

Article

Detection Method for Rice Seedling Planting Conditions Based on Image Processing and an Improved YOLOv8n Model

Bo Zhao ^{1,2,*}, Qifan Zhang ^{1,2} , Yangchun Liu ², Yongzhi Cui ² and Baixue Zhou ²

¹ College of Mechanical and Electrical Engineering, Qingdao Agricultural University, Qingdao 266109, China; zhangqifan1224@163.com

² National Key Laboratory of Agricultural Equipment Technology, Beijing 100083, China; lyc327@163.com (Y.L.); 18787171287@163.com (Y.C.); baixue@163.com (B.Z.)

* Correspondence: zhaoboshi@126.com

Abstract: In response to the need for precision and intelligence in the assessment of transplanting machine operation quality, this study addresses challenges such as low accuracy and efficiency associated with manual observation and random field sampling for the evaluation of rice seedling planting conditions. Therefore, in order to build a seedling insertion condition detection system, this study proposes an approach based on the combination of image processing and deep learning. The image processing stage is primarily applied to seedling absence detection, utilizing the centroid detection method to obtain precise coordinates of missing seedlings with an accuracy of 93.7%. In the target recognition stage, an improved YOLOv8 Nano network model is introduced, leveraging deep learning algorithms to detect qualified and misplaced seedlings. This model incorporates ASPP (atrous spatial pyramid pooling) to enhance the network's multiscale feature extraction capabilities, integrates SimAM (Simple, Parameter-free Attention Module) to improve the model's ability to extract detailed seedling features, and introduces AFPN (Asymptotic Feature Pyramid Network) to facilitate direct interaction between non-adjacent hierarchical levels, thereby enhancing feature fusion efficiency. Experimental results demonstrate that the enhanced YOLOv8n model achieves precision (P), recall (R), and mean average precision (mAP) of 95.5%, 92.7%, and 95.2%, respectively. Compared to the original YOLOv8n model, the enhanced model shows improvements of 3.6%, 0.9%, and 1.7% in P, R, and mAP, respectively. This research provides data support for the efficiency and quality of transplanting machine operations, contributing to the further development and application of unmanned field management in subsequent rice seedling cultivation.

Keywords: YOLOv8n; rice seeding planting conditions; deep learning; image processing



Citation: Zhao, B.; Zhang, Q.; Liu, Y.; Cui, Y.; Zhou, B. Detection Method for Rice Seedling Planting Conditions Based on Image Processing and an Improved YOLOv8n Model. *Appl. Sci.* **2024**, *14*, 2575. <https://doi.org/10.3390/app14062575>

Academic Editor: Norman Q. Arancon

Received: 14 February 2024
Revised: 7 March 2024
Accepted: 8 March 2024
Published: 19 March 2024



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

1. Introduction

Rice, as the largest in terms of planting area, highest in yield per unit area, and most abundant cereal crop in China, accounted for 34.58% of the total grain production in 2022, holding a dominant position in both grain consumption and production [1]. However, factors such as uneven field surfaces, insufficient sedimentation time in paddy fields, deep water layers in the field, exposed and hardened soil, harsh operating environments, and driver errors often lead to issues such as missed planting and floating seedlings during the operation of rice transplanting machines [2]. Currently, the assessment of transplanting machine operation quality heavily relies on manual judgment, resulting in low efficiency, poor working conditions, and increased labor costs [3]. In order to enhance rice yield and the efficiency of transplanting machine operations, it is crucial to identify the locations of abnormal seedling states and alert manual intervention. Timely alarms should be triggered in case of large-scale missed planting or floating seedlings, serving as reminders for improper driver operation or the need for machine maintenance. This approach aims to reduce the labor costs associated with subsequent reseeding efforts [4]. For the rapid

and accurate assessment of transplanting machine operation quality, achieving precise recognition of seedling morphology is of paramount importance.

Currently, scholars both domestically and internationally utilize image processing techniques to perform image segmentation and classification of seedlings based on crop features, including color and geometric morphological characteristics. Yuan Jiahong et al. [5], for instance, leverage color information from G-R and G-B components of seedlings against the water field background. They select the ExG factor suitable for most green crops and use the Otsu automatic threshold method to achieve optimal thresholding, effectively separating seedlings from the background. Chen Jie et al. [6] establish an HSV color model based on a rice seedling dataset and train it for the recognition of rice seedlings. Hayashi [7] employs both RGB cameras and TOF depth cameras to capture rice images. The author combines these two types of images, fusing pixel information related to color, texture, and shape features. A random forest model is constructed, utilizing multiple features for rice seedling recognition and achieving a 5% increase in accuracy. While these methods successfully segment seedlings from the water field background, the diverse morphologies of seedlings and the complex and variable paddy environments pose challenges. Furthermore, the effectiveness of these approaches is significantly influenced by the interaction between the target and background, leading to poor generalization ability and robustness of the models.

In recent years, as the application of machine vision in agriculture has continued to deepen, its technology has demonstrated the capability to handle complex patterns and large-scale data in agricultural production. With adaptability and real-time performance, methods for seedling object detection based on deep learning have become widespread [8]. Feng Huaiqu et al. [9], by improving deep learning algorithm models, proposed a Faster RCNN with VGG19 [10] model using transfer learning. This model achieved recognition and detection of corn seedlings under various conditions, including the full growth cycle, multiple weather conditions, and various angles. The recognition accuracy and recall rate were both above 95%. They also utilized the MobileNetV3-SSD with a pruning [11] model for distinguishing between seedlings and weeds, showing good detection results. Gao Zongyao et al. [12], focusing on detecting missing seedlings in rice seedlings, optimized the hyperparameters of the YOLOv5 model using the cosine annealing algorithm. By filtering out low-quality detection boxes based on the confidence and category of each output candidate region, the model accurately located the detection image containing the missing seedling position. The improved model achieved an mAP value of 93.6%, with a recognition accuracy of 96% for missing seedlings. Zhu Wei et al. [13] utilized low-altitude drone aerial photography, followed by training based on the GoogLeNet network model for the real-time classification and detection of single-hole seedlings. This enabled the timely judgment of floating seedlings, injured seedlings, and qualified seedling forms. Although the above methods are capable of accurately identifying targets, the studied network model has a large number of parameters, runs slowly, has low real-time performance, and fails to achieve an effective balance between identification efficiency and accuracy in paddy fields. Meanwhile, the target features in the image have small differences, high density, and complex backgrounds. Especially, the recognition environment is very harsh.

In summary, this paper focuses on the tasks of recognizing the morphology of floating seedlings, identifying qualified seedlings, and detecting missing seedlings. The proposed approach utilizes the YOLOv8 network as the foundational model for recognizing floating and qualified seedling targets. The original model is improved based on issues encountered during the training and prediction processes. Given the complex and variable nature of missing seedling situations, image processing techniques are more effective in accurately locating the positions of missing seedlings. Therefore, this paper integrates the algorithms for seedling planting condition recognition and missing seedling detection, ensuring real-time detection while enhancing the overall accuracy of the detection process. This integrated approach is more in line with practical scenario requirements, exhibits strong generalization, and provides a basis for optimizing the operation quality strategy for driving transplanting machines. The technology roadmap is shown in Figure 1.

