Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology) in press
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Research Article

https://doi.org/10.1631/jzus.B2500225

Improved lightweight convolutional neural network models for the detection and evaluation of Fusarium head blight in wheat

Wang ZHANG[1,2], Yi REN[1,2], Zidi GUO[1,2], Han LI[1,2], Man ZHANG[1,2], Jie LIU[3]✉, Ruicheng QIU[1,2]✉

[1]*Key Laboratory of Smart Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing 100083, China*
[2]*College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China*
[3]*Yantai Academy of Agricultural Sciences, Yantai 265500, China*

Abstract: Fusarium head blight (FHB), a frequent disease in wheat cultivation, can lead to substantial yield losses and the production of mycotoxins in grains. Therefore, the development of wheat varieties resistant to FHB is an important strategy to reduce related losses. In this respect, manual surveys of FHB are time-consuming and labor-intensive. To overcome this issue, this paper proposes a method for detecting and evaluating wheat FHB using color imaging and deep learning. Initially, a lightweight convolutional neural network model based on the YOLOv8s was designed to detect wheat spikes from color images. Testing showed that the mean average precision of the model in spike detection reached 0.964. Moreover, another lightweight model was developed for detecting wheat spikelets and FHB. To enhance the detection capability of the model for small objects, space-to-depth convolution and BiFormer attention modules were integrated. The results indicated that the model can accurately detect spikelet and FHB, with a mean average precision of 0.936. Finally, based on the wheat spikelet detection results, the rate of diseased wheat spikes (RD_S) and the disease index (DI_W) were calculated, in order to evaluate the severity of wheat FHB. For RD_S and DI_W, the coefficients of determination between phytologists' evaluations and the estimates derived from the proposed method were 0.71 and 0.93, respectively. These results demonstrate that the proposed method facilitates the accurate and efficient detection of wheat FHB and contributes to the quantitative evaluation of FHB in the field.

Key words: Fusarium head blight; Lightweight neural network; Disease detection; YOLO; Deep learning

1 Introduction

As one of the world's most widely cultivated crops, wheat plays an essential role in ensuring global food security. However, wheat is highly susceptibale to certain biotic stresses during its growth stages. Fusarium head blight (FHB) is a prevalent wheat disease induced by various *Fusarium* species, which is particularly prevalent in regions where relative air humidity and rainfall are high (Figueroa et al., 2018). Wheat FHB can cause spike rot symptoms in parts of wheat spikes and induce the production of toxic deoxynivalenol in wheat grains, leading to annual wheat yield losses of over 20% (Hellemans et al., 2018). Consequently, the development of resistant varieties is critical for wheat productivity and food security.

Infection with FHB pathogens induces color changes in wheat florets and spikelets: FHB outbreaks show the uniform yellowing of wheat spikes and a pink gelatinous fungal layer on the diseased spots. These symptoms can be visually observed and recorded under visible light (Qiu et al., 2019). Conventionally, the detection and evaluation of wheat FHB are mainly conducted by experienced phytologists, who determine the infection severity based on the color characteristics of wheat spikes (Saccon et al., 2017). However, this method has several limitations, including substantial time and labor costs, low survey efficiency, and high susceptibility to

✉ Ruicheng QIU, qrc@cau.edu.cn
Jie LIU, liujie@yt.shandong.cn
Ruicheng QIU, https://orcid.org/0000-0002-3113-5712

Received Apr. 30, 2025; Revision accepted Sept. 6, 2025;
Crosschecked xxx. xx, 20xx; Published online xxx. xx, 20xx

error due to subjective factors. Furthermore, the limited scope of individual quantitative surveys hampers the timely and comprehensive implementation of FHB management measures across large areas. Therefore, many researchers have explored approaches to effectively evaluate the damage severity of wheat FHB using non-destructive detection techniques, such as hyperspectral, multispectral, and color imaging.

Hyperspectral and multispectral imaging techniques are capable of capturing both the spectral data and images of wheat plants, and as such have been widely used in the detection and evaluation of wheat FHB. Specifically, Vincke et al. (2023) collected hyperspectral data of wheat spikes, developed partial least squares discriminant analysis models to discriminate distinct spike regions at the pixel level, and found that wavelengths associated with water and nitrogen are critical for FHB detection. Mustafa et al. (2024) used hyperspectral images and indices to develop classification models and detect the progression of wheat FHB. The results indicated that spectral changes were consistent with the development of wheat FHB. Chen Y. et al. (2023) proposed a spectral analysis model to assess FHB infection using wheat canopy visible spectral data. This model integrated 35 effective wavelengths and achieved a coefficient of determination (R^2) of 0.9490 for the modeling dataset and 0.9312 for the validation dataset. Moreover, Almoujahed et al. (2022) employed a hyperspectral camera to detect wheat FHB at the ripening stage for yield loss estimation. They segmented spikes from the background using normalized difference vegetation index images, then estimated the FHB infected area percentage via developed models. However, acquiring spectral data with a hyperspectral imager not only requires specialized operator expertise but also entails considerable costs. Furthermore, complex field conditions (such as variable lighting or wind) adversely affect the qualities of acquired spectral data, and extracting useful information from high-dimensional spectral datasets is also challenging. Collectively, these issues limit the widespread applications of hyperspectral and multispectral imaging in the field.

