MSA-YOLO: A Lightweight Detection Model for Wheat Spikelet Fusarium Head Blight Based on YOLO11

Lei Shi
Henan Agricultural University
Zilong Shang
Henan Agricultural University
Bing Bai
China University of Geosciences (Wuhan)
Yingyu Ma
Henan Agricultural University
Fei Yin
Henan Agricultural University
Wei Guo
Henan Agricultural University
Yumin Chen
Xiamen University of Technology
Jingkai Lei
Chengdu Leyou Technology Co., Ltd
Shufeng Xiong
Henan Agricultural University
Yong Li
liyong@henau.edu.cn

Henan Agricultural University

Research Article

Keywords: wheat Fusarium head blight, YOLO11, lightweight design, adaptive threshold focal loss, MobileOne

Posted Date: March 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9198074/v1

Additional Declarations: No competing interests reported.

MSA-YOLO: A Lightweight Detection Model for Wheat Spikelet Fusarium Head Blight Based on YOLO11

Lei Shi[1†], Zilong Shang[1†], Bing Bai[2], Yingyu Ma[1], Fei Yin[1], Wei Guo[1], Yumin Chen[3], Jingkai Lei[4], Shufeng Xiong[1], Yong Li[1*]

[1*]Henan Agricultural University, Zhengzhou, 450046, China.
[2]School of Computer Science, China University of Geosciences (Wuhan), Wuhan, 430078, China.
[3]College of Computer and Information Engineering, Xiamen University of Technology, Xiamen, 361024, China.
[4]Chengdu Leyou Technology Co., Ltd, Chengdu, 610041, China.

*Corresponding author(s). E-mail(s): liyong@henau.edu.cn;
†These authors contributed equally to this work.

Abstract

Fusarium Head Blight (FHB) is one of the most destructive fungal diseases in global wheat production. Traditional methods for FHB detection face limitations such as high technical expertise requirements, limited coverage scope, and insufficient timeliness, making them inadequate for modern precision agriculture management demands. To address this challenge, this study proposes a lightweight MSA-YOLO detection model based on the YOLO11 deep learning framework. The proposed model achieves a favorable balance between performance and efficiency through three innovative design aspects: first, it replaces the original backbone network with the MobileOne network, establishing a foundation for model lightweight design; second, it substitutes the multi-head attention mechanism in the C2PSA module's PSABlock with a more computationally efficient SE module, further reducing model complexity while maintaining detection performance; finally, it introduces an Adaptive Threshold Focal Loss (ATFL) function to address class imbalance issues, enhancing the model's recognition capability for minority classes. The experimental data comprise 629 photographs of wheat spikelets covering various growth and development stages. Results demonstrate that the improved MSA-YOLO model reduces parameter count from 2.58M to 1.65M and computational complexity from 6.4 GFLOPs to

3.9 GFLOPs. Furthermore, comparative analysis with YOLOv10, YOLOv9, YOLOv8, and YOLOv5 models shows that MSA-YOLO exhibits an exceptional balance between speed and accuracy, making it well suited for practical applications in precision agriculture monitoring systems.

Keywords: wheat Fusarium head blight, YOLO11, lightweight design, adaptive threshold focal loss, MobileOne

1 Introduction

Wheat is one of China's most important economic and food crops, with critical significance for national development across multiple dimensions. As a primary staple food for the Chinese population, wheat plays a vital role in ensuring national food security. Simultaneously, as a fundamental raw material for the food industry, wheat generates significant economic benefits for the country. Therefore, ensuring the healthy growth of wheat crops is crucial, not only for meeting the basic living needs of the population but also for the country's economic development and social stability.

However, Fusarium Head Blight (FHB), caused by the Fusarium graminearum species complex, has emerged as a serious threat to global wheat production, with particularly severe impacts in China. This disease not only causes wheat spikes to yellow and wither, resulting in 10%-20% yield losses, but also secretes highly toxic substances such as deoxynivalenol (DON), posing a dual threat to food safety and livestock farming Alisaac and Mahlein (2023); Moonjely et al. (2023); Al-Hashimi et al. (2025); Hamila et al. (2024). In recent years, with the intensification of global climate change, shifts in cultivation patterns, and continuous pathogen mutations, the frequency, transmission speed, and control difficulty of FHB in China have increased substantially, creating an increasingly severe situation. Given the sudden onset and rapid spread characteristics of FHB, traditional monitoring methods such as manual field surveys and laboratory analyses are no longer adequate to meet current needs, exhibiting significant limitations including high technical thresholds, poor timeliness, and limited coverage Liu et al. (2025); Morellos et al. (2025); Li et al. (2025). Therefore, timely and accurate assessment of wheat FHB incidence is of significant importance, enabling agricultural workers to formulate differentiated control measures that can effectively protect wheat yields and ensure wheat quality, thereby safeguarding national food security and sustainable agricultural development.

In recent years, plant disease phenotyping has rapidly evolved from traditional handcrafted-feature pipelines to deep learning and computer-vision-based frameworks. Recent reviews have shown that current research has expanded from simple leaf classification to object detection, severity estimation, segmentation, and field-deployable precision agriculture systems, while lightweight deployment, domain generalization, and limited annotated datasets remain major challenges Pacal et al. (2024); Upadhyay et al. (2025); Shoaib et al. (2025); Feng et al. (2024). Earlier studies based on feature engineering and shallow classifiers provided important foundations for crop disease recognition Kumar and Kukreja (2024, 2025); Mustafa et al. (2024); Liang et al. (2025), but their reliance on manual feature extraction limits robustness and efficiency under complex field conditions.

In recent years, deep learning technologies, especially Convolutional Neural Networks (CNNs), have achieved remarkable results in the field of image recognition. CNNs learn features in a manner similar to human vision, capturing various features from basic to complex. Just as traditional machine learning requires manually designed features, low-level features can be learned by CNNs, but CNNs are more powerful in that they can also learn more complex high-level features. Moreover, CNNs have a unique capability to transform image features through layer-by-layer feature transformations, converting the original feature representation into a representation in a completely new feature space. This transformation makes deep learning methods more accurate than traditional machine learning methods in wheat spike disease recognition. Currently, deep learning is widely applied in crop disease classification, severity recognition, and lesion segmentation Upadhyay et al. (2025); Shoaib et al. (2025). An improved U-Net-based model, Octave-UNet, has been proposed to segment lesions on wheat leaves damaged by stripe rust, achieving a segmentation accuracy of 96.06% and outperforming other semantic segmentation models in testing, thereby providing an effective solution for identifying wheat stripe rust lesions Li et al. (2022). An integrated deep learning algorithm combining Residual Channel Attention Blocks (RCAB), Feedback Blocks (FB), Elliptical Metric Learning (EML), and Convolutional Neural Networks (CNN) has also been proposed for wheat leaf disease detection, achieving an accuracy of 99.95% on public datasets with reduced time consumption, high recognition accuracy, and strong adaptability Xu et al. (2023). More recently, an improved lightweight deep learning model has been developed for identifying common wheat diseases in complex real-field environments, showing that attention-enhanced lightweight architectures can improve practical recognition performance under realistic field backgrounds Wen et al. (2024).

Furthermore, recent studies on wheat Fusarium Head Blight (FHB) have increasingly emphasized field-oriented, lightweight, and deployable frameworks. Technical reviews have summarized the rapid progress of multi-scale imaging and automated FHB assessment, while specialized studies have explored high-throughput spike detection and refined lesion segmentation, lightweight severity recognition through compression and distillation, UAV-based remote sensing, and lightweight YOLO-based mobile deployment Feng et al. (2024); Zhou et al. (2024); Gong et al. (2024); Bao et al. (2024); Yang et al. (2024). More recently, real-time and edge-oriented frameworks such as SCS-YOLO and EBS-YOLO further demonstrated that accurate in-field wheat FHB detection can be achieved with reduced computational complexity and stronger deployment feasibility Gao et al. (2025); Mao et al. (2025). These advances indicate that lightweight yet accurate detection models are increasingly required for practical field monitoring, epidemic assessment, and decision support in precision agriculture. In summary, the current application of deep learning technology provides a strong foundation for wheat FHB severity recognition, and this research therefore focuses on individual wheat spikes and adopts YOLO11n as the baseline model for lightweight improvement research.