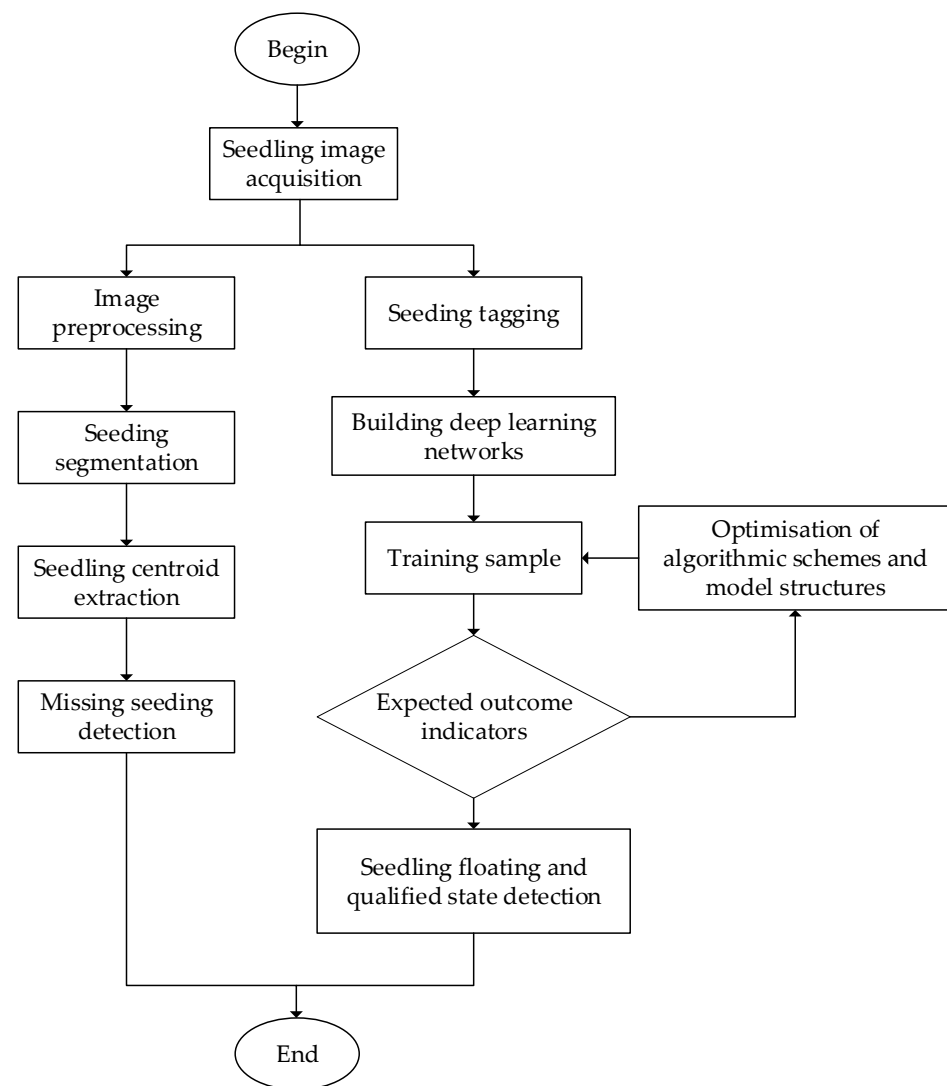


Figure 1. Technology roadmap for seedling planting condition.

2. Materials and Data

2.1. Seedling Image Acquisition

To ensure the diversity of seedling varieties and the variability of transplanting environments, the study area includes paddy fields in Jiangsu and gravel fields in the northeast region. The image collection took place in May 2023 in Dongming Town, Shuangliao City, Jilin Province, and Wujiaohu Village, Fanhe Town, Tieling City, Liaoning Province. In June 2023, additional images were captured in Sunjialu Village and Liuchen Village, Xinnan District, Changzhou City, Jiangsu Province, and Sunjialu Village, Liuchen Village, and Taihuang Village, Taixing City, Taizhou City. The subjects of the collection were seedlings of different varieties at each experimental site. The seedlings from the northeast region were characterized by taller plant height, broader sword-shaped leaves, and darker leaf color, while the seedlings from Jiangsu had shorter plant height, narrower strap-shaped leaves, and lighter leaf color.

Data collection was performed using a Panasonic Lumix V385 digital camera, equipped with optical stabilization, horizontal correction, and autofocus capabilities. This camera could capture high-quality images under the vibration of transplanting machine operation and outdoor natural light conditions. The pixel resolution of the data was 1920×1080 . The image collection involved both static and dynamic approaches. Static data collection involved manually capturing images by holding the camera vertically about 1.5 m above the paddy field. Dynamic data collection involved fixing the camera on the transplanting

machine's seedling tray frame and capturing images during the machine's operation. In total, 2732 seedling images were collected, including 10,992 qualified seedlings and 2547 floating seedling samples. Figure 2 illustrates the morphology of a single seedling, encompassing images of qualified and floating seedlings under different backgrounds and lighting conditions.

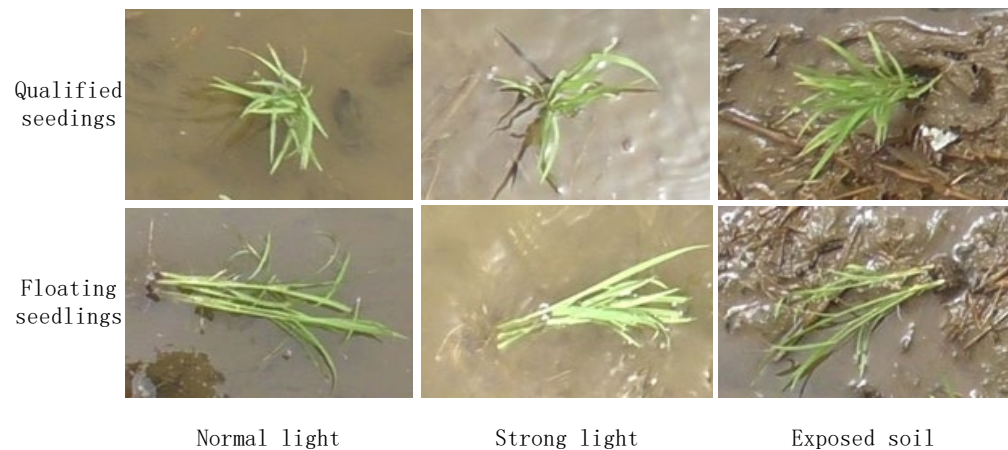


Figure 2. Rice seedling samples in different scenarios.

2.2. Dataset Construction

LabellImg was used as the annotation tool for creating the seedling detection dataset. To avoid missing any seedlings, each image was manually annotated to obtain the positional information of seedlings in both qualified and floating states, completing the construction of the seedling planting condition dataset. The dataset was divided randomly into training, validation, and test sets in a ratio of 7:1:2. This division helps address overfitting and underfitting issues. The dataset consists of 1913 training images, 273 validation images, and 546 test images.

3. Approach

3.1. Missing Seedling Detection

3.1.1. Image Preprocessing

The paddy field environment is diverse and complex, making the detection of missing seedlings susceptible to interference from floating straw, exposed soil, strong light reflection, and machine vibrations. Therefore, image preprocessing is necessary to extract seedling information from the background. Image preprocessing includes grayscale conversion, threshold segmentation, and labeling connected regions [14]. Since the green component of seedlings exhibits significant differences from other areas in the background, the ExG (Excess green) feature value [15] is employed for grayscale conversion. This effectively suppresses background information such as shadows, straw, and soil. The grayscale image is illustrated in Figure 3.

The Otsu method [16] is utilized to obtain the binary image of seedlings. Due to the influence of the actual shooting environment, the captured images often contain irrelevant information such as the sky, human reflections, and water ripples formed by light [17]. These interferences can create significant noise, as illustrated in Figure 4a, where non-agricultural information generates large isolated pixel blocks. To ensure that the detection of missing seedlings is not affected by this noise, the connected region labeling method [18] is employed to eliminate small connected regions that are not connected to the characteristic region. This ensures complete separation between seedlings and the background, as shown in Figure 4b. As rice seedlings are non-rigid body deformations, there may be gaps between adjacent leaves after image preprocessing. This can result in one seedling having multiple contours and centroids. To address this, a dilation operation [19] is performed to fill small

gaps between adjacent regions. Although the individual seedling morphology expands, it does not affect the centroid position of the seedlings, as shown in Figure 4c.

