

Article

YOLOv12-RSLW: An Efficient Detection and Severity Grading Framework for Rice False Smut via Count-Area Calibration

Xiao Liang ¹, Weijian Zhang ^{2,3}, Zixin Zhang ^{2,3}, Lulu Yang ⁴, Hongli Lian ¹  and Yingli Cao ^{2,3,*}

¹ College of Agriculture, Liaodong University, Dandong 118003, China; liangxiao@liaodongu.edu.cn (X.L.); lianhongli@liaodongu.edu.cn (H.L.)

² College of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang 110866, China; 2025200009@stu.syau.edu.cn (W.Z.); zzzxin@stu.syau.edu.cn (Z.Z.)

³ Liaoning Key Laboratory of Intelligent Agricultural Technology, Shenyang 110866, China

⁴ College of Information Engineering, Liaoning Vocational University of Technology, Jinzhou 121007, China; yangll_208918@lnuut.edu.cn

* Correspondence: caoyingli@syau.edu.cn

Abstract

Rice false smut is a major panicle disease that affects rice yield and grain quality and is an important target for resistance evaluation in breeding programs. Accurate field phenotyping is important for disease assessment and resistance screening, yet current assessment relies heavily on manual visual scoring and smut ball counting, which are laborious and subject to evaluator variation. In close-range single-panicle images, false smut balls are often small, dense, occluded, adhered, making automatic detection and severity grading difficult. To address these challenges, we developed a YOLOv12-RSLW detector by integrating RepGhost, SimAM, LSCD, and WIoU into YOLOv12. Detection boxes were then used to guide the Segment Anything Model for panicle and lesion mask extraction, allowing calculation of the lesion-to-panicle area ratio as a supplementary indicator for count-based severity grading. A total of 1911 original field images were collected. After augmentation, the dataset contained 5663 images, including 4531 training images, 566 validation images, and 566 test images. Detection performance was evaluated on the test set, while SAM segmentation was assessed using 80 manually annotated original images. YOLOv12-RSLW achieved 92.06% mAP@0.5, 91.46% precision, and 87.01% recall, with 3.45 M parameters and 6.0 GFLOPs. Compared with the baseline YOLOv12, mAP@0.5 and recall increased by 3.60 and 4.37 percentage points, respectively. Within the augmented dataset, 41.2% of samples initially assigned to Grade 1 and 28.1% of those assigned to Grade 2 met the area-ratio criteria for potential reassignment to higher grades. The framework provides a quantitative approach to rice false smut severity phenotyping and may support future resistance breeding after further validation.

Keywords: rice false smut; YOLOv12-RSLW; efficient object detection; segment anything model; severity grading calibration



Academic Editor: Gniewko Niedbała

Received: 7 July 2026

Revised: 14 August 2026

Accepted: 27 August 2026

Published: 31 August 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Rice false smut (RFS), caused by the fungal pathogen *Ustilaginoidea virens*, is an important panicle disease affecting rice production in many rice-growing regions [1,2]. Severe infections reduce grain weight and increase the proportion of unfilled grains, resulting in yield losses [3–5]. Under favorable climatic conditions and in susceptible cultivars, reported yield losses caused by RFS range from 3% to 80% [6]. In addition to yield reduction,

U. virens produces mycotoxins, including ustilaginoidins and ustiloxins, which may contaminate rice and feed products and raise concerns regarding food safety [7,8]. At present, field diagnosis and surveillance of RFS still rely mainly on manual visual scouting. Plant protection specialists and technicians assess disease severity by estimating the number and distribution of false smut balls on panicles [9]. However, manual scouting and visual rating require substantial labor and are affected by evaluator experience, resulting in variation among assessments [10]. For large-scale outbreaks, manual monitoring may be insufficient for rapid disease assessment. Traditional diagnostic methods, such as microscopic examination and reagent-based assays, are also limited by cost and processing time, restricting their application in large-scale field evaluation [11]. Therefore, intelligent detection methods are needed to assist manual assessment and enable more consistent and quantitative evaluation of RFS in field conditions [12].

In recent years, deep learning has been increasingly applied to crop disease detection. Compared with traditional image-processing and conventional machine-learning methods, convolutional neural networks (CNNs) can automatically learn discriminative lesion features and perform end-to-end classification and detection across diverse plant disease datasets [13–15]. Among these methods, the YOLO (You Only Look Once) family has been widely adopted because it offers a favorable balance between detection speed and accuracy [16]. Previous studies have used YOLO-based models to localize disease symptoms and identify disease categories in field images, including multispecies, tomato, and maize leaf diseases, as well as lesions in orchard crops [17–20]. YOLO-based detectors have also been used to recognize and count panicle-scale targets in cluttered field scenes [21], which is relevant to the quantitative assessment of panicle diseases such as rice false smut. In addition, YOLO has been applied to apple, cherry, and lemon detection in unconstrained environments [22–24].

Despite these advances, automated disease severity assessment and grading remain less developed than disease detection and classification tasks [25]. Most existing intelligent diagnosis systems focus on disease presence detection or category identification, while quantitative severity assessment, such as lesion area proportion or symptom extent, is less frequently considered [26]. Therefore, integrating severity-related measurements into automated detection frameworks has become an important direction for quantitative disease analysis. Plant pathology has established several severity evaluation indicators, including lesion area ratio, lesion count, and Disease Severity Index (DSI). Among these indicators, DSI summarizes disease intensity by combining severity grades across a population. In rice false smut assessment, disease severity is commonly evaluated according to the number of infected spikelets per panicle, and field-level DSI can be calculated to describe overall disease status [27]. The International Rice Research Institute (IRRI) further standardized disease-rating terminology and scoring scales through the Standard Evaluation System for Rice (SES), facilitating comparisons of disease severity across locations and evaluation groups [28].

Although quantitative severity indicators have been established in plant pathology, their integration into intelligent analysis of panicle diseases remains limited. For rice false smut, the transformation from object detection results to quantitative severity assessment is still challenging because infected regions may exhibit small-scale targets, dense distribution, and variable appearance in field images. In addition, many existing studies still focus on identifying whether a disease occurs, whereas efforts that transform detection outputs into interpretable and reproducible severity grades are relatively limited. For rice false smut specifically, grading annotation protocols and quantitative indicator systems have not yet been standardized, which hinders precision management decisions and cross-region comparability.

Motivated by these limitations, this study targets field detection and severity grading of rice false smut and makes the following contributions:

1. Multi-year, multi-site dataset construction and standardized preprocessing: Rice false smut panicle images collected from different years and sites were integrated to construct the dataset. Before model training, image preprocessing procedures were standardized, and class distribution and target-scale characteristics were analyzed to describe the dataset used for detection and severity assessment.
2. YOLOv12-RSLW detector development for rice false smut detection: An improved YOLOv12 detector was developed to improve the detection of small and densely distributed false smut targets in field images. RepGhost, SimAM, LSCD, and WIoU were incorporated into different components of the network to enhance feature representation, lesion-related feature responses, multi-scale prediction, and bounding-box regression, respectively.
3. A detection–segmentation framework for single-panicle severity assessment with count-area calibration: Because adhesion and coalescence of false smut lesions may cause count-based grading to underestimate visible lesion extent, detection boxes generated by YOLOv12-RSLW were used as prompts for SAM to obtain panicle and lesion masks, and the lesion-to-panicle area ratio was calculated. The initial severity grade was assigned according to the number of false smut balls, while the area ratio was introduced as a supplementary indicator to identify samples whose visible lesion extent corresponded to higher count-based severity categories. Based on the relationship between lesion area ratio and count-based grades, the Count–Area Calibration Strategy (CACS) was developed as a complementary strategy for severity characterization.

2. Materials and Methods

2.1. Data Acquisition and Preprocessing

2.1.1. Experimental Design and Close-Range Image Collection of Rice False Smut

Field experiments and close-range imaging of rice false smut were conducted in Liaoning Province, China, to construct a panicle-scale image dataset. Data were collected in two stages. In Stage I, images were acquired from August to September 2023, covering the heading-to-maturity period, at the Scientific Experiment Base of Shenyang Agricultural University in Shenyang. The rice cultivars included Liaojing 9, Beijing 202, Beijing 204, and Beijing 1705. Images were captured every three days between 10:00 and 15:00 using a Canon EOS 1500D digital camera (Canon Inc., Tokyo, Japan). This stage yielded 900 close-range field images of rice panicles affected by false smut, which constituted the initial dataset reported in our previous study [29]. In Stage II (2024–2025), additional images were collected at two experimental bases of Shenyang Agricultural University in Shenyang and Haicheng. The same sampling interval and daily acquisition period were used as in Stage I, with images captured every three days between 10:00 and 15:00 using an iPhone 12 Pro Max (Apple Inc., Cupertino, CA, USA). The Haicheng site primarily included Shennong 9903 and Huarun 2, whereas the Shenyang site included Beijing 2501, Beijing 2502, Beijing 2503, Liaojing 9, Beijing 202, Beijing 204, and Beijing 1705. This stage yielded 1011 close-range field images. By integrating images collected over three consecutive years (2023–2025), a total of 1911 close-range field images of rice false smut were assembled. The dataset covered multiple experimental sites, rice cultivars, field backgrounds, and image-acquisition conditions, thereby incorporating interannual and inter-site variation. These images were subsequently used for panicle and lesion detection and severity assessment in Table 1.

Table 1. Original images by year, site, and cultivar.

Year	Site	Cultivar	Images (<i>n</i>)
2023	Shenyang	Liaojing 9, Beijing 202, Beijing 204, Beijing 1705	900
2024	Haicheng	Shennong 9903, Huarun 2	500
2025	Shenyang	Beijing 2501, Beijing 2502, Beijing 2503, Liaojing 9, Beijing 202, Beijing 204, Beijing 1705	511

2.1.2. Data Preprocessing

Because field images collected in different years varied in resolution and imaging scale, all original images were standardized before model training. Each image was resized to a width of 640 pixels while preserving its aspect ratio, and the corresponding bounding-box coordinates were scaled accordingly. The resized images were then processed to 640×640 pixels using letterbox padding, with the same scaling factor and padding offsets applied to the bounding boxes. Random scaling, as described below, was performed offline before model training. These steps provided a consistent input scale while preserving the original image aspect ratio.

Before offline augmentation, the 1911 original image–annotation pairs were divided at the source-image level into training, validation, and test subsets containing 1520, 206, and 185 images, respectively. Augmentation was then performed separately within each subset. All augmented images derived from the same source image remained in the same subset, preventing derivatives of a single source image from appearing across different subsets. The original dataset contained 6324 annotated false smut balls and 3236 annotated panicle instances.

To increase sample diversity and reduce imbalance between the two annotation classes, six offline augmentation operations were applied. Random scaling resized images to 50–80% of their original scale, while random rotation was limited to -30° to 30° . Brightness was adjusted by modifying the V channel in the HSV color space, and Gaussian blur was used to simulate defocus or motion blur. Random contrast and saturation adjustments were also applied. Bounding-box coordinates were recalculated after geometric transformations.

Because false smut balls and coalesced lesions were more numerous than panicle instances, images containing panicle targets were selectively oversampled. Here, a coalesced lesion was defined as a continuous lesion region in which adjacent false smut balls could not be reliably distinguished individually. The number of additional panicle-containing images was estimated from the difference between the two class-instance counts and the average number of panicle annotations per image. For each sampled image, augmentation operations were randomly selected from the six methods described above. Consequently, the number of augmented versions varied among source images.

After augmentation, the dataset comprised 5663 images, including 4531 training images, 566 validation images, and 566 test images. It contained 11,191 panicle instances, including the original annotations and their transformed counterparts in the augmented images. The final dataset composition is summarized in Table 2, and representative augmented samples are shown in Figure 1.

Table 2. Dataset composition.