Alternatively, the color imaging processing technique supports non-destructive, rapid crop disease detection, with relatively simple data processing and low computational costs (Khan et al., 2019; Pantazi et al., 2019). Color imaging-based crop disease detection has been widely applied and achieved significant progress. Basavaiah and Anthony (2020) proposed a machine learning framework that integrates multimodal image features for tomato leaf disease identification. Their approach employed feature fusion techniques to enhance classification performance, and ultimately achieved 94% accuracy with an optimized random forest classifier. Abdu et al. (2020) developed an automated plant disease detection method that utilizes optimized features extracted from localized lesion regions of infected plants. Sampathkumar and Rajeswari (2022) used the fuzzy C-mean algorithm to distinguish symptomatic and healthy regions, with good performance in rice and apple disease detection. Nevertheless, traditional machine learning methods still have inherent drawbacks in crop disease detection, especially in extracting features from complex backgrounds. As such, improving the robustness of image feature extraction under field conditions is an urgent task. Currently, researchers have extensively adopted deep learning techniques to enhance both feature extraction and crop disease detection. These techniques enable automated feature extraction from color images, significantly improving model generalization. For crop disease detection, convolutional neural networks (CNNs) are extensively used as the primary deep learning approach (Ferentinos, 2018). Agarwal et al. (2020) proposed a simplified CNN model for

tomato disease detection, which demonstrated superior detection accuracy compared to conventional machine learning approaches. Fang et al. (2020) modified the loss function of the ResNet50 model and used an adaptive adjustment algorithm to process leaf images with non-uniform illumination, achieving a disease detection accuracy of 95.61%. Zhao et al. (2020) developed a new deep learning system for in-field crop disease identification, achieving 97.5% detection accuracy. Chen et al. (2021) adopted MobileNet-V2 as the backbone network of their CNN model, incorporated an attention mechanism to capture inter-channel feature relationships, optimized the loss function, and achieved 99.67% average detection accuracy on a public dataset. However, despite that deep learning models perform well in crop disease detection, their large parameter sizes and high computational resource requirements hinder deployment on resource-constrained devices like mobile and embedded systems. In recent years, researchers have focused on developing efficient lightweight models for crop disease detection. Hong et al. (2022) developed a lightweight model by integrating a lightweight network, non-maximum suppression, and complete intersection over union. This model can be deployed on unmanned aerial vehicles for wheat FHB identification. Mao et al. (2023) proposed a lightweight network based on an improved YOLO model for the severity evaluation of wheat FHB. While lightweight designs effectively reduce parameter and model sizes, they often lead to compromised detection accuracy. Consequently, further optimizations are needed to minimize the dimensionality and parameter sizes of advanced lightweight backbones and modules without compromising accuracy.

The accurate field evaluation of FHB severity is crucial for assessing wheat disease resistance in breeding and validating field management efficacy. Notably, quantitative FHB evaluation is particularly essential. A sufficient number of wheat spikes are required to be examined, with classification and quantification of diseased versus healthy spikelets and florets. Previous studies have employed color imaging and deep learning techniques for FHB severity evaluation, while these approaches often graded FHB severity by detecting wheat spikes and calculating the percentage of diseased pixels (Gao et al., 2024; Liu et al., 2024; Zhang et al., 2023). For this purpose, the precise segmentation of wheat spikes and diseased areas is required. Moreover, pixel-based estimation approaches are inconsistent with traditional visual observation methods used in field evaluations (Rößle et al., 2023). Thus, the quantitative detection and evaluation of wheat FHB needs to be further improved.

Therefore, the main purpose of this study was to develop lightweight models based on employing color imaging and deep learning techniques for wheat FHB detection and evaluation. The sub-objectives of this study were: (1) to develop lightweight models for detecting wheat spikes and spikelets in field color images; and (2) to propose a novel evaluation method for quantitative FHB assessment at both scales of individual spikes and the entire population.

2 Material and methods

To support deep learning-based wheat FHB detection and severity evaluation, an image processing system was developed. The system workflow, as illustrated in Fig. 1, comprises color image acquisition, image preprocessing, model development, and FHB detection coupled with quantitative severity evaluation. Wheat images were collected under field environments and processed to construct datasets specifically for spike and spikelet detection. However, the complex background of wheat fields poses challenges to the accurate detection of wheat spikes. As the primary infection sites of FHB, spikelets are characterized by small physical size, minimal image proportion, and low resolution in wheat images, making the direct, accurate detection and evaluation of wheat spikelets a challenging task. Therefore, two specialized models were developed separately to improve the detection performance of wheat spikes and spikelets, respectively. First, one lightweight model was designed to detect and extract wheat spikes from wheat images. Second, the extracted spikes were processed by the other developed model to detect individual spikelets and classify them. Finally, taking the detection results of spikes and spikelets as a basis, the severity of wheat FHB was quantitatively evaluated.