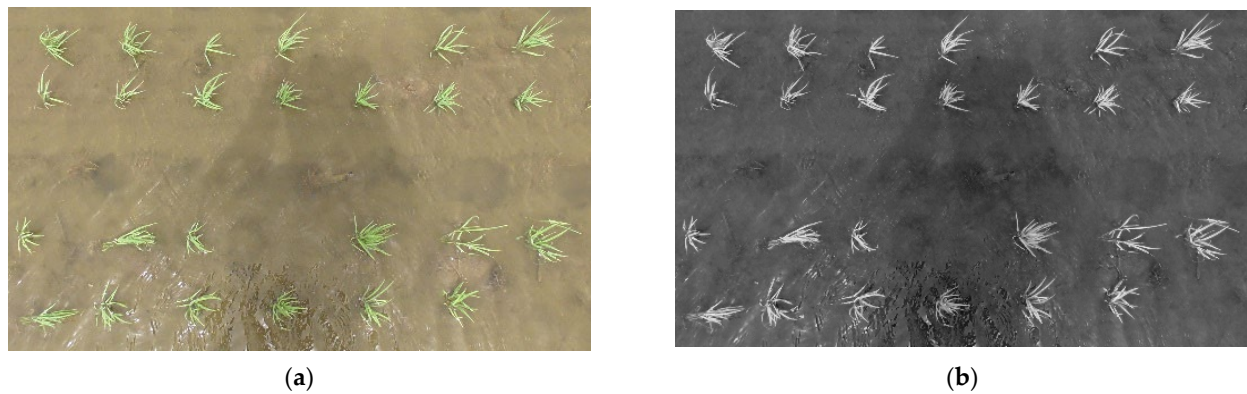


Figure 3. ExG (Excess green) grayscale. (a) Original image; (b) gray image.

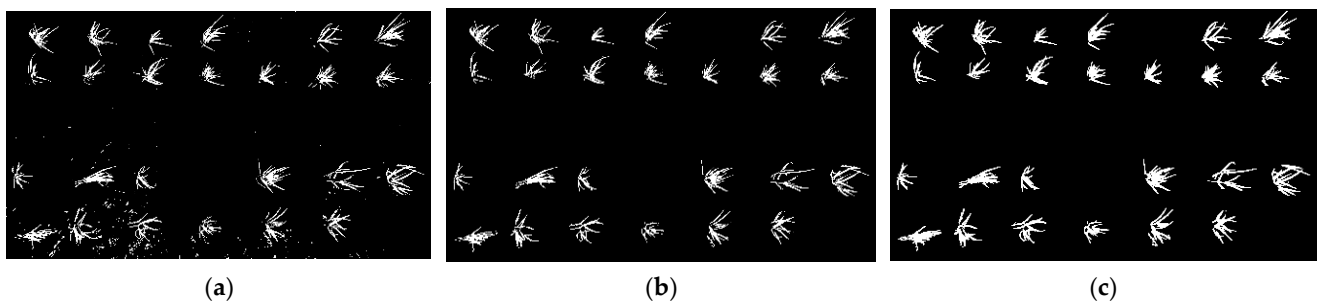


Figure 4. Separation of seedlings from the background. (a) Otsu thresholding; (b) connected component labeling; (c) seedling dilation processing.

3.1.2. Seedling Centroid Extraction

The specific steps for extracting seedling centroids after image preprocessing are as follows:

- (1) **Contour Detection:** Perform Canny edge detection [20] on the dilated seedling image. This aims to retain as few edge points as possible while reducing the impact of noise and false positives on the seedling boundaries. This step detects precise and stable contours of seedlings, as shown in Figure 5a.
- (2) **Centroid Extraction:** After detecting the seedling contours, calculate the minimum bounding rectangle for each contour and determine the centroid coordinates of the bounding rectangle. These centroid coordinates represent the centroids of the seedlings, as illustrated in Figure 5b.

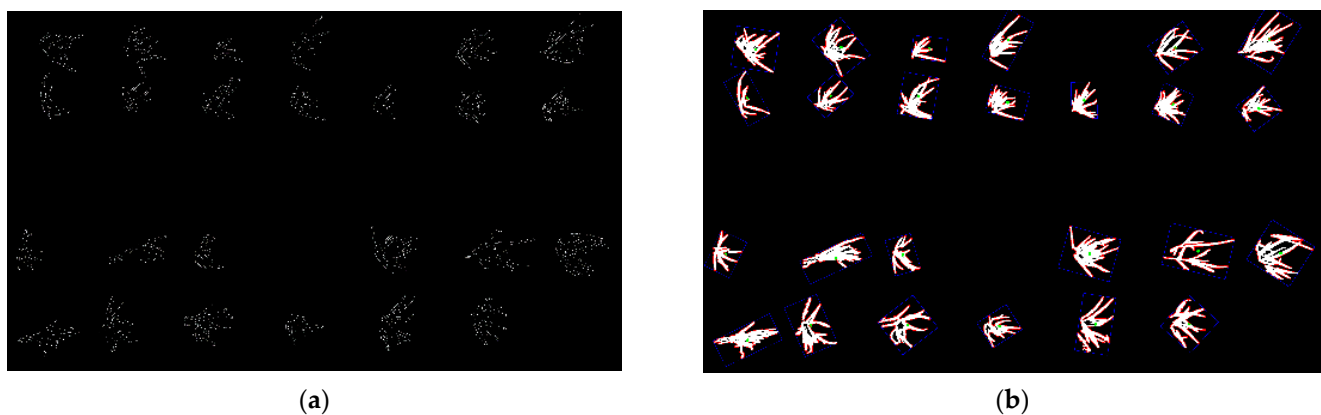


Figure 5. Extraction of seedling centroid. (a) Canny edge detection; (b) centroid retrieval.

3.1.3. Missing Seedling Detection

The detection of missing seedlings is determined by calculating whether the centroid distance of seedlings falls within the maximum plant spacing range of adjacent seedlings. Taking 673 seedling images from Wujiaohu Village, Fanhe Town, Tieling City, Liaoning Province, as an example, the average normal plant spacing in the image processing is 581.68 mm. When there is a floating seedling situation, the distance between adjacent seedlings is greater than the average plant spacing, and the maximum adjacent plant spacing is taken, which is 719.97 mm. Therefore, let d represent the maximum adjacent plant spacing, and D represents the sum of the adjacent centroid distances and the distances from the first and last seedlings to the boundaries. When $D > d$, it indicates a missing seedling phenomenon [21]. The detection results are shown in Figure 6.

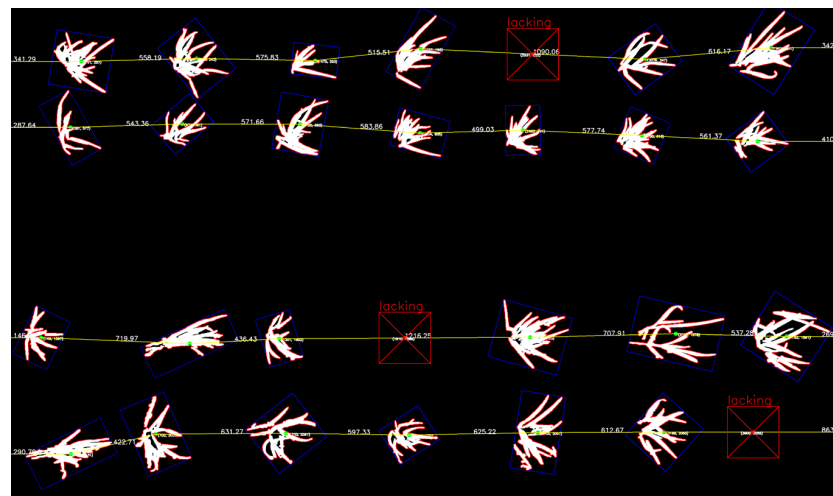


Figure 6. Detection of missing seedling centroids.

Simultaneously, during the operation of the transplanting machine, there may be continuous missing and floating seedling situations, including the occurrence of missing seedlings at the beginning and end of the collected images. Therefore, dividing D by the maximum plant spacing d yields the result n , as shown in Formula (1). As illustrated in Figure 7, the centroids' connecting line is divided into $n + 1$ segments, and the breakpoints are considered as the coordinates of the missing seedlings.

$$n = D \mid d \quad (1)$$

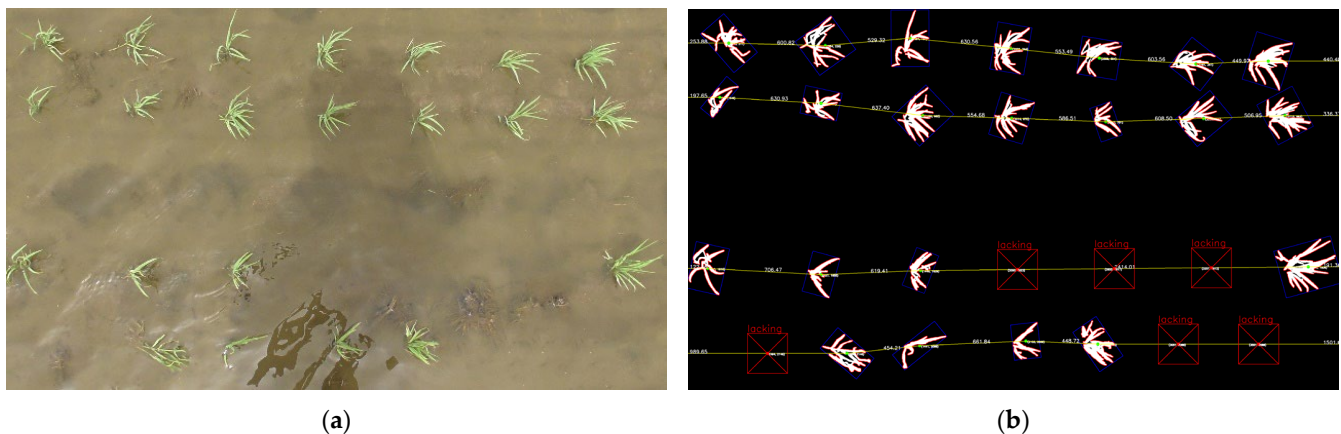


Figure 7. Detection of missing seedlings in different scenarios. (a) Original image; (b) missing seedling detection.