Dataset	Total Images	Original Images	Augmented Images	Original Panicle	Augmented Panicle	Original False Smut Balls	Augmented False Smut Balls
Training	4531	1520	3011	2531	8856	5026	11,236
Validation	566	206	360	388	1195	688	1462
Test	566	185	381	317	1140	610	1506
Total	5663	1911	3752	3236	11,191	6324	14,204

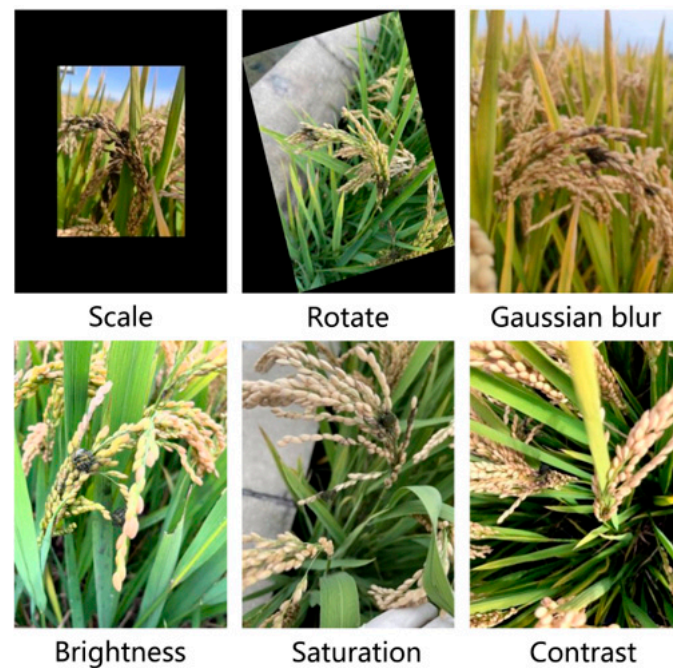


Figure 1. Data augmentation.

2.2. Count-Based Severity Grading and Exploratory Area-Ratio Supplementary Strategy

Several severity-grading systems for rice false smut are used in China, reflecting differences in disease forecasting, plant-protection statistics, field efficacy evaluation, and resistance assessment [30]. Their grading criteria and severity intervals are therefore not fully uniform. Deng et al. used six categories including Grade 0, corresponding to 0, 1, 2–5, 6–10, 11–15, and ≥ 16 false smut balls per panicle [31]. Tang et al. used intervals of 0, 1, 2, 3–5, 6–9, and ≥ 10 false smut balls per panicle [32]. The Chinese agricultural industry standard NY/T 1464.54-2014 [33], Pesticide Guidelines for the Field Efficacy Trials—Part 54: Fungicides against False Smut of Rice, uses the same count intervals as Tang et al. for field efficacy trials, represented by disease grades 0, 1, 3, 5, 7, and 9. Regional standards have also been developed for different purposes. DB43/T 504-2009 [34] specifies procedures for rice false smut forecasting surveys, whereas DB43/T 505-2009 [35] addresses the identification and evaluation of varietal resistance to rice false smut. Despite differences in grade numbering and application purposes, these count-based systems show substantial agreement in the low and moderate severity ranges. In particular, 0 and 1 false smut ball per panicle are consistently distinguished, while the interval of 2–5 false smut balls is either defined directly or represented by the adjacent intervals of 2 and 3–5. Based on these shared count boundaries, the present study retained Grade 0 and Grade 1 as separate categories and defined Grade 2 as 2–5 false smut balls per panicle. For higher severity levels, the existing systems further subdivide samples above this range, but such samples were relatively limited in the present dataset. We therefore grouped samples with ≥ 6 false smut balls per panicle into Grade 3. Accordingly, Grade 0 denotes a healthy panicle without visible false smut balls or coalesced lesions, Grade 1 indicates one false smut ball per panicle, Grade 2 corresponds to 2–5 false smut balls per panicle, and Grade 3 represents ≥ 6 false smut balls per panicle. Representative field images of Grades 0–3 are shown in Figure 2.

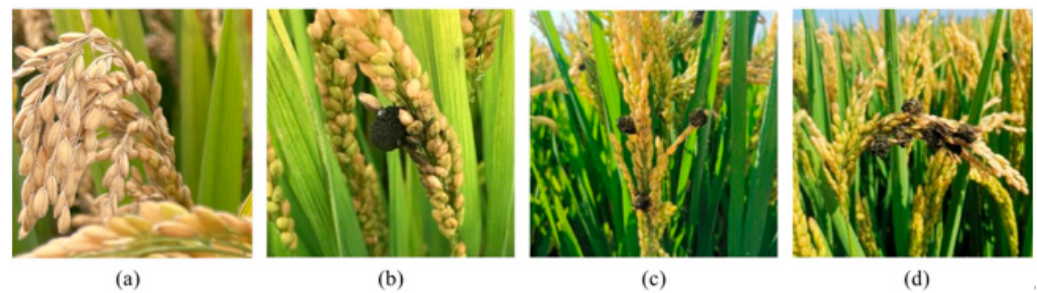


Figure 2. Representative field images of count-based rice false smut severity grades. (a) Grade 0, a healthy panicle without visible false smut balls or coalesced lesions; (b) Grade 1, with one false smut ball per panicle; (c) Grade 2, with two to five false smut balls per panicle; and (d) Grade 3, with six or more false smut balls per panicle.

However, the field manifestation of rice false smut is often complex and uncertain. Multiple smut balls may adhere to and merge with each other, forming continuous, patch-like lesions. As illustrated in Figure 3, the symptoms in such cases are closer to a single coalesced lesion rather than several clearly separable individual smut balls, making it difficult to distinguish severity solely by counting smut balls per panicle. Therefore, for these samples, we annotated the entire coalesced lesion region as a whole. If severity is assessed solely according to the detected number of smut balls per panicle, the resulting count-based grade may not fully represent the affected lesion area in samples with adhesion or coalescence, thereby creating a potential discrepancy between count-based and area-based disease characteristics.

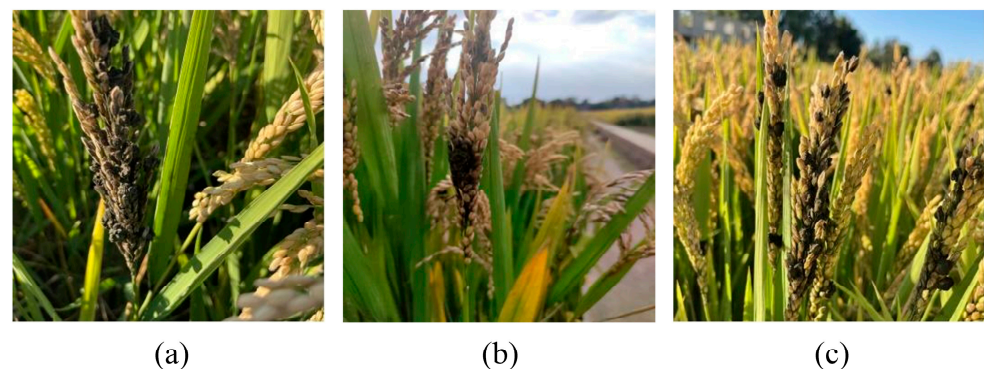


Figure 3. Representative field examples of adhered and coalesced lesions. (a–c) Different lesion adhesion and coalescence patterns that obscure the boundaries of individual false smut balls and may lead to underestimation by count-based severity grading.

To address this practical issue, we augment the count-based grading rule (Grades 1–3) by incorporating a lesion-area indicator and propose incorporating a lesion-area indicator and propose an exploratory Count–Area Calibration Strategy (CACS). Using a detection–segmentation pipeline composed of YOLOv12-RSLW and SAM, we obtain pixel-level masks for both the panicle region and the lesion region. The pixel areas (i.e., the number of mask pixels) are then computed for each mask, and the lesion-to-panicle area ratio is calculated as the proportion of lesion area relative to the panicle area.

2.3. Design of the YOLOv12-RSLW Detector for Rice False Smut

Under field conditions, false smut targets are often small and densely distributed, while scale variation, blurred boundaries, adhesion, and background interference further complicate detection. To improve the representation and localization of these targets, YOLOv12 was adopted as the baseline and modified at the backbone feature-extraction,

attention-modeling, detection-head, and bounding-box regression stages to form YOLOv12-RSLW. YOLOv12 is an attention-centric one-stage real-time detector that retains the end-to-end detection paradigm of the YOLO family while incorporating attention-based feature modeling [36]. As shown in Figure 4, five RepGhost modules are arranged at different stages of the backbone and are labeled RepGhost1 to RepGhost5 according to their positions; all five modules share the same internal structure.

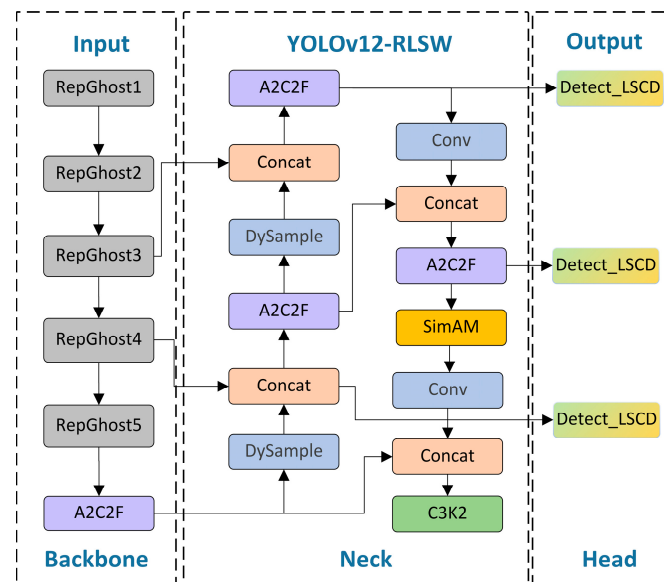


Figure 4. YOLOv12-RSLW network structure.

The A2C2F structure was used for cross-stage feature modeling to facilitate feature aggregation and gradient propagation across different feature levels. This structure supports information interaction across scales and reduces the effect of target-size variation on feature representation. SimAM was embedded into key feature output stages to provide parameter-free attention weighting. It constructs an energy function from within-channel statistics, including the mean and variance of the feature maps, and generates element-wise attention weights through normalization and sigmoid mapping. This adaptive weighting emphasizes lesion-related responses while suppressing background activations.

To improve the prediction of small and densely distributed targets, LSCD was incorporated into the detection head to strengthen multi-scale feature utilization. Its shared detection structure was designed to support prediction across different feature resolutions, including closely adjacent or partially adhered false smut targets. WIoU was introduced into the regression branch for bounding-box localization. Its dynamic scaling and non-monotonic focusing mechanism assigns different gradient weights according to sample difficulty, allowing the regression loss to adapt to targets with varying localization quality. Together, these components modify feature aggregation, attention weighting, multi-scale prediction, and bounding-box regression within the YOLOv12 framework.

2.3.1. RepGhost

RepGhost was introduced into selected stages of the backbone network to improve feature extraction while retaining a simple deployment structure [37]. As shown in Figure 5a, RepGhost uses a multi-branch configuration during training, consisting of a main convolutional branch and a batch normalization (BN) bypass branch. Figure 5b shows the equivalent single convolutional path obtained after structural reparameterization for deployment. The outputs of these branches are combined by element-wise addition. During deployment, structural re-parameterization folds the training-time branches into an equivalent single

convolutional path, reducing the structural complexity of the inference network while preserving the corresponding transformation. This configuration was used to strengthen local feature extraction for the joint detection of rice panicles and false smut targets.

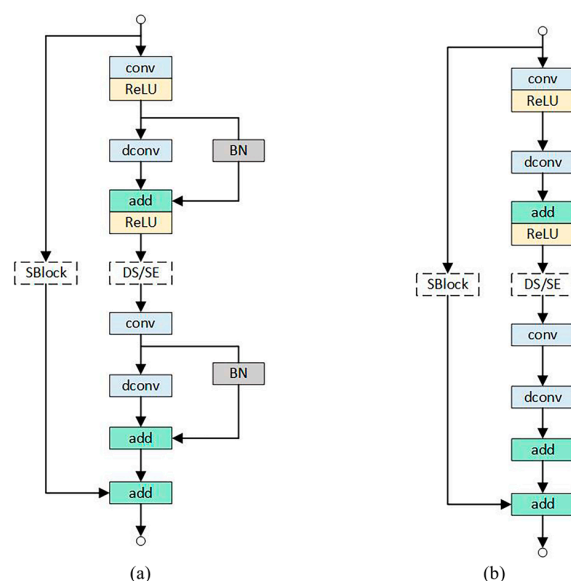


Figure 5. RepGhost module structure. (a) Re-parameterized multi-branch configuration during training. (b) Equivalent single convolutional path during deployment.

