



Article

A Precise Plot-Level Rice Yield Prediction Method Based on Panicle Detection

Junshuo Wei ^{1,2}, Xin Tian ^{1,2}, Weiqi Ren ^{1,2}, Rui Gao ^{1,2}, Zeguang Ji ^{1,2}, Qingming Kong ^{1,2,*} and Zhongbin Su ^{1,2,*} ¹ Key Laboratory of Northeast Smart Agricultural Technology, Ministry of Agriculture and Rural Affairs, Harbin 150030, China² School of Electrical and Information, Northeast Agricultural University, Harbin 150030, China

* Correspondence: kkqmmmm@neau.edu.cn (Q.K.); suzb001@163.com (Z.S.)

Abstract: Accurately estimating rice yield is essential for ensuring global food security, enhancing agricultural productivity, and promoting agricultural economic growth. This study constructed a dataset of rice panicles at different growth stages and combined it with an attention mechanism and the YOLOv8 network to propose the YOLOv8s+LSKA+HorNet rice panicle detection and counting model, based on a drone remote sensing platform. Using the panicle count data collected by this model, along with the thousand-grain weight, number of grains per panicle, and actual yield data from a rice nitrogen gradient experimental field, various machine learning models were trained to ultimately propose a field-level rapid rice yield estimation model, RFYOLO. The experimental results show that the rice panicle detection and counting model can achieve an average precision (AP) of 98.0% and a detection speed of 20.3 milliseconds. The final yield estimation model achieved a prediction R^2 value of 0.84. The detection and counting model significantly reduced missed and duplicate detections of rice panicles. Additionally, this study not only enhanced the model's generalization ability and practicality through algorithmic innovation but also verified the impact of yield data range on the stability of the estimation model through the rice nitrogen gradient experiment. This is significant for early rice yield estimation and helping agricultural producers make more informed planting decisions.



Citation: Wei, J.; Tian, X.; Ren, W.; Gao, R.; Ji, Z.; Kong, Q.; Su, Z. A Precise Plot-Level Rice Yield Prediction Method Based on Panicle Detection. *Agronomy* **2024**, *14*, 1618. <https://doi.org/10.3390/agronomy14081618>

Academic Editors: Vjekoslav Tadić, Dorijan Radočaj and Mladen Jurišić
Received: 25 June 2024
Revised: 17 July 2024
Accepted: 22 July 2024
Published: 24 July 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

The estimation of plot-level rice yield is crucial for the development and practical needs of precision agriculture [1]. Precision agriculture aims to accurately manage and predict the growth conditions, status, and yield of crops through the integration of advanced information technology and agricultural techniques. Plot-level yield estimation is particularly important in this context, as it provides precise agricultural decision support for farmers and farm managers for specific plots [2,3]. With the development of remote sensing technology, UAV aerial photography, GIS applications, and machine learning algorithms, plot-level rice yield estimation has become more feasible and effective [4]. These technologies provide high-resolution, high-spatiotemporal-density crop growth monitoring data, enabling more accurate assessments of crop growth conditions at different growth stages. Plot-level rice yield estimation is significant for refining the crop management process and improving agricultural production efficiency [5].

Early rice yield estimation methods primarily utilized remote sensing technology. Wang et al. explored the use of hyperspectral and multispectral remote sensing technology to estimate rice yields [6]. Their study aimed to improve the accuracy and efficiency of rice yield estimation by analyzing vegetation indices, chlorophyll content, and other spectral features related to rice yield. A study by Duan et al. developed an innovative method for extracting rice growth parameters that combined the spectral and texture features of

UAV multispectral imagery [7]. However, these methods faced challenges such as similar spectra for different substances, different spectra for the same substance, cloud and fog cover, and the lack of ideal adaptability and stability of estimation accuracy [8]. Therefore, although these methods have improved the accuracy, efficiency, and reliability of rice yield estimation to some extent, further research and improvement are needed to overcome these challenges.

As the application of deep learning in object detection and counting within agriculture becomes increasingly widespread, significant advancements have been made [9]. Yu et al. proposed a method utilizing convolutional neural networks (CNNs) to analyze over 22,000 vertically downward digital images taken at the rice maturity stage and harvest time, successfully predicting 68% of the yield variation with a relatively small error (relative root mean square error of 0.22) [10]. Plot-level rice yield estimation has thus become a feasible option for improving precision agriculture. The yield of cereal crops is related to the number of panicles per square meter, the number of grains per panicle, and the weight of 1000 grains [11,12]. Therefore, counting the number of panicles is an effective method for estimating predicted rice yield. Detecting objects and drawing bounding boxes around them is a common method for object counting. Besides the number of panicles, object detection can also obtain information about the size and location of panicles [13]. Xu et al. proposed the MHW-PD model, which enhances the expression of rice panicle features by dynamically selecting feature learning networks and constructing an adaptive Multi-Scale Hybrid Window (MHW) to improve the detection and counting accuracy of small rice panicles [14]. Yang et al. proposed an improved YOLOv4 object detection model for wheat ear detection, achieving accuracy levels of 94%, 96.04%, and 93.11% for wheat ears of three different density distributions [15]. Tan et al. designed the RiceRes2Net model, introducing several technical innovations. By improving the Cascade R-CNN architecture and applying Soft-NMS with four different IoU thresholds, the model reduces false suppression of overlapping targets and improves detection accuracy [16]. Han et al. combined deep learning with remote sensing technology, using high-resolution remote sensing images and a CNN for feature extraction, achieving pixel-level image segmentation through a fully convolutional network (FCN) to accurately identify rice panicle areas [17]. The constructed multivariable regression model incorporated key factors such as the number of panicles per unit area, the number of grains per panicle, the seed setting rate, and the thousand-grain weight, thereby improving prediction accuracy. Stepwise linear regression was used to select feature variables, simplifying the model structure. With the improvement of rice panicle object detection and counting technology, Jia et al. combined deep neural networks—especially the optimized YOLO v5 model—to identify and count rice panicles in the field and estimate rice yield [18]. This method enhances the estimation process by augmenting the rice panicle image dataset, using deep learning models for more accurate identification and counting, and building a rice yield estimation model through regression analysis. Although this represents a complete process from rice panicle counting to yield estimation, the current yield estimation regression equation is too simplistic, with the use of a single parameter limiting the model's large-scale applicability.

In the field of plot-level rice yield estimation research, traditional remote sensing technology faces challenges due to data collection conditions and processing complexity, which reduce the precision and efficiency of estimation. To address these issues, this study proposes a new model that combines object detection and regression algorithms. The goal is to quickly collect data on multiple rice varieties and conditions using drone technology to ensure data diversity and quality. The main contributions are as follows:

- (1) The model adopts an innovative convolutional neural network structure, YOLOv8s+LSKA+HorNet, which significantly improves the accuracy of rice panicle identification by integrating two attention mechanisms and reducing detection time through algorithm optimization, thereby increasing work efficiency. Additionally, adjustments to the loss function reduce the model's dependence on specific quality

- samples, enhancing its generalization ability and effectively reducing the occurrence of missed and false detections.
- (2) The model uses intelligent image segmentation and recombination technology, which not only maintains the integrity of images but also achieves and outputs a more precise count of detected rice panicles.
 - (3) By analyzing the number of panicles in fixed-size pictures and combining multi-dimensional agricultural parameters such as actual per mu rice yield data (where “mu” refers to a traditional Chinese unit of area equivalent to approximately 0.0667 hectares), the weight of 1000 grains, and the number of grains per panicle, the model uses the Random Forest algorithm to train a stable plot-level rice yield estimation model.

This innovative approach improves the accuracy and efficiency of rice yield estimation, effectively resolving the shortcomings of previous methods. These improvements have practical and theoretical significance in advancing the development of precision agriculture.

2. Materials and Methods

2.1. Experimental Location

The experimental site for this study was the Agricultural Extension Center of Northeast Agricultural University in Harbin, Heilongjiang Province, China. It is located in the southwestern part of Heilongjiang Province (approximately $45^{\circ}30'59.95''$ N to $45^{\circ}31'21.75''$ N, $127^{\circ}01'34.28''$ E to $127^{\circ}01'58.39''$ E), at an elevation of about 112 m, covering an area of approximately 330,000 m². We selected 4263 m² of land as our experimental field.

This area is characterized by a cold-temperate continental monsoon climate with distinct seasonal features, an annual average precipitation of 591–783 mm, fertile soil, low terrain, and a soil organic matter content of 3–6%, indicating high natural fertility. Generally, crops are harvested once a year. The experimental station’s cultivated land is primarily paddy fields, with rice as the main crop. The location of the research area is shown in Figure 1.

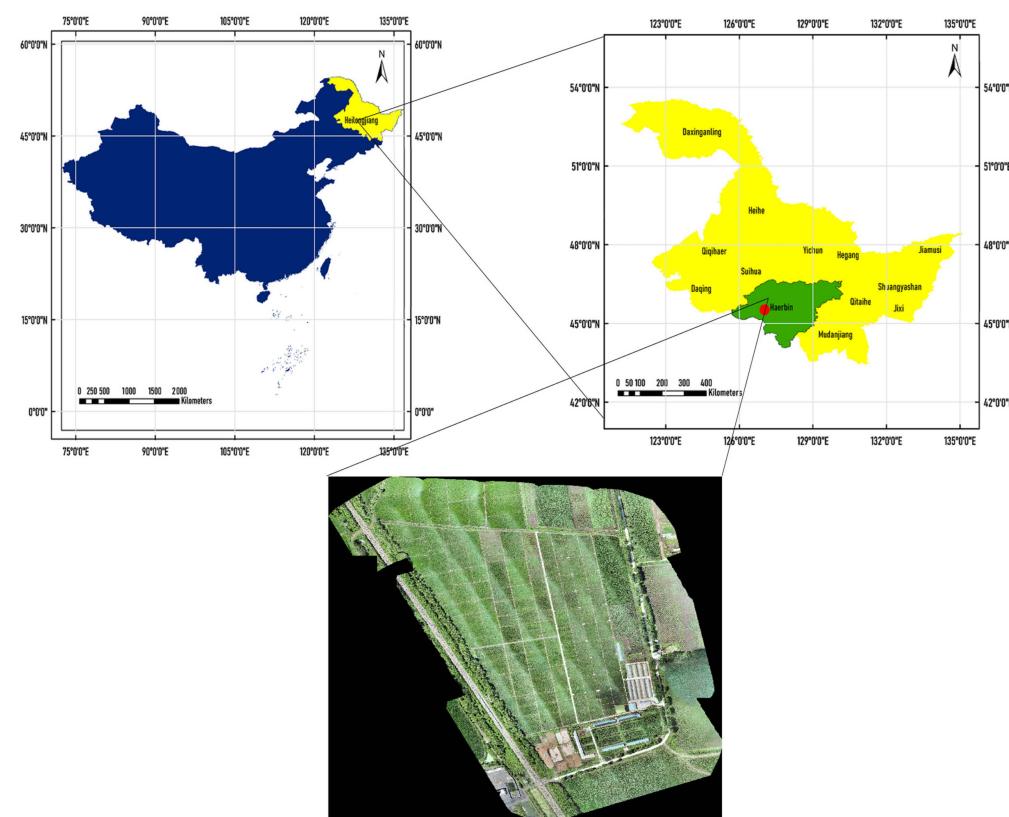


Figure 1. Schematic diagram of the experimental site.

2.2. Experimental Design

This experiment aimed to improve the accuracy and practicality of plot-level rice yield estimation. Using low-altitude remote sensing by drones to collect rice image information, computer vision technology accurately counted the number of rice panicles, thereby achieving precise plot-level rice yield prediction. The experimental design was divided into two main steps: first, accurate identification and counting of rice panicles were achieved using the YOLOv8 object detection algorithm; second, by combining key indicators such as the average number of panicles per area, thousand-grain weight of the rice variety, number of grains per panicle, and actual per mu yield data, a stable yield estimation model was formed by establishing a regression model between these variables and actual yield. To accurately collect the required multi-dimensional data, rice test fields with multiple varieties and conditions were arranged in advance. During the rice heading stage, drones were used for aerial photography to efficiently collect data. Then, during the rice maturity stage, the actual yield data from each test field were collected as important data for model training.

2.3. Experimental Treatment

2.3.1. Selection of Multiple Rice Varieties

The selection of multiple rice varieties aimed to enhance the adaptability of the yield estimation model and enrich the diversity of experimental data. This study selected four representative hybrid rice varieties for in-depth exploration: Longjing 301, Longjing 3010, Lianyu 124, and Suijing 18. These varieties were all planted in the same experimental area to ensure data consistency and comparability. The experimental area was carefully divided into several plots of equal size, each planting only one variety of rice, with each plot being 0.1 acre (about 66.7 m²). Buffers were set between each plot, and signs were placed at the edges to distinguish the different rice varieties. The rice sowing and planting work began in mid-April 2023, and transplantation was carried out in late May. These rice fields adopted the local conventional rice planting methods. The experimental field trial area and the broader planting area followed the traditional management practices of local farmers, ensuring that experimental conditions were consistent with the actual agricultural production environment.