3.2. Seedling Floating and Qualified State Recognition and Improvement

3.2.1. YOLOv8n Network Model

In the realm of one-stage object detection models, YOLOv8 is known for its efficiency and flexibility [22], excelling in both speed and accuracy. YOLOv8 is the optimal model in the YOLO series for detection speed and precision. Specifically, the YOLOv8n [23] network model is compact, fast in inference speed, and provides a powerful choice for lightweight detection within the YOLO series. It is well suited for scenarios with real-time requirements. Therefore, this paper selects the YOLOv8n model as the foundational model for real-time detection of seedling planting conditions. The improved YOLOv8n network structure is divided into four parts, as shown in Figure 8: input, backbone, neck, and prediction.

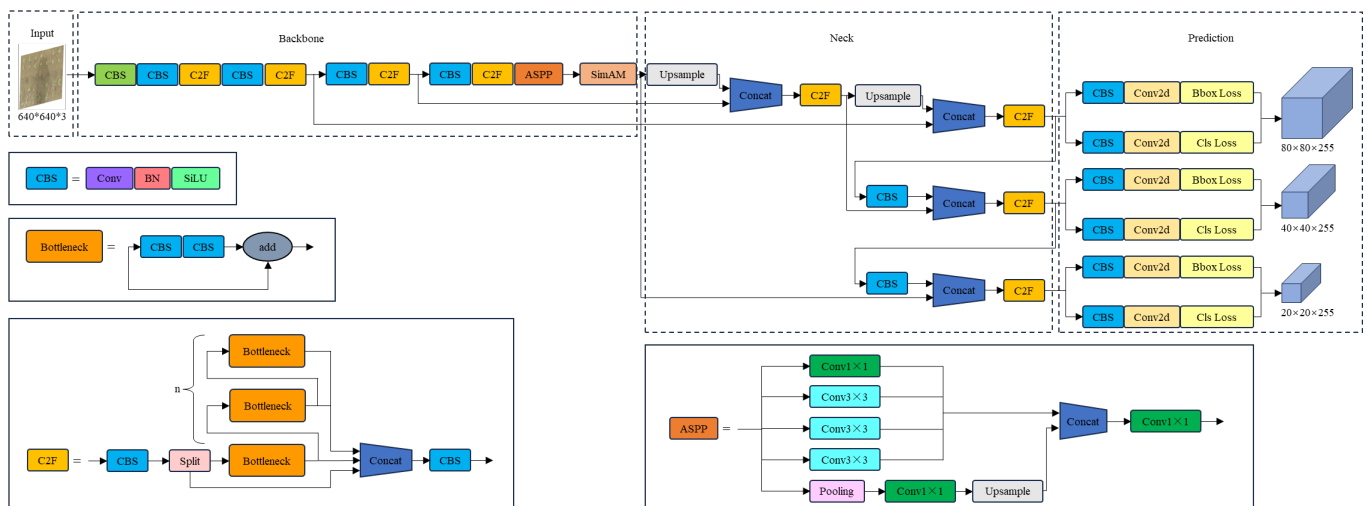


Figure 8. The network structure of improved YOLOv8n. Note: Conv represents the convolutional layer, Concat denotes the fusion of multi-scale features, Bbox Loss refers to the predicted bounding box loss, Cls loss is the loss of the target category. The CBS module is composed of a Conv convolutional layer, BN layer, and SiLU activation function. Bottleneck represents the bottleneck residual structure. C2f is a convolutional structure integrated with the CBS module and n Bottleneck modules. ASPP stands for the atrous spatial pyramid pooling layer, and pooling indicates the pooling operation.

In the YOLOv8n model, the original seedling images are resized to 640×640 , maintaining the aspect ratio, with the remaining areas filled using background padding. These resized images are then input into the backbone section. Within the main network, feature extraction is performed through three effective feature layers. Subsequently, the AFPN (Asymptotic Feature Pyramid Network) in the neck section is employed, utilizing upsampling operations to segment and concatenate features, achieving feature fusion at different stages [24]. In the prediction section, a decoupled head structure is utilized for separate classification and regression tasks. This not only reduces the number of channels but also enhances detection speed and accuracy. The output includes seedling detection categories, bounding box coordinates, and predicted confidence scores.

3.2.2. Improvement of YOLOv8n Model

This study involves recognizing two states of single seedling targets: qualified and floating states. To enhance the detection performance of the model, an improved YOLOv8n algorithm is proposed based on the YOLOv8n algorithm. In this improvement, the SPPF module in the original backbone network of the YOLOv8n algorithm is replaced with the ASPP module. Additionally, a SimAM attention mechanism is added during the feature extraction process in the backbone network. The neck network introduces the AFPN structure to optimize the pyramid structure, ensuring applicability to both qualified and floating seedling detection states while improving the real-time detection accuracy of the model.

3.2.3. Backbone Network Optimization

The backbone is the main feature extraction network of the YOLOv8n model. The input image undergoes convolution and activation functions in the backbone network to extract features, resulting in three feature layers. The YOLOv8n model continues to use the YOLOv5 model's SPPF (Spatial Pyramid Pooling Fast) to concatenate convolution and pooling results, extracting information from different receptive fields [25]. However, the pooling operation in the SPPF structure may lose some local feature information of seedlings, failing to fully represent the detailed root features of seedlings in the floating state. This affects subsequent feature extraction operations in the neck network. Therefore, this paper replaces the original SPPF structure with ASPP (atrous spatial pyramid pooling) [26] in the YOLOv8n model. ASPP introduces multiple atrous convolutions with different sampling rates to control the size of the convolutional kernel's receptive field. The pooling operation allows the network to adaptively determine the receptive field size of the feature map, extracting higher-level feature information. This enhances the network's ability to recognize and detect the same target in different forms while maintaining the model's lightweight and efficiency. The ASPP structure is illustrated in Figure 9.

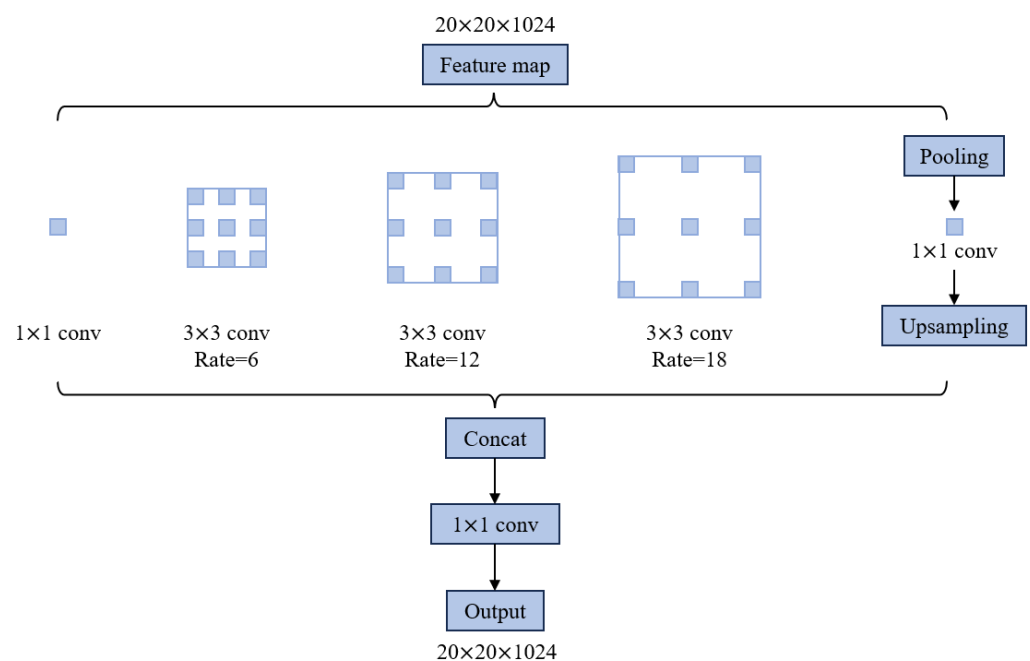


Figure 9. Structure of ASPP (atrous spatial pyramid pooling).