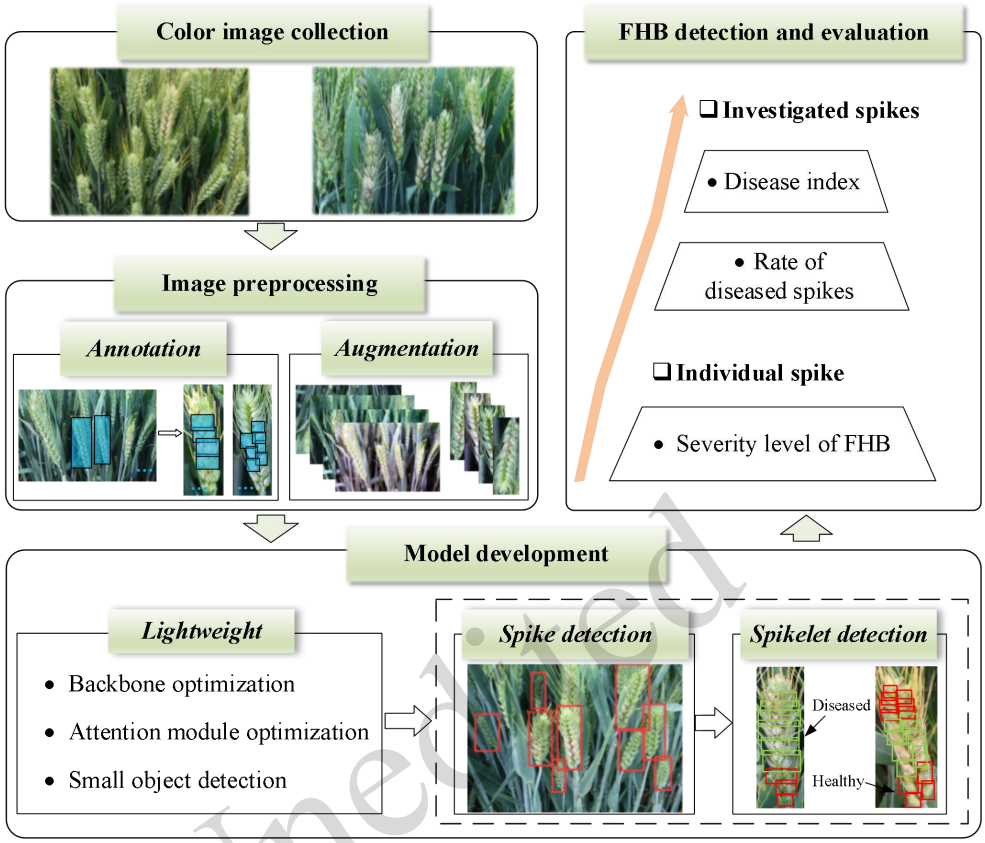


Fig. 1 Workflow diagram of the proposed wheat FHB detection and evaluation model.

2.1 Data collection

Field experiments were conducted at Yantai Academy of Agricultural Sciences (121°17'1.113"E, 37°29'39.358"N), Yantai, China. Multiple wheat lines with varying levels of resistance to FHB were cultivated. Wheat spikes at the flowering stage were inoculated with FHB pathogens via the grain spawn method, then bagged and kept under high humidity for 24 h. Wheat images were collected from 22 to 26 May, 2023, between 9:00am and 4:00pm. At the time of collection, most wheat lines reached the milk growth stage. A handheld digital camera (EOS R8, Canon, Japan) with a resolution of 6000×4000-pixels was used to acquire horizontal-view color images of different wheat lines. The distance between the camera and spikes was maintained at approximately 35 cm. The focal length and exposure time of the camera were set to 50 mm and 1/80 s, respectively. The wheat sampling areas were shaded with an umbrella to minimize interference from ambient sunlight. A total of 2,143 color images were collected from 89 wheat lines, including 1,172 images of FHB-infected wheat at various severity levels and 971 images of healthy wheat samples.