2 Wheat Spikelet Fusarium Head Blight Image Dataset

2.1 Data Source

This study utilized a wheat Fusarium Head Blight image dataset collected from the National Wheat Technology Engineering Research Base of Henan Agricultural University (113°40'E, 35°1'N), located in Yuanyang County, Henan Province. Yuanyang County has a temperate continental monsoon climate with distinct seasons and concentrated rainfall. The experimental field was divided into multiple small planting areas, with wheat experimental varieties planted using autumn sowing methods, arranged in two rows per variety and randomly distributed. The experimental varieties included Ningmai 9, Yangmai 158, Sumai 3, and Zhoumai 18, with Fusarium graminearum as the pathogen.

To obtain wheat Fusarium Head Blight disease data at different infection stages, data collection was conducted from May 2 to May 15, 2022, and from May 11 to May 24, 2023. During the collection process, smartphone cameras were used to capture the lateral sides of wheat spikes at a 90°vertical angle, ensuring that each image contained only one wheat spike. The images were saved in JPG format, with pixel dimensions primarily of 1200 × 1600 pixels, while some images were 1920 × 1920 pixels. After screening the original images and eliminating redundant and blurry pictures, a total of 629 effective wheat spike images were ultimately obtained. Examples of some dataset samples are shown in Fig. 1.

Fig. 1: Examples of original wheat spikelet images

2.2 Data Processing

To enhance the model's ability to learn image features from the dataset and to account for real wheat field environments, we performed appropriate data augmentation operations on the collected original images, including random rotation and random adjustment of image brightness to simulate actual field shooting conditions. The data augmentation effects are shown in Fig. 2. We defined healthy spikelets in wheat spikes as HS and diseased spikelets as DS. After data augmentation, the original dataset was doubled to obtain a total of 1,258 wheat Fusarium Head Blight images. The dataset was divided in an 8:1:1 ratio, resulting in 1,008 images for the training set, 125 images for the test set, and 125 images for the validation set. The training set was used to train the object detection model, the validation set was used to evaluate model performance, and the test set was used to test the model's actual detection effectiveness for wheat Fusarium Head Blight.

Table 1: Distribution of labels by category in the dataset

Dataset	No. of Images	Total No. of Spikelets	No. of Healthy Spikelets	No. of Diseased Spikelets
Training Set	1008	21324	17025	4299
Validation Set	125	2647	2074	573
Test Set	125	2639	2129	510

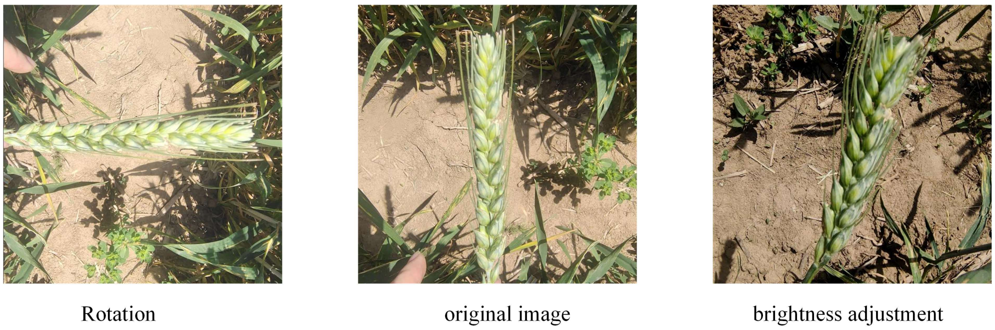


Fig. 2: Examples of data augmentation

3 Object Detection Model

3.1 Introduction to the YOLO11 Model

The YOLO11 Ultralytics (2024) model, introduced by Ultralytics, represents the latest iteration in the YOLO series and maintains the classic architecture consisting of four components: Input layer, Backbone network, Neck network, and Head network. The Input layer feeds image data into the model; the Backbone network extracts features from input images; the Neck network fuses features extracted by the backbone; and the Head network performs predictions on the fused features, outputting location and category information of targets. Compared to YOLOv8 Jocher et al. (2023), YOLO11 achieves a significant improvement in mean Average Precision (mAP) while reducing parameter count by 22%. This improvement enables YOLO11 to maintain high accuracy while substantially enhancing computational efficiency, making it particularly suitable for deployment on resource-constrained devices. As shown in Fig. 3.

The YOLO11 series includes five versions based on model scale: YOLO11n, YOLO11s, YOLO11m, YOLO11l, and YOLO11x. As model depth and width increase sequentially, detection accuracy improves, though training time extends accordingly. Considering practical agricultural production requirements, this study selected YOLO11n as the baseline model. This model achieves a good balance between detection accuracy and recognition speed, meeting the requirements for real-time detection of wheat Fusarium Head Blight.

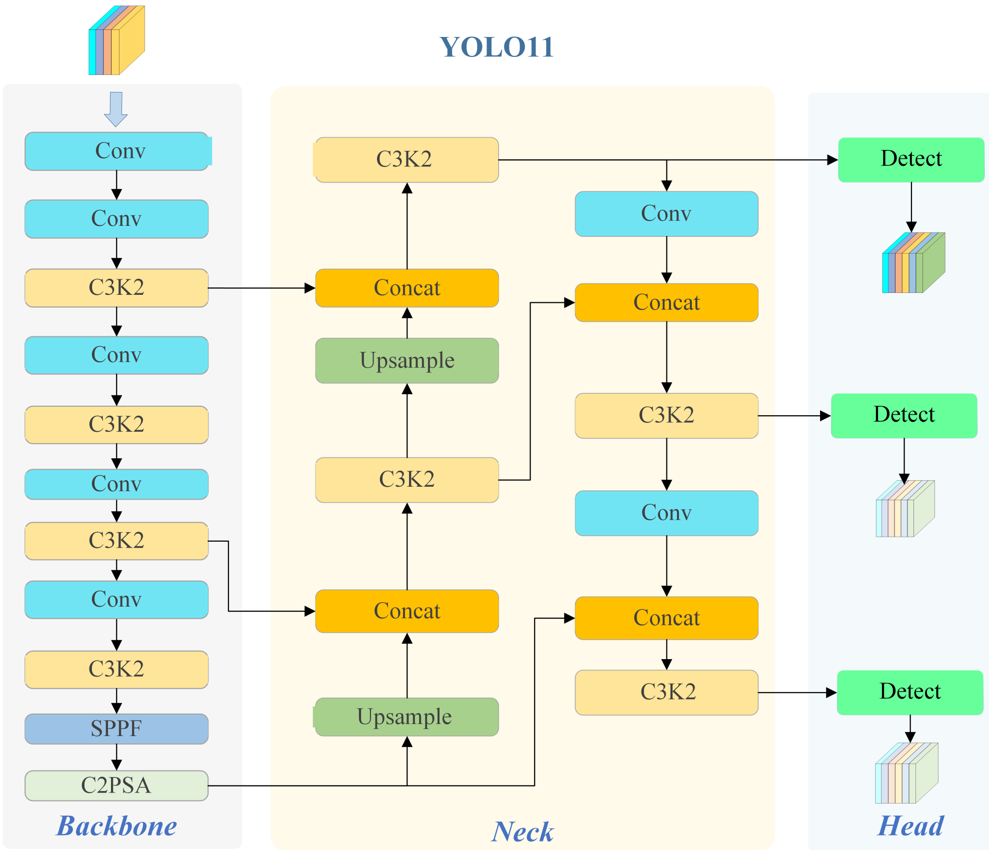


Fig. 3: Architecture of YOLO11

3.2 Introduction to the Improved YOLO Model

To improve the adaptability of YOLO11n model for Fusarium head blight detection in complex field environments and enable efficient deployment on edge devices, this study enhanced the model in three aspects: optimizing model structure, reducing computational complexity, and enhancing adaptability to field environments. First, in the Backbone part, we replaced the original network with MobileOne Yu et al. (2023), effectively reducing parameters and computational cost while maintaining inference speed. Second, in the C2PSA module, we substituted the multi-head attention mechanism of PSABlock with an SE Hu et al. (2020) module, achieving further lightweight effects while maintaining or even improving detection performance. Finally, we replaced the binary cross-entropy loss function (BCE With Logits Loss) with Adaptive Threshold Focal Loss (ATFL) Lin et al. (2020), further improving detection accuracy for healthy and diseased wheat spikelets. These improvements enable YOLO11 model to demonstrate stronger adaptability and higher efficiency in Fusarium head blight detection in complex field environments while meeting deployment requirements for edge devices.