The propagation process of the ASPP structure is as follows: The feature map of the third layer with a size of $20 \times 20 \times 1024$ is input into the ASPP module. The first branch uses a 1×1 convolution to reduce the dimensionality of the feature map obtained in the previous step while preserving the original receptive field. The second to fourth branches employ dilated convolutions with a kernel size of 3×3 and dilation rates of 6, 12, and 18, respectively. These branches stack corresponding dilated convolutional layers to extract features at different scales. The fifth branch directly applies global average pooling to the input feature map, followed by a 1×1 convolution for integration, normalization, and activation functions. Subsequently, upsampling is used to obtain a feature layer of the same size as the branches. Finally, the five branches are integrated and stacked to enhance the network's ability to capture multi-scale context, resulting in a feature map of size $20 \times 20 \times 1024$.

3.2.4. Adding SimAM Attention Mechanism

The paddy field environment is complex and variable, with factors such as exposed soil, deep water layers, water ripples, strong light reflections, and floating straw.

Especially in windy conditions, rice seedling leaves may exhibit varying degrees of lodging in the same direction, resembling the floating state. To avoid the omission or misjudgment of the model, and at the same time to further capture the feature information of normal seedlings and the key details of floating seedlings, in addition to improving the recognition performance of targets with inconspicuous features. Therefore, this study chooses to apply SimAM (Simple, Parameter-free Attention Module) [27] to the main feature extraction network. It is added after the SPPF module in the backbone network.

The studied BAM and CBAM attention modules combine spatial attention and channel attention in parallel or serial combinations, which can only refine features in one dimension [28]. But, the SimAM attention mechanism introduces a three-dimensional modular architecture in space and channel without adding any extra parameters. It also enhances the flexibility and effectiveness of the module in terms of its ability to improve the representation of convolutional networks in space and channels. The computation of attention weights is accelerated and a lightweight attention module is obtained, as shown in Figure 10.

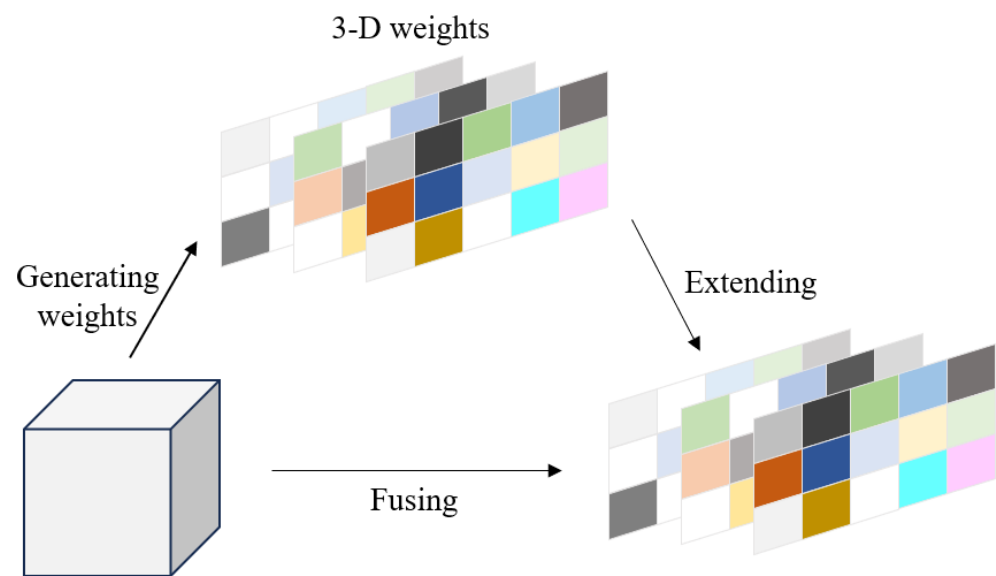


Figure 10. Simple, parameter-free attention module.

This module draws inspiration from the phenomenon in visual neuroscience where active neurons can inhibit the activity of surrounding neurons. It evaluates the importance of each neuron based on priority, and its energy function is shown in Formula (2).

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{u})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (2)$$

In this context, $u_t = \frac{1}{M-1} \sum_{i=1}^{M-1} x_i$, $\sigma_t^2 = \frac{1}{M-1} \sum_{i=1}^{M-1} (x_i - u_t)^2$, t represents the size of the neuron, u_t and σ_t^2 are mean and variance calculated over all neurons except t in that channel, and the importance of the neuron can be evaluated through $1/e_t^*$. The lower the energy, the greater the difference between the neuron and the surrounding neurons, indicating a higher level of importance.

$$\tilde{\mathbf{X}} = \text{sigmoid}\left(\frac{1}{\mathbf{E}}\right) \odot \mathbf{X} \quad (3)$$

In the equation, $\frac{1}{\mathbf{E}}$ represents the summation of e_t^* across all channels and spatial dimensions, $\odot$ is the logical operator for XOR operation, and sigmoid is used to constrain and suppress excessively large values in $\mathbf{E}$ without affecting the relative importance of each neuron. Through this operation, an enhanced feature matrix $\tilde{\mathbf{X}}$ is obtained.

3.2.5. Neck Network Optimization

The neck network of the YOLOv8n model consists of FPN (Feature Pyramid Network) and PAN (Path Aggregation Network), forming a pyramid structure named PAFPN (Path Aggregation Network with Feature Pyramid Network). However, this structure is relatively simple, limited to the simple addition of feature information between adjacent levels, and there still exists a semantic gap between non-adjacent levels. To address this issue, this study introduces AFPN (Asymptotic Feature Pyramid Network) [29] to enhance the interaction capability between non-adjacent levels. The two pyramid structures are illustrated in Figure 11.

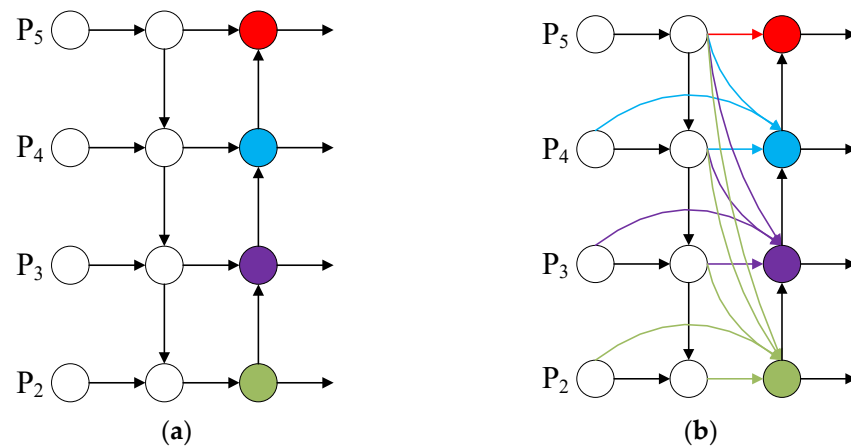


Figure 11. Two types of neck network structures. (a) PANet; (b) AFPN.