2.3.2. Rice Nitrogen Gradient Experiment

This study implemented gradient experiments with different nitrogen fertilizer concentrations to simulate and capture rice yield variations under various complex environmental conditions, thereby enhancing the data foundation and applicability of the yield estimation model. Specifically, four nitrogen application levels were set for each rice variety: 0, 5, 10, and 15 kg/mu. These four scenarios were referred to as N₀, N₅, N₁₀, and N₁₅. The fertilization strategy was executed as follows: 40% of the total nitrogen as basal fertilizer, 30% as tillering fertilizer, 10% as regulating fertilizer, 20% as panicle fertilizer, and 10% as grain fertilizer. Four identical experimental plots were planted for each scenario, totaling 64 plots. Their distribution is shown in Figure 2. Through this meticulous design, the final yield data for each experimental unit were successfully collected, aiming to further optimize and refine the yield estimation model.

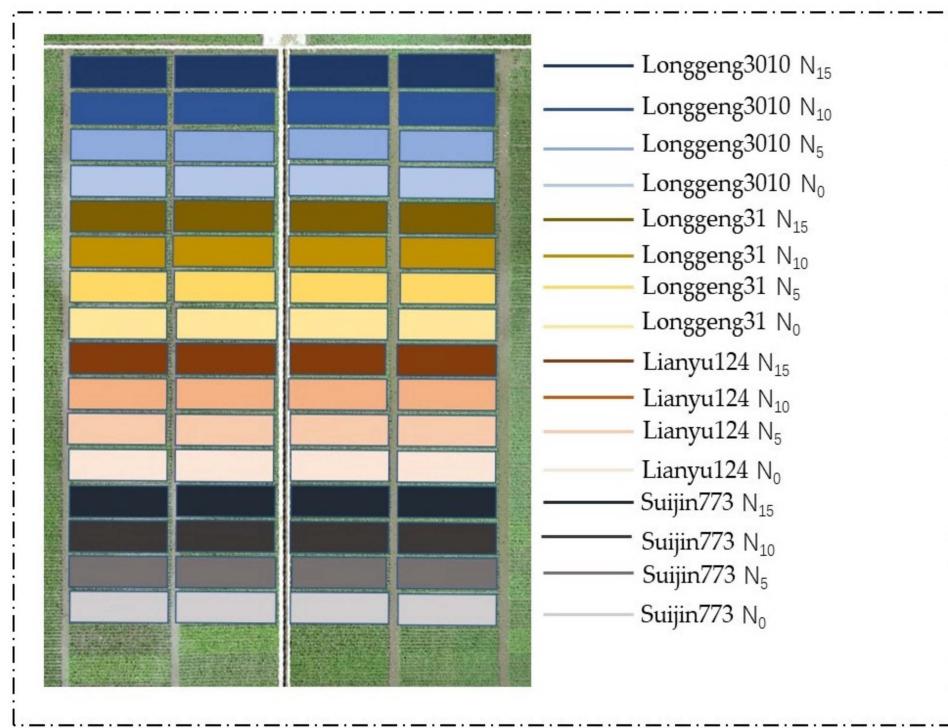


Figure 2. Schematic diagram of the nitrogen gradient experimental field.

2.3.3. Data Collection

Data collection was carried out on 20 July, 31 July, and 11 August 2023, at 9 a.m., and precise image acquisition tasks were executed using the Matrice 300 RTK (Matrice 300 RTK, SZ DJI Technology Co., Ltd., Shenzhen, China) drone equipped with the Zenmuse H20N (Zenmuse H20N, SZ DJI Technology Co., Ltd., Shenzhen, China) multifunctional gimbal camera. To accurately capture images of the rice fields, the camera was set to point directly downward at 90 degrees, ensuring clear and detailed images were obtained from the optimal viewpoint. The core objective of the task was to obtain photographs from a fixed height of 3 m, selecting multiple locations above each experimental field to ensure image diversity. This strategy not only ensured comprehensive coverage of the entire field but also guaranteed high-resolution captured images. During the three days of flight tasks, a total of 1280 RGB images were collected, each with a resolution of 3860×2140 pixels. The images included all 64 experimental fields, with 20 images collected for each field, documenting in detail the changes in rice panicles at different growth stages and under various lighting conditions, providing a rich data resource for the study. The number of rice panicle images is shown in Table 1.

Table 1. Light conditions and number of images at three growth stages.

Growth Period	Illumination Factor	Image Quantity
Early Stage	Weak Light (9 AM)	235
	Strong Light (1 PM)	216
Middle Stage	Weak Light (9 AM)	234
	Strong Light (1 PM)	198
Late Stage	Weak Light (9 AM)	201
	Strong Light (1 PM)	196

2.4. Construction of the Target Detection Dataset and the Rice Panicle Counting Process

To construct an efficient rice panicle image database, various complex rice images were selected during the data collection phase. This database includes clear, blurry, partially or heavily occluded images, and photos taken under various lighting conditions. Such

image diversity helps capture the various states of rice panicles in real-world environments, thereby enhancing the model's performance in practical applications. To enrich the dataset, various image enhancement techniques were employed, including image flipping and rotations of 90°, 180°, and 270°. Furthermore, techniques such as color adjustment and noise addition were introduced. The number of images processed by each data augmentation technique accounted for 10% of the total dataset. These methods not only significantly increased the diversity of the data but also notably enhanced the model's ability to recognize all types of rice panicle forms, especially in the detection of small targets.

During the data processing and preparation phase, the open-source tool “labelimg” was used to precisely annotate the locations of rice panicles, and the images were uniformly resized to a resolution of 512 × 512 pixels. There are several reasons and advantages for dividing images into 512 × 512-pixel sizes. Firstly, this segmentation significantly improves computational efficiency and reduces computational costs and memory consumption, making it possible to process large datasets in limited resource environments, while also speeding up model training and inference. Secondly, it enhances detail capture capabilities, allowing the model to more accurately identify small targets. Additionally, smaller image blocks are more convenient for data augmentation operations, thereby increasing the diversity of training samples and improving the model's robustness and generalization ability. A total of 1016 images under various conditions were ultimately selected as the complete dataset. Finally, the images were divided into training, validation, and test sets in an 8:1:1 ratio. This allocation strategy was intended to ensure balanced training and evaluation of the model across data subsets, enhancing dataset accuracy and robustness in practical applications (Figures 3 and 4).

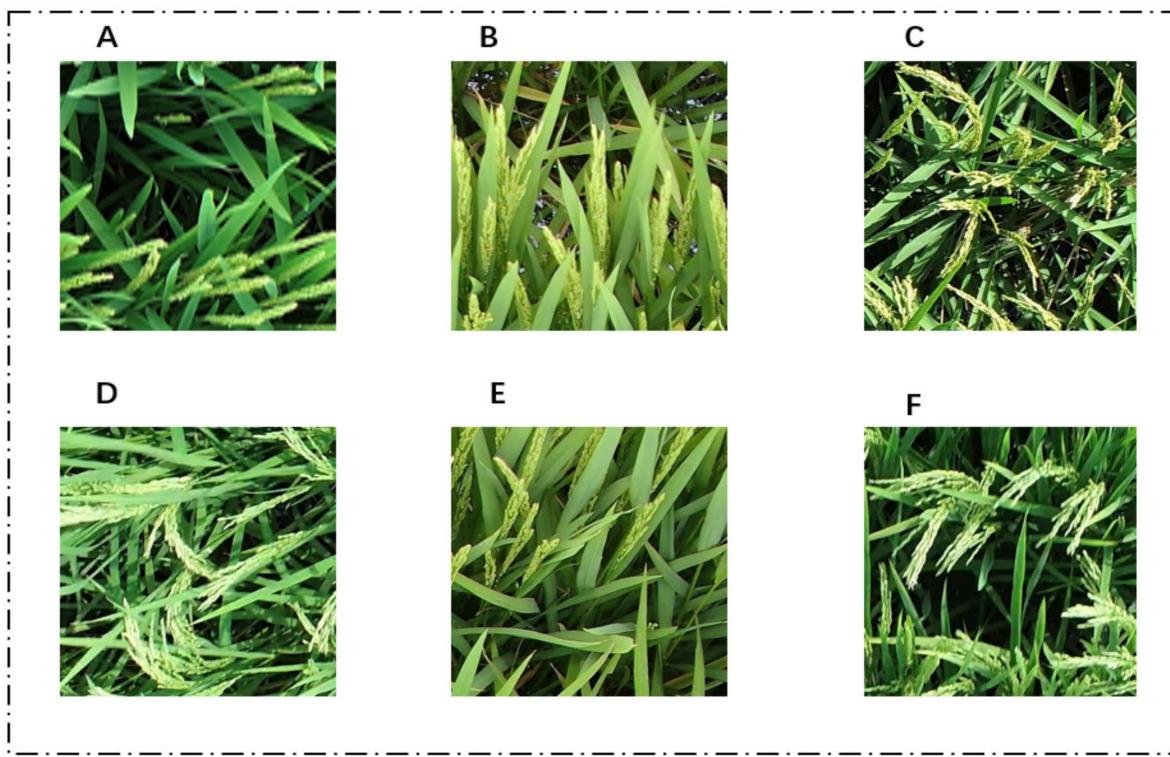


Figure 3. Field rice panicle images. (A) Clear; (B) slight obstruction; (C) strong illumination; (D) high curvature; (E) severe obstruction; (F) severe occlusion.

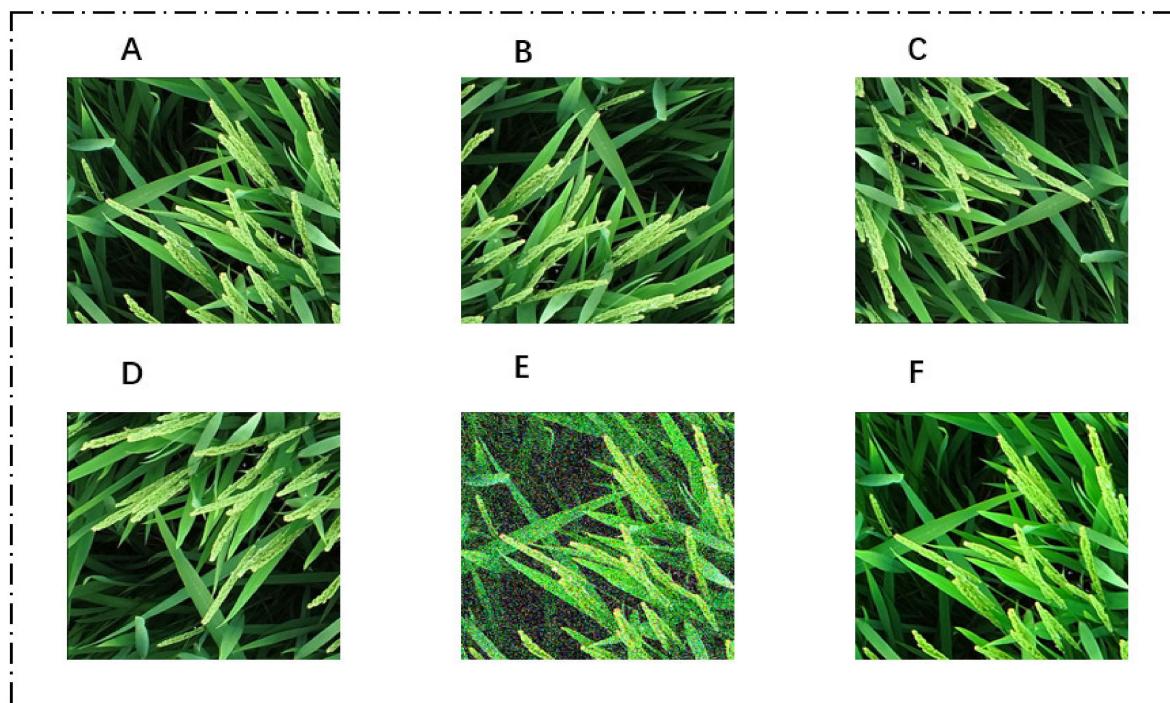


Figure 4. Enhanced field rice panicle images. (A) Original rice panicle image; (B) rotated 90° clockwise; (C) rotated 180° clockwise; (D) rotated 270° clockwise; (E) noise addition; (F) color enhancement.

After the stable detection model was trained, we prepared to detect rice panicle images and collect the number of rice panicles contained in each image. After detection, these small image blocks with detection boxes were reassembled into the complete original large image with detection boxes, as detailed in Figure 5. Finally, after detection and adjustment, we obtained the number of rice panicles contained in each detected image. The comparison of this method with traditional large image detection results can be seen in Figure 6.

