domain-Adaptive NWD-YOLOv5 (基于域自适应 NWD-YOLOv5 的复杂环境下水稻幼苗计数), *Computer Engineering*, Vol. 51, no.3, pp. 320-333.

[4] Fan, S., Li, J., Zhang, Y., Tian, X., Wang, Q., He, X., Zhang C., & Huang, W., (2020). On line detection of defective apples using computer vision system combined with deep learning methods, *Journal of Food Engineering*, Vol.286, pp. 10102.
[5] Gejiletu, H., Ulaankhuu, H., & Nyandag, B.., (2025). Detection and counting of grazing cattle from aerial images using CNN, *INMATEH-Agricultural Engineering*, Vol.75, no.1, pp. 456-468.
[6] Jeyaraj, P. R., Asokan, S. P., & Nadar, E. R. S., (2022). Computer-Assisted Real-Time Rice Variety Learning Using Deep Learning Network, *Rice Science*, Vol.29, no.5, pp. 489-498.
[7] 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, no.5, pp. 302-311.
[8] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S., (2017). Feature pyramid networks for object detection, In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117-2125.
[9] Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J., (2018). Path aggregation network for instance segmentation, In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8759-8768.
[10] Liu, W., Lu, H., Fu, H., & Cao, Z., (2023). Learning to Upsample by Learning to Sample, In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6027-6037.
[11] Oraif I., (2023). Natural Language Processing (NLP) and EFL Learning: A Case Study Based on Deep Learning, *Journal of Language Teaching and Research*, Vol.15, no.1, pp. 201-208.
[12] Qin J., Wan J., Song G., Yao H., Guo L., & Wang X., (2024). Image Dataset of Wheat, Corn, and Rice Seedlings in Heilongjiang Province in 2022 (2022 年黑龙江小麦、玉米、水稻苗期*图像数据集), Journal of Agricultural Big Data*, Vol.6, no.4, pp. 558-563.
[13] Tan, M., Pang, R., & Le, Q. V., (2020). EfficientDet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10781-10790.
[14] Tang Y., (2022). *Research and Development of Rapid Monitoring Technology and System for Wheat Emergance Quality (小麦出苗质量快速监测技术及系统构建),* MSc Thesi, Nanjing Agricultural University, Nanjing/China.
[15] Tian, Y., Ye, Q., & Doermann, D., (2025). YOLOv12: Attention-Centric Real-Time Object Detectors*, arXiv preprint arXiv:2502*.12524, New York/USA. DOI: https://doi.org/10.48550/arXiv.
[16] Wang B., Li L., Li S., & Yang H., (2023). Research on apple leaf disease segmentation and classification based on semantic segmentation network, *INMATEH-Agricultural Engineering*, Vol.69, no.1, pp. 295-304.
[17] Wang, J., Xu, C., Yang, W., & Yu, L., (2021). A normalized Gaussian Wasserstein distance for tiny object detection, *arXiv preprint arXiv:2110.13389*. DOI: https://doi.org/10.48550/arXiv.2110.13389
[18] Wang Y., Li Q., Dai Z., & Xu Y., (2024). Current status and trends in large language modeling research (大语言模型研究现状与趋势), *Chinese Journal of Engineering*, Vol.46, no.8, pp. 1411-1425.
[19] Xu P., Zhou C., Zhao X., & Wu C. (2025). Rice leaf disease recognition model based on improved ResNet34 (基于改进 ResNet34 的水稻叶片病害识别模型), *Journal of Chinese Agricultural Mechanization*, Vol. 46, no.12, pp. 86-93.
[20] Xu W., (2024). Design of an English Corpus for Agricultural Machines Based on Deep Learning (基于深度学习的农机机器英语语料库的设计), *Journal of Agricultural Mechanization Research*, Vol.46, no.10, pp. 208-212.
[21] Xue L., Zhang Z., Qi Z., Han Y., Xu D., Zhang Z., & Zhou X., (2024). Effects of Straw Returning on N2O Emission and Yield under Water-saving Irrigation in Black Soil Paddy Field (节水灌溉下秸秆还田形式对黑土区稻田 N2O 排放与产量的影响), *Transactions of the Chinese Society for Agricultural Machinery*, Vol.55, no.4, pp. 280-289.