According to the architecture diagram of AFPN, progressive feature pyramids undergo multiple upsampling operations. Through multiple rounds of feature fusion, integrating feature information across hierarchical levels effectively avoids the loss of channel features while reducing semantic gaps between non-adjacent levels. In addition, the AFPN structure utilizes the ASFF module (Adaptively Spatial Feature Fusion) [30] for adaptive spatial fusion in multi-level feature integration. It allocates different spatial weights, emphasizing key hierarchical information and reducing conflicting information in the fusion process at each stage. Finally, the model achieves effective feature fusion from bottom to top, enhancing the distinctive morphological and local detailed features of seedlings in two states. Enriching semantic information improves the image resolution features, further strengthening the robustness and generalization of seedling detection in different paddy field environments.

4. Experiments

4.1. Experimental Setup

The hardware configuration for the network training in this experiment includes an Intel Core i7-12700H processor, 16 GB of RAM with a clock frequency of 4800 MHz, and an NVIDIA GeForce RTX 3060 graphics card. PyTorch is employed as the deep learning framework, complemented by the General-purpose computing on Graphics Processing Units (GPGPU) architecture CUDA 11.7 and the deep neural network GPU acceleration library CuDNN 8.9.2. The convergence of the improvement algorithm is analyzed and assessed through the convergence level of the loss curve. The model training is conducted with 300 epochs, a batch size of 16, and an initial learning rate set to 0.01 by default.

4.2. Model Evaluation Metrics

For the evaluation of seedling detection results, this paper adopts precision, recall, X_{AP} (average precision), and X_{mAP} (mean of average precision) as indicators to evaluate the

performance of the detection model. The model's operational recognition speed is judged using Frames Per Second (FPS). The calculation formulas are as follows:

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

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

$$X_{AP} = \int_0^1 P(R) dR \quad (6)$$

$$X_{mAP} = \frac{\sum_{C=1}^C X_{AP}(C)}{C} \times 100\% \quad (7)$$

In the equation, X_{TP} (True Positive) represents the number of samples correctly detected as qualified seedlings and correctly recognized as floating seedlings, X_{FP} (False Positive) represents the number of samples incorrectly detected as qualified seedlings and incorrectly recognized as floating seedlings, X_{FN} (False Negative) represents the number of seedlings missed in the image, C denotes the number of data categories, and, in this study, the detection targets include qualified seedlings and floating seedlings, hence $C = 2$.

5. Results and Analysis

5.1. Experimental Results

A total of 273 images from the test set were randomly selected for the detection of floating seedlings, missing seedlings, and qualified seedlings. Simultaneously, the improved YOLOv8n model was compared with the original model using experimental data. Three scenarios were selected for seedling morphological detection: normal seedlings, continuous missing seedlings, and floating seedlings with missing seedlings, and exposed soil in paddy fields, as shown in Figure 12.

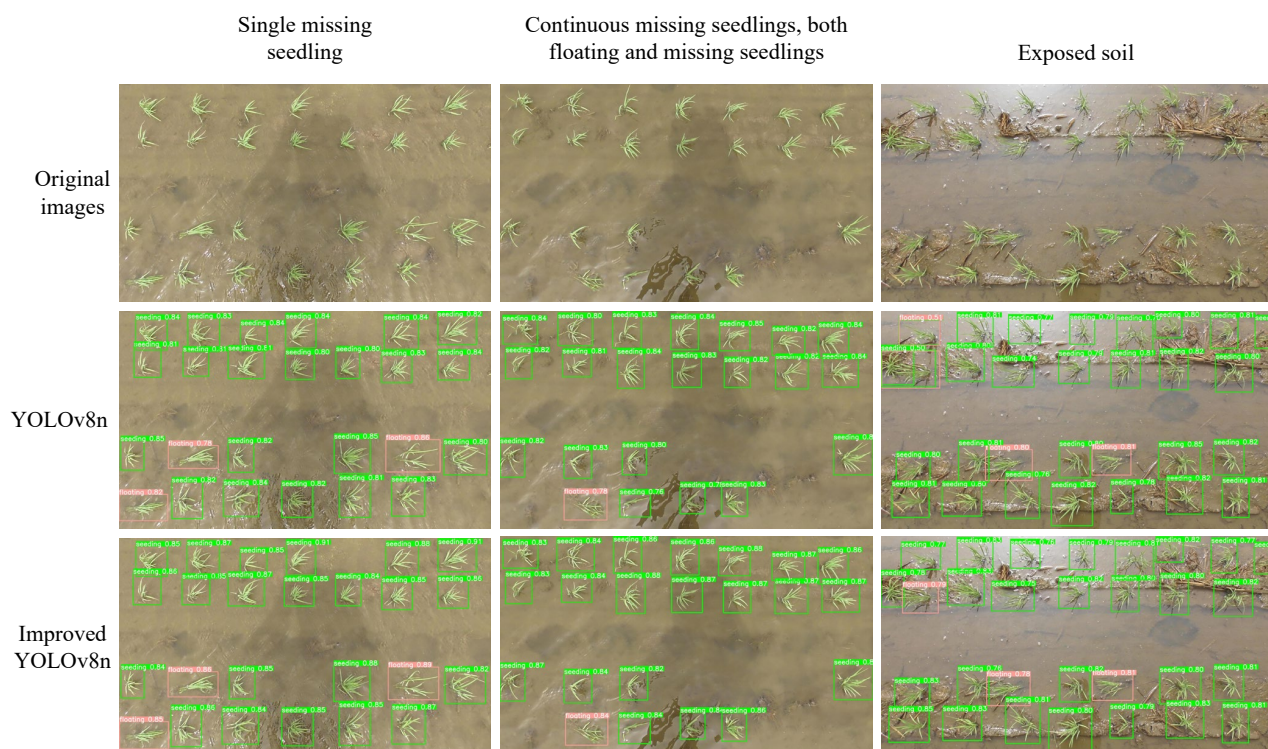


Figure 12. Comparison of seedling quality and floating seedling detection before and after the improvement of YOLOv8n.

As shown in Figure 11, due to the similarity in features between qualified seedlings and floating seedling states, the original YOLOv8n model is prone to false positives and false negatives when encountering complex scenes resembling exposed soil. Additionally, it may produce confused bounding box predictions when detecting indistinct targets, with lower confidence scores. In contrast, the improved YOLOv8n model demonstrates precise and rapid identification of seedling targets during detection, with slightly enhanced confidence scores. The utilization of the enhanced model along with the missing seedling detection algorithm accurately assesses the planting status of seedlings, as depicted in Figure 13.

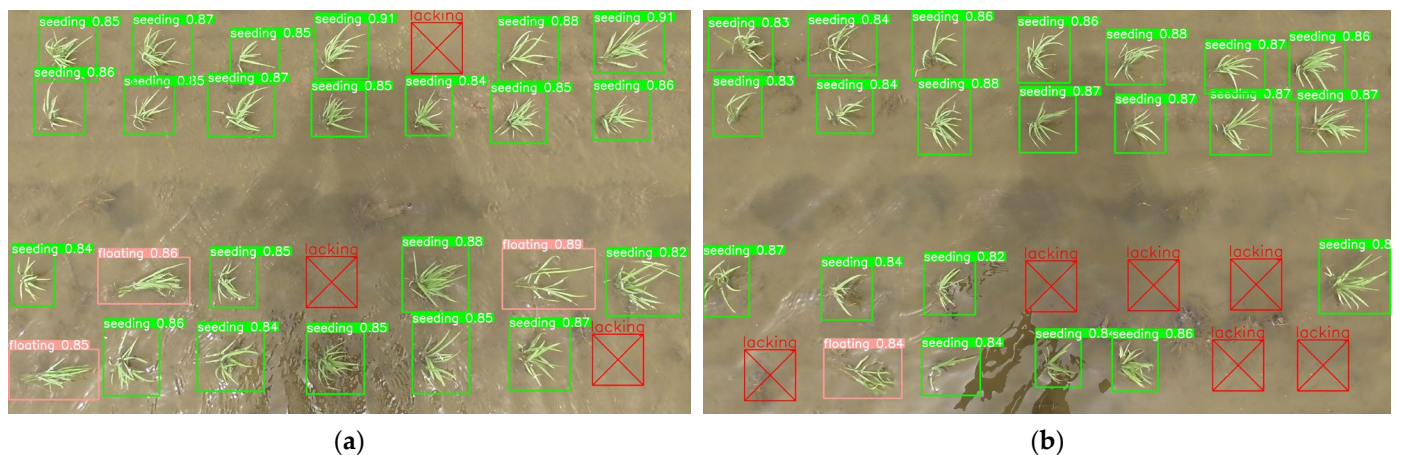


Figure 13. Seedling floating and qualified state detection. (a) Single missing seedling; (b) continuous missing seedlings, both floating and missing seedlings.