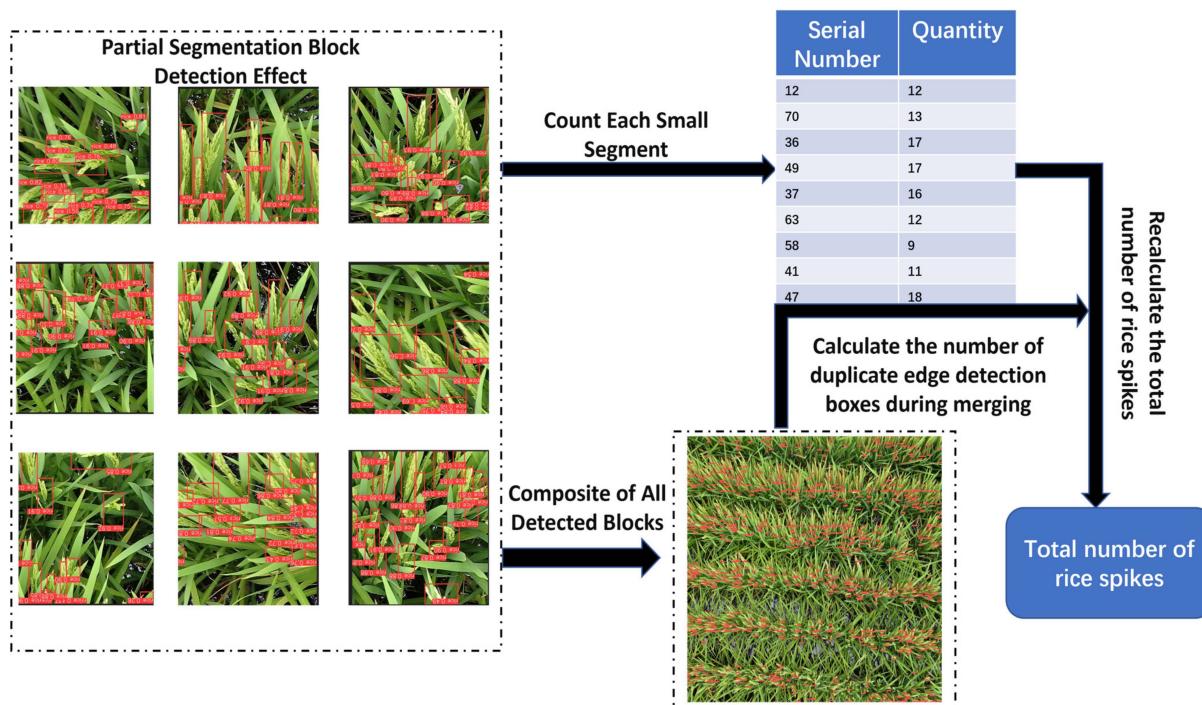


Figure 5. Results of rice panicle prediction and accurate counting.

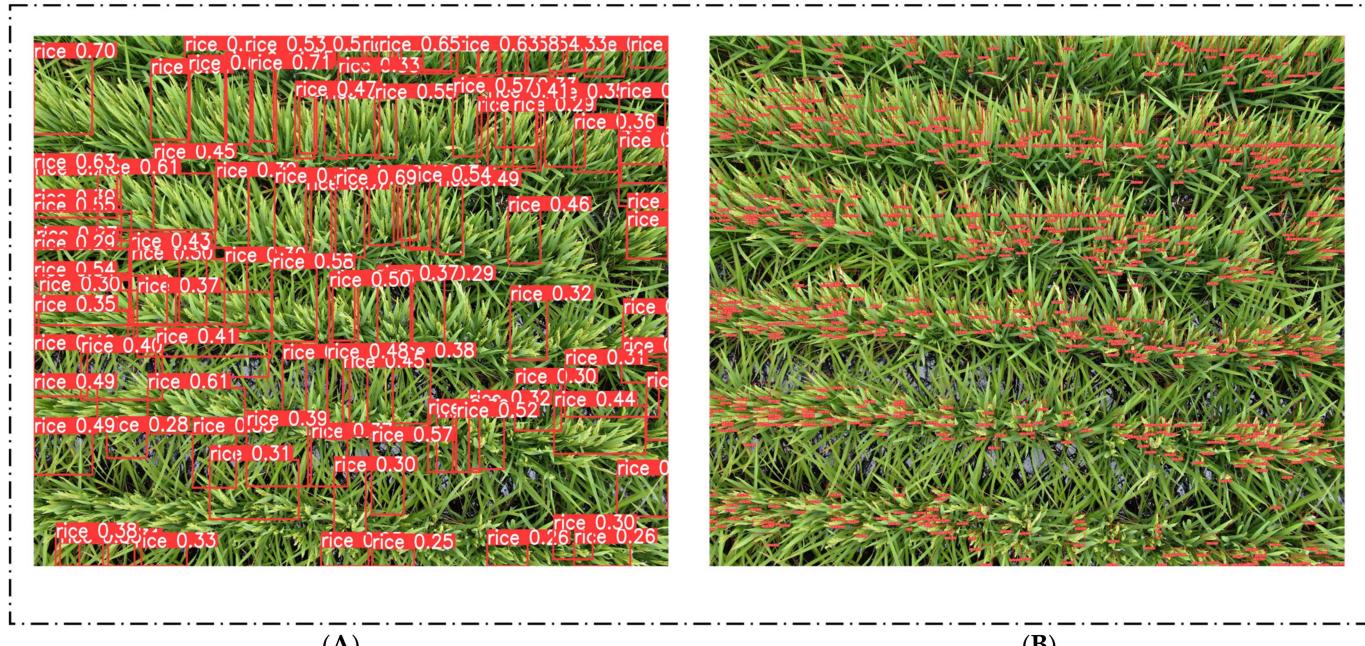


Figure 6. Comparison of detection effects between large image and segmented small image detection. (A) Detection effect on the original image; (B) detection effect after segmenting into small images.

2.5. Object Detection Module

2.5.1. YOLOv8 Detection Model

YOLOv8 is an updated version of YOLOv5, released by Ultralytics on 10 January 2023, supporting image classification, object detection, and instance segmentation [19]. It introduces a state-of-the-art model that improves the backbone network and head structure, adopts an instance segmentation model based on YOLACT, transitions from anchor-based to anchor-free detection, and includes detailed improvements in model structure, loss computation, data augmentation, and training strategy [20]. Specifically, YOLOv8's technical improvements include a backbone network utilizing a gradient-rich C2f structure, an improved loss calculation method, and new task-aligned assigner and distribution focal loss classes, along with adjustments in data augmentation strategies toward the end of training [21].

2.5.2. HorNet

The HorNet model, developed jointly by Tsinghua University and Meta AI, is an advanced deep learning model for complex computer vision tasks [22]. It leverages recursive gated convolution (g^n Conv) structures that circumvent the quadratic complexity problem of self-attention mechanisms and achieves efficient spatial interactions through a pyramid-like design. HorNet expands the modeling capabilities of self-attention to any order and is compatible with multiple types of convolutional kernels, enhancing the model's flexibility and application range. This model is particularly suitable for precise identification and localization tasks such as target detection of rice panicles, making significant contributions to agricultural automation and precision agriculture. The structure diagram of the HorNet module is shown in Figure 7, which includes the following components: MLP (Multi-Layer Perceptron), FFN (Feed-Forward Network), gnConv (Gated Non-linear Convolution), DWCconv (Depthwise Separable Convolution), Proj (Projection), Mul (Multiplication), and Layer Norm (Layer Normalization).

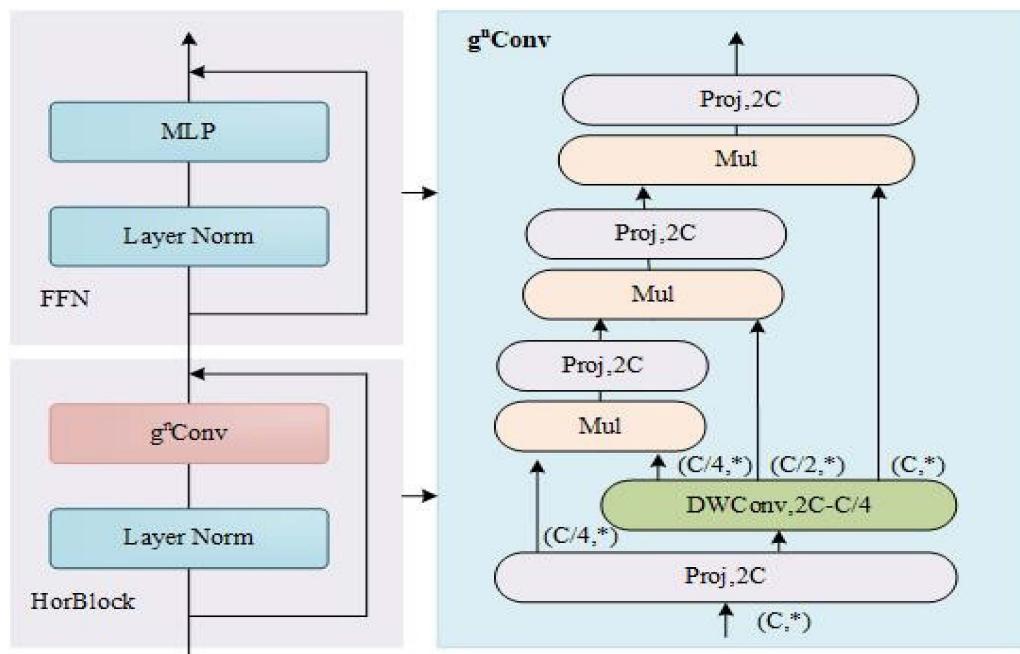


Figure 7. HorNet module structure diagram (* indicates unspecified dimensions in the tensor shape description).

2.5.3. LSKA

The LSKA (Large Separable Kernel Attention) module is an attention mechanism noted for its high computational and memory efficiency, exceptional performance, and superior performance across a variety of visual tasks [23]. By decomposing 2D convolutional kernels into sequences of 1D kernels, LSKA significantly reduces computational complexity and memory requirements, facilitating the use of large kernels. Moreover, with the increase in kernel size, it tends to more effectively capture the shape features of objects. The design philosophy of LSKA has been successfully implemented in YOLOv8, enhancing detection accuracy while lowering the computational and memory costs associated with increased kernel size. The structure diagram of the LSKA module is shown in Figure 8, which includes the following components: DW-Con (Depthwise Convolution), Conv (Convolution).

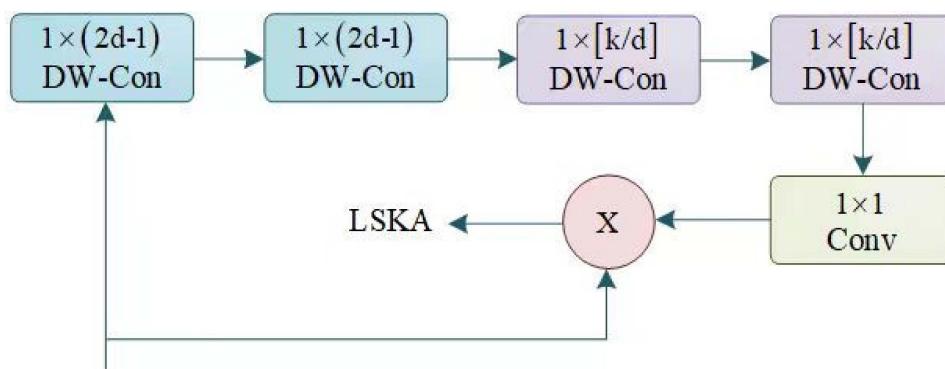


Figure 8. LSKA module structure diagram.

2.5.4. Adding the Loss Function

The Wise-IOU loss function significantly improves the accuracy and generalization ability of object detection through a series of carefully formulated equations [24]. This mechanism allows the model to more intelligently allocate gradient gains, reducing competitiveness among high-quality anchor boxes while also decreasing the harmful gradients

produced by low-quality samples. In the initial version of Wise-IoU (v1), the formula used is

$$LWIoUv1 = RWIoU \times LIoU \quad (1)$$

where $RWIoU$ is a weighting factor used to adjust the traditional IoU loss.

This weighting factor is calculated as

$$RWIoU = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{w^2 + H^2}\right) \quad (2)$$

taking into account the distance between the centers of the box and the target box, where x and y are the center coordinates of the anchor box, x_{gt} and y_{gt} are the center coordinates of the target box, and Wg and Hg are the width and height of the minimum enclosing rectangle. Version 2 of Wise-IoU introduces a monotonic focusing coefficient as

$$L\gamma * IoU \quad (3)$$

where γ is a positive parameter used to adjust the gradient gain.

In version 3, a dynamic non-monotonic focusing coefficient is introduced as

$$r = \frac{\beta}{\delta a \beta - \delta} \quad (4)$$

where IoU represents the formula for the anomaly of the anchor box as

$$\beta = \frac{L * IoU}{LIoU} \quad (5)$$

which is the ratio of the current anchor box's IoU to its historical average IoU value, and δ and α are hyperparameters used to control the sensitivity of the focusing mechanism and adjust the range. The combined effect of these formulas allows Wise-IoU to dynamically adjust gradient gains based on the quality of anchor boxes, thereby improving the accuracy of object detection.

2.5.5. Final Network Structure

Building on YOLOv8, this model achieves seamless integration with the network framework by incorporating three HorNet attention mechanisms into the backbone, replacing the convolutional layers at the 2nd, 4th, and 8th layers, and introducing the LSKA attention mechanism into the network head, replacing the convolutional layers at the 15th and 21st layers. The model's performance is further optimized by replacing the original CIOU loss function with the Wise-IoU loss function. Combined, these improvements result in a more efficient and accurate rice panicle detection, identification, and counting model, utilizing deep learning technology to enhance the recognition of complex rice panicle morphologies. By incorporating structurally complex layers, the model can capture subtle features crucial for accurately identifying small-sized rice panicles in dense fields. The extended receptive field and integrated attention mechanism enhance the capturing of panicle features, especially improving recognition accuracy in cases where the panicle and background colors or textures are similar. The feature fusion strategy effectively integrates information from different layers to enhance multi-scale processing capabilities, providing an accurate and efficient rice panicle detection solution, demonstrating excellent real-world performance. The final network architecture is shown in Figure 9.