2.2 Dataset construction

LabelImg software was used to annotate spikes and spikelets in the collected wheat images and to construct two datasets (designated as Dataset 1 and Dataset 2), which were then used for developing spike and spikelet detection models, respectively. First, the complete and unobstructed spikes in the images were selected, and each selected spike was marked with a rectangular bounding box, as shown in Fig. 2, to construct Dataset 1. Second, spikes were extracted from wheat images, and the corresponding spikelets within these extracted spikes were then selected and annotated. Notably, apical spikelets were not labeled. In addition, since wheat spikelets exhibit different color and shape characteristics in the frontal versus lateral views, spikelets were labeled using two different methods to construct Dataset 2 (Fig. 2). For Dataset 2, all wheat spikelets were classified into

healthy and infected classes. Subsequently, all labeled images in Datasets 1 and 2 were reviewed and corrected by professional phytologists to ensure the accuracy of annotations. Finally, Dataset 1 and Dataset 2 contained 1,082 and 1,244 labeled images, respectively. Images in each dataset were randomly divided into training, validation, and testing subsets at an approximate ratio of 6:2:2. To enhance the diversity of the datasets, several data augmentation techniques were applied to the images in each dataset, including horizontal and vertical flipping, contrast enhancement, and histogram equalization operations. In the end, Dataset 1 and Dataset 2 contained 5,291 and 6,124 annotated images, respectively. As detailed in Table 1, the training, validation, and testing subsets of Dataset 1 comprised 3175, 1058, and 1058 images, while those of Dataset 2 included 3674, 1225, and 1225 images, respectively.

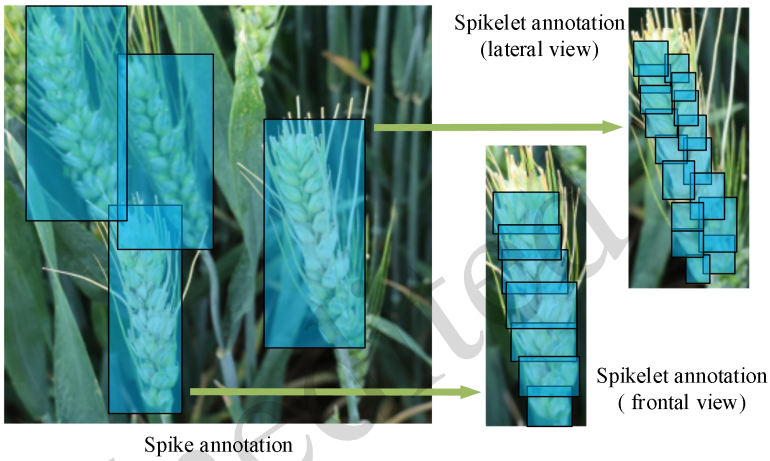


Fig. 2 Annotation of wheat spike images.

Table 1 Details of the datasets.

Dataset	Image number			
	Training set	Validation set	Testing set	Total
Dataset 1	3175	1058	1058	5291
Dataset 2	3674	1225	1225	6124

2.3 Wheat spike detection model

2.3.1 Network structure

To quantitatively evaluate FHB severity, spike detection is a fundamental step. The YOLO model, a widely used object detection framework, is known for the simultaneous localization and classification ability of objects within images. It treats object detection as a unified regression task. Unlike conventional object detection methods, the YOLO model directly predicts bounding boxes and class probabilities for objects across entire images, enabling a significant improvement in detection speed (Redmon et al., 2016). In this study, a lightweight model derived from the YOLOv8s model was designed to improve the detection performance of wheat spikes. The structure of this model is shown in Fig. 3. In this model, a lightweight backbone network, namely FasterNet (Chen J. et al., 2023), an efficient attention module - efficient multi-scale attention (EMA) (Ouyang et al., 2023), and the asymptotic feature pyramid network (AFPN) (Hu et al., 2022) for feature fusion were applied and optimized. Their details will be described in the following sections.

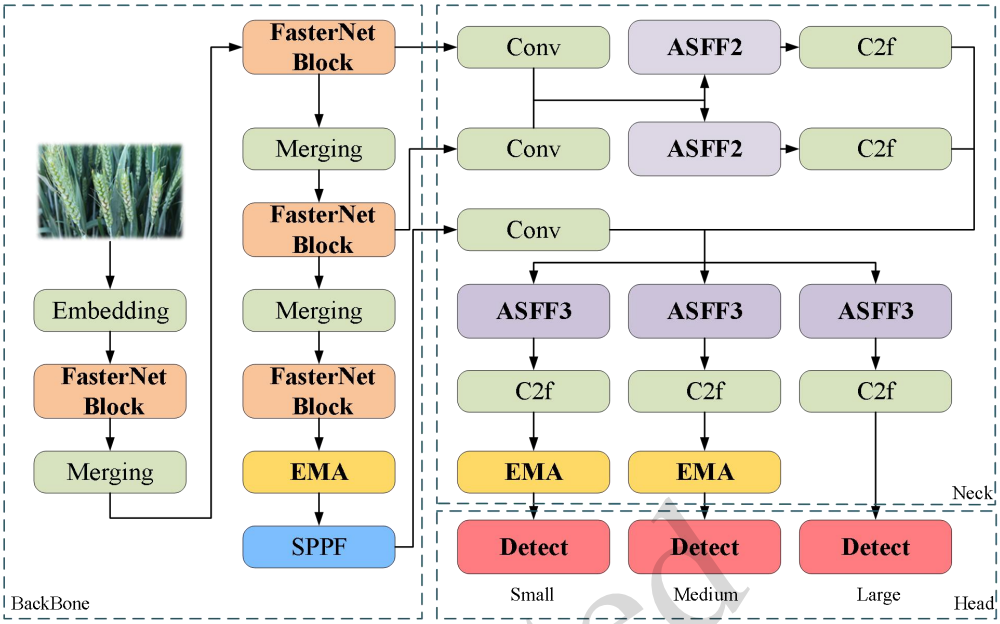


Fig. 3 Structure of the lightweight model for wheat spike detection. SPPF means spatial pyramid pooling fast module, and ASFF means adaptively spatial feature fusion module (Liu et al., 2019).