2.3.2. SimAM

SimAM (Simple, Parameter-Free Attention Module) is a parameter-free attention mechanism that adaptively assigns a weight to each neuron in a feature map. Its key idea is to derive a closed-form solution to an energy function constructed from within-channel statistics, thereby estimating the importance of each spatial location and rescaling feature responses accordingly. As a result, SimAM enhances feature representation while introducing virtually no additional learnable parameters [38]. As shown in Figure 6, unlike attention modules that generate only 1D channel-wise weights or 2D spatial weights, SimAM aims to infer a full 3D attention map of size $C \times H \times W$, enabling joint modulation across both channels and spatial positions. After integrating SimAM, target-related activations can be strengthened and redundant background responses can be suppressed at the feature level, providing more stable and discriminative representations for detecting rice panicles and false smut balls in complex field scenes.

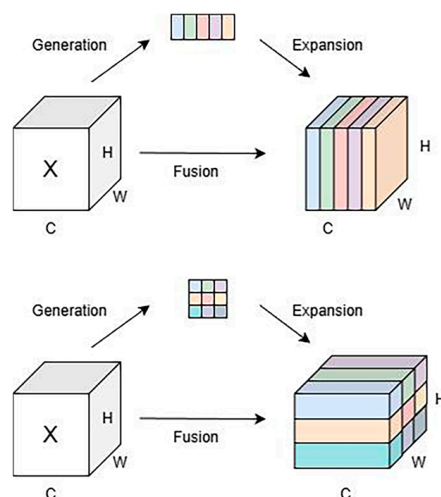


Figure 6. Cont.

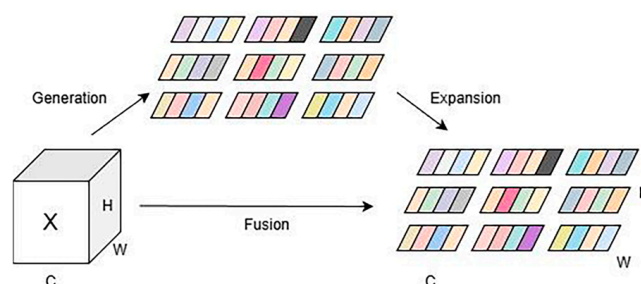


Figure 6. SimAM module structure.

2.3.3. Loss Function Design

To address the difficulty of precise localization for hard samples, we redesign the regression loss in YOLOv12 by replacing the conventional CIoU-based formulation with Wise-IoU (WIoU) in the proposed model [39]. Compared with CIoU, WIoU introduces two key improvements. First, it explicitly increases the model's focus on hard-to-localize cases—such as targets embedded in cluttered backgrounds or severely blurred regions—thereby improving bounding-box localization accuracy. Second, it incorporates a dynamic penalty mechanism that adaptively adjusts the strength of the penalty according to sample difficulty.

In rice false smut detection, CIoU exhibits notable limitations. The target scales of false smut balls and rice panicles differ substantially, and in some samples the symptoms appear as coalesced or block-like lesions. However, CIoU assigns a uniform gradient weighting across samples, which is suboptimal for such multi-scale and morphologically uncertain targets. Moreover, field imagery often contains low-quality samples due to occlusion and non-uniform illumination; naively forcing stronger regression on these samples may degrade localization performance rather than improve it. To overcome these issues, we adopt WIoU to enhance both multi-scale adaptability and bounding-box regression precision.

WIoU employs a Dynamic Non-monotonic Focusing Mechanism, where the outlier degree (instead of IoU alone) is used to assess anchor quality and to provide an adaptive strategy for gradient gain allocation. The WIoU loss is defined as:

$$\mathcal{L}_{WIoU} = \frac{\beta}{\delta \alpha^{\beta-\delta}} \cdot R_{WIoU} \cdot \mathcal{L}_{IoU}, \mathcal{L}_{IoU} = (1 - IoU) \quad (1)$$

where δ and α are hyperparameters controlling the focusing mechanism. R_{WIoU} denotes the distance-aware attention factor, which measures the normalized distance between the center points of the predicted box and the ground-truth box, and is computed as:

$$R_{WIoU} = \exp \left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)} \right) \quad (2)$$

In addition, β represents the outlier degree, which quantifies how much the current anchor quality deviates from the overall sample distribution. This term guides the model to allocate different learning weights to samples of varying difficulty and serves as the core component for intelligent gradient assignment. Here, $\mathcal{L}_{IoU}^*$ and $\overline{\mathcal{L}_{IoU}}$ denote two forms of the $\mathcal{L}_{IoU}$ loss term, where $\mathcal{L}_{IoU}^*$ does not participate in gradient backpropagation, and $\overline{\mathcal{L}_{IoU}}$ is taken as the mean $\mathcal{L}_{IoU}$ loss used for optimization.

$$\beta = \frac{\mathcal{L}_{IoU}^*}{\overline{\mathcal{L}_{IoU}}} \in [0, +\infty) \quad (3)$$

2.3.4. LSCD Detection Head

The LSCD detection head was introduced into YOLOv12-RSLW to reduce redundant computation and improve multi-scale prediction of false smut targets [40]. As shown in Figure 7, LSCD receives three feature maps, P3, P4, and P5, from the neck network. P3 preserves higher-resolution spatial information for small-target detection, whereas P4 and P5 provide stronger semantic and contextual information for distinguishing panicles and coalesced lesions. Each feature map is first processed by a 3×3 Conv_GN block for feature alignment. The aligned features are then passed through a shared prediction structure composed of 3×3 and 1×1 Conv_GN layers. Unlike the original detection head, which uses separate prediction branches for different scales, LSCD shares the main convolutional parameters across feature levels to reduce repeated computation while retaining scale-specific outputs. In the regression branch, a Scale layer adjusts the bounding-box regression outputs according to feature resolution. The resulting detection boxes are subsequently used as prompts for SAM-based mask extraction and count-area severity calibration.

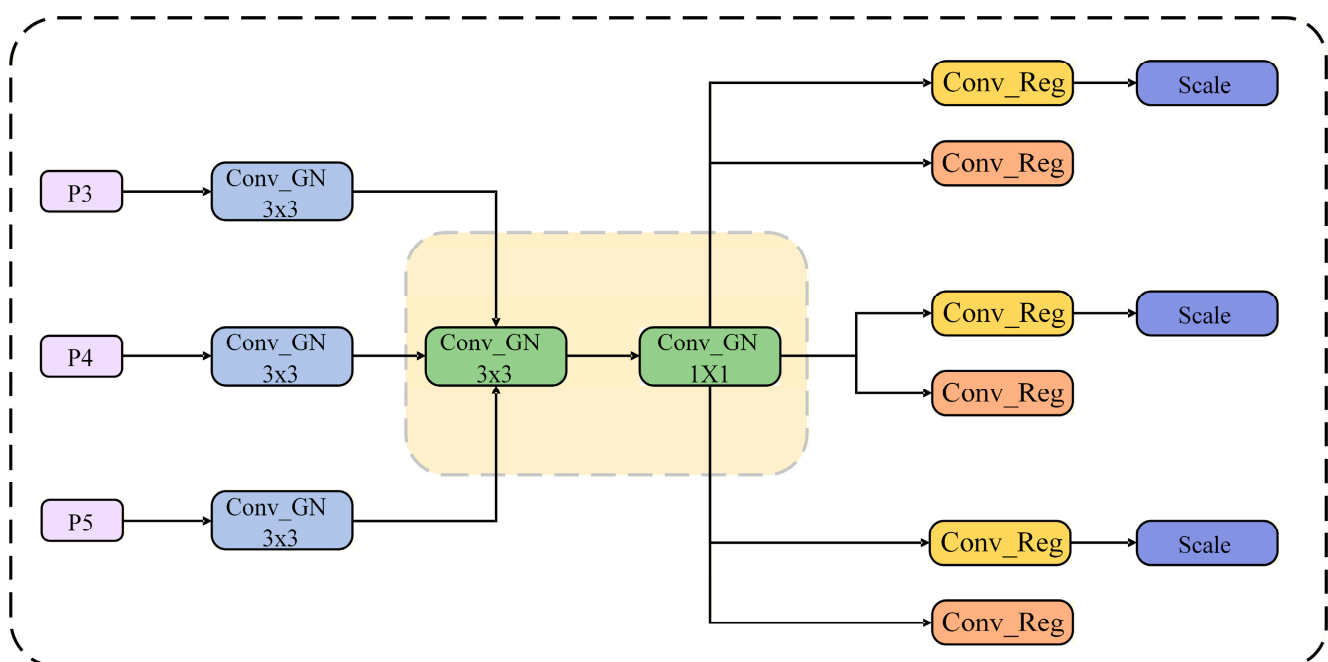


Figure 7. LSCD module structure.

2.3.5. Intelligent Detection and Severity Grading of Rice False Smut

The overall methodological framework is shown in Figure 8. It comprises three main stages: dataset construction, detector development and validation, and inference-based severity assessment. Field images collected from 2023 to 2025 were annotated for rice panicles and false smut regions and processed according to the procedures described in Section 2.1. After preprocessing and offline augmentation, the dataset contained 5663 images, including 4531 training images, 566 validation images, and 566 test images. For detector development, YOLOv12 was used as the baseline. RepGhost, SimAM, LSCD, and WIoU were incorporated to improve feature representation, lesion-related feature responses, multi-scale prediction, and bounding-box regression, respectively, forming the YOLOv12-RSLW detector. The model was trained and selected using the settings described in Section 2.4, and its detection performance was evaluated using the metrics defined in Section 2.5. During inference, YOLOv12-RSLW generated bounding boxes for rice panicles and false smut regions. The detected boxes were then used as prompts for SAM to obtain the corresponding pixel-level masks. Based on these masks, the lesion-to-panicle area ratio

was calculated for each panicle. Severity assessment combined the detected number of false smut balls with the lesion-to-panicle area ratio. The initial grade was determined using the count-based rule defined in Section 2.2. For samples with lesion adhesion or coalescence, the area ratio was used as a supplementary criterion to characterize lesion extent. Samples whose area ratios reached the thresholds derived from higher-grade distributions were assigned the corresponding CACS grades for subsequent analysis. The workflow therefore links object detection, mask extraction, quantitative lesion measurement, and severity grading at the single-panicle level.

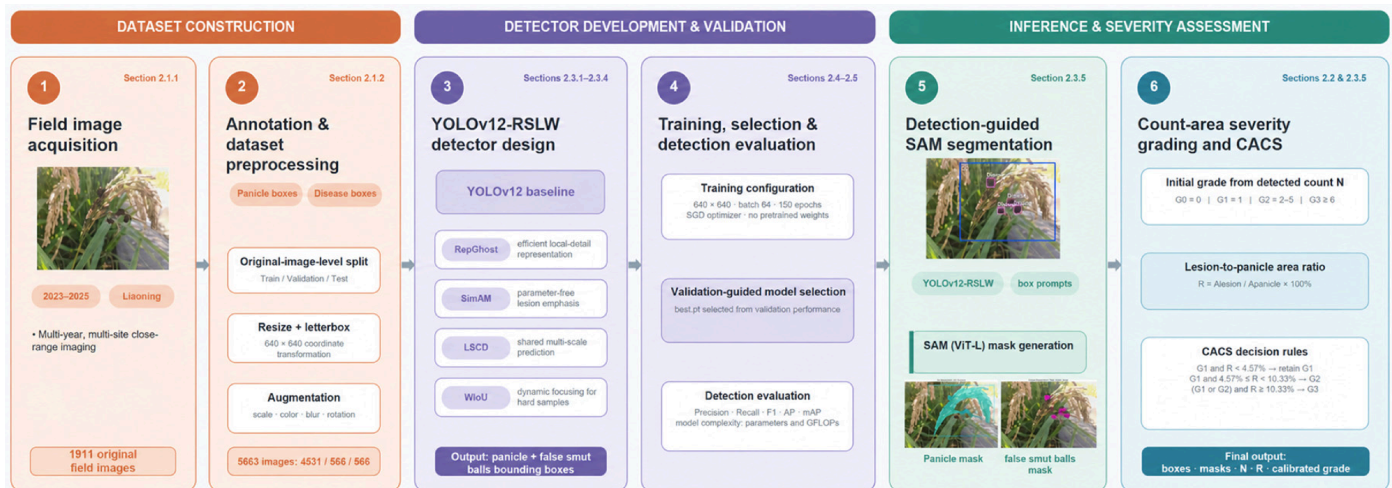


Figure 8. Methodological Framework.