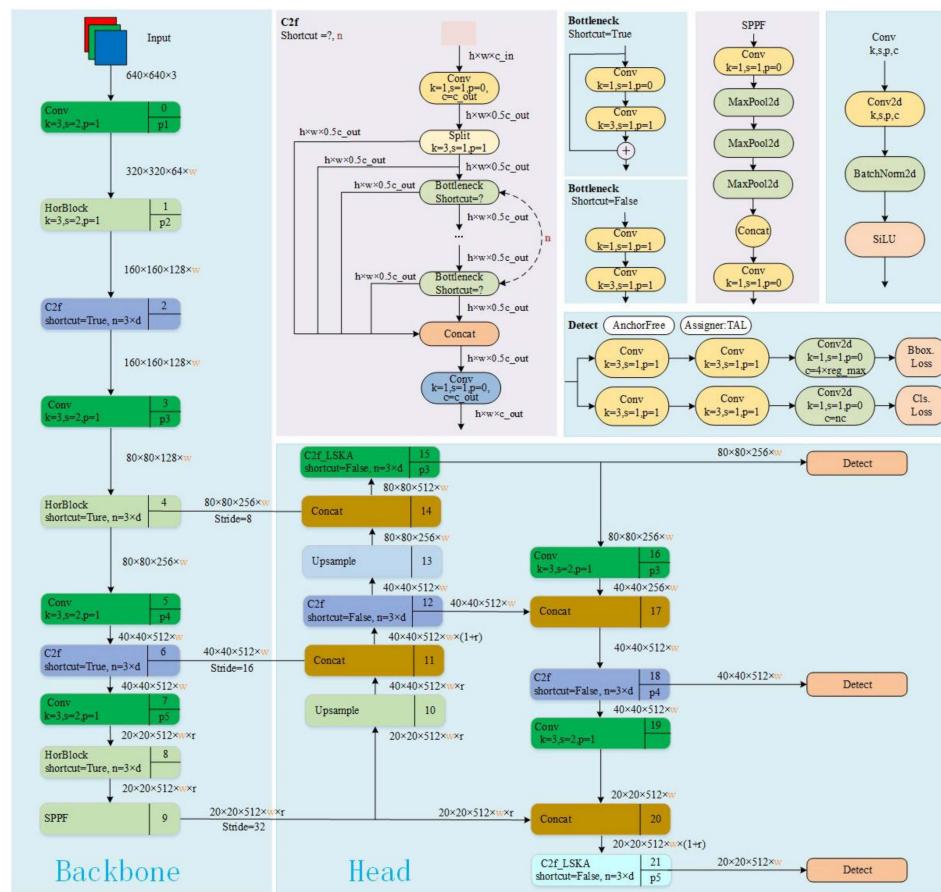


Figure 9. Schematic diagram of the network structure.

2.5.6. Rice Panicle Counting Technique

During the prediction phase, the panicle counting method begins by segmenting the original images into smaller blocks of 512×512 pixels. This segmentation strategy not only optimizes the efficiency of target detection but also ensures accurate identification of rice panicles in small-sized images. After detection, these smaller images are reassembled into the complete original image, annotated with precise detection box markings. To address the issue of duplicate detection boxes at the edges, several key steps were taken. First, an overlapping area was added to each block during segmentation to ensure that targets at the edges are not completely divided into different blocks, thereby improving detection integrity. Next, the model performs object detection on each block containing an overlapping area, saving the detected boxes and their coordinates in the original image. If a detection box is located within the overlapping area, it is marked as a potential duplicate detection box. To eliminate duplicates, we traverse all detection boxes and calculate the IoU value between each pair. If the IoU value of two boxes exceeds the set threshold of 0.9, they are merged into a new detection box using the minimum top-left coordinate and the maximum bottom-right coordinate of the two boxes. Finally, the merged detection boxes are drawn on the original image, and the final image is saved. This method ensures that all duplicate detection boxes are removed, leaving only unique detection boxes, thus providing accurate detection results and statistical data. Through these steps, we effectively address the issue of duplicate detection boxes at the edges, ensuring the accuracy of the statistical results.

2.6. Yield Estimation Model

2.6.1. Yield Estimation Dataset

After establishing a stable model for detecting and counting rice panicles, the photo images collected from the 64 experimental plots are analyzed. For each plot, 10 images are selected, and the number of rice panicles is detected and counted for each image. The average number of panicles in these 10 images is calculated, representing the average number of rice panicles within the 2.23 m^2 of each experimental plot. This method is applied sequentially to monitor and count the average number of rice panicles in the 2.23 m^2 of each of the 64 plots [25]. Statistical analysis of these 64 data points is conducted, followed by consultation of the national rice information database for statistics on the thousand-grain weight and the number of grains per panicle for each variety. After the rice harvest season, the actual yield data from these 64 plots are collected. The final dataset for each plot includes the number of rice panicles within the plot area (2.23 m^2), thousand-grain weight, number of grains per panicle, and actual yield data. This dataset is then split into a training set and a test set at a 6:4 ratio, providing a comprehensive basis for further analysis and evaluation.

2.6.2. Yield Estimation Information and Methods

After calculating the average number of rice panicles per 2.23 m^2 , this study incorporates factors that significantly affect rice yield, such as thousand-grain weight, number of grains per panicle, and actual yield information for the corresponding year. These factors are used as yield estimation data to generate the yield estimation model. The statistical information format is shown in Table 2.

Table 2. Statistical table of yield estimation factors.

Variety	Number of Spikes per 2.23 m^2 (Units)	1000-Grain Weight (g)	Number of Grains per Spike (Units)	Yield per Mu (kg)
Longgeng3010	683	25.6	110	533
Longgeng31	671	26.3	86	586
Lianyu124	638	28	90	555
Suijin773	641	26	106	549

For the yield estimation model, we selected five classic algorithms to train the model: Random Forest, Huber Loss Function, Least Squares Method (LSM), Total Least Squares Method (TLS), and XGBoost. Random Forest is a powerful ensemble learning algorithm that achieves higher accuracy and generalization ability than a single decision tree by combining the prediction results of multiple decision trees [26]. The core of this algorithm lies in its “wisdom of the crowd” concept—each individual decision tree contributes its predictive power, together forming a stronger, more robust, and overfitting-resistant predictive model [27]. The Huber Loss Function combines the advantages of mean squared error (MSE) and mean absolute error (MAE), exhibiting greater robustness when handling outliers. For noisy datasets, the Huber Loss Function can provide more stable estimation results, reducing sensitivity to outliers and thus improving the model’s prediction accuracy. The Least Squares Method (LSM) is a classic regression analysis method that finds the best-fit model by minimizing the sum of squared errors between predicted and actual values. Despite its simplicity, LSM performs excellently when handling linear relationship data, providing effective estimation results quickly. Total Least Squares Method (TLS) is similar to the traditional Least Squares Method but considers errors in both dependent and independent variables. It provides more accurate estimation results when dealing with large measurement errors or uncertainties in independent variables. XGBoost is an efficient and flexible gradient-boosting framework that enhances model accuracy and robustness by integrating multiple trees. XGBoost excels in handling large-scale datasets and complex models, offering strong parallel computing capabilities and high flexibility, effectively

capturing non-linear relationships in the data. By comparing these algorithms, we selected the best-performing yield estimation model. Information on the input parameters of the models and the sample size statistics can be found in Table 3.

Table 3. Detailed information on input parameters and sample sizes for yield estimation models.

Parameter Name	Description	Sample Size
Number of Panicles	Number of panicles per 2.23 m^2	64
Thousand Grain Weight	Weight of one thousand grains of rice	64
Number of Grains per Panicle	Number of grains per panicle	64
Yield per Mu	Actual yield per Mu of rice	64

For the selected five models, we designed five additional yield estimation datasets. The creation process is the same as the final yield estimation data creation process mentioned above, except the sample size is changed to 16, as each nitrogen gradient includes 16 experimental plots. These datasets consist of data collected within each of the four nitrogen gradients, forming separate datasets for each gradient and an integrated dataset from all four gradients. These datasets are the N_0 dataset, N_5 dataset, N_{10} dataset, N_{15} dataset, and the comprehensive dataset. Each dataset contains 16 samples. Based on this, we trained the models to observe how nitrogen gradient influences rice yield and, consequently, how it affects the models.

2.7. Experimental Procedure

First, the drone-captured images of rice fields with a resolution of 3840×2160 were cropped to obtain images of size 512×512 . A total of 1016 images were selected as the dataset. These images were manually annotated to generate the training label files, which were then divided into training, validation, and test sets in an 8:1:1 ratio. Next, the training dataset was used to train various improved YOLOv8 network models. Each model was trained for 2000 epochs, with a batch size of 16 for both training and validation, ultimately obtaining the optimal network weights. After training, the testing set was used to evaluate the performance of different network models, comparing the results with those of the original YOLOv8 network and the YOLOv8s+LSKA+HorNet network to determine the best rice panicle detection model. The target detection technique was applied to 10 standard-sized images from each experimental field to perform panicle detection and counting, and then to obtain the average number of rice panicles per 2.23 m^2 of rice field. Images were extracted from 64 experimental fields of this size, and the average number of rice panicles per 2.23 m^2 was tallied.

After collecting these image data, it was also necessary to obtain the thousand-grain weight and panicle grain count of the corresponding rice varieties from statistical sources. These key parameters were later used to train the yield estimation model. Ultimately, the best-performing model was selected, providing strong support for rice panicle identification and yield estimation in natural environments. The experimental procedure is illustrated in Figure 10.

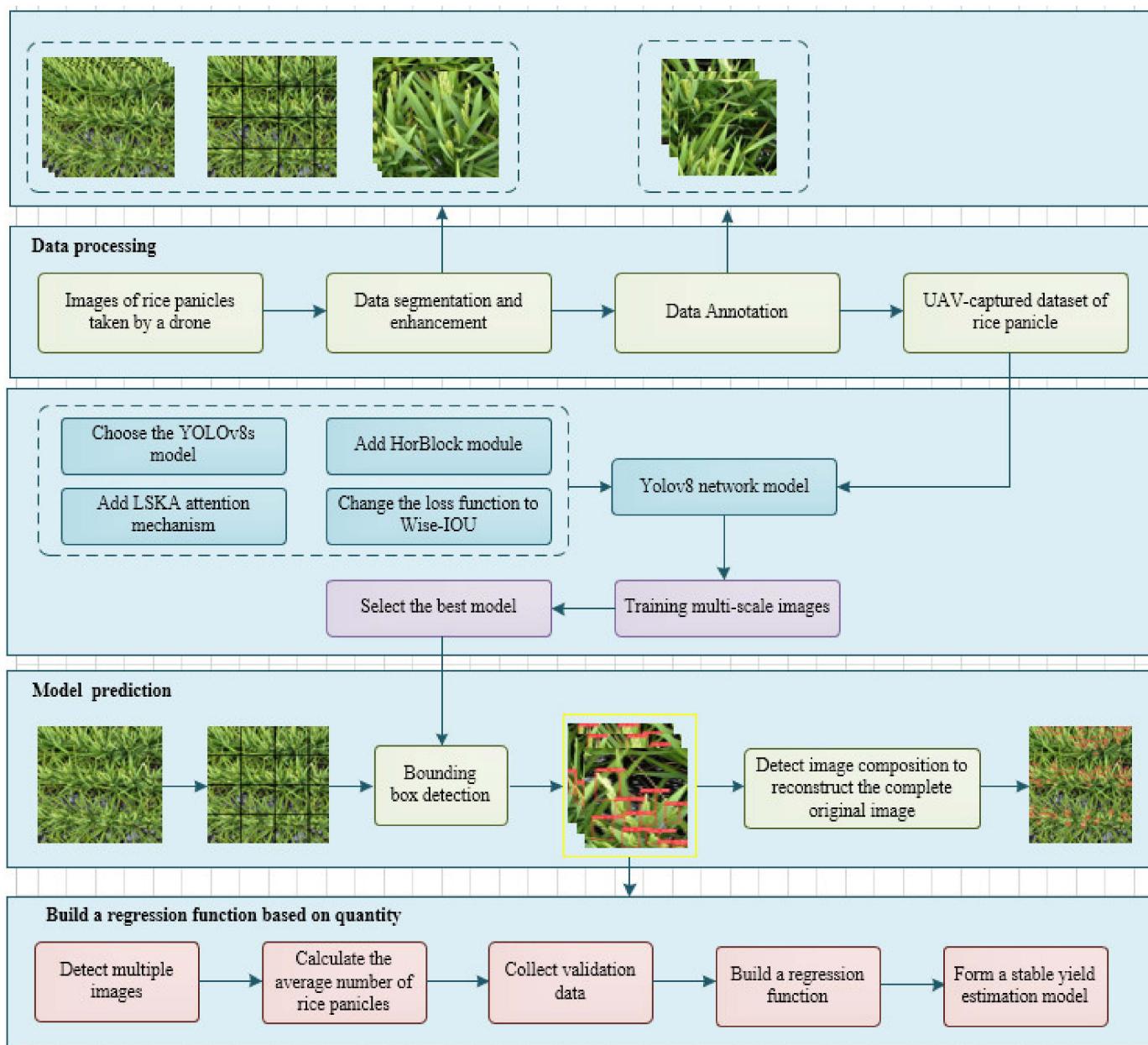


Figure 10. Schematic diagram of the experimental procedure.