2.3.2 Backbone optimization

The lightweight model applied YOLOv8s as its main structure, and the backbone network of YOLOv8s was replaced with FasterNet, as illustrated in Fig. 4. The innovation of FasterNet is to reduce the floating-point operations (FLOPs) while simultaneously increasing the numbers of floating-point operations per second. This results in faster processing speeds and lower computational complexity. The structure incorporates the partial convolution (PConv) technique (Liu et al., 2018), which performs standard convolution only on selected input channels while keeping the rest intact. This approach effectively minimizes computational redundancy and reduces memory access requirements.

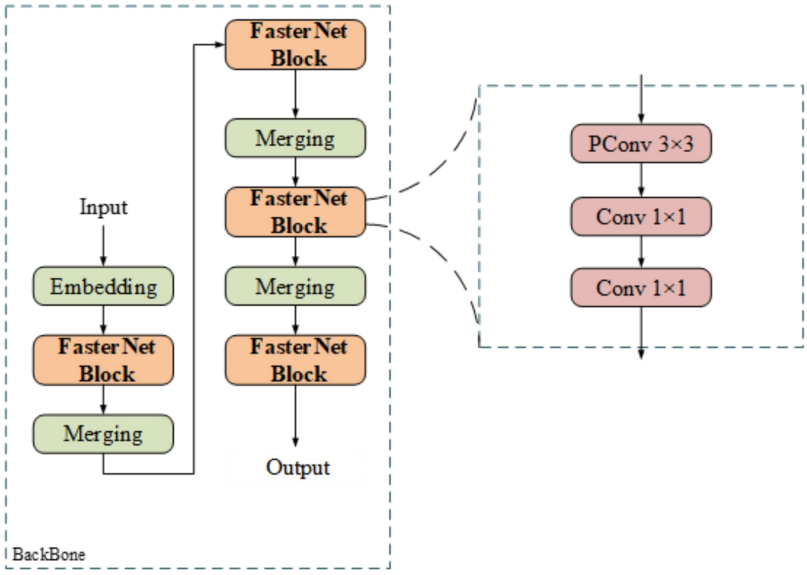


Fig. 4 Backbone network of the lightweight model for wheat spike detection.

2.3.3 Neck optimization

To improve wheat spike detection accuracy, the designed model incorporated an efficient attention module EMA. This module utilizes a distinctive approach to enable cross-channel and cross-space information interaction through grouped structures and multiscale convolutions while preserving channel dimensionality. Its primary purpose is to preserve channel information while minimizing computational complexity, achieved by converting select channels into batch dimensions and partitioning channels into multiple sub-feature sets (Wang et al., 2024). This approach ensures the uniform distribution of spatial semantic features across each feature set. A lightweight and generalized module was developed, which is robust and highly compatible with the adopted backbone network.

Additionally, AFPN was used for feature fusion, enabling direct interaction between non-adjacent levels by fusing two adjacent low-level features while progressively integrating high-level features. This strategy alleviates the inherent semantic gaps in non-adjacent level interactions (Chen Z. et al., 2023). The ASFF module was also implemented to address multi-object information conflicts that arise during feature fusion at each spatial location.

2.4 Spikelet and FHB detection model

2.4.1 Network structure

In Section 2.3, the complete and unobstructed spikes were detected and segmented from the background. In this section, the other model was designed to process these spikes for detecting their spikelets and evaluating FHB severity. The structure of this model is shown in Fig. 5. First, the MobileViT (Mehta and Rastegari, 2021) was used as the backbone network of the YOLOv8s model for feature extraction. Second, the space-to-depth convolution (SPD-Conv) module (Sunkara and Luo, 2022) was integrated to enhance the detection capability for small objects. This module addresses the limitations of traditional CNN structures through refined information processing. Finally, the BiFormer attention module (Zhu et al., 2023) was incorporated to enhance the efficiency and accuracy of the designed model in dense prediction tasks.