3.2.1 Backbone network improvement

MobileOne is a lightweight convolutional neural network optimized for mobile devices, designed to achieve millisecond-level inference speed while maintaining high model accuracy. This model inherits the efficient architectural design of MobileNetV1 Howard et al. (2017) and integrates the reparameterization concept from RepVGG Ding et al. (2022), innovatively proposing the "complex training, simple inference" design philosophy. MobileOne employs a multi-branch structure during the training phase to enhance feature representation capabilities. During inference, these branches are merged into an equivalent single convolutional layer through reparameterization techniques, significantly improving inference efficiency.

As illustrated in the figure, the left side of Fig. 4 shows MobileOne's training structure. MobileOne's backbone network is designed based on MobileNetV1's 3×3 convolution and 1×1 pointwise convolution, primarily consisting of two basic units: depthwise convolution modules and pointwise convolution modules. The left side features a 1×1 convolution scaling branch, the middle contains k parallel over-parameterized 3×3 depthwise convolution branches (each comprising a convolution layer and batch normalization layer), and the right side is a skip connection branch with a batch normalization layer. The pointwise convolution module consists of k parallel over-parameterized 1×1 convolution branches (each similarly containing convolution and batch normalization layers) and a skip connection branch with a batch normalization layer. In this study, k is set to 4. The right side of Fig. 4 displays MobileOne's inference structure, where branches from the training phase are removed through reparameterization, with convolution and batch normalization operations folded into a single convolutional layer. Specifically, for convolution layers, weights W and biases b are calculated by summing the corresponding parameters from various branches. For batch normalization in skip connections, they are folded into a convolution layer with a 1×1 identity kernel and implemented by padding K-1 zeros. At this point, the model possesses a simple feedforward structure without any branches or skip connections, thereby reducing memory access costs. This research adopts MobileOne as the backbone network, ensuring the model can thoroughly learn features of both healthy and diseased wheat spike spikelets while achieving faster inference speed.

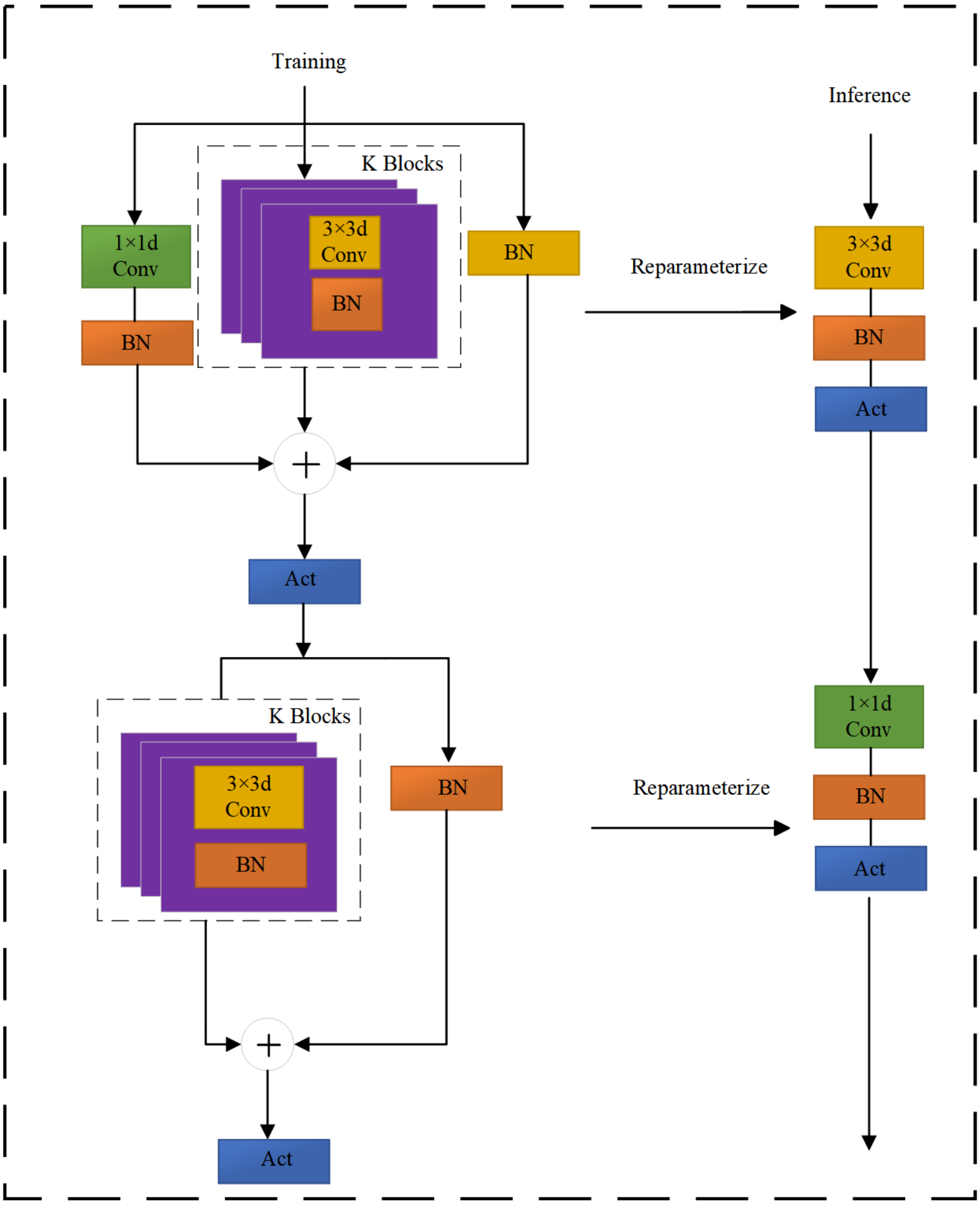


Fig. 4: Architecture of the MobileOne network

It is well-known that the quality of activation functions can determine the performance of neural network models. While reviewing literature on model lightweighting, we discovered that the PP-LCNet Cui et al. (2021) network model proposed by Baidu replaced the commonly used ReLU activation function with the Hard-swish activation function, which improved detection performance without increasing inference time. Considering that model lightweighting efforts often lead to accuracy degradation issues, this study replaced MobileOne's original ReLU Nair and Hinton (2010) activation function with the Hard-swish activation function to ensure that the model's detection performance for wheat spikelet Fusarium head blight would not significantly decrease. The Hard-swish activation function was introduced in the MobileNetV3 Howard et al. (2019) model, modified from the Swish activation function. Although the Swish activation function features no upper bound, a lower bound, smoothness, and non-monotonicity, outperforming ReLU in deeper models, the sigmoid(x) function involves complex calculations with numerous exponential computations. In the MobileNetV3 model, an approximate function was adopted to approach the Swish function, thereby avoiding extensive exponential operations and achieving better results in lightweight models. This improvement is particularly important for diseases like wheat spikelet Fusarium head blight that require high-precision detection, as the disease features are often subtle and

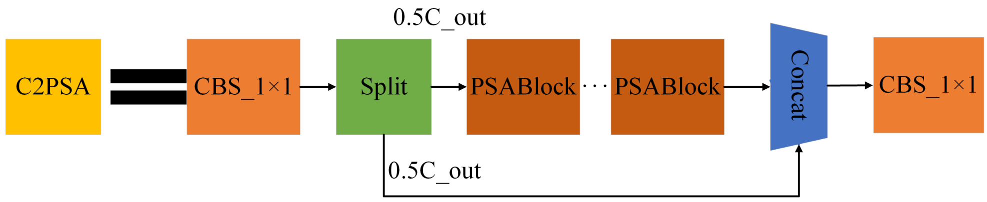
$$\mathrm{Swish}(x) = x \cdot \sigma(x) \tag{1}$$

$$\mathrm{sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

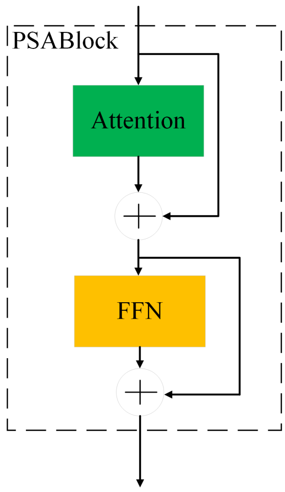
$$\text{Hard-swish}(x) = \begin{cases} 0 & \text{if } x \leq -3 \\ \frac{x(x+3)}{6} & \text{if } -3 < x < 3 \\ x & \text{if } x \geq 3 \end{cases} \tag{3}$$

3.2.2 Improvement of the C2PSA module

In the YOLO11 model, a C2PSA module was proposed based on the C2f foundation, implementing a convolutional block with attention mechanism to enhance feature extraction and processing of images. The structural flow of C2PSA is shown in Fig. 5, where input features first pass through a convolutional module, then the resulting feature map is divided into two parts: one retaining the original features, and the other processing features through multiple PSA modules. Finally, the features from both parts are concatenated. This approach enhances feature processing and extraction capabilities. The PSA module, serving as the core component of the C2PSA module, further augments feature processing capabilities. Its detailed structure is illustrated in the figure below. The PSA module aims to enhance feature extraction capabilities through multi-head attention mechanisms and feed-forward neural networks. It can selectively incorporate residual structures (shortcut) to optimize gradient propagation and network training effectiveness. Additionally, using FFN allows input features to be mapped to higher-dimensional spaces, capturing complex non-linear relationships in input features and enabling the model to learn richer feature representations. While this design can improve the model's feature extraction and processing capabilities, it also increases computational complexity.