2.8. Evaluation Metrics for Object Detection and Yield Estimation

In this paper, we selected precision (P), recall (R), harmonic mean ($F1$), and mean average precision (mAP) as the evaluation metrics for object detection. These metrics are based on the following definitions: the number of rice panicles correctly identified (TP), the number of objects misclassified as rice panicles (FP), rice panicles not detected (FN), the total number of samples (N), and the number of detected target categories (N_t). The average precision (AP) is essentially the average calculation of precision (P) values across the precision-recall ($P-R$) curve. In this study, since detection is solely focused on rice panicles, the mean average precision (mAP) across all categories in the dataset is equivalent to the average precision (AP). The formulas for P , R , AP , mAP , and $F1$ are shown in Equations (6)–(10), respectively.

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

$$AP = \frac{\sum precision}{N_t} \quad (8)$$

$$mAP = \frac{AP}{N} \quad (9)$$

$$F_1 = \frac{2 \times precision \times recall}{precision + recall} \quad (10)$$

To ensure the accuracy and effectiveness of the rice yield estimation model, this paper describes several key statistical metrics used to evaluate the model's performance. These metrics include the residual sum of squares (RSS), mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE) [28]. These metric values provide a quantified performance overview for the yield estimation model, ensuring adjustments and optimizations can be made based on actual performance. The introduction of these metrics not only intuitively evaluates the quality of the model's fit, but also provides a scientific basis for the final model selection. Below are the specific formulas for these metrics, where y_i represents the actual values of the samples and \hat{y}_i represents the estimated values of the samples.

Sum of Squared Residuals (RSS): This metric calculates the sum of the squares of the differences between the predicted and actual values, reflecting the total amount of model prediction error. The formula is

$$RSS = \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (11)$$

Mean Squared Error (MSE): This metric represents the average of the squares of the residuals, offering a comprehensive measure of the magnitude of errors. The formula is

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (12)$$

Mean Absolute Error (MAE): This metric calculates the average of the absolute differences between the predicted values and the actual values. It serves as an intuitive indicator of prediction accuracy. The formula is

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (13)$$

Root Mean Squared Error (RMSE): This is the square root of the mean squared error (MSE), providing another measure of the magnitude of errors. It is particularly useful in dealing with data that have extreme values. The formula is

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (14)$$

Coefficient of Determination (R^2): This is 1 minus the ratio of the sum of squared residuals to the total sum of squares, indicating the proportion of the variance explained by the model relative to the total variance. The closer the R^2 value is to 1, the better the model fits the data. The formula for calculating R^2 is

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (15)$$

By calculating and analyzing these statistical metrics, a deeper understanding of the model's performance can be achieved, and appropriate optimizations can then be made to enhance the accuracy and reliability of the yield estimation model.

2.9. Experimental Equipment and Platform

The image acquisition system consists of a Matrice 300 RTK drone (Matrice 300 RTK, SZ DJI Technology Co., Ltd., Shenzhen, China) equipped with a Zenmuse H20N (Zenmuse H20N, SZ DJI Technology Co., Ltd., Shenzhen, China) multifunctional gimbal camera, which was used for the image data collection of field-grown rice. The drone is equipped with advanced starlight sensors and high-definition cameras capable of capturing high-quality images under various lighting conditions, ensuring comprehensive coverage of a vast field area. The Zenmuse H20N (Zenmuse H20N, SZ DJI Technology Co., Ltd., Shenzhen, China) gimbal camera integrates zoom and wide-angle cameras, providing high-resolution and multi-angle image capture capabilities.

A high-performance desktop computer was used as the main processing platform in this experiment. The computer ran on the Ubuntu 20.04.3 operating system (Canonical Co., Ltd., Mann, UK) and was equipped with a powerful Intel i9-12900k processor (Intel Corporation, Santa Clara, CA, USA) (with 16 cores). For processing graphics and video tasks, an NVIDIA GeForce RTX 3090 graphics card (NVIDIA Corporation, Santa Clara, CA, USA) with 24 GB of video memory was selected. The programming work for the entire experiment was completed in the Pycharm environment (JetBrains s.r.o., Prague, Czech Republic), using the Python 3.9.7 language (Meta Platform Inc., Menlo Park, CA, USA). Moreover, a development environment based on the Pytorch 1.12.1 (Meta Platforms, Inc., Menlo Park, CA, USA) framework and CUDA version 11.4 (NVIDIA Corporation, Santa Clara, CA, USA) was built and configured in Anaconda3 (Anaconda Inc., Austin, TX, USA). The specific software and hardware configurations used in the experiment are referenced in Table 4.

Table 4. Detailed table of software and hardware configurations used in the experiment.

Configuration	Parameter
CPU	Intel i9-12900k
Memory	64 G
GPU	Nvidia RTX3090
Operating System	Ubuntu 20.04.3
Deep Learning Framework	Pytorch 1.11.0
Programming Language	Python 3.9.7
GPU Computing Platform	CUDA 11.4

3. Results

3.1. Comparison of YOLOv8s-LSKA-HorNet Model with Other Mainstream Networks

This experiment was conducted using the same dataset used for training, primarily aimed at demonstrating the advantages and characteristics of the YOLOv8s-LSKA-HorNet model in rice panicle detection. The final results of the experiment are presented in Table 5 and Figures 11 and 12. In Figure 11, red boxes represent correctly detected targets, while yellow and blue boxes indicate duplicate and missed detections, respectively.

Table 5. Comparison of network performance trained on the same dataset among different networks.

Model	Precision (%)	Recall (%)	mAP@.50 (%)	F1-Score (%)	Detection Times (ms)
YOLOv8s+LSKA+HorNet	98.0	96.8	99.2	97.3	20.3
YOLOv8s	94.3	90.0	95.6	92.1	18.9
YOLOv7	90.2	87.8	89.4	88.9	21.6
YOLOv5s	89.6	87.9	88.6	88.7	41.9
DDQ	91.3	90.1	91.6	90.2	42.5
DINO	92.4	91.8	92.7	92.1	32.6



Figure 11. Prediction results of various models on the same set of photos.

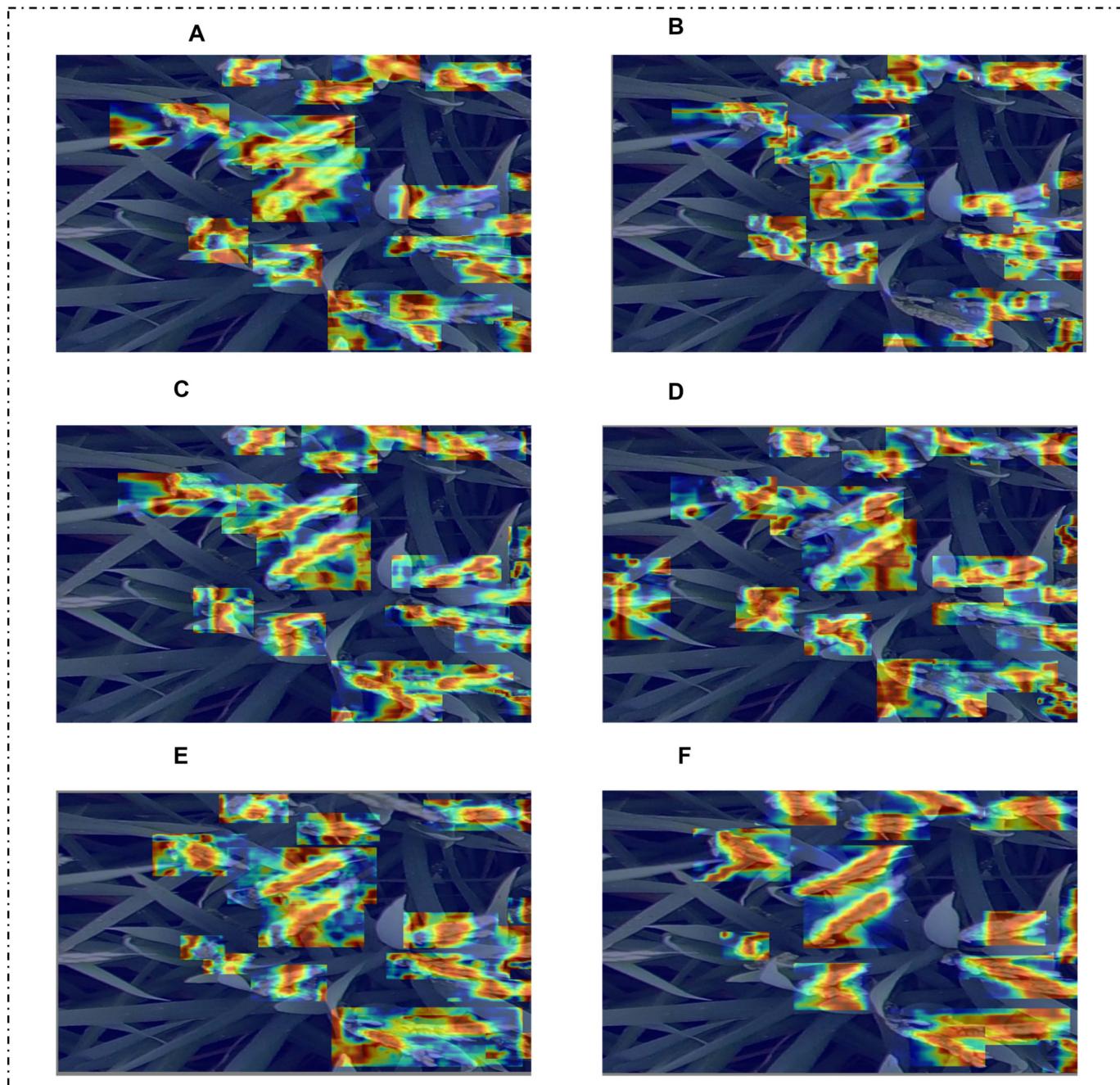


Figure 12. Heat maps of model predictions. (A) YOLOv8s+LSKA+HorNet; (B) YOLOv8s; (C) YOLOv7; (D) YOLOv5s; (E) DDQ; (F) DINO.

The experimental data showcased the performance of various advanced object detection models, including YOLOv8s+LSKA+HorNet, YOLOv8s, YOLOv7, YOLOv5s, DDQ, and DINO, across multiple key performance metrics. Among these models, the YOLOv8s+LSKA+HorNet combination demonstrated superior performance, achieving excellent results in precision, recall, mAP, and F1 scores while maintaining a fast detection speed. This indicates a highly effective balance between identification accuracy and processing speed in rice detection tasks.

3.2. Impact of Composite Scaling Factors on Effectiveness of Different Sizes of YOLOv8 Model Variants

This section explores the impact of composite scaling parameters on different scaled versions of the YOLOv8 model (n , s , m , l , x) through comparative experiments. Composite scaling parameters represent a systematic method for adjusting the size of neural network architectures. This approach involves coordinated and balanced adjustments to three key dimensions of the network: depth (i.e., the number of layers), width (i.e., the number of channels or units per layer), and resolution or maximum number of channels. The goal of this method is to enhance model performance while effectively controlling computational costs and model complexity. The performance of different scaled version models is presented in Figure 13 and Table 6.

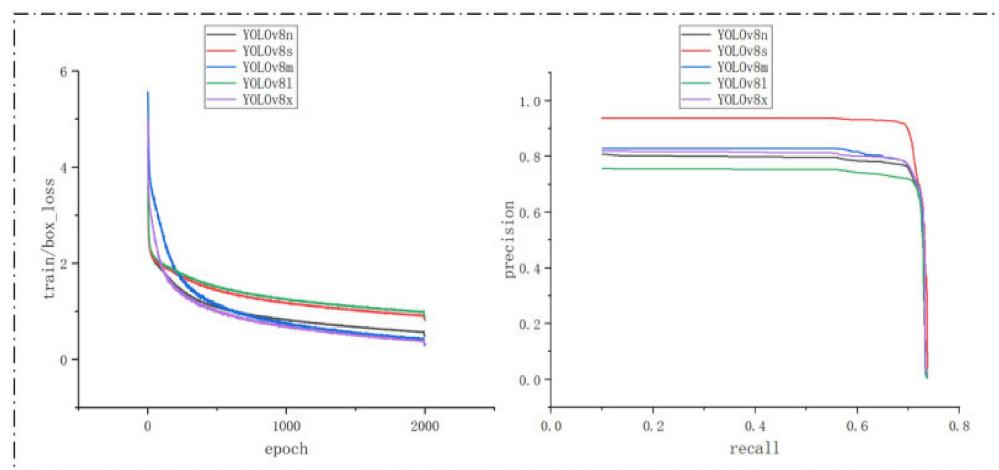


Figure 13. Loss curves and P-R graphs for different scaled version models trained on the same dataset.

Table 6. Comparison of network performance for different scaled version models trained on the same dataset.

Model	Precision (%)	Recall (%)	mAP@.50 (%)	F1-Score (%)	Detection Times (ms)
YOLOv8n	78.4	77.9	80.8	78.1	25.0
YOLOv8s	94.3	90.0	95.6	92.1	18.9
YOLOv8m	94.2	87.0	93.2	90.4	19.9
YOLOv8l	83.0	79.2	85.4	81.0	39.6
YOLOv8x	69.3	67.4	70.0	68.3	32.0

Among the variants of YOLOv8, the YOLOv8s model achieved an mAP@.50 of 95.6% with specific settings, outperforming the variants YOLOv8n (80.8%), YOLOv8m (93.2%), and YOLOv8x (70.0%). The loss function graph indicates that YOLOv8s had a faster loss reduction in the early stages of training compared to other models, and its loss curve remained relatively lower across different versions. This suggests that it could achieve a lower total loss in the same number of epochs, potentially indicating better performance. Specifically, lower loss values are usually associated with better prediction accuracy and generalization capability.