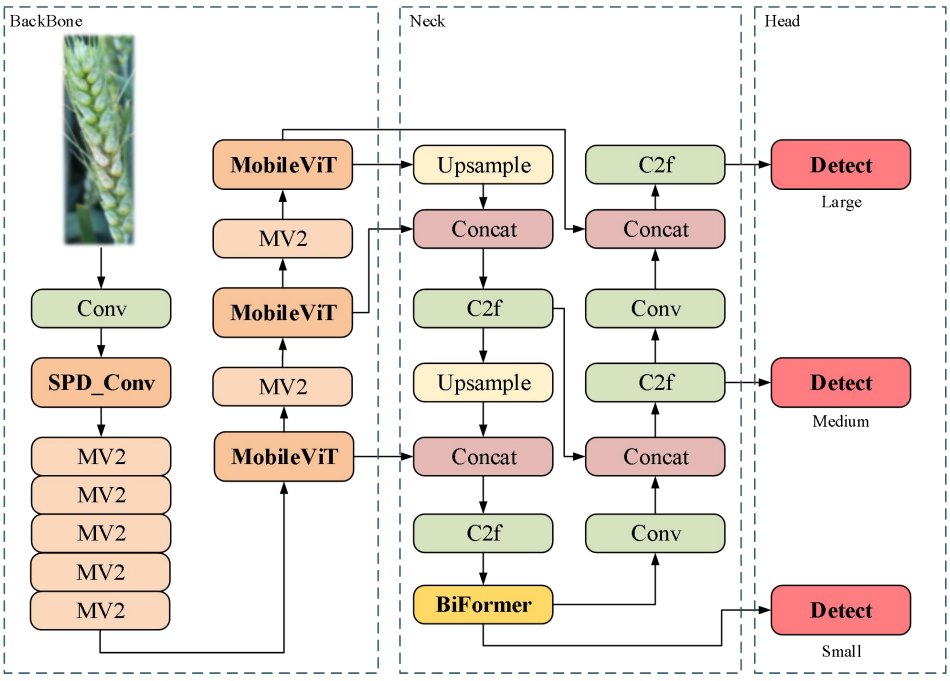


Fig. 5 Structure of the lightweight model for wheat spikelet and FHB detection.

2.4.2 Backbone optimization

Compared to the entire wheat spike, the wheat spikelet is relatively small and occupy only minimal areas in the image. To improve small object detection, the SPD-Conv module, shown in Fig. 6, was introduced. This module consists of an SPD layer and a non-spanning convolution layer, which exhibits good compatibility and can be integrated into most CNN architectures. It also shows good performance for both small object detection and low-resolution image analysis, making it a particularly powerful tool for spikelet detection. To improve the lightweight performance of YOLOv8s, MobileViT was adopted to replace the model's backbone network. As depicted in Fig. 6, MobileViT is a lightweight vision transformer especially designed for mobile devices. Integrating the advantages of CNN and vision transformer, it provides a new approach on global information processing. This approach treats the transformer as a convolution and replaces local processing in the convolution, thereby realizing a more efficient encoding of global information (Howard et al., 2019). The MobileViT block comprises standard convolution, pointwise convolution and a transformer component, which work together to preserve the spatial orders of pixels and blocks. As a result, this model outperforms conventional CNN and vision transformer models on multiple tasks and datasets.

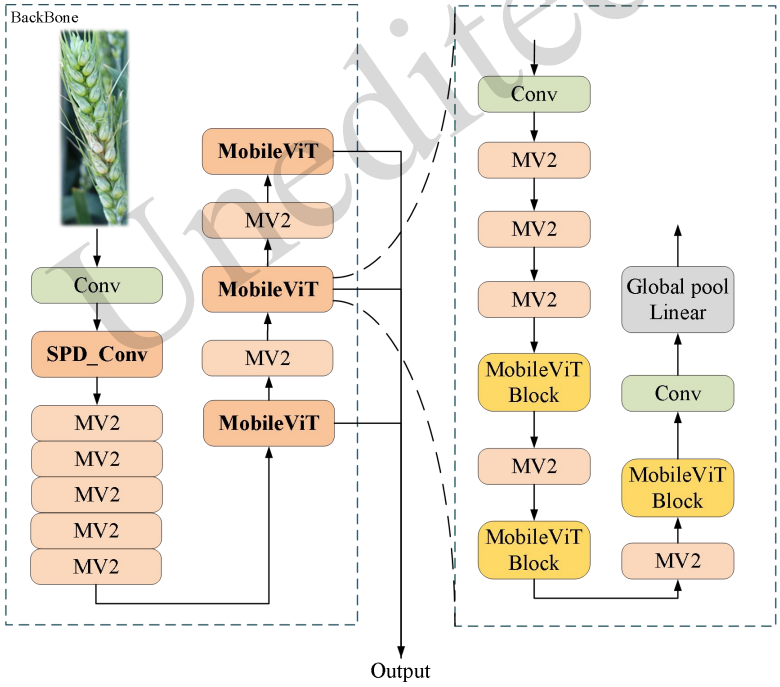


Fig. 6 Backbone network of the lightweight model for wheat spike and FHB detection.