2.4. Experimental Environment and Parameter Settings

All experiments were conducted on a workstation running Microsoft Windows 11 Pro for Workstations and equipped with an Intel Xeon W5-2455X CPU (12 cores/24 threads, 3.2 GHz), 128 GB RAM, and an NVIDIA RTX A4000 GPU with 16 GB memory. Model training and evaluation were implemented in Python 3.10 using PyTorch 2.2.2 and Ultralytics 8.3.9.

Standard YOLOv12, defined by the local yolov12.yaml configuration, was used as the baseline, while YOLOv12-RSLW incorporated RepGhost, SimAM, LSCD, and WIoU. No pretrained weights were used; all models were initialized from their corresponding YAML configurations with pretrained = False. The dataset configuration was specified in data_yolov12main.yaml and contained two target classes, Panicle and Disease. All models used the same training, validation, and test subsets. The input size was 640×640 pixels, the batch size was 64, and the maximum number of training epochs was 150. Early stopping was configured with patience = 100, although all reported models completed the full 150 epochs. The checkpoint best.pt was selected according to validation performance. SGD was used with an initial learning rate of 0.01, a final learning-rate factor of 0.01, momentum of 0.937, and weight decay of 0.0005. A three-epoch warm-up period was applied with a warm-up momentum of 0.8 and a warm-up bias learning rate of 0.1. Automatic mixed precision was enabled, the random seed was fixed at 0, and deterministic training was applied. During validation, the confidence threshold was left unspecified (conf = None) so that the Ultralytics default setting was used for AP calculation. The NMS IoU threshold was 0.7, the maximum number of detections per image was 300, and class-agnostic NMS was disabled.

For segmentation, the original Segment Anything Model (SAM, version 1.0) released by Meta AI was used with the ViT-L image encoder and the pretrained checkpoint sam_vit_l_0b3195.pth, without task-specific fine-tuning. Input images were converted

to RGB format, resized while preserving their aspect ratio, and then processed using the normalization and padding procedures of SAM. For the 80-image segmentation evaluation, manually annotated bounding boxes were used as SAM prompts and transformed to the SAM input coordinate space. Multiple candidate masks were generated for each prompt, and the mask with the highest predicted IoU score was retained. The selected mask was restricted to the corresponding prompt box, after which only the largest 8-connected component was kept. No additional hole filling or morphological opening or closing was applied.

2.5. Evaluation Metrics

To comprehensively evaluate the performance of YOLOv12-RSLW, we adopted standard metrics widely used in object detection, including Precision, Recall, Average Precision (AP), and Mean Average Precision (mAP). Precision measures the proportion of true positives among all predicted positives, reflecting the model's tendency to produce false detections. Recall measures the proportion of true positives among all ground-truth positives, reflecting the model's ability to avoid missed detections. These two metrics are complementary and jointly characterize the accuracy and completeness of detection. The F1-score is the harmonic mean of Precision and Recall and provides a single measure of their balance, particularly when both false detections and missed detections need to be considered. For mAP@0.5 evaluation, a prediction was counted as a true positive (TP) when its predicted class matched the ground-truth class and the bounding-box IoU with the matched ground-truth object was at least 0.5. Unmatched predictions were counted as false positives (FP), while unmatched ground-truth objects were counted as false negatives (FN). The reported Precision and Recall values correspond to the operating point at which the mean F1-score is maximized.

AP summarizes the trade-off between precision and recall across different confidence thresholds and reflects the overall detection performance for a single class. mAP is computed as the mean of AP values across all classes and therefore indicates the overall multi-class detection capability for rice panicles and false smut balls. AP was calculated from the confidence-ranked precision–recall curve using the precision envelope and 101-point interpolation over recall values from 0 to 1, followed by trapezoidal integration. The reported mAP@0.5 was obtained by averaging the AP values of the Panicle and Disease classes at an IoU threshold of 0.5. The formulations of Precision, Recall, AP, and mAP are given in Equations (4)–(8).

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = \frac{2PR}{P + R} \quad (6)$$

$$AP_c = \sum_{k=0}^{99} \frac{r_{k+1} - r_k}{2} \left[\tilde{P}_c(r_k) + \tilde{P}_c(r_{k+1}) \right] \quad (7)$$

$$mAP@0.5 = \frac{1}{C} \sum_{c=1}^C AP_c@0.5 \quad (8)$$

Segmentation performance was evaluated using Intersection over Union (IoU) and the Dice coefficient. For the i -th evaluation sample, S_i denotes the SAM-generated mask and G_i denotes the corresponding manually annotated ground-truth mask. IoU measures the overlap relative to the union of the predicted and manual masks, whereas Dice measures their overlap relative to their combined area. The reported IoU and Dice values were

averaged over the 80 evaluation images for each segmentation target. Because severity assessment subsequently uses the lesion-to-panicle area ratio, the ratios obtained from automatic segmentation were also compared with those calculated from manual masks. Let M_i and A_i denote the manually measured and automatically estimated lesion-to-panicle area ratios for the i -th sample, respectively, and let N denote the number of evaluated samples ($N = 80$). $\bar{M}$ and $\bar{A}$ denote the mean manual and automatic area ratios, respectively. MAE measures the mean absolute difference between the two estimates, while RMSE gives greater weight to larger deviations. R^2 describes the agreement of the automatic estimates with the manual measurements, and Pearson's r quantifies their linear association. The corresponding formulations are given in Equations (9)–(14).

$$IoU_i = \frac{|S_i \cap G_i|}{|S_i \cup G_i|} \quad (9)$$

$$Dice_i = \frac{2|S_i \cap G_i|}{|S_i| + |G_i|} \quad (10)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |A_i - M_i| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - M_i)^2} \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (M_i - A_i)^2}{\sum_{i=1}^N (M_i - \bar{M})^2} \quad (13)$$

$$r = \frac{\sum_{i=1}^N (M_i - \bar{M})(A_i - \bar{A})}{\sqrt{\sum_{i=1}^N (M_i - \bar{M})^2 \sum_{i=1}^N (A_i - \bar{A})^2}} \quad (14)$$

3. Results and Analysis

3.1. Evaluation of the Standard YOLOv12 Model and Dataset Characteristics

The baseline performance of YOLOv12 was evaluated using the precision–recall (PR) and F1–confidence curves shown in Figure 9a,b. The AP values for the Disease and Panicle classes were 0.813 and 0.956, respectively, yielding an overall mAP@0.5 of 0.885. The lower AP for the Disease class shows that false smut targets were more difficult to detect than panicles. The maximum F1-score was 0.85 at a confidence threshold of 0.374, which was used to characterize the balance between Precision and Recall for the baseline model.

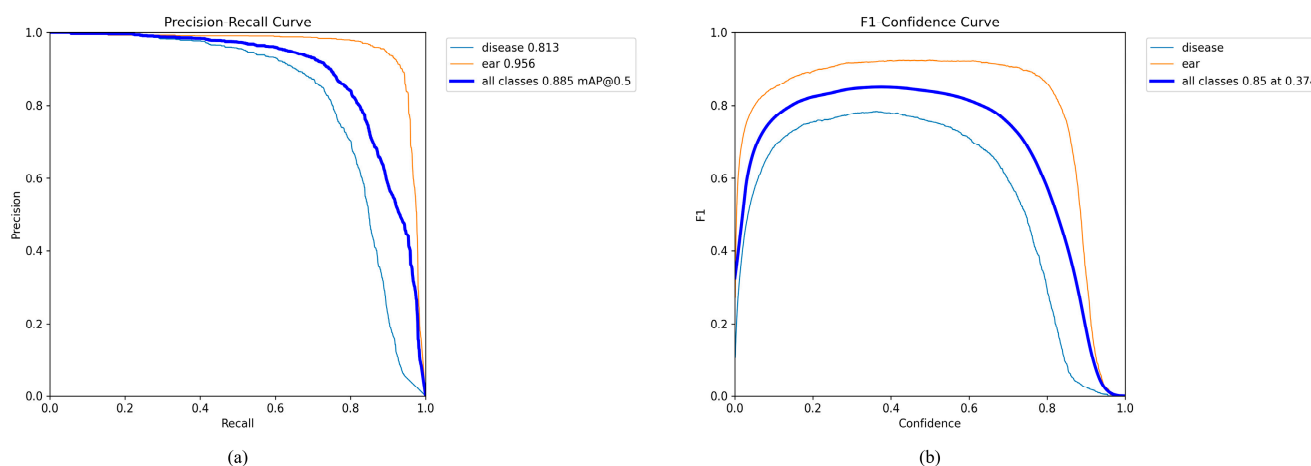


Figure 9. Results of the rice false smut detection task. (a) PR curve. (b) F1-confidence curve.

At the 640×640 model-input scale, the median bounding-box area was 922.34 pixels² for Disease targets and 38,266.56 pixels² for Panicle targets. Disease targets had median width and height values of 30.25 and 30.48 pixels, respectively, with a median aspect ratio of 0.98. Their spatial distribution was also locally dense: each image contained an average of 2.51 Disease targets and a maximum of 17, while 38.16% and 13.72% of the images contained at least three and five Disease targets, respectively. The median nearest-neighbor distance between Disease centroids was 58.85 pixels, and 18.15% of the targets had another Disease target within 32 pixels. As shown in Figure 10, the target centroids were concentrated mainly around panicle regions.

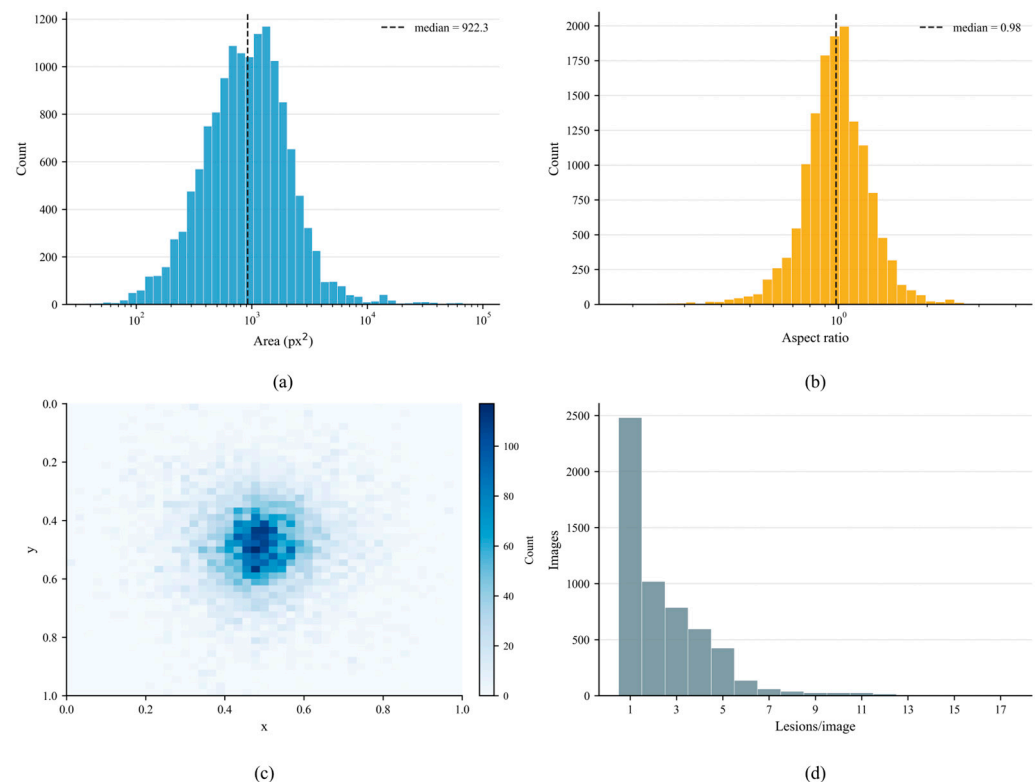


Figure 10. Quantitative characteristics of false smut targets in the dataset. (a) Bounding-box area distribution; (b) aspect ratio distribution; (c) centroid heatmap; (d) number of false smut targets per image.

At the median target size, a Disease bounding box occupies approximately 3.8×3.8 cells on a stride-8 feature map and 1.9×1.9 cells on a stride-16 feature map. This limited spatial representation may reduce the preservation of local boundary and texture information after successive downsampling. In addition, short inter-target distances may increase feature interference between adjacent false smut balls and make individual targets more difficult to separate. These characteristics may partly explain the lower detection performance observed for small Disease targets. Their effect is examined further in the size-sensitivity analysis in Section Sensitivity of Detection Performance to False Smut Ball.