5.2. Ablation Experiment

In order to achieve precise and efficient detection of seedling planting status, this study replaced the original YOLOv8n model with ASPP, added the SimAM attention module, and introduced AFPN. Through ablation comparative experiments, the changes and improvements of various parameters before and after the improvement can be more intuitively analyzed. The experimental results are shown in Table 1.

Table 1. Comparison of ablation experiment performance.

Model	Precision/%	Recall/%	mAP/%	FPS	Model Size/MB
YOLOv8n	91.9	91.8	93.5	120	6.3
①	93.9	90.6	94.1	104	4.6
②	89.3	92.1	95.4	90	9.0
③	92.0	92.6	94.5	92	4.8
Improved YOLOv8n	95.5	92.7	95.2	113	8.8

Note: Model ① refers to the introduction of AFPN; Model ② introduces ASPP and AFPN; Model ③ incorporates the SimAM attention mechanism and AFPN.

As shown in the table, the introduction of the AFPN pyramid structure effectively improves the efficiency of network feature extraction, with an increase in accuracy by 2%. Meanwhile, there is a significant decrease in model weight size and computational complexity, maintaining the lightweight principle of the model. When using the ASPP module as the new model with the AFPN structure, there is a slight increase in recall rate and mAP. If the SimAM attention module is directly added to the end of the backbone network, there is an improvement in accuracy and recall rate, but the FPS decreases to 92. Ultimately, the improved YOLOv8n model simultaneously introduces three modules,

resulting in improvements in accuracy, recall rate, and mAP by 3.6%, 0.9%, and 1.7%, respectively. However, compared to the original model, there is a slight decrease in inference speed, an increase in weight size, and computational complexity. This indicates that the introduction of each structure increases the complexity of the model but enhances its information extraction capability and feature fusion properties.

The four optimization methods in the table above all show improvements in parameter metrics compared to the original model. The accuracy, recall rate, and mAP curves of the five models are depicted in Figure 14. Through ablation experiments, it is concluded that optimizing both the backbone network and the neck network of the YOLOv8n model simultaneously, while slightly decreasing detection speed, does not compromise the real-time detection requirements for seedlings. Additionally, it enhances the model's ability to recognize the planting status of seedlings.

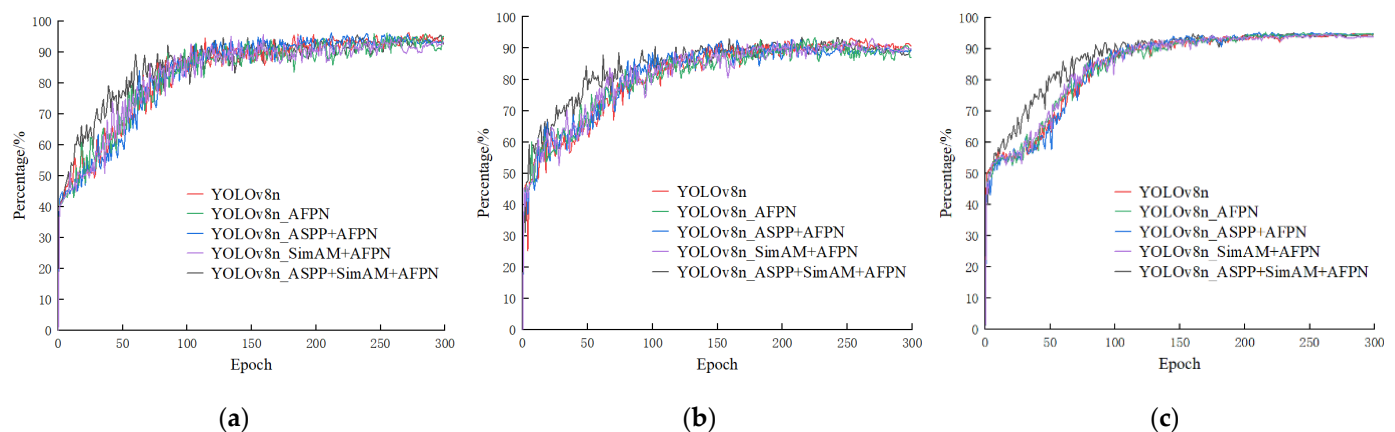


Figure 14. Performance parameter variation curve based on different improved algorithms for YOLOv8n. (a) Precision; (b) recall; (c) mAP.

5.3. Comparison Experiment of Different Models

The improved YOLOv8n model is experimentally compared with other mainstream two-stage detection models such as Faster-RCNN and single-stage detection models including SSD, YOLOv5, YOLOv7, and YOLOv8n. The experimental results are presented in Table 2.

Table 2. Comparison of experimental results of different network models.

Models	mAP/%	FPS	Model Size/MB
Faster-RCNN	83.9	15	124.2
SSD	67.6	34	92.1
YOLOv5	90.3	125	13.8
YOLOv7	91.6	103	74.8
YOLOv8n	91.9	120	6.3
Improved YOLOv8n	95.2	113	8.8

From the table, it can be observed that Faster-RCNN has the lowest average precision, slowest inference speed, and largest weight, making it unsuitable for seedling planting status detection. The original YOLOv8n model exhibits the highest overall performance, with an inference speed of 120, slightly lower than the YOLOv5 model, but still meets the real-time detection requirements, demonstrating the feasibility of using this model for seedling planting status detection.

The average precision of the improved YOLOv8 model surpasses other models, with improvements of 11.3%, 27.6%, 4.9%, and 3.3%, respectively. Due to the increased complexity of the improved model, the inference speed decreases from 120 to 113, and the weight

size increases from 6.3 MB to 8.8 MB. However, the focus on improving the recognition accuracy of qualified seedlings and floating seedlings detection in the improved model helps avoid issues such as missed detections and false alarms. Moreover, the slight decrease in frame rate and minor increase in weight do not affect the real-time detection requirements or the practical application of the model in real-world environments.

6. Conclusions

- (1) Addressing the impact of complex paddy field environments and other interfering factors on seedling missing detection, as well as various scenarios such as isolated missing seedlings, continuous missing seedlings, and floating seedlings with missing seedlings, this paper proposes a centroid detection method. By calculating the distance between adjacent seedlings and selecting appropriate thresholds, it can swiftly locate and assess the missing seedling situation between adjacent seedlings, achieving an overall accuracy of 93.7%.
- (2) For the detection of seedling qualification and floating status, this paper improves upon the YOLOv8n model. Extraction of multi-scale feature information of seedling status using ASPP structure to replace SPPF structure. Additionally, a SimAM attention module is added to the backbone network to enhance the extraction of crucial information. Moreover, the introduction of AFPN enhances feature fusion efficiency while maintaining network lightweight, thereby further improving detection performance.
- (3) Based on the actual conditions of paddy fields, this paper constructs datasets for seedling qualification and floating status to train, validate, and test the model. Compared with the original model, the improved YOLOv8n model in this study shows enhancements in accuracy, recall rate, and average precision of seedling recognition by 3.6%, 0.9%, and 1.7%, respectively. Despite the increase in model complexity, the weight size has increased by 2.5 MB, and the FPS has decreased to 113. However, it still meets the requirements for lightweight and real-time detection. The seedling planting status recognition model consists of centroid detection and the YOLOv8 network. This model demonstrates good performance in seedling detection under actual field conditions and can provide a reference for monitoring the quality of transplanting operations and the development of unmanned rice field management in the future.
- (4) The study in this paper shows that this assay has high accuracy and stability in detecting the rice seedling planting status. Crops such as corn, wheat, and soybean, for example, also require an insertion process, and the method could be considered for application to these crops in the future. It can also be used for intelligent applications such as detecting the growth status and health of crops, farm monitoring, and management.

Overall, different agricultural environments may have conditions such as light, soil type, humidity, and temperature, which may affect the performance of the detection algorithm. Therefore, applying the method to different agricultural environments and evaluating its performance and stability are important directions for future research.