2.4.3 Neck optimization

A wheat spike contains numerous spikelets, which are densely distributed. To improve the efficiency and accuracy of intensive spikelet detection, the BiFormer attention module was integrated into the model. This novel mechanism facilitates dynamic query-aware sparsity in visual transformers, which incorporates region-to-region routing and token-to-token attention (Park et al., 2023). This module adaptively focuses queries on relevant tokens while eliminating irrelevant ones and delivers exceptional performance and computational efficiency in intensive prediction tasks. These advantages significantly improve the performance of wheat spikelet detection under actual field conditions.

2.5 Model training

The designed models for wheat spike detection, and spikelet detection and FHB evaluation were developed using the PyTorch framework and trained on Dataset 1 and Dataset 2, respectively. Model training was conducted using an AutoDL cloud server computing platform, which operates on the Ubuntu 20.04 operating system and PyTorch version 2.0.0.

When training the models for wheat spike detection, spikelet detection and FHB evaluation, all images were initially resized to 640×640 pixels. To ensure sufficient model training, the maximum number of iterations was set to 200, and the momentum was set to 0.9. The AutoBatch technique was employed to dynamically determine the batch size, which was done automatically according to the available GPU memory to achieve a maximized value. To minimize the computational costs, stochastic gradient descent was selected as the optimizer. The initial learning rate was set to 0.01 to accelerate model convergence, and a learning rate decay of 0.0005 was incorporated to prevent overfitting. In addition, an early stop mechanism was employed to keep track of the model performance during the training process.

2.6 Wheat FHB evaluation

2.6.1 Evaluation of the lightweight models

Several parameters were used to evaluate the detection performance of the designed models, including precision (*P*), recall (*R*) (Ahlgren and Kekalainen, 2007), and mean average precision (mAP) (Zheng et al., 2015). Specifically, *P*, *R*, and mAP were calculated using the following Equations 1-3:

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$\mathrm{mAP} = \frac{1}{N}\sum_{n=1}^{N} AP_n \tag{3}$$

where *TP* stands for true positive and represents the number of correctly predicted wheat spike or spikelet samples; *FP* stands for false positive and represents the number of non-diseased samples incorrectly identified as diseased; and *FN* stands for false negative and represents the number of diseased samples mistakenly classified as non-diseased. *N* and *n* represent the numbers of samples and classes, respectively. *AP* means the average precision for a specific category, which was calculated using *P* and *R*. The mAP represents the average prediction performance across all categories, with a higher mAP indicating better overall model performance. Intersection over union (IoU) is defined as the ratio of the intersection area to the union area of the predicted and labeled bounding boxes. The mAP50 was calculated at IoU threshold 0.50. The mAP50:95 refers to the average precision when the IoU threshold ranges from 0.50 to 0.95 at intervals of 0.05. In addition, FLOPs was applied to assess the computational complexity of the model (Tan and Le, 2019).

Gradient-weighted class activation mapping (Grad-CAM) can provide interpretability for deep learning models in image classification and object detection tasks. Herein, Grad-CAM was used to analyze the feature extraction in both the original YOLOv8s and the designed detection models, visualizing model decision-making processes through feature-expression heatmaps. High-precision detection models typically generate feature heatmaps that cover entire objects. Thus, heatmap coverage acts as an effective indicator for evaluating the feature extraction capability and detection accuracy of the model.

2.6.2 Evaluation of wheat FHB severity

As specified in Chinese national standard GB/T15796-2011, three indices were employed to evaluate wheat FHB severity, including the rate of diseased spikes (RD_S), the FHB severity level of individual wheat

spikes (SL_S), and the FHB disease index for wheat (DI_W). These indices were calculated by Equations 4-6:

$$\mathrm{RD\_S} = \frac{AD}{AH} \times 100\% \tag{4}$$

$$\mathrm{SL\_S} = \frac{AX}{AY} \tag{5}$$

$$\mathrm{DI\_W} = \frac{\sum (h_i \times i)}{H \times 4} \times 100\% \tag{6}$$

where *AD* represents the number of diseased spikes; *AH* represents the total number of spikes; *AX* represents the number of diseased spikelets per spike; *AY* represents the total number of spikelets in the investigated spike; h_i represents the number of diseased spikes corresponding to each FHB severity level; *i* is the value of each severity level; and *H* denotes the total number of investigated spikes. The severity levels or values of *i* are determined by the SL_S indices. As defined in Table 2, these indices are categorized into 5 levels based on the number of diseased spikelets per spike. DI_W is a composite index used to assess the prevalence and FHB severity, representing the average level of disease incidence.