3.2. Ablation Study

Ablation experiments were conducted under identical training and evaluation settings to assess the contribution of RepGhost, SimAM, LSCD, and WIoU. Standard YOLOv12 served as the baseline, and each module was first evaluated independently before testing the complete YOLOv12-RSLW configuration. As summarized in Table 3, the baseline achieved 88.46% mAP@0.5 and 62.23% mAP@0.5:0.95.

Table 3. Results of the ablation study.

Model	RepGhost	SimAM	LSCD	WIoU	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Precision (%)	Recall (%)	F1 (%)	Params	GFLOPs
YOLOv12					88.46	62.23	88.14	82.64	85.30	2,527,166	5.8
+RepGhost	✓				91.51	67.17	91.10	86.91	88.96	3,568,941	6.2
+SimAM		✓			88.53	63.27	89.25	81.13	85.00	2,538,486	6.0
+LSCD			✓		89.93	63.91	90.12	82.47	86.13	2,365,181	5.1
+WIoU				✓	89.58	64.02	90.04	83.17	86.47	2,527,166	5.8
YOLOv12-RSLW	✓	✓	✓	✓	92.06	68.02	91.46	87.01	89.18	3,454,097	6.0

Among the individual modules, RepGhost produced the largest performance gain, increasing mAP@0.5 to 91.51% and mAP@0.5:0.95 to 67.17%, although the parameter count increased from 2.53 M to 3.57 M. SimAM produced only a marginal change in mAP@0.5 but increased Precision while reducing Recall, indicating a different Precision–Recall trade-off when used alone. LSCD increased mAP@0.5 to 89.93% while reducing the parameter count to 2.37 M and computational cost to 5.1 GFLOPs. WIoU increased mAP@0.5 to 89.58% without changing the parameter count or GFLOPs relative to the baseline.

The complete YOLOv12-RSLW configuration achieved the highest overall performance, with 92.06% mAP@0.5 and 68.02% mAP@0.5:0.95. Precision, Recall, and F1-score reached 91.46%, 87.01%, and 89.18%, respectively, corresponding to increases of 3.32, 4.37, and 3.88 percentage points over YOLOv12. The complete model contained 3.45 M parameters and required 6.0 GFLOPs. Although RepGhost increased model complexity, the shared convolutional design of LSCD partly offset this increase, resulting in fewer parameters and lower computational cost than the RepGhost-only configuration.

Figure 11 compares the training curves of mAP@0.5 and Precision for YOLOv12 and the different ablation configurations. During the later stages of training, YOLOv12-RSLW maintained higher mAP@0.5 and Precision than the other configurations, consistent with the final results reported in Table 3.

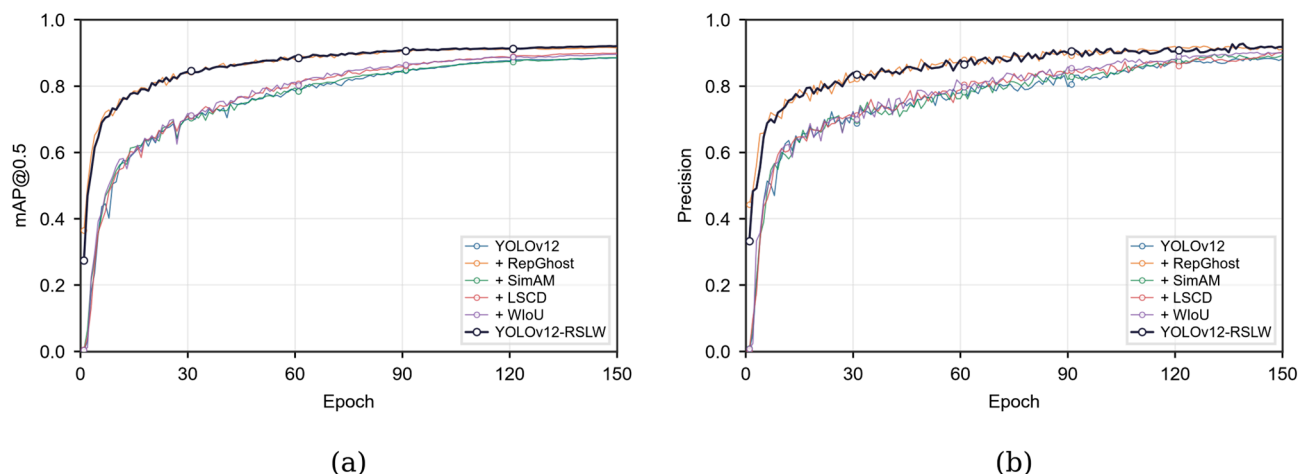
**Figure 11.** Training curves of different models in the ablation study. (a) mAP@0.5. (b) Precision.

Table 4 compares the class-specific performance of the different model configurations. Standard YOLOv12 showed a clear gap between the two target classes: Precision, Recall, and AP@0.5 were 85.42%, 72.16%, and 81.32% for Disease, compared with 90.93%, 93.14%, and 95.56% for Panicle, respectively. Among the single-module configurations, RepGhost produced the largest improvement for Disease, increasing Recall from 72.16% to 79.00% and AP@0.5 from 81.32% to 85.98%. SimAM produced little change in Disease AP@0.5 and

reduced Recall to 70.66%, whereas LSCD and WIoU mainly increased Disease Precision, reaching 88.69% and 87.62%, respectively.

Table 4. Class-Specific Detection Performance of Different Model Configurations.

Model	DiseaseP (%)	DiseaseR (%)	DiseaseAP@0.5 (%)	EarP (%)	EarR (%)	EarAP@0.5 (%)
YOLOv12	85.52	72.16	81.32	90.93	93.14	95.56
+RepGhost	86.20	79.00	85.98	95.88	95.06	97.56
+SimAM	86.48	70.66	81.49	91.45	92.80	95.78
+LSCD	88.69	71.87	83.64	91.24	93.39	96.34
+WIoU	87.72	72.64	82.56	93.74	92.89	96.74
YOLOv12-RSLW	88.10	78.79	86.67	95.52	94.64	97.40

The complete YOLOv12-RSLW model achieved a Disease AP@0.5 of 86.67%, with Precision of 88.01% and Recall of 78.79%. Relative to YOLOv12, these values increased by 5.35, 2.59, and 6.63 percentage points for AP@0.5, Precision, and Recall, respectively. For Panicle, YOLOv12-RSLW achieved 97.40% AP@0.5, with Precision and Recall of 95.52% and 94.64%. The larger improvement observed for Disease indicates that the performance gains of the complete model were concentrated more strongly on the more difficult target class.

Sensitivity of Detection Performance to False Smut Ball

To examine the effect of target size on false smut detection, a size-based sensitivity analysis was conducted on the test set using individual target instances as the evaluation unit. The 566 test images contained 1506 ground-truth Disease boxes. Because the original images differed in resolution, each bounding box was mapped to the 640×640 inference scale using $s = \min(640/W, 640/H)$, where W and H denote the original image width and height. The scaled bounding-box areas were ranked and divided into three equal-sized groups according to the tertiles of the area distribution, with 502 instances per group. The resulting thresholds were ≤ 625.82 pixels² for small targets, 625.82–1295.55 pixels² for medium targets, and >1295.55 pixels² for relatively large targets. These thresholds were dataset-specific and were not based on the COCO object-size definitions. The trained YOLOv12-RSLW model was evaluated separately on each group without retraining or parameter adjustment.

As shown in Table 5, detection performance increased with target size. AP@0.5 was 59.78% for small targets, 85.38% for medium targets, and 88.84% for relatively large targets. The corresponding maximum F1-scores were 63.08%, 81.56%, and 85.34%. Relative to the relatively large group, the small group showed decreases of 29.06 percentage points in AP@0.5 and 22.26 percentage points in maximum F1-score.

Table 5. Detection performance of YOLOv12-RSLW across different false smut ball size groups.

Size Group	Bounding-Box Area (Pixels ²)	n	Median Bbox Size (Pixels)	AP@0.5 (%)	Max F1 (%)
Small	≤ 625.82	502	23.34×20.77	59.78	63.08
Medium	625.82–1295.55	502	37.78×32.46	85.38	81.56
Relatively large	>1295.55	502	55.43×48.06	88.84	85.34

The median bounding-box size of the small group was 23.34×20.77 pixels at the model-input scale. After successive downsampling, targets of this size occupy relatively few spatial positions on the feature maps, which may limit the preservation of local boundary and texture information. This size-dependent loss of spatial detail may partly account for the lower detection performance observed for the small-target group.

3.3. Comparative Analysis of Detection Performance Across Models

YOLOv12-RSLW was compared with YOLOv5, YOLOv6, YOLOv8n, YOLOv10n, YOLO11, and the baseline YOLOv12 under the same evaluation settings. Table 6 summarizes mAP@0.5, Precision, Recall, parameter count, and model size for all models. Among the comparison models, YOLOv12 achieved the highest baseline mAP@0.5 of 88.46%, with Precision of 88.14% and Recall of 82.64%. YOLOv12-RSLW increased these values to 92.06%, 91.46%, and 87.01%, corresponding to gains of 3.60, 3.32, and 4.37 percentage points, respectively. The proposed model also exceeded YOLOv8n by 4.84 percentage points in mAP@0.5 and 8.06 percentage points in Recall.

Table 6. Comparison of different detection models.

Model	mAP@0.5 (%)	Precision (%)	Recall (%)	Params	Model Size (MB)
YOLOv5	85.82	85.45	77.35	2,508,854	7.2
YOLOv6	86.86	86.54	80.05	4,238,342	11.8
YOLOv8n	87.22	86.78	78.95	3,011,238	8.2
YOLOv10n	86.87	86.72	79.89	2,707,820	8.4
YOLO11	86.00	85.97	77.79	2,590,230	6.4
YOLOv12	88.46	88.14	82.64	2,527,166	5.2
YOLOv12-RSLW	92.06	91.46	87.01	3,454,097	7.2

YOLOv12-RSLW contained 3.45 M parameters and had a model size of 7.2 MB, which was smaller than those of YOLOv6, YOLOv8n, and YOLOv10n. Under identical hardware and inference settings, however, its inference speed decreased from 55.18 FPS for YOLOv12 to 37.89 FPS. Thus, the improvement in detection performance was accompanied by a reduction in inference speed, reflecting the computational cost introduced by the modified network components.

3.4. Results of Segmentation and Severity Grading for Rice False Smut

Based on the count-based severity grading rule defined above, grade-wise statistics were performed for the panicle instances in the augmented dataset. As shown in Figure 11, the dataset contained 11,191 panicle instances, including 4883 healthy and 6308 diseased panicles, accounting for 43.63% and 56.37%, respectively. Among the diseased instances, 3359 were assigned to Grade 1 (53.25%), 2648 to Grade 2 (41.98%), and 301 to Grade 3 (4.77%). Grade 1 and Grade 2 were more frequent than Grade 3 within the augmented dataset (Figure 12).

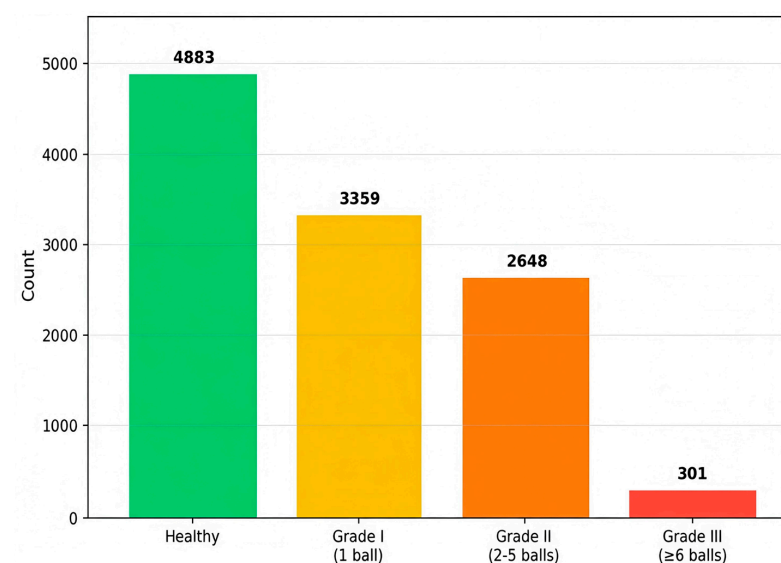


Figure 12. Sample counts across severity grades for rice false smut.