(a)



(b)

Fig. 5: Core architecture of the C2PSA module

To this end, this paper intends to replace the multi-head attention mechanism with the SE (Squeeze-and-Excitation) module from SENet (Squeeze-and-Excitation Networks). The multi-head attention mechanism obtains the query, key, and value matrices by applying linear transformations to the input sequence, then divides these matrices into multiple heads to perform parallel scaled dot-product attention operations. The outputs of all heads are then concatenated and passed through a linear transformation to obtain the final output. In contrast, the core idea of SENet is to enhance the modeling of inter-channel dependencies through the Squeeze-and-Excitation (SE) module. As shown in Fig. 6, the SE module adaptively recalibrates channel feature responses by explicitly modeling the interdependencies between channels, thereby significantly improving the network's feature representation capability.

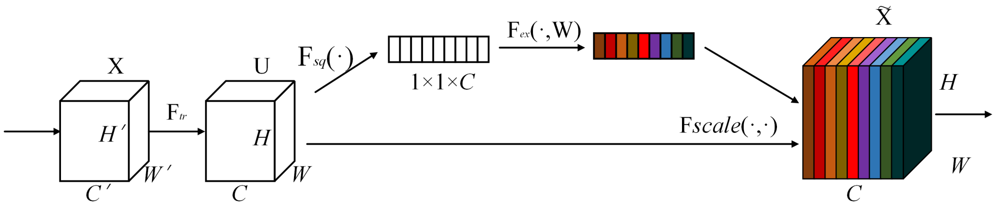


Fig. 6: Architecture of the SE module

The input feature map X is first transformed by F_{tr} to produce the feature map U. Then, U is compressed into a global channel descriptor through global average pooling F_{sq}. This descriptor is passed through two fully connected layers F_{ex} (for dimensionality reduction and expansion), and activated by the ReLU and Sigmoid functions. Finally, the original feature map U is multiplied by the learned channel weights F_{scale} to obtain the recalibrated feature map $\tilde{X}$. This structure helps the network learn inter-channel dependencies and adaptively strengthen or suppress certain feature channels.

3.2.3 Improvement of the loss function

For a single spikelet, there is often an imbalance between the number of Fusarium-infected spikelets and healthy spikelets, and this class imbalance can significantly reduce the detection performance of the model. The original YOLO11 model employs BCEWithLogitsLoss, which does not adequately address the issue of class imbalance. To solve this problem, we replace the original binary cross-entropy loss function (BCEWithLogitsLoss) with the Adaptive Threshold Focal Loss (ATFL). ATFL inherits the core advantages of Focal Loss by reducing the weight of easily classified samples and increasing the weight of hard-to-classify samples, thereby enhancing the model's ability to learn features of Fusarium head blight. In addition, ATFL can avoid manual hyperparameter tuning through its adaptive mechanism, saving time, and can automatically adjust hyperparameters according to the training data, thus improving the model's generalization ability and its capacity to learn from difficult samples.

$$\mathrm{ATFL} = \begin{cases} -(\lambda - p_t)^{-\ln(p_t)} \log(p_t), & p_t \leq 0.5 \\ -(1 - p_t)^{-\ln(p_t)} \log(p_t), & p_t > 0.5 \end{cases} \tag{4}$$

Here, λ is a tunable parameter, and p_t denotes the predicted probability of the target class.

3.3 Experimental configuration

All experiments in this study were conducted using Python version 3.10 as the programming language. PyTorch version 2.1.2 was selected as the deep learning framework, with CUDA version 11.8. The GPU used was an NVIDIA GeForce RTX 3080. The remaining network model hyperparameter settings are shown in Table 2.

Table 2: Network model hyperparameter settings

Parameter	Value
Image input size (pixel)	640×640
Initial learning rate	0.01
Optimizer	Adam
Momentum	0.937
Weight decay	0.0005
Batch size	8
Epoch	500

3.4 Evaluation Metrics

In order to evaluate the detection results of the network for Fusarium head blight on single wheat spikelets, we selected Precision, Recall, Average Precision (AP), and mean Average Precision (mAP) as evaluation metrics. To assess the lightweight performance of the improved model, we used the number of parameters, computational cost (FLOPs), and inference speed (FPS) as lightweight evaluation metrics, where FPS is calculated on the CPU. The calculation formulas are as follows:

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

Taking the Fusarium-infected spikelet (DS) category as an example, in the above formulas, TP denotes the number of samples that are actually of the DS category and are also detected as DS, FN denotes the number of samples that are actually of the DS category but are detected as other categories, and FP denotes the number of samples that are actually not of the DS category but are detected as DS.

$$AP = \int_0^1 P(R)\, dR \tag{7}$$

$$mAP = \frac{1}{n} \sum_{i=1}^{n} AP_i \tag{8}$$

where $P(R)$ is the precision-recall (PR) curve, AP is the average precision over all points on the PR curve, and mAP is the mean of AP values across different categories.

4 Results and Analysis

4.1 The influence of the Hard-swish function on the model

In order to achieve a lightweight model, the MobileOne network was adopted and improved in the Backbone part. Subsequently, we attempted to replace the ReLU activation function in the MobileOne network with the Hard-swish activation function. To investigate the impact of the Hard-swish activation function on model detection performance, we conducted ablation experiments, and the specific results are shown in Table 3.

Table 3: Analysis of results for different activation functions

Activation	Precision (%)	Recall (%)	mAP (%)	FPS	Params (M)	FLOPs (G)
ReLU	92.7	93.4	95.9	27.62	1.7	4.0
Hard-swish	92.0	93.8	96.2	25.97	1.7	4.0

According to the results in the table, it can be observed that, considering the Hard-swish activation function is computationally more complex than the ReLU activation function, the FPS decreases slightly. However, the Hard-swish activation function possesses better nonlinear expressive capabilities. Although the precision decreases slightly from 92.7 to 92.0, both mAP and recall show improvements. In terms of model speed and detection performance, Hard-swish can provide better detection performance than ReLU while maintaining the same model complexity. Although it introduces a slight computational overhead, this trade-off is worthwhile in scenarios such as wheat Fusarium head blight detection, where high-precision detection is required.

4.2 Ablation Experiment Result Analysis

Based on the replacement of the activation function with Hard-swish, a series of ablation experiments were conducted to verify the effectiveness of other improvements proposed in this paper. The results of the ablation experiments are shown in Table 4.

Table 4: Ablation experiment results

MobileOne	SE	ATFL	Precision (%)	Recall (%)	mAP (%)	FPS	Params (M)	FLOPs (G)
×	×	×	92.1	93.0	96.1	22.47	2.58	6.40
✓	×	×	92.0	93.8	96.2	25.97	1.70	4.00
✓	✓	×	92.6	92.8	96.3	25.64	1.65	3.90
✓	✓	✓	93.2	93.6	96.7	27.93	1.65	3.90

As can be seen from the data in the table, compared with the baseline YOLO11n model, the MobileOne backbone with the Hard-swish activation function increases the recall by 0.8 percentage points, slightly improves the mAP by 0.1 percentage points, and only reduces the precision by 0.1 percentage points. Meanwhile, the inference speed increases from 22.47 FPS to 25.97 FPS, the number of parameters decreases from 2.58M to 1.70M (a reduction of about 34%), and the computational cost decreases from 6.4 GFLOPs to 4.0 GFLOPs (a reduction of about 37.5%). This demonstrates that the combination of the MobileOne architecture and the Hard-swish activation function significantly improves computational efficiency and reduces storage and hardware resource requirements while maintaining detection performance, making it more suitable for deployment on edge devices and mobile terminals.