Precision and recall also reflected a similar performance trend, with YOLOv8s reaching 94.3% and 90.0%, respectively, further proving its superiority over other versions. YOLOv8m performed comparably to YOLOv8s with a precision of 94.2% and a recall of 87%, only slightly lower in mAP@.50.

3.3. Results and Analysis of Ablation Experiments

This study improved the YOLOv8 algorithm by integrating the LSKA and HorNet modules to further enhance the model's performance in object detection tasks. To assess the effectiveness of these two modules, particularly in addressing the issues of missed and duplicate detections of rice panicles, a series of ablation experiments were designed, and comparative analyses were conducted on a specific dataset. The improved YOLOv8 algorithm and the original YOLOv8 algorithm were both tested under identical conditions to ensure fairness and comparability of results. During the ablation studies, the performance of the YOLOv8 algorithm with the LSKA module introduced was first assessed, followed by the variant with only the HorNet module added, and finally, the complete model containing both LSKA and HorNet modules. The results of the ablation experiments for each model trained on a unified dataset and the comparative detection results for the same set of images are shown in Figures 14 and 15 and Table 7. In Figure 15, red boxes represent correctly detected targets, while yellow and blue boxes indicate duplicate and missed detections, respectively.

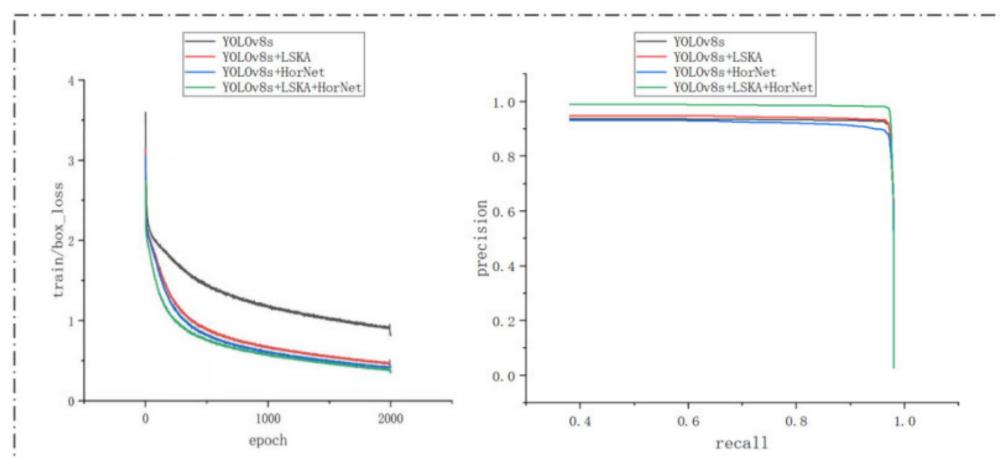


Figure 14. Loss curves and P-R graphs for models in ablation experiments trained on the same dataset.

The loss function curves in Figure 14 indicate that among all the compared model configurations, the combination of YOLOv8s with the LSKA and HorNet algorithms showed the lowest rate of loss decrease during training, suggesting that this model could optimize more effectively and reduce prediction errors during the learning process. Additionally, the precision-recall (P-R) curve in the same figure further confirms the advantage of the YOLOv8s+LSKA+HorNet configuration. Its position near the top-right corner of the graph indicates the model maintains high recall while also achieving high precision, which is particularly important in object detection tasks. According to the performance indicator comparison in Table 7, the YOLOv8s+LSKA+HorNet configuration achieved a precision of 98.0% and an mAP@.50 of up to 99.29%, significantly surpassing other model configurations that use the LSKA or HorNet algorithms individually. Clearly, this configuration has an advantage in feature extraction and object detection accuracy. From the prediction results of the four models in Figure 15, it can be seen that the YOLOv8s+LSKA+HorNet model has a strong recognition capability for heavily obscured rice panicles and can also identify rice panicles in the deep background of photos. It has improved in avoiding duplicate detections, and the detection results are superior to other models.



Figure 15. Prediction results of the trained models. (A) YOLOv8s; (B) YOLOv8s+LSKA; (C) YOLOv8s+HorNet; (D) YOLOv8s+LSKA+HorNet.

Table 7. Comparison of network performance for different models in ablation studies trained on the same dataset.

Model	Precision (%)	Recall (%)	mAP@.50 (%)	F1-Score (%)
YOLOv8s	94.3	90.0	95.6	92.1
YOLOv8s+LSKA	94.4	89.3	96.1	95.8
YOLOv8s+HorNet	93.6	89.6	94.5	91.6
YOLOv8s+LSKA+HorNet	98.0	96.8	99.2	97.3

3.4. Impact of the Loss Function on Results

This section will explore the impact of the Wise-IoU loss function designed in YOLOv8, including aspects of recognition rate and convergence speed. Experiments were conducted to study the effects of different loss function ratios on recognition results within this experimental method. The research findings are shown in Figure 16 and Table 8.

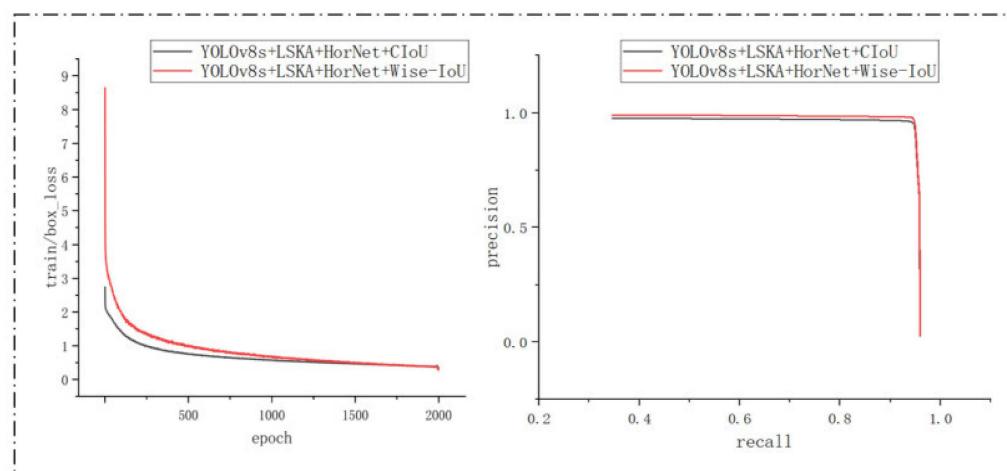


Figure 16. Loss curves and P-R graphs for models with different loss functions trained on the same dataset.

Table 8. Comparison of network performance for models with different loss functions trained on the same dataset.

Model	Precision (%)	Recall (%)	mAP@.50 (%)	F1-Score (%)	mAP50-95 (%)
YOLOv8s+LSKA+HorNet+CIoU	97.3	94.9	98.6	96.1	74.9
YOLOv8s+LSKA+HorNet+Wise-IoU	98.0	96.8	99.2	97.3	85.4

The introduction of the Wise-IoU loss function brought significant improvements to the model's performance, which is depicted in the charts. The loss function curve on the left shows how the model more rapidly adapted to the characteristics of rice panicles during training. The incorporation of the Wise-IoU method led to a quicker reduction in loss values, accelerating the model's convergence.

Table 8 provides a more detailed numerical comparison of these functions. Across all given performance metrics, Wise-IoU performed better, with a 0.7% increase in precision, a 1.9% increase in recall, a 0.6% increase in mAP@0.50, a 1.3% increase in F1 score, and a significant 10.5% increase in mAP50-95.

3.5. The Impact of Different Nitrogen Gradients on the Yield Estimation Model

We used five different models to train and test the rice data collected under different nitrogen gradients, ultimately obtaining the results of each model's training on different datasets. We observed the prediction performance using the R^2 values, and the experimental data are shown in Table 9. In the table, we use $N_0 \cdot R^2$ to represent the R^2 values achieved

by each model on the N₀ dataset, N₅_R² for the R² values on the N₅ dataset, N₁₀_R² for the R² values on the N₁₀ dataset, N₁₅_R² for the R² values on the N₁₅ dataset, and C_R² for the R² values on the combined dataset.

Table 9. Prediction performance of different network models on various datasets.

Model	N ₀ _R ²	N ₅ _R ²	N ₁₀ _R ²	N ₁₅ _R ²	C_R ²
Random Forest	0.53	0.61	0.65	0.72	0.79
Huber Loss	0.32	0.45	0.44	0.54	0.68
LSM	0.44	0.56	0.61	0.65	0.69
TLS	0.23	0.34	0.39	0.41	0.46
XGBoost	0.48	0.63	0.58	0.64	0.73

3.6. Effectiveness of the Yield Estimation Model

The training and prediction performance of the yield estimation models, trained on the collected data, are displayed using R² charts, providing a basis for comparing the performance of different yield estimation models. The training and prediction results of the models are shown in Figure 17 and Tables 10 and 11.

In this paper, we conducted an in-depth analysis and comparison of four different yield estimation models, focusing on key metrics such as mean squared error (MSE), residual sum of squares (RSS), root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R²). These metrics quantify the degree of prediction error and the relationship between predicted and actual values.

We evaluated the performance of five different yield estimation models on the training and validation sets: Random Forest, Huber Loss, Least Squares Method (LSM), Total Least Squares (TLS), and XGBoost. As shown in Figure 17, the fit between each model's predicted and actual values can be observed. The black dashed line represents the ideal prediction, where the predicted values are exactly equal to the actual values. We judged the fit of each model by observing these scatter plots.

Random Forest performed exceptionally well on the training set, with an R² value close to 1, indicating a very good fit. However, on the test set, although the R² value decreased, it still remained at a high level, demonstrating good generalization ability. The Huber Loss model also performed well on the training set, but the R² value dropped significantly on the test set, indicating poorer generalization compared to Random Forest. The Least Squares Method model showed stable performance on both the training and test sets, with an R² value slightly lower than Random Forest but higher than Huber Loss. The TLS model had low R² values on both the training and test sets, showing poor performance. The XGBoost model performed very well on both sets, second only to Random Forest.

Tables 10 and 11 provide the specific performance metrics for each model on the training and validation sets, including MSE, RSS, RMSE, and MAE. On the training set, Random Forest had the best metrics, demonstrating its excellent fit. The Huber Loss and Least Squares Method models followed, while the TLS model had significantly worse metrics. The XGBoost model also performed very well, second only to Random Forest.

On the test set, Random Forest maintained low MSE and MAE, indicating strong generalization ability. The Huber Loss model showed significant increases in all metrics, indicating weaker generalization. The Least Squares Method model remained stable, with small increases in metrics. The TLS model performed the worst on the test set, with significant increases in all metrics. The XGBoost model had stable metrics on the test set, indicating strong generalization ability.

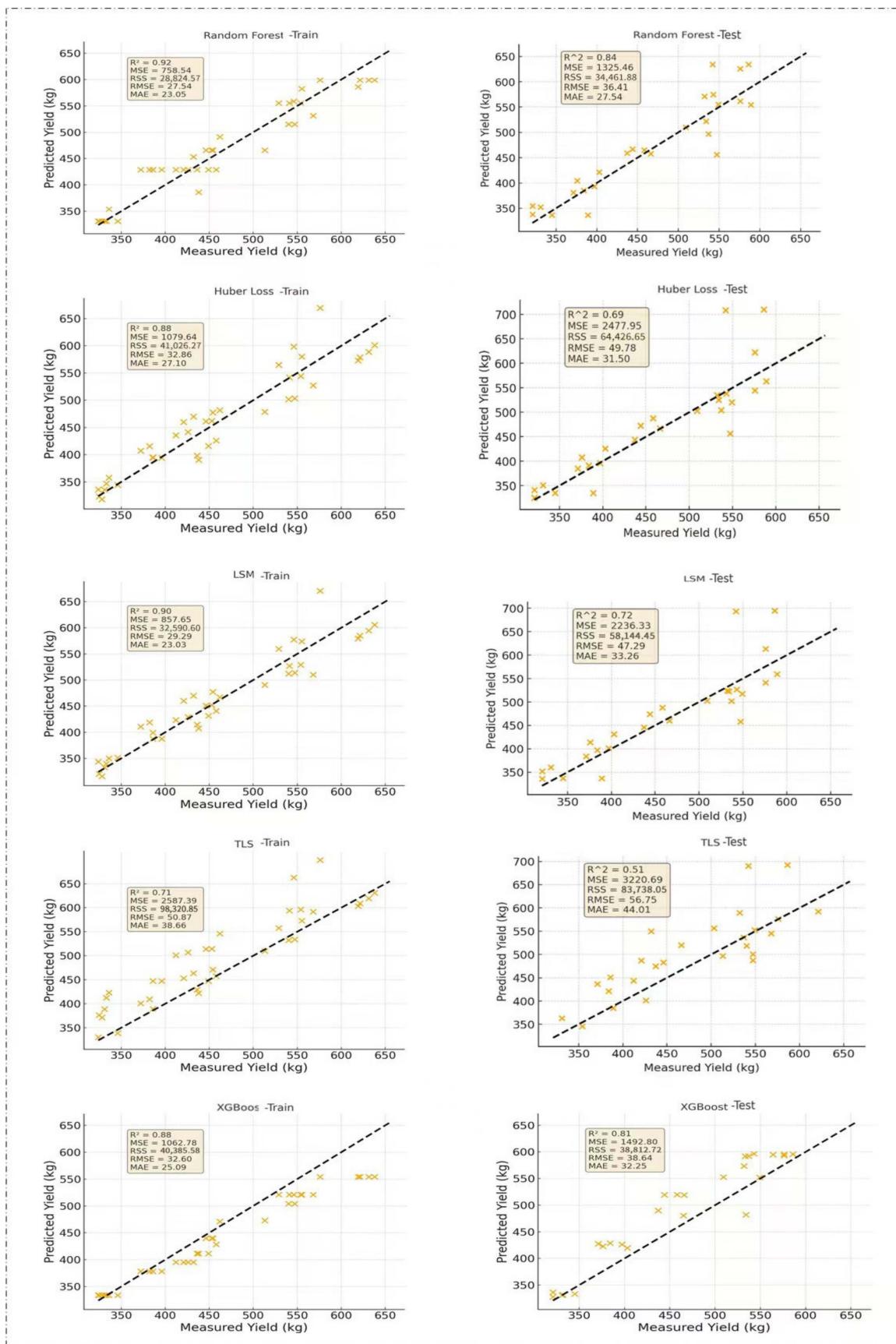


Figure 17. R^2 graphs of training and prediction for different yield estimation models (The yellow X indicates the intersection of the actual and estimated yields on the coordinates).