SAM segmentation was quantitatively evaluated using 80 manually annotated images (Table 7). The images were stratified by acquisition year, experimental site, and count-based severity grade, with random sampling performed within each stratum. Sampling quotas for the year–site combinations were assigned approximately in proportion to the number of eligible original images. For Panicle segmentation, Dice, IoU, Precision, and Recall were 90.22%, 82.71%, 92.59%, and 88.94%, respectively. The corresponding values for Disease segmentation were 84.36%, 73.70%, 91.17%, and 79.81%. The higher Precision than Recall for Disease indicates that segmentation errors were more frequently associated with omitted lesion regions than with the inclusion of irrelevant background.

Table 7. Quantitative evaluation of SAM segmentation results.

Segmentation Target	Dice (%)	IoU (%)	Precision (%)	Recall (%)
Panicle	90.22	82.71	92.59	88.94
False smut balls	84.36	73.70	91.17	79.81

Figure 13 further shows the Dice distributions for Panicle and Disease segmentation across the 80 evaluation images. Panicle segmentation had a mean Dice of 0.902 and a relatively narrow interquartile range, whereas Disease segmentation had a lower mean Dice of 0.844 and a broader distribution with several low-value samples. The greater variation in Disease segmentation may be related to the smaller size and less distinct boundaries of false smut lesions, with adhesion and partial occlusion further increasing segmentation difficulty.

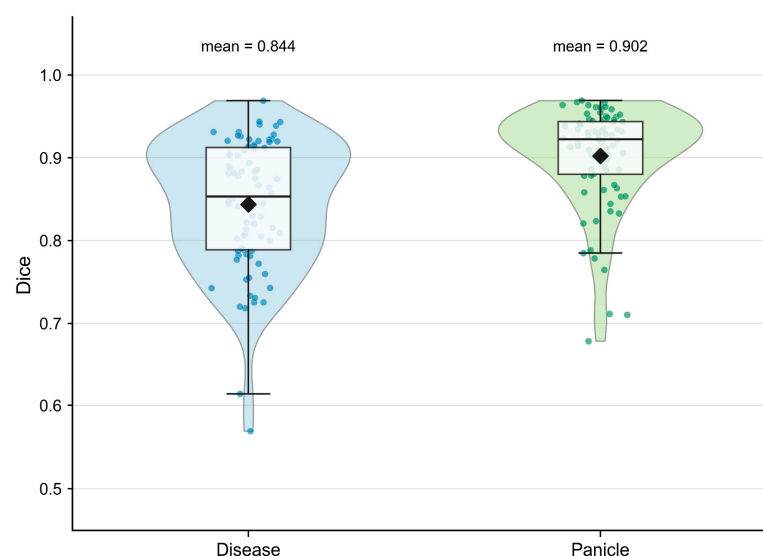


Figure 13. Distribution of Dice values for SAM segmentation of Disease and Panicle targets across the 80-image evaluation set. Black diamonds indicate the mean Dice values.

Table 8 compares the lesion-to-panicle area-ratio estimates obtained under two SAM prompt settings. When bounding boxes derived from manual LabelMe v5.10.1 segmentation annotations were used as SAM prompts, the mean manual and SAM-derived area ratios were 11.31% and 10.74%, respectively. The corresponding MAE and RMSE were 1.88 and 3.03 percentage points, with an R^2 of 0.946 and a Pearson's r of 0.979. When YOLOv12-RSLW predicted boxes were used as SAM prompts, the mean estimated area ratio was 9.55%, while the MAE and RMSE increased to 3.05 and 5.48 percentage points, respectively. The R^2 and Pearson's r were 0.823 and 0.918. The decrease in agreement under predicted-box prompts indicates that detection errors can propagate to subsequent SAM

segmentation and area-ratio estimation. Nevertheless, the end-to-end estimates remained correlated with the manual measurements.

Table 8. Evaluation of Lesion Area Relative to Panicle Area.

Evaluation Setting	SAM Prompt Boxes	Manual Area Ratio, Mean (%)	Automatic Area Ratio, Mean (%)	MAE (%)	RMSE (%)	R2	Pearson r
SAM with manual boxes	Manually annotated boxes	11.31	10.74	1.88	3.03	0.946	0.979
YOLOv12-RSLW + SAM with predicted boxes	YOLOv12-RSLW predicted boxes	11.31	9.55	3.05	5.48	0.823	0.918

As shown in Figure 14, panels (a–c) present the detection results produced by YOLOv12-RSLW, while panels (d–f) show the corresponding SAM segmentation results. Representative examples from Grade 1 and Grade 2 samples indicate that, due to the complex field manifestations of rice false smut, lesion adhesion, merging, and coalesced expansion can cause noticeable errors when severity is assessed solely based on the count of smut balls.

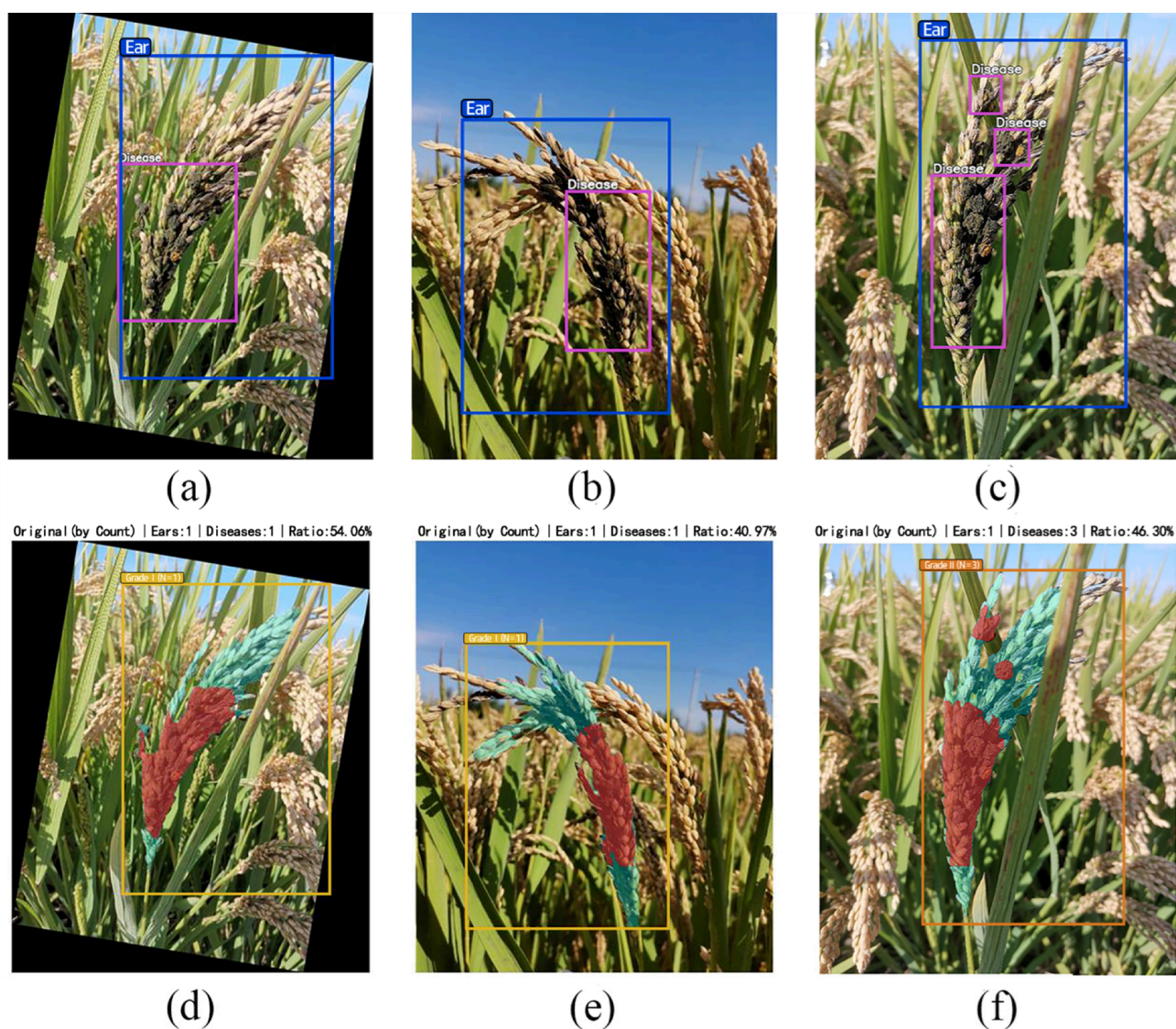


Figure 14. Detection and segmentation results for rice false smut. (a–c) YOLOv12-RSLW detection results. (d–f) Corresponding SAM segmentation results.

Furthermore, we performed a statistical analysis of the lesion area-ratio distributions for samples graded based on counts of false smut balls. To incorporate area-based calibration on top of the count-based grading scheme, we first define a single-panicle lesion area-ratio metric as follows:

$$R = \frac{A_d}{A_p} \times 100\% \quad (15)$$

where the numerator denotes the total pixel area of all lesion masks within a panicle, and the denominator denotes the pixel area of the corresponding panicle mask.

Because the area-ratio distributions overlapped across severity grades, the first quartile (Q1) of the higher-grade distribution was used as the calibration threshold for lower-grade samples. A lower-grade sample reaching or exceeding this threshold was considered to have an area ratio within the distribution range of the corresponding higher count-based grade. Accordingly, the Q1 of Grade 2 (4.57%) was used as the Grade 1 → 2 threshold, while the Q1 of Grade 3 (10.33%) was used as the Grade 2 → 3 threshold.

Figure 15a shows the area-ratio distribution of the 3359 Grade 1 samples. Of these, 1975 (58.8%) had ratios below 4.57%, 954 (28.4%) were between 4.57% and 10.33%, and 430 (12.8%) were at or above 10.33%. Among the 2648 Grade 2 samples shown in Figure 15b, 1905 (71.9%) had ratios below 10.33%, whereas 743 (28.1%) reached or exceeded this threshold. Figure 15c summarizes the resulting CACS assignments. In total, 41.2% of Grade 1 samples and 28.1% of Grade 2 samples met the area-ratio criteria for potential reassignment to higher count-area categories. The overlap in area-ratio distributions shows that samples assigned to the same count-based grade can differ substantially in visible lesion extent.

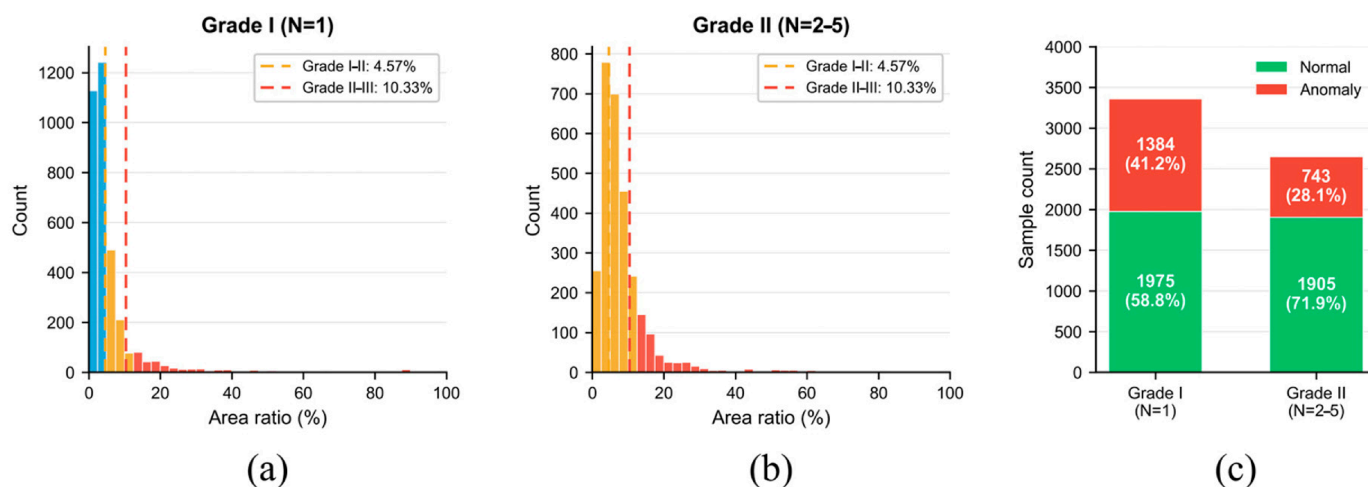


Figure 15. Lesion-to-panicle area-ratio distributions and Count–Area Calibration Strategy (CACS) outcomes. (a) Initially assigned Grade 1 samples ($N = 1$); the vertical dashed lines indicate the 4.57% Grade 1 → 2 threshold and the 10.33% Grade 2 → 3 threshold. (b) Initially assigned Grade 2 samples ($N = 2-5$); the vertical dashed line at 10.33% indicates the Grade 2 → 3 threshold. (c) Numbers and proportions of samples retained at their initial grade or upgraded after CACS.