When the SE attention mechanism is introduced on the basis of MobileOne and Hard-swish, the model's precision increases to 92.6% and the mAP rises to 96.3%. Although the recall decreases by 1 percentage point compared to the second group of experiments (a decrease of only 0.2 percentage points compared to the original model), the FPS only slightly drops to 25.64. This confirms that the SE module enhances the model's ability to recognize key features by adjusting the weights of feature channels. Notably, the number of parameters further decreases from 1.70M to 1.65M, and the computational cost decreases from 4.0 GFLOPs to 3.9 GFLOPs, indicating that the introduction of the SE module does not increase computational redundancy.

With the further adoption of Adaptive Threshold Focal Loss (ATFL) on top of the above optimizations, all model metrics are improved: precision increases to 93.2%, recall rises to 93.6%, mAP reaches 96.7%, and FPS increases to 27.93. These results indicate that ATFL effectively addresses the class imbalance problem and enhances the model's ability to learn from hard-to-classify samples by dynamically adjusting sample weights. ATFL not only compensates for the recall reduction caused by the SE module but also further improves precision, achieving overall optimization of detection performance. Such improvements in the loss function do not increase the computational cost during inference, yet can significantly enhance model performance.

Through the systematic combination of backbone optimization (MobileOne), activation function improvement (Hard-swish), attention mechanism modification (SE), and loss function replacement (ATFL), this study provides a feasible solution for efficient object detection in resource-constrained environments, offering valuable reference for practical applications such as wheat Fusarium head blight identification. Experimental results confirm that the optimized model improves detection accuracy while reducing computational resource consumption, meeting the dual requirements of high efficiency and high performance for object detection in edge computing environments.

4.3 Analysis of Experimental Results for Different Models

To further validate the lightweight effectiveness and detection performance of the model, other models from the YOLO series were selected for comparative experiments. The training, validation, and testing of all models were conducted under the same environment, and the FPS of each model was calculated on the CPU.

As shown in Table 5, different object detection models exhibit varying performance in terms of both accuracy and efficiency. Firstly, according to the results in Fig. 7, the

Table 5: Results of different models

Model	Precision (%)	Recall (%)	mAP (%)	FPS	Params (M)	FLOPs (G)
YOLOv5n	93.8	92.2	96.0	23.47	2.18	5.9
YOLOv8n	93.3	93.1	96.3	23.86	2.68	6.9
YOLOv9t	94.8	93.1	96.6	18.62	1.73	6.5
YOLOv10n	94.3	95.0	97.0	20.40	2.70	8.3
MSA-YOLO	93.2	93.6	96.7	27.93	1.65	3.9

mAP values of all models are relatively close. MSA-YOLO, in particular, demonstrates a significant advantage in computational efficiency while maintaining high detection accuracy. Specifically, MSA-YOLO achieves an mAP of 96.7%, only 0.3 percentage points lower than YOLOv10n, but its number of parameters (1.65M) and computational cost (3.9 GFLOPs) are 38.9% and 53.0% lower than those of YOLOv10n, respectively. In terms of inference speed, MSA-YOLO reaches 27.93 FPS, which is significantly higher than other models and 36.9% faster than YOLOv10n.

YOLOv10n achieves the best performance in terms of precision (94.3%) and recall (95.0%), with an mAP of 97.0%, but its high computational cost (8.3 GFLOPs) and parameter count (2.70M) limit its application in resource-constrained environments. Although YOLOv9t has a relatively small number of parameters (1.73M), close to that of MSA-YOLO, its inference speed (18.62 FPS) and computational cost (6.5 GFLOPs) do not reflect the efficiency advantages expected of a lightweight model. YOLOv5n and YOLOv8n show balanced performance in terms of accuracy and speed, but their computational resource requirements are not outstanding relative to their performance.

Considering the balance between detection accuracy and computational efficiency, MSA-YOLO provides a more reasonable trade-off between performance and efficiency in resource-constrained scenarios, offering an effective reference for the edge deployment of wheat Fusarium head blight identification.

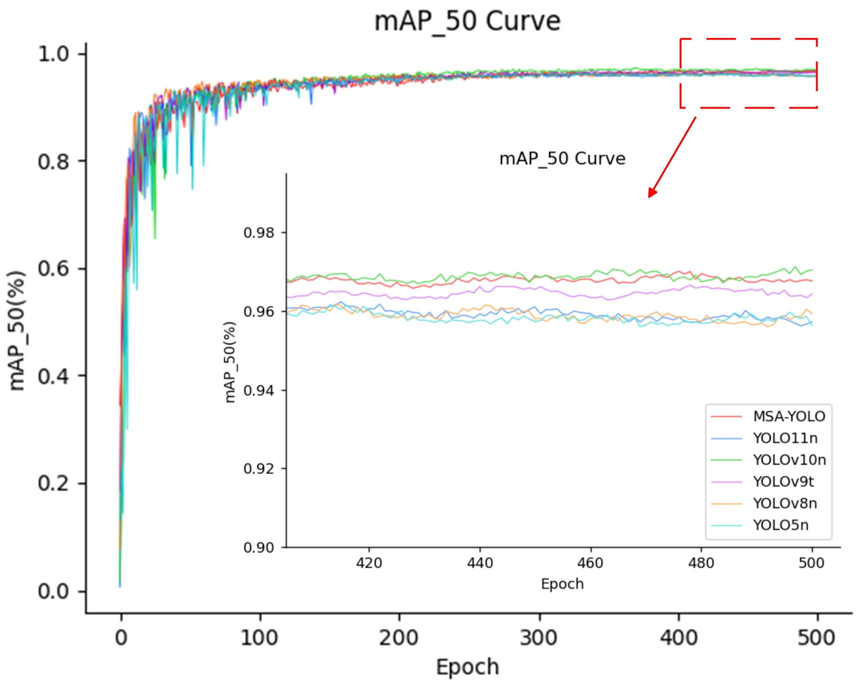


Fig. 7: Comparison of mAP_{50} curves for different YOLO-based models.

Next, in order to better evaluate the real-time inference time of the models, the average inference time of each model was assessed using the same set of 125 wheat test images on the same training device.

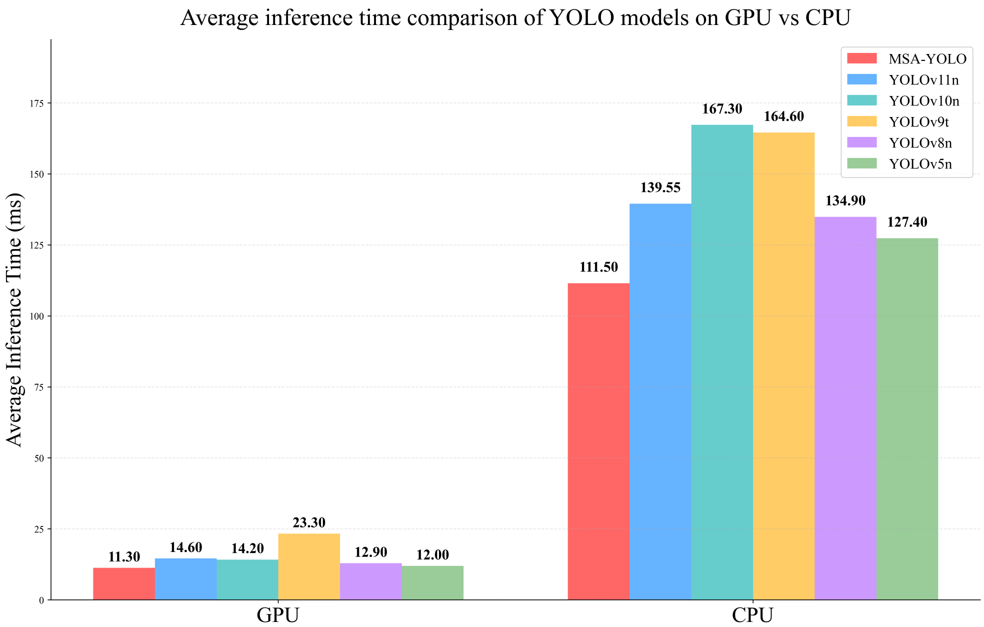


Fig. 8: Average inference time comparison of different models on GPU and CPU.