Table 10. Comparison of training performance of yield estimation models trained on the same dataset using different models.

Method	MSE (kg^2)	RSS (kg^2)	RMSE (kg)	MAE (kg)
Random Forest	758.54	28,824.57	27.54	23.05
Huber Loss	1079.64	41,026.27	32.86	27.10
LSM	857.65	32,590.60	29.29	23.03
TLS	2587.39	98,320.85	50.87	38.66
XGBoost	1062.78	40,385.58	32.60	25.09

Table 11. Comparison of prediction performance of yield estimation models trained on the same dataset using different models.

Method	MSE (kg^2)	RSS (kg^2)	RMSE (kg)	MAE (kg)
Random Forest	1325.46	34,461.88	36.41	27.54
Huber Loss	2477.95	64,426.65	49.78	31.50
LSM	2236.33	58,144.45	47.29	33.26
TLS	3220.69	83,738.05	56.75	44.01
XGBoost	1492.80	38,812.72	38.64	32.25

Overall, based on the analysis, the Random Forest and XGBoost models performed the best on this dataset, showing strong generalization ability and making them the preferred choice for yield estimation models.

4. Discussion

4.1. Rice Panicle Detection and Counting Techniques

The detection and counting of rice panicles present a key technical challenge in paddy rice cultivation. However, in recent years, significant developments have been made in this technology. Desai et al. utilized convolutional neural networks (CNNs), a type of deep learning model, to analyze time-series RGB images of rice fields for detecting flowering panicles [29]. By counting the number of flowering areas detected in each image, the number of flowering panicles at different time points can be estimated. However, manually counting detection boxes is cumbersome and greatly reduces detection efficiency. Wang et al. selected YOLOv5 as the base model for this task, customizing and optimizing it to meet the demands of detecting rice panicles in large-scale field images. Considering the potential for duplicate detections in overlapping areas of field images, the paper proposed a method beyond the traditional non-maximum suppression (NMS) by using two indicators (IOB and BOU) to quantitatively measure the overlap patterns and degrees between overlapping boxes and more accurately remove duplicate detections [30]. This experiment showed good detection results for rice in static single-case scenarios but did not demonstrate sufficient detection of rice panicles in complex environments. Chen et al. utilized an algorithm called Refined Feature Fusion (RFF-PC) to automatically count rice panicles by analyzing high-density and multi-scale images of rice fields captured by drones through CNNs [31]. Their algorithm enhances feature extraction and fusion through multi-scale convolution, feature pyramid fusion, and channel attention mechanisms. Additionally, it uses a refined Gaussian kernel to generate ground truth that closely approximates the real shape, thereby significantly improving detection efficiency and accuracy.

This experiment aimed to achieve plot-level yield estimation for paddy rice, thus requiring the detection of rice panicles in complex, real-world agricultural environments. Previous panicle detection experiments mainly relied on images of rice fields captured with a single camera, a method limited by the singularity of its data. This experiment, through the use of drones for aerial photography and image collection under different lighting conditions and rice growth forms, ensured the diversity and richness of data sources. The model designed in this study introduces attention mechanisms and optimizations to the loss function based on YOLOv8, aiming to overcome the challenges of detecting rice

panicles in complex environments and thereby improving the accuracy of detection under various conditions and the reliability of yield estimation.

4.2. Comparison with Other Deep Learning Models

The results of this experiment demonstrate the advantages of the YOLOv8s+LSKA+HorNet model in terms of its superior performance metrics. The model excels in key performance indicators such as precision, recall, mAP, and F1 scores, ensuring high accuracy and reliability in object detection tasks. Specifically for rice detection, the model covers most of the rice areas with dense and precise bounding boxes and displays high confidence levels, indicating its ability to accurately distinguish between rice and background imagery, thus reducing false detections. Additionally, in the comparison of prediction results for the same set of photos in Figure 11, it can be seen that this model performs well in terms of missed detections and duplicate detections, exhibiting a high level of confidence. The model's fast detection speed offers a practical advantage in real-world applications. The heat maps generated by the YOLOv8s+LSKA+HorNet model also significantly demonstrate its advantages. These heat maps provide an intuitive representation of the model's recognition process, using color changes to indicate the model's focus on different areas. The positions of rice panicles are clearly and accurately marked in the heat maps, with highlighted areas closely and clearly confined to the contours of the panicles, rarely extending into irrelevant background areas. Overall, YOLOv8s+LSKA+HorNet achieves a good balance between precision, speed, and stability, demonstrating its leading position in complex image processing tasks.

Zhang et al. proposed an optimized YOLOv8 algorithm to improve the accuracy of small object detection on water surfaces while maintaining rapid detection speed [32]. This improvement effectively reduced noise interference and minimized missed detections. Experimental results confirmed that the algorithm could effectively solve the issue of missed detections in small object detection and significantly increase processing speed. In a related study, Pan et al. addressed the challenge of detecting small objects from a drone perspective by proposing an improved YOLOv8 model [33]. By addressing issues with shared attention weight parameters and introducing a large, separable convolutional attention mechanism, this experiment enhanced the Spatial Pyramid Pooling layer (SPPF), increasing interaction between features at different levels, thereby improving the model's ability to recognize complex samples. This experiment demonstrated the YOLOv8 model's exceptional ability to accurately identify small objects captured by drones. Yang et al. proposed a lightweight YOLOv8 tomato detection algorithm that effectively improves the accuracy and efficiency of tomato detection by combining feature enhancement and attention mechanisms [34]. First, depth separable convolution (DSConv) is used instead of standard convolution, which significantly reduces the computational complexity and the number of parameters while maintaining high detection accuracy. Secondly, the Dual Path Attention Gate (DPAG) module is introduced to enhance the feature extraction capability, which improves the detection accuracy of the model in complex environments and enhances the model's ability to distinguish between tomato and background. The above also verifies that YOLOv8 has strong plasticity in small target detection and can be used as an improved base model.

4.3. Impact of Attention Mechanisms

The LSKA (Large Separable Convolutional Attention) mechanism, introduced by Kin Wai Lau, Lai-Man Po, and Yasar Abbas Ur Rehman, demonstrates significant advantages in object detection. Hu and colleagues developed an enhanced lightweight rebar detection network to improve accuracy and efficiency in complex scenes [35]. LSKA was applied to refine the Spatial Pyramid Pooling—Fast (SPPF) structure, enhancing the network's spatial awareness and computational efficiency. This allows for more effective processing of features with complex spatial distributions, such as densely packed and multilayered rebars, thereby improving model accuracy while reducing computational resource consumption,

making it better suited for real-time detection applications. Experiments show that LSKA significantly improves target detection against complex backgrounds. Qing et al. proposed an improved YOLO-FastestV2 model, introducing the LSKA mechanism, which reduces references and computational complexity by decomposing the large convolution kernel into horizontal and vertical one-dimensional convolution kernels while improving the ability to capture key image features. LSKA significantly enhances the network's spatial perception, for example, for high-density and multi-layer stacked wheat ears [36].

In this experiment, the challenge in detecting rice panicles lies in their small pixel area within images and variable shapes, making subtle feature capture difficult. This task demands advanced feature extraction capability, as rice leaves may severely obscure panicles, increasing the difficulty of accurately distinguishing rice panicles from the background. The model needs strong information integration and noise suppression abilities. To meet real-time detection needs while maintaining high accuracy, improving detection speed and efficiency is necessary. These factors collectively pose the primary challenge for rice panicle detection tasks. To address these issues, we incorporated the LSKA attention mechanism. LSKA expands the receptive field by breaking down large convolutional kernels into smaller, separable ones, enabling more detailed capture of small target rice panicle features and effectively enhancing feature extraction capability. Strengthening the interactions between features at different levels improves the model's accuracy in complex backgrounds and enhances environmental noise suppression.

Compared to traditional models, HorNet's main technical advantages include processing high-order spatial interactions at a lower computational cost and effectively expanding the model's representational capacity through Gated Non-linear Convolution (gnConv). This not only improves computational efficiency but also ensures high performance in complex visual tasks such as image classification, object detection, and semantic segmentation. Yu et al. designed the YOLOv7 algorithm HB-YOLO for tracking dim, small objects in satellite remote sensing videos [37]. By integrating HorNet to enhance high-order spatial interaction capabilities and replacing traditional convolution operations, HorNet optimizes high-order spatial interactions, significantly boosting feature extraction and recognition capabilities. This mechanism allows HorNet to more effectively capture details and patterns in complex images, improving the accuracy and efficiency of visual tasks.

In this experiment, rice panicle images were severely affected by sunlight with varying degrees of bending, leading to multilayered rice panicle detection. To overcome this issue, we incorporated the HorNet attention mechanism. By focusing on key features and areas within images, this model enhances the ability to capture details of rice panicles, especially in images with complex or changing field backgrounds. Its high sensitivity to small targets and effective suppression of background noise further enhance recognition accuracy and robustness. Additionally, the HorNet attention mechanism improves the model's processing speed and efficiency, enabling rapid processing of large volumes of images. The adaptability and generalizability of the HorNet model may also be enhanced, allowing it to meet the detection needs of different growth environments and rice varieties.

4.4. Impact of Loss Functions

In this experiment, the core challenge of rice detection lies in addressing the overlap and occlusion among dense panicles. In large field environments, panicles are closely packed and frequently obstruct each other, necessitating precise and clear generation of detection bounding boxes. To enhance detection accuracy and provide reliable data for yield estimation, the experiment utilized the Wise-IoU loss function. The introduction of Wise-IoU significantly improved the model's sensitivity to small field targets like rice panicles through a dynamic, non-monotonic focusing mechanism, optimizing the similarity measurement between detection boxes and actual annotations. This optimization not only reduced false positives but also enhanced the model's ability to recognize rice panicles against complex field backgrounds. Detection boxes optimized with Wise-IoU were more

accurately aligned with rice panicles, significantly improving detection accuracy and robustness, particularly in resolving small targets and occlusion issues.

Liu et al. proposed a YOLOv7 method optimized with PConv, SE attention mechanisms, and Wise-IoU, aimed at addressing the complexity and high computational demands faced by deep convolutional neural networks in practical applications [38]. By reducing the adverse effects of low-quality samples through a dynamic, non-monotonic focusing mechanism, this method optimized the similarity measurement between detection boxes and true annotations. This experiment effectively validated that the Wise-IoU algorithm can significantly lower the false detection rate. Xiao and colleagues introduced the YOLOv8-GAM-Wise-IoU model for the automatic detection of bridge surface cracks. By incorporating the Wise-IoU loss function, this model enhanced detection accuracy and generalization [39]. The Wise-IoU loss function, by finely adjusting the boundaries of object detection, effectively reduced errors when addressing overlapping images or varying sizes of cracks as determined by the evaluation metrics of precision, recall, F1 score, and mAP. This experiment also showed that the Wise-IoU loss function significantly aids in generating precise detection bounding boxes and optimizing the model's ability to recognize crack boundaries. This advantage also applies to rice panicle detection, as the model improves the accuracy of bounding box generation.

4.5. Effect of Nitrogen Gradient on Yield Estimation Models

The prediction effect of different estimation models on various nitrogen gradient datasets and comprehensive datasets reveals that as the nitrogen gradient increases, the R^2 values of the models generally rise. This is because, in actual production, rice crops grown with normal fertilization account for the majority, and higher nitrogen gradients result in rice growth that closely resembles real growth conditions, thus improving model prediction accuracy. Additionally, in the comprehensive dataset, we have enhanced the span of rice yield data, thereby enriching data diversity and more realistically simulating actual rice yields. By considering multiple aspects of real-world conditions, it is evident from the experiments that increasing the nitrogen gradient to broaden the yield data range significantly impacts the robustness of the model. Therefore, when training the yield estimation model, controlling the yield through nitrogen gradients and enriching the data span is crucial [40].