The CACS calibration outcomes are summarized in Table 9, including the initial count-based grade, lesion-to-panicle area-ratio interval, calibrated grade, sample number, and corresponding proportion.

Figure 16 illustrates the YOLOv12-RSLW detection results, the initial count-based grades, and the grades obtained after applying CACS. The first and fourth columns show the detection results; the second and fifth columns show the initial grades determined by counting false smut balls; and the third and sixth columns show the corresponding CACS-based grades. Samples with area ratios below the relevant threshold retained their

initial grades, whereas those meeting the threshold of a higher grade were assigned the corresponding CACS grade. These examples demonstrate how lesion-area information is incorporated into CACS to supplement count-based grading in samples with lesion adhesion, coalescence, or patch-like expansion.

Table 9. CACS-based severity-grade calibration results.

Initial Grade	Area Ratio	Calibrated Grade	n	Proportion (%)
Grade 1	<4.57%	Grade 1	1975	58.8
Grade 1	4.57–10.33%	Grade 2	954	28.4
Grade 1	≥10.33%	Grade 3	430	12.8
Grade 2	<10.33%	Grade 2	1905	71.9
Grade 2	≥10.33%	Grade 3	743	28.1

Note: R denotes the lesion-to-panicle area ratio, and n denotes the number of panicle samples.



Figure 16. Visualization of detection, segmentation, and CACS-based severity calibration results.

4. Discussion

The results of this study indicate that target size remains an important factor limiting the detection performance of rice false smut. AP@0.5 reached 88.84% for relatively large false smut targets but decreased to 59.78% for the smallest group. At an input resolution of 640×640 pixels, very small targets retain limited spatial information after successive downsampling, which may weaken boundary and texture cues and increase confusion with surrounding panicle structures. Similar challenges associated with small objects, scale variation, and complex agricultural backgrounds have also been reported in previous studies on rice disease recognition and crop disease detection [41–43]. The ablation results further revealed clear differences in the contributions of individual modules. RepGhost produced the largest single-module improvement for the Disease class, whereas LSCD improved detection performance while reducing computational redundancy. WIoU provided a moderate gain without increasing model complexity. SimAM showed a different pattern, producing only a slight change in mAP@0.5 when introduced alone, with no improvement in Recall. Its contribution in the complete model is therefore more reasonably interpreted as complementary to the other components rather than as an independent source of substantial performance gain. Even with the complete YOLOv12-RSLW model, detection performance remained lower for the smallest targets, indicating that fine-scale lesion representation remains an important source of detection error.

Detection confidence is another practical factor affecting the balance between false-positive and false-negative detections in field applications. Under real field conditions, the appropriate confidence threshold depends on the operational objective. In resistance-screening scenarios where missing diseased panicles is considered particularly costly, a lower threshold may be preferred to increase Recall, although this would also increase the number of false-positive detections requiring manual verification. In large-scale surveys where reducing manual workload is a higher priority, a higher threshold may be selected to maintain greater Precision. The appropriate operating point may also shift when the model is deployed under different disease prevalence or background complexity. These trade-offs should therefore be reassessed when the framework is transferred to new sites or imaging conditions.

Segmentation introduces another source of uncertainty. In this study, the Disease masks showed lower Dice, IoU, and Recall values than the Panicle masks. This difference is consistent with the morphological characteristics of the two targets. Panicles generally form larger and more continuous regions, whereas false smut lesions are smaller, have irregular boundaries, and are frequently affected by adhesion and partial occlusion. Similar difficulties in pixel-level lesion delineation have been reported in previous crop disease segmentation studies [44–46]. The area-ratio evaluation also revealed the effect of prompt-box accuracy. With annotation-derived prompts, the area-ratio estimate achieved an R^2 of 0.946 and an MAE of 1.88 percentage points. Using YOLOv12-RSLW predicted boxes reduced R^2 to 0.823 and increased MAE to 3.05 percentage points, indicating measurable error propagation from detection to subsequent segmentation and area-ratio estimation. Pearson's r remained 0.918 under the end-to-end setting, showing that the estimates retained a clear association with the manual measurements. Lesions with severe adhesion or indistinct boundaries remain more susceptible to segmentation error. Lesion-area information becomes particularly relevant when similar numbers of false smut balls correspond to different visible lesion extents. Previous studies have used lesion segmentation and area-related measurements to derive quantitative disease phenotypes from crop images [47–49]. In rice false smut, enlargement, adhesion, or coalescence of smut balls may result in markedly different visible lesion areas even when the detected counts are similar. The lesion-to-panicle area ratio therefore provides information that cannot be

fully captured by count alone. Direct numerical comparison with previous studies is not appropriate because the diseases, target organs, imaging scales, annotation protocols, and evaluation datasets differ substantially.

For adhered or coalesced rice false smut lesions, a widely accepted reference based on the lesion-to-panicle area ratio has not yet been established. CACS was therefore used to supplement the initial count-based grade with information on visible lesion extent. In the present dataset, the thresholds of 4.57% and 10.33% corresponded to the first quartiles of the Grade 2 and Grade 3 area-ratio distributions, respectively. The position of these thresholds directly affects the number of samples assigned to higher CACS grades. Lowering a threshold would make the rule more sensitive to enlarged or coalesced lesions, but overestimation of lesion area caused by segmentation error, imaging geometry, or background interference could also increase unnecessary upward grade reassignment. Raising the threshold would make the rule more conservative, although some samples with relatively extensive visible lesions might remain in their original count-based grade. Because these thresholds were not optimized against an independent severity reference, the present data do not provide an empirical error rate for CACS-based grade reassignment. The thresholds are therefore more appropriately interpreted as distribution-based decision boundaries. Crossing a threshold indicates that the area characteristics of a sample overlap with those of a higher count-defined grade; it does not demonstrate that the original count-based grade was incorrect.

The lesion-to-panicle area ratio also has an inherent physical limitation. It represents the visible two-dimensional projected area of the diseased region in a single image rather than the true diseased surface area of the entire three-dimensional panicle. Rice panicles have a complex spatial structure, and branches, grains, and false smut balls may occlude one another. Lesions on the rear side of a panicle may not be visible, and changes in camera viewpoint can alter the projected areas of both the lesion and the panicle. Previous three-dimensional plant phenotyping studies have similarly shown that viewpoint and self-occlusion can affect image-based structural measurements [50]. Normalizing lesion area by panicle area can reduce the influence of differences in absolute image scale, but it cannot fully eliminate these geometric effects. Accordingly, *R* should be interpreted as the proportion of visible lesion area relative to visible panicle area under the acquired view. Multi-view imaging or three-dimensional reconstruction would be more appropriate when a more complete estimate of total lesion surface area is required.

Generalization across field conditions also requires further consideration. The dataset spans 2023–2025 and includes different experimental sites, rice cultivars, illumination conditions, field backgrounds, and imaging devices. This design broadens the range of field and imaging conditions represented in the current dataset, while also introducing variation in image resolution, color response, shooting distance, illumination, and target scale. Previous field disease studies have likewise reported sensitivity to changes in image acquisition conditions [51,52]. Image standardization and data augmentation may reduce part of the variation across acquisition conditions. However, selective oversampling produced unequal numbers of augmented instances from the 3236 original panicles. The resulting 11,191 instances are therefore image-level computational samples rather than independent biological observations. This sampling structure may have influenced the observed Grade 1–3 proportions and, to some extent, the Q1-based CACS thresholds. The present experimental design also does not demonstrate stable generalization to completely unseen sites, years, cultivars, or devices. Independent evaluation across sites or imaging devices would provide stronger evidence of external robustness. Severity-grade imbalance should also be considered when interpreting CACS. Grade 3 contained only 301 panicles, accounting for 4.77% of the diseased samples. Because severity grade is not a training class of the detector, this imbalance does not

indicate a shortage of a separate Grade 3 detection class. Its more direct effect is on estimation of the Grade 3 area-ratio distribution, which may make the Q1-derived 10.33% threshold more sensitive to changes in sample composition.

The scope of the conclusions should be distinguished from potential applications that remain to be validated. The current experiments directly support rice panicle and false smut detection, SAM-based segmentation, and lesion-to-panicle area-ratio estimation under the conditions represented in the current dataset. The end-to-end evaluation further quantified the effect of detection errors on subsequent segmentation and area-ratio estimation. The CACS analysis further shows how grade assignments change when lesion-area information is incorporated into count-based grading. The present results do not demonstrate that the resulting CACS grades are more accurate than an independent pathological reference. Resistance screening, breeding selection, genomic prediction, and large-scale field deployment were not directly evaluated in this study. Image-derived disease measurements have been used as quantitative phenotypes in previous crop studies [53], providing a relevant precedent for considering such downstream applications. Whether the lesion count, area ratio, or CACS-derived grade developed here can serve as a useful phenotypic variable for rice breeding will require further evaluation in independent germplasm populations and multi-environment trials.

5. Conclusions

This study developed YOLOv12-RSLW for rice false smut detection and quantitative severity assessment under complex field conditions. The detector achieved an mAP@0.5 of 92.06% and an mAP@0.5:0.95 of 68.02%, outperforming the baseline YOLOv12. Coupling YOLOv12-RSLW with SAM enabled pixel-level lesion extraction and lesion-to-panicle area-ratio estimation. End-to-end evaluation using predicted boxes achieved an R^2 of 0.823 and an MAE of 3.05 percentage points, showing that detection errors introduced measurable but limited degradation in subsequent area estimation. The area ratio provided complementary information for count-based severity assessment, particularly for adhered or coalesced lesions. Further validation is required before the CACS thresholds can be generalized beyond the conditions represented in the current dataset.

Author Contributions: X.L.: Conceptualization, methodology, software, formal analysis, investigation, data curation, visualization, writing—original draft. W.Z.: Methodology, software, validation, investigation, writing—original draft. Z.Z.: Validation, visualization. L.Y.: Data curation. H.L.: Writing—review and editing. Y.C.: Conceptualization, resources, writing—review and editing, supervision, project administration, funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Liaoning Provincial Future Industry Frontier Technology Project, “Intelligent Coupling Analysis of Genotype-Phenotype and Digital Breeding Screening Decision-Making Research for Rice Germplasm Resources” (grant number 2025JH2/101330182).

Data Availability Statement: The original dataset and annotation files presented in this study are openly available in the public GitHub repository at <https://github.com/caibird0525/yolov12-dataset> (accessed on 6 July 2026). The repository includes the rice false smut image dataset, annotation files, and related information supporting the findings of this study.

Acknowledgments: The authors declare that no generative AI or AI-assisted technologies were used in the preparation of this manuscript.