4.4 Analysis of the Results of Wheat Spike Scab

In this study, a systematic analysis was conducted on the comparison between the model predictions and the ground truth using 125 independent test images. For each image, we recorded the true values of key indicators, including the number of healthy spikelets, the number of diseased spikelets, the diseased spikelet rate, and the severity of the disease, and quantitatively compared them with the model predictions. As shown in the fitting plots in Fig. 9, the fitting performance for the number of healthy spikelets, the number of diseased spikelets, and the diseased spikelet rate was excellent, with R^2 values reaching 0.87, 0.92, and 0.93, respectively. Meanwhile, the RMSE values for all indicators remained at low levels, fully demonstrating that the MSA-YOLO model possesses high detection accuracy and reliable recognition capability.

It is worth noting that the R^2 value for the number of healthy spikelets (0.87) was slightly lower than that of the other indicators. We believe this is mainly because the number of healthy spikelets is much greater than that of diseased spikelets, resulting in a more pronounced absolute deviation even when the detection algorithm has the same proportional error. Nevertheless, an R^2 value of 0.87 still indicates that the model maintains high accuracy in detecting healthy spikelets, which is sufficient to meet practical application requirements.

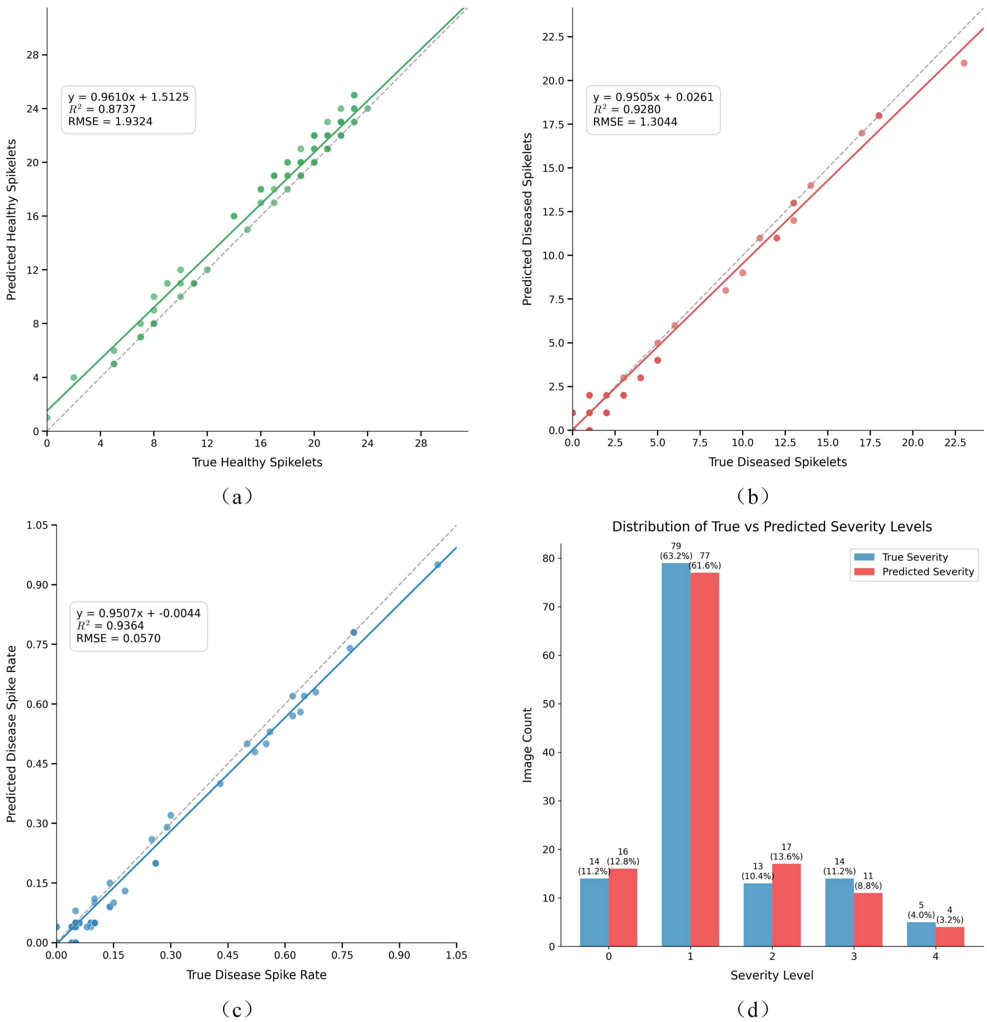
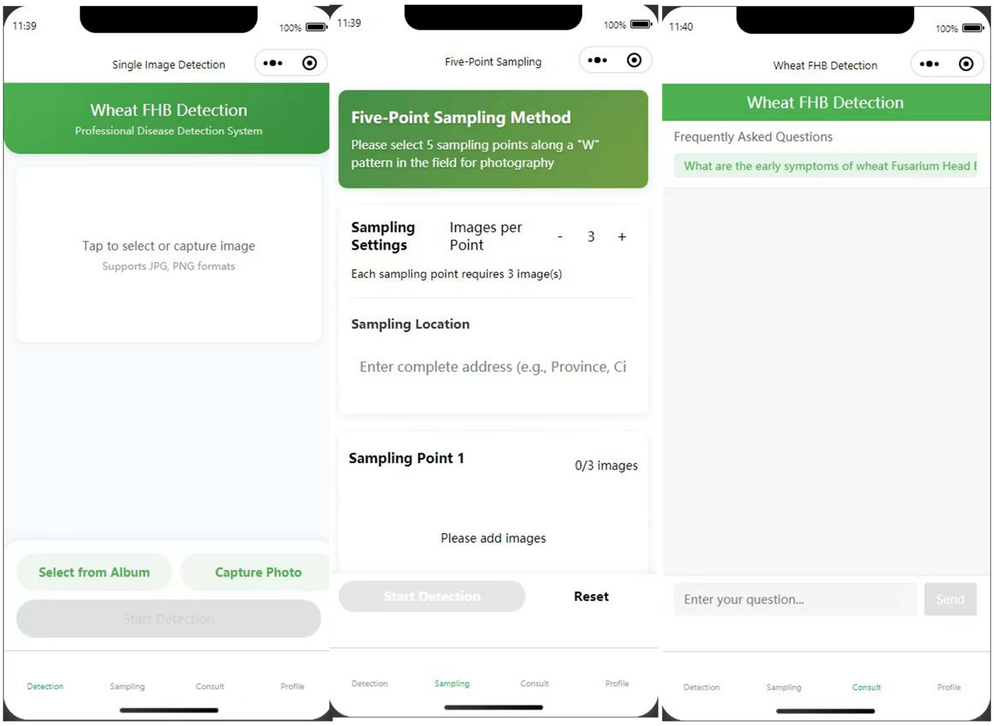


Fig. 9: Fitting results between ground truth and model predictions for (a) the number of healthy spikelets, (b) the number of diseased spikelets, (c) the diseased spikelet rate, and (d) the disease severity level distribution.

In this study, the assessment of wheat Fusarium head blight severity was conducted using the grading method specified in the national standard GB/T15796-2011 "Technical Specification for Monitoring and Forecasting Wheat Fusarium Head Blight." According to this specification, disease severity is defined as the proportion of spikelets infected with Fusarium head blight to the total number of spikelets, and the severity is classified into five levels: Level 0 (no diseased spikelets); Level 1 (the number of diseased spikelets accounts for less than 1/4 of all spikelets); Level 2 (the number of diseased spikelets accounts for 1/4 to 1/2 of all spikelets); Level 3 (the number of diseased spikelets accounts for 1/2 to 3/4 of all spikelets); and Level 4 (the number of diseased spikelets accounts for more than 3/4 of all spikelets). As shown in Fig. 9(d), there is little difference between the distribution of actual disease severity and the predicted disease severity, which further demonstrates the excellent detection performance of the model. In addition, we found that the disease severity of the test set data is mainly concentrated at Level 1. This is likely because the Fusarium pathogen was artificially inoculated into the middle and upper parts of the wheat spikes, resulting in most diseased spikelets being distributed in the middle and upper parts of the spike, and consequently, the disease severity is mainly at Level 1.