Author Contributions: Conceptualization, B.Z. (Bo Zhao); methodology, B.Z. (Bo Zhao) and Q.Z.; validation, Q.Z.; writing—original draft preparation, Q.Z.; writing—review and editing, B.Z. (Bo Zhao), Q.Z., Y.L., Y.C. and B.Z. (Baixue Zhou); investigation, B.Z. (Bo Zhao); supervision, Y.L. and Y.C.; funding acquisition, B.Z. (Bo Zhao) and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China, Grant No. 2021YFD2000601.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data of the experimental images used to support the findings of this research are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy.

Acknowledgments: The authors would like to extend their appreciation to National Key Laboratory of Agricultural Equipment Technology, Beijing and Qingdao Agricultural University, Qingdao, for supporting the project possible.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Singh, P.K.; Kumar, A.; Naresh, R.K.; Singh, V.K.; Tiwari, H.; Bhatt, R.; Tomar, S.S.; Chandra, M.S.; Singh, P.K.; Mahajan, N.C.; et al. Mechanized Paddy Transplanting Impacts on the Productivity and Sustainability of the Rice Cultivation System in North West IGP: A Review. *Agric. Mech. Asia* **2023**, *54*, 13529–13544.
2. Yin, X.; Du, J.; Noguchi, N.; Yang, T.; Jin, C. Development of autonomous navigation system for rice transplanter. *Int. J. Agric. Biol. Eng.* **2018**, *11*, 89–94. [\[CrossRef\]](#)
3. Wang, S.; Wang, X. An Identification Algorithm of Rice Seedling Rows with Seedlings Absence Based on CS-Yolov5 Model. In Proceedings of the 2023 5th International Conference on Robotics Systems and Automation Engineering (ICRSAE), Prague, Czech Republic, 20–23 April 2023; IEEE: Prague, Czech Republic, 2023; pp. 41–48.
4. Jiang, F.; Lu, Y.; Chen, Y.; Cai, D.; Li, G. Image recognition of four rice leaf diseases based on deep learning and support vector machine. *Comput. Electron. Agric.* **2020**, *179*, 105824. [\[CrossRef\]](#)
5. Tseng, H.H.; Yang, M.D.; Saminathan, R.; Hsu, Y.C.; Yang, C.Y.; Wu, D.H. Rice seedling detection in UAV images using transfer learning and machine learning. *Remote Sens.* **2022**, *14*, 2837. [\[CrossRef\]](#)
6. Chen, J.; Tao, Y.; Li, R.; Xiao, Y.; Liu, D.; Zhou, Z.; Zhu, Y.; Wang, L.; Wang, J.; Cao, Q. A vision-based partial paddy field sampling method for rice seedling detection. *Biosyst. Eng.* **2017**, *155*, 21–32.
7. Hayashi, S.; Kobayashi, K.; Yamamoto, S.; Watanabe, M.; Amano, Y.; Kamata, J.; Sakai, K.; Umeda, M. Rice-seedling planting robot using Enrique visual feedback. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS), Detroit, MI, USA, 1–5 October 2013; IEEE: Tokyo, Japan, 2013; pp. 4106–4112.
8. Xu, B.; Chai, L.; Zhang, C. Research and application on corn crop identification and positioning method based on Machine vision. *Inf. Process. Agric.* **2023**, *10*, 106–113. [\[CrossRef\]](#)
9. Quan, L.; Feng, H.; Lv, Y.; Wang, Q.; Zhang, C.; Liu, J.; Yuan, Z. Maize Seedling detection under different growth stages and complex field environments based on an improved Faster R-CNN. *Biosyst. Eng.* **2019**, *184*, 1–23. [\[CrossRef\]](#)
10. Nawaz, M.; Nazir, T.; Khan, M.A.; Rajinikanth, V.; Kadry, S. *Plant Disease Classification Using VGG-19 Based Faster-RCNN International Conference on Advances in Computing and Data Sciences*; Springer Nature: Cham, Switzerland, 2023; pp. 277–289.
11. Yu, G.; Fan, H.; Zhou, H.; Wu, T.; Zhu, H. Vehicle target detection method based on improved SSD model. *J. Artif. Intell.* **2020**, *2*, 125. [\[CrossRef\]](#)
12. Gao, Z.; Wu, B.; Zheng, C.; Wang, Y.; Xie, C.; Wang, D.; Li, Y.; Zhang, Q.; Chen, Q. Intelligent Detection System for the Missing Hill of Rice Seedling Based on Improved YOLOv5. *Comput. Electron. Agric.* **2022**, *193*, 107293.
13. Zhu, W.; Ma, L.; Zhang, P.; Liu, D.Y. Morphological recognition of rice seedlings based on GoogLeNet and UAV images. *J. South China Agric. Univ.* **2022**, *43*, 99–106.
14. Liao, J.; Wang, Y.; Yin, J.; Zhang, S.; Liu, L.; Zhu, D. Detection of seedling row centerlines based on sub-regional feature points clustering. *Nongye Jixie Xuebao/Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 134–141.
15. Li, B.; Xu, X.; Zhang, L.; Han, J.; Bian, C.; Li, J.; Liu, J.; Jin, L. Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 161–172. [\[CrossRef\]](#)
16. Onur, T.Ö. Improved image denoising using wavelet edge detection based on Otsu's thresholding. *Acta Polytech. Hung.* **2022**, *19*, 79–92. [\[CrossRef\]](#)
17. Gao, K.; Chen, Z.D.; Weng, S.; Zhu, H.P.; Wu, L.Y. Detection of multi-type data anomaly for structural health monitoring using pattern recognition neural network. *Smart Struct. Syst.* **2022**, *29*, 129–140.
18. Mo, Y.; Wu, Y.; Yang, X.; Liu, F.; Liao, Y. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing* **2022**, *493*, 626–646. [\[CrossRef\]](#)
19. Atkinson, D.; Becker, T. A 117 line 2D digital image correlation code written in MATLAB. *Remote Sens.* **2020**, *12*, 2906. [\[CrossRef\]](#)
20. Sekehravani, E.A.; Babulak, E.; Masoodi, M. Implementing canny edge detection algorithm for noisy image. *Bull. Electr. Eng. Inform.* **2020**, *9*, 1404–1410. [\[CrossRef\]](#)
21. Lee, Y.J.; Shin, B.S. Development of potato yield monitoring system using machine vision. *J. Biosyst. Eng.* **2020**, *45*, 282–290. [\[CrossRef\]](#)
22. Terven, J.; Cordova-Esparza, D. A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond. *arXiv* **2023**, arXiv:2304.00501.
23. Wang, X.; Gao, H.; Jia, Z.; Li, Z. BL-YOLOv8: An improved road defect detection model based on YOLOv8. *Sensors* **2023**, *23*, 8361. [\[CrossRef\]](#)

24. Li, B.; Zhang, J.; Liang, Y. PaFPN-SOLO: A SOLO-based Image Instance Segmentation Algorithm. In Proceedings of the IEEE 2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML), Hangzhou, China, 25–27 March 2022; IEEE: Hangzhou, China, 2022; pp. 557–564.
25. Wan, J.; Fu, Y.; Jiang, W.; Pu, S.; Yu, J. Tackling Over-Smoothing: Graph Hollow Convolution Network with Topological Layer Fusion. In Proceedings of the IEEE 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 18–23 July 2022; IEEE: Padua, Italy, 2022; pp. 1–9.
26. Li, R.; Wu, Y. Improved YOLO v5 wheat ear detection algorithm based on attention mechanism. *Electronics* **2022**, *11*, 1673. [\[CrossRef\]](#)
27. Yang, L.; Zhang, R.Y.; Li, L.; Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In *International Conference on Machine Learning, Virtual Event, 18–24 July 2021*; PMLR: Cambridge, MA, USA, 2021; pp. 11863–11874.
28. Zhang, Y.; Song, C.; Zhang, D. Deep learning-based object detection improvement for tomato disease. *IEEE Access* **2020**, *8*, 56607–56614. [\[CrossRef\]](#)
29. Huang, Z.; Li, L.; Krizek, G.C.; Sun, L. Research on Traffic Sign Detection Based on Improved YOLOv8. *J. Comput. Commun.* **2023**, *11*, 226–232. [\[CrossRef\]](#)
30. Yang, G.; Wang, J.; Nie, Z.; Yang, H.; Yu, S. A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention. *Agronomy* **2023**, *13*, 1824. [\[CrossRef\]](#)

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