4.6. Yield Prediction Models

The selection of yield estimation models is a vital part of the yield estimation process. Jeong et al. applied the Random Forest (RF) machine learning method to global and regional crop yield predictions and compared its performance with Multivariate Linear Regression (MLR) [41]. The results showed that RF significantly outperformed MLR in accuracy, demonstrating its effectiveness and versatility in predicting yields for crops of different scales, as well as its capability to automatically handle missing values and provide assessments of variable importance. Basha et al. explored how the Random Forest algorithm can be used to predict crop yields to promote sustainable agriculture. By comparing the prediction results of the Thomson model with those of the Random Forest model, the effectiveness of the Random Forest model in crop yield prediction was proved. Early experiments on rice yield estimation mainly relied on aerial photography in field environments and biomass data, a method that often led to considerable errors [42]. With technological advancements, although target detection, rice panicle counting, and yield model development have been improved, most yield models are linear equations with a single influencing factor, resulting in poor adaptability [43].

To improve the adaptability of previous linear yield estimation models and to enrich the yield dataset, this experiment adopted a design involving multiple rice varieties and gradients of nitrogen fertilizer treatment to obtain data under various conditions. We employed the Random Forest algorithm and introduced several elements related to rice yield. By combining the prediction results of multiple decision trees and applying a voting

mechanism, the accuracy of the yield estimation results was improved, and thereby, a more stable and precise yield estimation model was constructed.

5. Conclusions

This study introduced a method for plot-level rice yield estimation utilizing drone-captured photographs adapted for complex field environments. The main conclusions are as follows:

- (1) Performance tests of the model demonstrated that the method achieved a rice panicle detection accuracy of 98.0%, a recall rate of 96.8%, an mAP@.5 of 99.29%, and an F1 score of 97.39%, which is an improvement of 3.7% over the original YOLOv8 network. Overall, the model demonstrated high detection accuracy levels for rice panicle photos collected in complex field environments.
- (2) In comparison to yield estimation results, the YOLOv8s+LSKA+HorNet network model exhibited higher recognition precision, stability, and execution efficiency compared to standard models such as YOLOv8. The detected number of rice panicles, combined with a stable yield estimation model trained using the Random Forest algorithm, was more accurate than other yield estimation models, achieving an R^2 value of 0.85.

The RFYOLO method proposed in this study facilitates the use of drone-captured imagery to enhance the accuracy and continuity of yield estimation work, significantly improving yield estimation efficiency. This has important implications for enhancing the efficiency of rice yield estimation using drone photography.

Author Contributions: Software, X.T.; Resources, W.R. and R.G.; Data curation, Z.J.; Writing—original draft, J.W.; Project administration, Z.S.; Funding acquisition, Q.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [National Key Research and Development Program of China] grant number [2023YFD230050502] And the APC was funded by [National Key Research and Development Program of China].

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest

References

1. Sishodia, R.P.; Ray, R.L.; Singh, S. Applications of remote sensing in precision agriculture: A review. *Remote Sens.* **2020**, *12*, 3136. [[CrossRef](#)]
2. Nyéki, A.; Neményi, M. Crop Yield Prediction in Precision Agriculture. *Agronomy* **2022**, *12*, 2460. [[CrossRef](#)]
3. Shafi, U.; Mumtaz, R.; García-Nieto, J.; Hassan, S.A.; Zaidi, S.A.R.; Iqbal, N. Precision agriculture techniques and practices: From considerations to applications. *Sensors* **2019**, *19*, 3796. [[CrossRef](#)] [[PubMed](#)]
4. Ma, L.; Li, M.; Tong, L.; Wang, Y.; Cheng, L. Using unmanned aerial vehicle for remote sensing application. In Proceedings of the 2013 21st International Conference on Geoinformatics, Kaifeng, China, 20–22 June 2013; pp. 1–5.
5. Stuart, A.M.; Pame, A.R.P.; Silva, J.V.; Dikitanan, R.C.; Rutsaert, P.; Malabayabas, A.J.B.; Lampayan, R.M.; Radanielson, A.M.; Singleton, G.R. Yield gaps in rice-based farming systems: Insights from local studies and prospects for future analysis. *Field Crops Res.* **2016**, *194*, 43–56. [[CrossRef](#)]
6. Wang, F.; Yao, X.; Xie, L.; Zheng, J.; Xu, T. Rice yield estimation based on vegetation index and fluorescence spectral information from UAV hyperspectral remote sensing. *Remote Sens.* **2021**, *13*, 3390. [[CrossRef](#)]
7. Duan, B.; Fang, S.; Zhu, R.; Wu, X.; Wang, S.; Gong, Y.; Peng, Y. Remote estimation of rice yield with unmanned aerial vehicle (UAV) data and spectral mixture analysis. *Front. Plant Sci.* **2019**, *10*, 204. [[CrossRef](#)] [[PubMed](#)]
8. Chang, C.-I. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Trans. Inf. Theory* **2000**, *46*, 1927–1932. [[CrossRef](#)]
9. Sanaeifar, A.; Guindo, M.L.; Bakhshipour, A.; Fazayeli, H.; Li, X.; Yang, C. Advancing precision agriculture: The potential of deep learning for cereal plant head detection. *Comput. Electron. Agric.* **2023**, *209*, 107875. [[CrossRef](#)]
10. Tanaka, Y.; Watanabe, T.; Katsura, K.; Tsujimoto, Y.; Takai, T.; Tanaka, T.S.T.; Kawamura, K.; Saito, H.; Homma, K.; Mairoua, S.G. Deep learning enables instant and versatile estimation of rice yield using ground-based RGB images. *Plant Phenomics* **2023**, *5*, 0073. [[CrossRef](#)] [[PubMed](#)]

11. Ashfaq, M.; Khan, A.S.; Ullah Khan, S.H.; Ahmad, R. Association of Various Morphological Traits with Yield and Genetic Divergence in Rice (*Oryza sativa*). *Int. J. Agric. Biol.* **2012**, *14*, 55–62.
12. Li, R.; Li, Z.; Ye, J.; Yang, Y.; Ye, J.; Xu, S.; Liu, J.; Yuan, X.; Wang, Y.; Zhang, M. Identification of SMG3, a QTL coordinately controls grain size, grain number per panicle, and grain weight in rice. *Front. Plant Sci.* **2022**, *13*, 880919. [[CrossRef](#)] [[PubMed](#)]
13. Deng, R.; Tao, M.; Huang, X.; Bangura, K.; Jiang, Q.; Jiang, Y.; Qi, L.S. Automated counting grains on the rice panicle based on deep learning method. *Sensors* **2021**, *21*, 281. [[CrossRef](#)] [[PubMed](#)]
14. Xu, C.; Jiang, H.; Yuen, P.; Ahmad, K.Z.; Chen, Y. MHW-PD: A robust rice panicles counting algorithm based on deep learning and multi-scale hybrid window. *Comput. Electron. Agric.* **2020**, *173*, 105375. [[CrossRef](#)]
15. Yang, B.; Gao, Z.; Gao, Y.; Zhu, Y. Rapid detection and counting of wheat ears in the field using YOLOv4 with attention module. *Agronomy* **2021**, *11*, 1202. [[CrossRef](#)]
16. Tan, S.; Lu, H.; Yu, J.; Lan, M.; Hu, X.; Zheng, H.; Peng, Y.; Wang, Y.; Li, Z.; Qi, L. In-field rice panicles detection and growth stages recognition based on RiceRes2Net. *Comput. Electron. Agric.* **2023**, *206*, 107704. [[CrossRef](#)]
17. Han, X.; Liu, F.; He, X.; Ling, F.J.C.I. Research on rice yield prediction model based on deep learning. *Comput. Intell. Neurosci.* **2022**, *2022*, 1922561. [[CrossRef](#)] [[PubMed](#)]
18. Qiong, J. Research on Large-scale Water Storage Estimation Method Based on YOLO V5 Network Using Deep Learning. Master's Thesis, Jilin Agricultural University, Changchun, China, 2024.
19. Terven, J.; Córdova-Esparza, D.-M.; Romero-González, J.-A. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [[CrossRef](#)]
20. Bolya, D.; Zhou, C.; Xiao, F.; Lee, Y.J. Yolact: Real-time instance segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9157–9166.
21. Wei, D.; Xu, X.; Shen, H.; Huang, K. C2f-fwn: Coarse-to-fine flow warping network for spatial-temporal consistent motion transfer. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; pp. 2852–2860.
22. Rao, Y.; Zhao, W.; Tang, Y.; Zhou, J.; Lim, S.-N.; Lu, J. Hornet: Efficient high-order spatial interactions with recursive gated convolutions. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 10353–10366.
23. Lau, K.W.; Po, L.-M.; Rehman, Y.A.U. Large separable kernel attention: Rethinking the large kernel attention design in cnn. *Expert Syst. Appl.* **2024**, *236*, 121352. [[CrossRef](#)]
24. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism. *arXiv* **2023**, arXiv:2301.10051.
25. Hu, H.; Li, H.; Li, C.; Wang, Q.; He, J.; Li, W.; Zhang, X. Design and experiment of broad width and precision minimal tillage wheat planter in rice stubble field. *Trans. Chin. Soc. Agric. Eng.* **2016**, *32*, 24–32.
26. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
27. Sakamoto, T. Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 208–228. [[CrossRef](#)]
28. Liu, J.; Huffman, T.; Qian, B.; Shang, J.; Li, Q.; Dong, T.; Davidson, A.; Jing, Q. Crop yield estimation in the Canadian Prairies using Terra/MODIS-derived crop metrics. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2685–2697. [[CrossRef](#)]
29. Desai, S.V.; Balasubramanian, V.N.; Fukatsu, T.; Ninomiya, S.; Guo, W. Automatic estimation of heading date of paddy rice using deep learning. *Plant Methods* **2019**, *15*, 76. [[CrossRef](#)]
30. Wang, X.; Yang, W.; Lv, Q.; Huang, C.; Liang, X.; Chen, G.; Xiong, L.; Duan, L. Field rice panicle detection and counting based on deep learning. *Field Crops Res.* **2022**, *13*, 966495. [[CrossRef](#)] [[PubMed](#)]
31. Chen, Y.; Xin, R.; Jiang, H.; Liu, Y.; Zhang, X.; Yu, J.J.C. Refined feature fusion for in-field high-density and multi-scale rice panicle counting in UAV images. *Comput. Electron. Agric.* **2023**, *211*, 108032. [[CrossRef](#)]
32. Zhang, M.; Wang, Z.; Song, W.; Zhao, D.; Zhao, H. Efficient Small-Object Detection in Underwater Images Using the Enhanced YOLOv8 Network. *Appl. Sci.* **2024**, *14*, 1095. [[CrossRef](#)]
33. Zhai, X.; Huang, Z.; Li, T.; Liu, H.; Wang, S. YOLO-Drone: An Optimized YOLOv8 Network for Tiny UAV Object Detection. *Electronics* **2023**, *12*, 3664. [[CrossRef](#)]
34. 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](#)]
35. Zhichao, H.; Yi, W.; Junping, W.; Wanli, X.; Bilian, L. Improved Lightweight Rebar Detection Network Based on YOLOv8s Algorithm. *Adv. Comput. Signals Syst.* **2023**, *7*, 107–117.
36. Qing, S.; Qiu, Z.; Wang, W.; Wang, F.; Jin, X.; Ji, J.; Zhao, L.; Shi, Y. Improved YOLO-FastestV2 wheat spike detection model based on a multi-stage attention mechanism with a LightFPN detection head. *Front. Plant Sci.* **2024**, *15*, 1411510. [[CrossRef](#)] [[PubMed](#)]
37. Yu, C.; Feng, Z.; Wu, Z.; Wei, R.; Song, B.; Cao, C. HB-YOLO: An Improved YOLOv7 Algorithm for Dim-Object Tracking in Satellite Remote Sensing Videos. *Remote Sens.* **2023**, *15*, 3551. [[CrossRef](#)]
38. Zhigang, L.; Baoshan, S.; Kaiyu, B. Optimization of YOLOv7 Based on PConv, SE Attention and Wise-IoU. *Int. J. Comput. Intell. Appl.* **2024**, *23*, 2350033. [[CrossRef](#)]
39. Xiong, C.; Zayed, T.; Abdelkader, E.M. A novel YOLOv8-GAM-Wise-IoU model for automated detection of bridge surface cracks. *Constr. Build. Mater.* **2024**, *414*, 135025. [[CrossRef](#)]

40. Linquist, B.A.; Liu, L.; van Kessel, C.; van Groenigen, K.J. Enhanced efficiency nitrogen fertilizers for rice systems: Meta-analysis of yield and nitrogen uptake. *Field Crops Res.* **2013**, *154*, 246–254. [[CrossRef](#)]
41. Jeong, J.H.; Resop, J.P.; Mueller, N.D.; Fleisher, D.H.; Yun, K.; Butler, E.E.; Timlin, D.J.; Shim, K.-M.; Gerber, J.S.; Reddy, V.R. Random forests for global and regional crop yield predictions. *PLoS ONE* **2016**, *11*, e0156571. [[CrossRef](#)]
42. Basha, S.M.; Rajput, D.S.; Janet, J.; Somula, R.S.; Ram, S. Principles and practices of making agriculture sustainable: Crop yield prediction using Random Forest. *Scalable Comput. Pract. Exp.* **2020**, *21*, 591–599. [[CrossRef](#)]
43. Muslim, M.; Romshoo, S.A.; Rather, A. Paddy crop yield estimation in Kashmir Himalayan rice bowl using remote sensing and simulation model. *Environ. Monit. Assess.* **2015**, *187*, 316. [[CrossRef](#)] [[PubMed](#)]

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.