4.5 Mini-Program Deployment and Field Application Demonstration

To further validate the practical applicability of MSA-YOLO in real agricultural scenarios, the proposed lightweight model was deployed in a WeChat mini program for wheat Fusarium Head Blight (FHB) diagnosis. The system was developed based on the Flask framework to enable efficient communication between the front-end user interface and the back-end inference server. The mini program integrates two main functional modules: image-based disease detection and information consultation. In the detection module, users can acquire wheat spike images either by taking photos on site or by uploading images from the mobile album, after which the system automatically performs disease severity prediction and returns diagnostic results in real time, significantly improving the timeliness of FHB assessment (Fig. 10). To support large-scale field investigations, a five-point sampling strategy is embedded in the system, guiding users to collect images at five representative locations along a W-shaped route within the field, thereby enhancing the robustness and representativeness of the diagnostic results. In addition, the consultation module provides educational materials and frequently asked questions related to wheat FHB, helping users better understand early symptoms and disease management practices. Overall, the deployment of this mini program demonstrates the feasibility of integrating lightweight deep learning models into mobile platforms for precision agriculture and real-time crop disease monitoring.



(a) (b) (c)

Fig. 10: Deployment workflow and user interfaces of the wheat FHB diagnosis mini program: (a) image acquisition interface supporting photo capture and album upload; (b) five-point sampling module along a W-shaped route in the field; (c) consultation and FAQ interface for disease knowledge.

5 Class-Specific Performance and Robustness Evaluation

In order to comprehensively evaluate the recognition performance of the model, this study conducted a detailed analysis of the identification results of MSA-YOLO for both diseased spikelets (DS) and healthy spikelets (HS). The specific results are shown in Table 6. In addition, the detection performance of the recognition model under different lighting conditions and at different growth stages was also examined.

As shown in Fig. 11, the recognition model proposed in this study demonstrates stable performance under diverse lighting conditions and at different growth stages. However, it is noteworthy that the number of healthy spikelet (HS) samples in the dataset is significantly greater than that of diseased spikelet (DS) samples. This class imbalance leads to better recognition performance for healthy spikelets. As shown in Table 6, the precision for healthy spikelets reaches 95.2%, while that for diseased

Table 6: Detailed detection results for different spikelet categories

Category	Precision (%)	Recall (%)	mAP (%)
HS	95.2	97.5	98.3
DS	91.3	89.8	95.1

spikelets is only 91.3%; additionally, the recall for healthy spikelets is 7.7 percentage points higher than that for diseased spikelets. Furthermore, experimental observations indicate that, during certain growth stages, the morphological differences between diseased and healthy spikelets are relatively subtle. This similarity in features inevitably introduces a certain degree of classification error, which affects the overall recognition performance of the model.

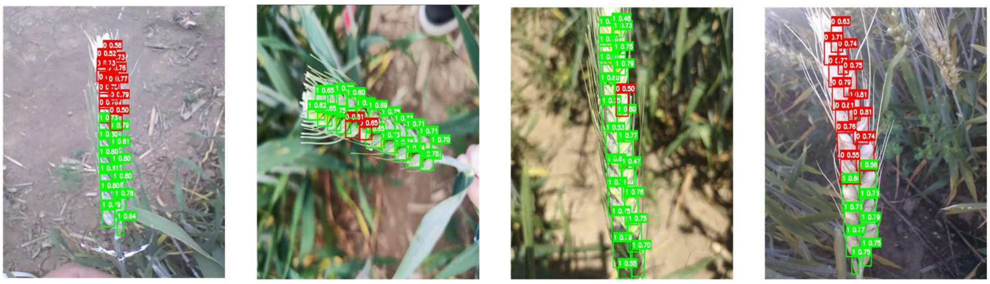


Fig. 11: Representative detection results of MSA-YOLO under different lighting conditions and growth stages.

From the confusion matrix in Fig. 12, it can also be observed that some diseased and healthy spikelets are detected as background, with the number of diseased spikelets being higher than that of healthy spikelets. We believe that, in such cases, the lighting conditions may cause some features of diseased spikelets to resemble background information. Therefore, in future research, we will consider using a more balanced training dataset and enhancing feature extraction capabilities to further improve the detection accuracy for diseased spikelets.

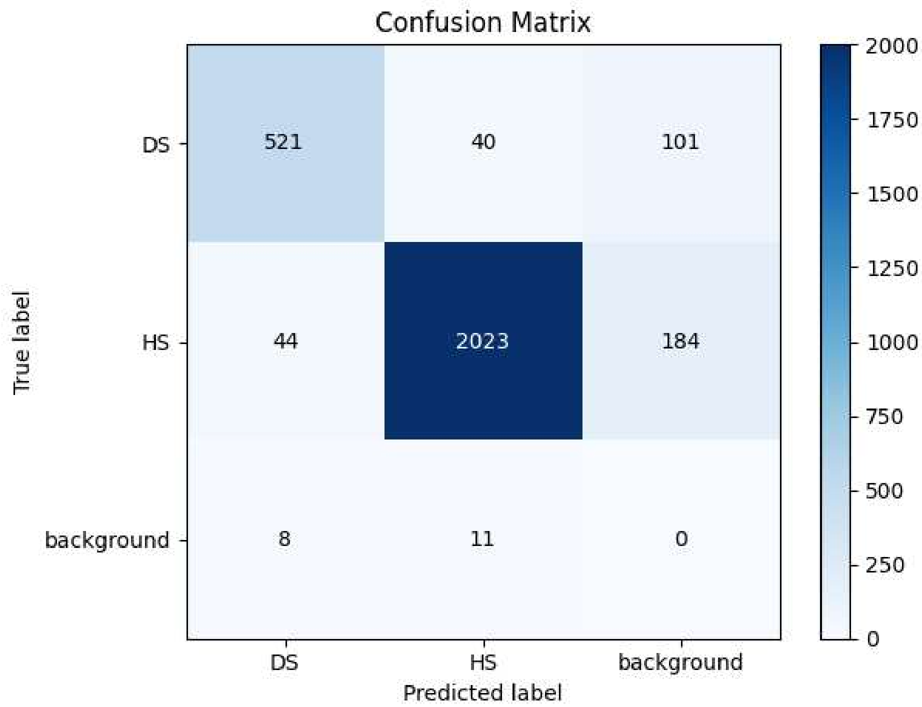


Fig. 12: Confusion matrix of MSA-YOLO for healthy spikelets (HS), diseased spikelets (DS), and background classes.

6 Discussion

To develop a timely and accurate monitoring method for wheat Fusarium head blight, recent studies have progressively shifted from single-scenario recognition pipelines toward field-oriented, lightweight, and multi-scale phenotyping frameworks. Technical reviews have highlighted the rapid development of automated FHB assessment across close-range imaging, UAV remote sensing, and multi-scale phenotyping Feng et al. (2024). In complex field environments, DeepFHB achieved high-throughput spike detection and refined lesion segmentation, whereas MGD-based severity recognition improved mobile deployment potential through network compression and distillation Zhou et al. (2024); Gong et al. (2024). UAV-based studies further extended FHB monitoring from close-range RGB images to broader field-scale disease assessment Bao et al. (2024). In addition, lightweight models such as YOLOv8s-CGF, SCS-YOLO, and EBS-YOLO demonstrated that high accuracy and real-time deployment are increasingly attainable for wheat FHB detection under resource-constrained conditions Yang et al. (2024); Gao et al. (2025); Mao et al. (2025). Compared with these recent studies, the proposed MSA-YOLO maintains competitive detection accuracy while providing a favorable balance among parameter count, computational cost, and inference speed, which is particularly important for practical agricultural deployment.

7 Conclusion

To address the challenge of wheat Fusarium head blight recognition in complex field environments, this study proposes an improved lightweight detection model, MSA-YOLO. Based on YOLO11n, the model incorporates three key optimizations: First, the original backbone is replaced with the MobileOne network, which adopts a re-parameterization design philosophy, laying the foundation for model lightweighting. Second, the PSABlock in the C2PSA module is improved by substituting the original multi-head attention mechanism with the more computationally efficient SE module, thereby further reducing computational cost while maintaining detection accuracy. Finally, considering the class imbalance between healthy and diseased spikelets in practical applications, the traditional binary cross-entropy loss function (BCEWithLogitsLoss) is replaced with the Adaptive Threshold Focal Loss (ATFL), enhancing the model's ability to recognize minority class samples.