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

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Full Term
RFS	Rice False Smut
CNN	Convolutional Neural Network
YOLO	You Only Look Once
SAM	Segment Anything Model
CACS	Count–Area Calibration Strategy
DSI	Disease Severity Index
mAP	Mean Average Precision
AP	Average Precision
WIoU	Wise Intersection over Union
SimAM	Simple, Parameter-Free Attention Module
LSCD	Lightweight Shared Convolutional Detection Head
BN	Batch Normalization
PR	Precision–Recall
GFLOPs	Giga Floating-Point Operations

References

1. Fang, A.; Zhang, R.; Qiao, W.; Peng, T.; Qin, Y.; Wang, J.; Tian, B.; Yu, Y.; Sun, W.; Yang, Y.; et al. Sensitivity baselines, resistance monitoring, and molecular mechanisms of the rice false smut pathogen *Ustilaginoidea virens* to prochloraz and azoxystrobin in four regions of southern China. *J. Fungi* **2023**, *9*, 832. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Sun, W.; Fan, J.; Fang, A.; Li, Y.; Tariqjaveed, M.; Li, D.; Hu, D.; Wang, W.-M. *Ustilaginoidea virens*: Insights into an emerging rice pathogen. *Annu. Rev. Phytopathol.* **2020**, *58*, 363–385. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Khanal, S.; Gaire, S.P.; Zhou, X.-G. Kernel smut and false smut: The old-emerging diseases of rice—A review. *Phytopathology* **2023**, *113*, 931–944. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Qiu, J.; Meng, S.; Deng, Y.; Huang, S.; Kou, Y. *Ustilaginoidea virens*: A fungus infects rice flower and threatens world rice production. *Rice Sci.* **2019**, *26*, 199–206. [\[CrossRef\]](#)
5. Arumugam Gopalakrishnan, M.; Chellappan, G.; Ayyanar, K.; Ramasamy, J.; Santhosh Ganapati, P.; Nagaraini Karupphasamy, S. Integrating spore trapping technology with loop-mediated isothermal amplification assay for surveillance and sustainable management of rice false smut disease. *Front. Microbiol.* **2024**, *15*, 1485275. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Baite, M.S.; Khokhar, M.K.; Meena, R.P. Management of false smut disease of rice: A review. In *Cereal Grains*; IntechOpen: London, UK, 2021. [\[CrossRef\]](#)
7. Meng, J.; Sun, W.; Mao, Z.; Xu, D.; Wang, X.; Lu, S.; Lai, D.; Liu, Y.; Zhou, L.; Zhang, G. Main ustilaginoidins and their distribution in rice false smut balls. *Toxins* **2015**, *7*, 4023–4034. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Wang, W.; Fu, X.; Li, Y.; Jing, M.; Yang, Y.; Xu, D.; Lai, D.; Wang, M.; Wang, B.; Zhou, L. Detection of ustilaginoidins in rice samples by immunoassay based on monoclonal antibodies prepared from hemiustilaginoidin-derived haptens. *Heliyon* **2023**, *9*, e23036. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Huded, S.; Pramesh, D.; Chittaragi, A.; Sridhara, S.; Chidanandappa, E.; Prasannakumar, M.K.; Manjunatha, C.; Patil, B.; Shil, S.; Pushpa, H.D.; et al. Spatial distribution patterns for identifying risk areas associated with false smut disease of rice in southern India. *Agronomy* **2022**, *12*, 2947. [\[CrossRef\]](#)
10. Shi, T.; Liu, Y.; Zheng, X.; Hu, K.; Huang, H.; Liu, H.; Huang, H. Recent advances in plant disease severity assessment using convolutional neural networks. *Sci. Rep.* **2023**, *13*, 2336. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Mahlein, A.-K. Plant disease detection by imaging sensors—Parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis.* **2016**, *100*, 241–251. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Sultan, T.; Chowdhury, M.S.; Jahan, N.; Mridha, M.F.; Alfarhood, S.; Safran, M.; Che, D. LeafDNet: Transforming leaf disease diagnosis through deep transfer learning. *Plant Direct* **2025**, *9*, e70047. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* **2016**, *7*, 1419. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [\[CrossRef\]](#)
15. Barbedo, J.G.A. Factors influencing the use of deep learning for plant disease recognition. *Biosyst. Eng.* **2018**, *172*, 84–91. [\[CrossRef\]](#)

16. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [\[CrossRef\]](#)
17. Kaur, R.; Mittal, U.; Wadhawan, A.; Almogren, A.; Singla, J.; Bharany, S.; Hussien, S.; Rehman, A.U.; Al-Huqail, A.A. YOLO-LeafNet: A robust deep learning framework for multispecies plant disease detection with data augmentation. *Sci. Rep.* **2025**, *15*, 28513. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Abulizi, A.; Ye, J.; Abudukelimu, H.; Guo, W. DM-YOLO: Improved YOLOv9 model for tomato leaf disease detection. *Front. Plant Sci.* **2025**, *15*, 1473928. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Meng, Y.; Zhan, J.; Li, K.; Yan, F.; Zhang, L. A rapid and precise algorithm for maize leaf disease detection based on YOLO MSM. *Sci. Rep.* **2025**, *15*, 6016. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Li, H.; Shi, L.; Fang, S.; Yin, F. Real-time detection of apple leaf diseases in natural scenes based on YOLOv5. *Agriculture* **2023**, *13*, 878. [\[CrossRef\]](#)
21. Song, Z.; Ban, S.; Hu, D.; Xu, M.; Yuan, T.; Zheng, X.; Sun, H.; Zhou, S.; Tian, M.; Li, L. A Lightweight YOLO model for rice panicle detection in fields based on UAV aerial images. *Drones* **2025**, *9*, 1. [\[CrossRef\]](#)
22. Kuznetsova, A.; Maleva, T.; Soloviev, V. YOLOv5 versus YOLOv3 for apple detection. In *Cyber-Physical Systems: Modelling and Intelligent Control*; Kravets, A.G., Bolshakov, A.A., Shcherbakov, M., Eds.; Studies in Systems, Decision and Control; Springer: Cham, Switzerland, 2021; Volume 338, pp. 349–358. [\[CrossRef\]](#)
23. Gai, R.; Chen, N.; Yuan, H. A detection algorithm for cherry fruits based on the improved YOLO-v4 model. *Neural Comput. Appl.* **2023**, *35*, 13895–13906. [\[CrossRef\]](#)
24. Li, G.; Huang, X.; Ai, J.; Yi, Z.; Xie, W. Lemon-YOLO: An efficient object detection method for lemons in the natural environment. *IET Image Process.* **2021**, *15*, 1998–2009. [\[CrossRef\]](#)
25. Leite, D.; Brito, A.; Faccioli, G. Advancements and outlooks in utilizing convolutional neural networks for plant disease severity assessment: A comprehensive review. *Smart Agric. Technol.* **2024**, *9*, 100573. [\[CrossRef\]](#)
26. Elfouly, M.K.; AbdelAziz, A.M.; Gomaa, W.H.; Abdalla, M. A deep learning-based framework for large-scale plant disease detection using big data analytics in precision agriculture. *J. Big Data* **2025**, *12*, 205. [\[CrossRef\]](#)
27. Chiang, K.S.; Liu, H.I.; Bock, C.H. A discussion on disease severity index values. Part I: Warning on inherent errors and suggestions to maximise accuracy. *Ann. Appl. Biol.* **2017**, *171*, 139–154. [\[CrossRef\]](#)
28. IRRI. *Standard Evaluation System for Rice (SES)*, 5th ed.; International Rice Research Institute: Los Baños, Philippines, 2013.
29. Yang, L.; Guo, F.; Zhang, H.; Cao, Y.; Feng, S. Research on lightweight rice false smut disease identification method based on improved YOLOv8n model. *Agronomy* **2024**, *14*, 1934. [\[CrossRef\]](#)
30. Zhang, Z.; Chen, X.; Shen, H.; Huang, S. Research and Application Status of Severity Grading Standards of the Rice False Smut Disease in China. *China Rice* **2020**, *26*, 18–21.
31. Deng, G.; Liu, Z.; Yang, Z. Study on the Grading Standard of Rice False Smut. *J. Agric. Sci.* **1989**, *35*, 23–25.
32. Tang, C.S.; Gao, J.Z.; Cao, G.P.; Huang, S.X.; Liu, E.M.; Liu, C.L. Studies on the Grading Standard of Rice False Smut Disease. *Plant Prot.* **2001**, *27*, 18–21.
33. NY/T 1464.54-2014; Pesticide Guidelines for the Field Efficacy Trials—Part 54: Fungicides against False Smut of Rice. Ministry of Agriculture of the People's Republic of China: Beijing, China, 2014.
34. DB43/T 504-2009; Specification for Survey and Forecast of Rice False Smut. Hunan Provincial Bureau of Quality and Technical Supervision: Changsha, China, 2009.
35. DB43/T 505-2009; Technical Specification for Identification and Evaluation of Rice Variety Resistance to Rice False Smut. Hunan Provincial Bureau of Quality and Technical Supervision: Changsha, China, 2009.
36. Tian, Y.; Ye, Q.; Doermann, D. YOLOv12: Attention-centric real-time object detectors. *arXiv* **2025**, arXiv:2502.12524. [\[CrossRef\]](#)
37. Chen, C.; Guo, Z.; Zeng, H.; Xiong, P.; Dong, J. RepGhost: A hardware-efficient Ghost module via re-parameterization. *arXiv* **2022**, arXiv:2211.06088. [\[CrossRef\]](#)
38. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. SimAM: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the 38th International Conference on Machine Learning (ICML), Online, 18–24 July 2021; pp. 11863–11874. Available online: <https://proceedings.mlr.press/v139/yang21o.html> (accessed on 6 July 2026).
39. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism. *arXiv* **2023**, arXiv:2301.10051. [\[CrossRef\]](#)
40. Zeng, J.; Zhong, H. YOLOv8-PD: An improved road damage detection algorithm based on YOLOv8n model. *Sci. Rep.* **2024**, *14*, 12052. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Deng, R.; Tao, M.; Xing, H.; Yang, X.; Liu, C.; Liao, K.; Qi, L. Automatic Diagnosis of Rice Diseases Using Deep Learning. *Front. Plant Sci.* **2021**, *12*, 701038. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Li, J.; Qiao, Y.; Liu, S.; Zhang, J.; Yang, Z.; Wang, M. An Improved YOLOv5-Based Vegetable Disease Detection Method. *Comput. Electron. Agric.* **2022**, *202*, 107345. [\[CrossRef\]](#)

43. Zhang, D.-Y.; Luo, H.-S.; Cheng, T.; Li, W.-F.; Zhou, X.-G.; Guo, W.; Gu, C.-Y.; Diao, Z. Enhancing Wheat Fusarium Head Blight Detection Using Rotation YOLO Wheat Detection Network and Simple Spatial Attention Network. *Comput. Electron. Agric.* **2023**, *211*, 107968. [\[CrossRef\]](#)
44. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.-Y.; et al. Segment Anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023; pp. 4015–4026.
45. Feng, C.; Jiang, M.; Huang, Q.; Zeng, L.; Zhang, C.; Fan, Y. A Lightweight Real-Time Rice Blast Disease Segmentation Method Based on DFFANet. *Agriculture* **2022**, *12*, 1543. [\[CrossRef\]](#)
46. Gao, J.; Westergaard, J.C.; Sundmark, E.H.R.; Bagge, M.; Liljeroth, E.; Alexandersson, E. Automatic Late Blight Lesion Recognition and Severity Quantification Based on Field Imagery of Diverse Potato Genotypes by Deep Learning. *Knowl.-Based Syst.* **2021**, *214*, 106723. [\[CrossRef\]](#)
47. Divyanth, L.G.; Ahmad, A.; Saraswat, D. A Two-Stage Deep-Learning Based Segmentation Model for Crop Disease Quantification Based on Corn Field Imagery. *Smart Agric. Technol.* **2023**, *3*, 100108. [\[CrossRef\]](#)
48. Patil, R.R.; Kumar, S.; Chiwhane, S.; Rani, R.; Pippal, S.K. An Artificial-Intelligence-Based Novel Rice Grade Model for Severity Estimation of Rice Diseases. *Agriculture* **2023**, *13*, 47. [\[CrossRef\]](#)
49. Lee, D.-Y.; Na, D.-Y.; Heo, Y.S.; Wang, G.-L. Digital Image Quantification of Rice Sheath Blight: Optimized Segmentation and Automatic Classification. *Agriculture* **2025**, *15*, 2478. [\[CrossRef\]](#)
50. Gibbs, J.A.; Pound, M.; French, A.P.; Wells, D.M.; Murchie, E.; Pridmore, T. Plant Phenotyping: An Active Vision Cell for Three-Dimensional Plant Shoot Reconstruction. *Plant Physiol.* **2018**, *178*, 524–534. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Gui, P.; Dang, W.; Zhu, F.; Zhao, Q. Towards Automatic Field Plant Disease Recognition. *Comput. Electron. Agric.* **2021**, *191*, 106523. [\[CrossRef\]](#)
52. Clohessy, J.W.; Sanjel, S.; O'Brien, G.K.; Barocco, R.; Kumar, S.; Adkins, S.; Tillman, B.; Wright, D.L.; Small, I.M. Development of a High-Throughput Plant Disease Symptom Severity Assessment Tool Using Machine Learning Image Analysis and Integrated Geolocation. *Comput. Electron. Agric.* **2021**, *184*, 106089. [\[CrossRef\]](#)
53. Stewart, E.L.; Wiesner-Hanks, T.; Kaczmar, N.; DeChant, C.; Wu, H.; Lipson, H.; Nelson, R.J.; Gore, M.A. Quantitative Phenotyping of Northern Leaf Blight in UAV Images Using Deep Learning. *Remote Sens.* **2019**, *11*, 2209. [\[CrossRef\]](#)

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.