Experimental results demonstrate that the improved MSA-YOLO model achieves significant increases in inference speed on both GPU and CPU platforms, while reducing the number of parameters and computational cost. In terms of detection performance, evaluation metrics such as precision, recall, and mAP_{50} all surpass those of the original model. The improved model also exhibits excellent recognition performance for key agricultural indicators such as diseased spikelet rate and disease severity.

In summary, the MSA-YOLO model not only demonstrates outstanding capability in wheat Fusarium head blight detection, but its lightweight characteristics also make it particularly suitable for deployment on edge computing devices or in resource-constrained environments, providing a feasible solution for real-time field disease monitoring.

Statements and Declarations

Funding

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31501225 and 31701309), the Natural Science Foundation of Henan Province (Grant No. 242301420143), the National Key Research and Development Program of China (Grant No. 2017YFD0301105), and the Key Research and Development Project of Henan Province (Grant No. 241111110080).

Author contributions

L.S.: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. Z.S.: Data curation, Validation, Investigation. B.B.: Resources, Methodology, Writing – review and editing. Y.M.: Data curation, Visualization, Investigation. F.Y.: Software, Validation, Formal analysis. W.G.: Investigation, Resources. Y.C.: Methodology, Validation, Writing – review and editing. J.L.: Investigation, Visualization. S.X.: Supervision, Project administration. Y.L.: Conceptualization, Supervision, Funding acquisition, Writing – review and editing. All authors read and approved the final manuscript.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

References

Al-Hashimi A, Aina O, Daniel AI, et al (2025) Critical review on characterization, management, and challenges of Fusarium head blight disease in wheat. Physiological and Molecular Plant Pathology 136:102557. https://doi.org/10.1016/j.pmpp.2024.102557

Alisaac E, Mahlein AK (2023) Fusarium head blight on wheat: Biology, modern detection and diagnosis and integrated disease management. Toxins 15(3):192. https://doi.org/10.3390/toxins15030192

Bao W, Huang C, Hu G, et al (2024) Detection of Fusarium head blight in wheat using UAV remote sensing based on parallel channel space attention. Computers and Electronics in Agriculture 217:108630

Cui C, Gao T, Wei S, et al (2021) PP-LCNet: A lightweight CPU convolutional neural network. URL https://arxiv.org/abs/2109.15099, arXiv:2109.15099

Ding X, Zhang X, Ma N, et al (2022) RepVGG: Making VGG-style convnets great again. IEEE Transactions on Pattern Analysis and Machine Intelligence 44:12642–12655

Feng G, Gu Y, Wang C, et al (2024) Wheat Fusarium head blight automatic non-destructive detection based on multi-scale imaging: A technical perspective. Plants 13(13):1722. https://doi.org/10.3390/plants13131722

Gao C, He B, Guo W, et al (2025) SCS-YOLO: A real-time detection model for agricultural diseases—a case study of wheat Fusarium head blight. Computers and Electronics in Agriculture 238:110794

Gong Z, Gao C, Feng Z, et al (2024) Integrating masked generative distillation and network compression to identify the severity of wheat Fusarium head blight. Computers and Electronics in Agriculture 227:109647

Hamila O, Henry CJ, Molina OI, et al (2024) Fusarium head blight detection, spikelet estimation, and severity assessment in wheat using 3D convolutional neural networks. Canadian Journal of Plant Science 104(4):358–374. https://doi.org/10.1139/cjps-2023-0127

Howard A, Sandler M, Chu G, et al (2019) Searching for MobileNetV3. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 1314–1324

Howard AG, Zhu M, Chen B, et al (2017) MobileNets: Efficient convolutional neural networks for mobile vision applications. URL https://arxiv.org/abs/1704.04861, arXiv:1704.04861

Hu J, Shen L, Sun G (2020) Squeeze-and-Excitation networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 42:2011–2023

Jocher G, Chaurasia A, Qiu J (2023) YOLOv8 by ultralytics. GitHub repository, URL https://github.com/ultralytics/ultralytics, accessed 21 Mar 2026

Kumar D, Kukreja V (2024) Image segmentation, classification, and recognition methods for wheat diseases: Two decades' systematic literature review. Computers and Electronics in Agriculture 221:109005. https://doi.org/10.1016/j.compag.2024.109005

Kumar D, Kukreja V (2025) Impact of image segmentation and feature sets in automated plant disease classification: A comprehensive review based on wheat plant images. Progress in Artificial Intelligence 14(4):451–504. https://doi.org/10.1007/s13748-025-00389-6

Li S, Dong P, Zhang H, et al (2025) Integrating ecological niche and epidemiological models to predict wheat Fusarium head blight using remote sensing and meteorological data. Computers and Electronics in Agriculture 234:110255. https://doi.org/10.1016/j.compag.2025.110255

Li Y, Wang H, Zhang J (2022) Segmentation of stripe rust on wheat leaves using octave-unet. Computers and Electronics in Agriculture 198:107013

Liang X, Yang S, Mu L, et al (2025) A non-destructive system using uve feature selection and lightweight deep learning to assess wheat Fusarium head blight severity levels. Agronomy 15(9):2051. https://doi.org/10.3390/agronomy15092051

Lin TY, Goyal P, Girshick R, et al (2020) Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 42:318–327

Liu S, Dai J, Huang J, et al (2025) Combining ultralow-altitude drone phenotyping with deep learning analytics to assess resistance and disease dynamics of Fusarium head blight in wheat. The Crop Journal 13(5):1372–1385. https://doi.org/10.1016/j.cj.2025.08.009

Mao R, Yuan H, Li F, et al (2025) EBS-YOLO: Edge-optimized bidirectional spatial feature augmentation for in-field detection of wheat Fusarium head blight epidemics. Plant Methods 21(1):1–16. https://doi.org/10.1186/s13007-025-01449-7

Moonjely S, Ebert M, Paton-Glassbrook D, et al (2023) Update on the state of research to manage Fusarium head blight. Fungal Genetics and Biology 169:103829. https://doi.org/10.1016/j.fgb.2023.103829

Morellos A, Pantazi XE, Almoujahed MB, et al (2025) Forecasting of fusarium head blight spatial distribution in winter wheat using machine learning. Computers and Electronics in Agriculture 231:109967. https://doi.org/10.1016/j.compag.2025.109967

Mustafa G, Zheng H, Khan IH, et al (2024) Enhancing fusarium head blight detection in wheat crops using hyperspectral indices and machine learning classifiers. Computers and Electronics in Agriculture 218:108663. https://doi.org/10.1016/j.compag.2024.108663

Nair V, Hinton GE (2010) Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning, pp 807–814

Pacal I, Kunduracioglu I, Alma MH, et al (2024) A systematic review of deep learning techniques for plant diseases. Artificial Intelligence Review 57(11):304. https://doi.org/10.1007/s10462-024-10944-7

Shoaib M, Sadeghi-Niaraki A, Ali F, et al (2025) Leveraging deep learning for plant disease and pest detection: a comprehensive review and future directions. Frontiers in Plant Science 16:1538163. https://doi.org/10.3389/fpls.2025.1538163

Ultralytics (2024) YOLO11: You only look once – version 11. Ultralytics documentation, URL https://docs.ultralytics.com/models/yolo11/, accessed 21 Mar 2026

Upadhyay A, Chandel NS, Singh KP, et al (2025) Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture. Artificial Intelligence Review 58(3):92. https://doi.org/10.1007/s10462-024-11100-x

Wen X, Maimaiti M, Liu Q, et al (2024) MnasNet-SimAM: An improved deep learning model for the identification of common wheat diseases in complex real-field environments. Plants 13(16):2334. https://doi.org/10.3390/plants13162334

Xu L, Cao B, Zhao F, et al (2023) Wheat leaf disease identification based on deep learning algorithms. Physiological and Molecular Plant Pathology

Yang C, Sun X, Wang J, et al (2024) YOLOv8s-CGF: A lightweight model for wheat ear Fusarium head blight detection. PeerJ Computer Science 10:e1948

Yu J, Li J, Zhou C (2023) MobileOne: An improved one millisecond mobile backbone. URL https://arxiv.org/abs/2306.01195, arXiv:2306.01195

Zhou Q, Huang Z, Liu L, et al (2024) High-throughput spike detection and refined segmentation for wheat Fusarium head blight in complex field environments. Computers and Electronics in Agriculture 227:109552