Table 2 Criteria for FHB severity levels

Severity level (*i*)	0	1	2	3	4
SL_S	0	0-1/4	1/4-1/2	1/2-3/4	3/4-1

To evaluate the performance of the proposed evaluation method, the statistics R^2, root mean square error (RMSE), and mean absolute error (MAE), as described in the following equations, were used:

$$R^2 = 1 - \frac{\sum_{i=1}^{Nd} (y_i - x_i)^2}{\sum_{i=1}^{Nd} (y_i - \bar{y}_i)^2} \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{Nd} \sum_{i=1}^{Nd} (y_i - x_i)^2} \tag{8}$$

$$\mathrm{MAE} = \frac{1}{Nd} \sum_{i=1}^{Nd} (y_i - x_i)^2 \tag{9}$$

where *Nd* represents the number of tested images; y_i represents the manually evaluated FHB level; x_i represents the level evaluated by the proposed method; and $\bar{y}_i$ is the mean value of y_i.

3 Results and discussion

3.1 Results for wheat spike detection

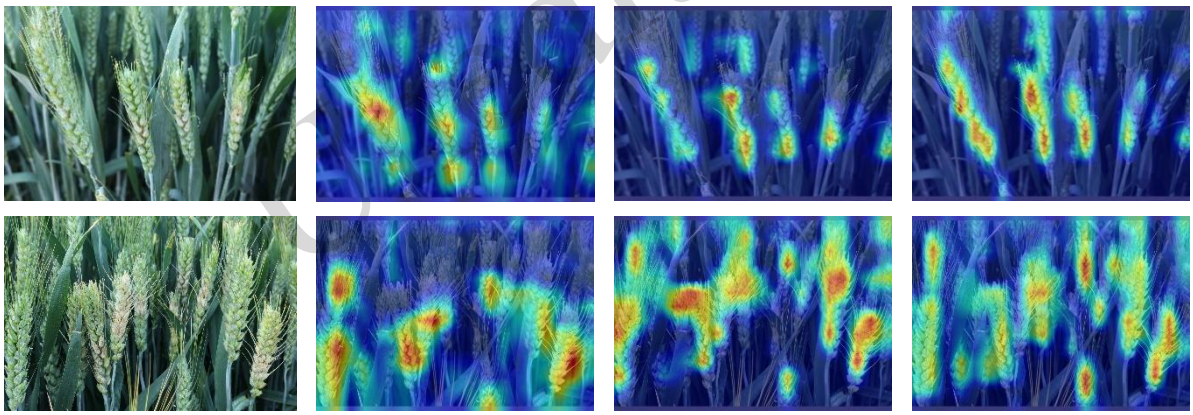
A lightweight deep learning model that integrates FasterNet, EMA, and AFPN modules was developed for the detection of complete and unobstructed spikes in wheat images. Ablation experiments were conducted to investigate the lightweighting effects of FasterNet and EMA modules on the model, with the results presented in Table 3. Compared with the original YOLOv8s model, the parameters and FLOPs of model C were reduced from 11.2 M and 28.6 G to 6.1 M and 16.8 G, respectively. This demonstrates that FasterNet and EMA can significantly decrease the model complexity while improving model's lightweight performance. While model C achieved a lower mAP50:95 than model A, its mAP50 showed a slight improvement. Furthermore, the AFPN feature fusion method was integrated into model D, leading to its establishment as the final wheat spike detection model in this study. The performance of model D was improved while maintaining its lightweightness

with the parameters of 6.8 M. Both the mAP50 and mAP50:95 of model D were higher than those of the original model A, reaching 0.964 and 0.822, respectively. These results indicate that the FasterNet, EMA, and AFPN modules can effectively reduce computational burden while facilitating high-precision wheat spike detection.

Table 3 Results of ablation experiments for wheat spike detection.

Model number	Models	Parameters (M)	FLOPs (G)	mAP50	mAP50:95
A	YOLOv8s	11.2	28.6	0.950	0.820
B	A+FasterNet	6.1	16.2	0.954	0.804
C	B+EMA	6.1	16.8	0.955	0.809
D	C+AFPN	6.8	22.1	0.964	0.822

Two color images with different spike densities were processed using both the original YOLOv8s and the designed wheat spike models, and the corresponding feature heatmaps are displayed in Fig. 7. The feature heatmaps generated by the designed model showed more precise coverage of wheat spike area and clearer boundary delineation than those of YOLOv8s. These results demonstrate that combining the EMA and AFPN modules can improve the feature extraction capability of the designed model, therefore enabling more accurate wheat spike detection.



(a) (b) (c) (d)

Fig. 7 Visualizations of spike features extracted by different detection models. (a) Color images of wheat spikes; (b) Grad-CAM visualizations for the YOLOv8s model; (c) Grad-CAM visualizations for the YOLOv8s-FasterNet model; (d) Grad-CAM visualizations for the designed model. Red regions mean they have high scores for class, and blue means evidence for class.

As shown in Figure 8, the designed model outperforms the YOLOv8s model in wheat spike detection. Many spikes are either small or have indistinct color features at the boundaries of adjacent spikes, such as the regions marked by yellow rectangles in Fig. 8, and both of these issues pose challenges for the YOLOv8s model. Moreover, the designed model can accurately detect the complete wheat spikes while excluding those that are only partially visible.

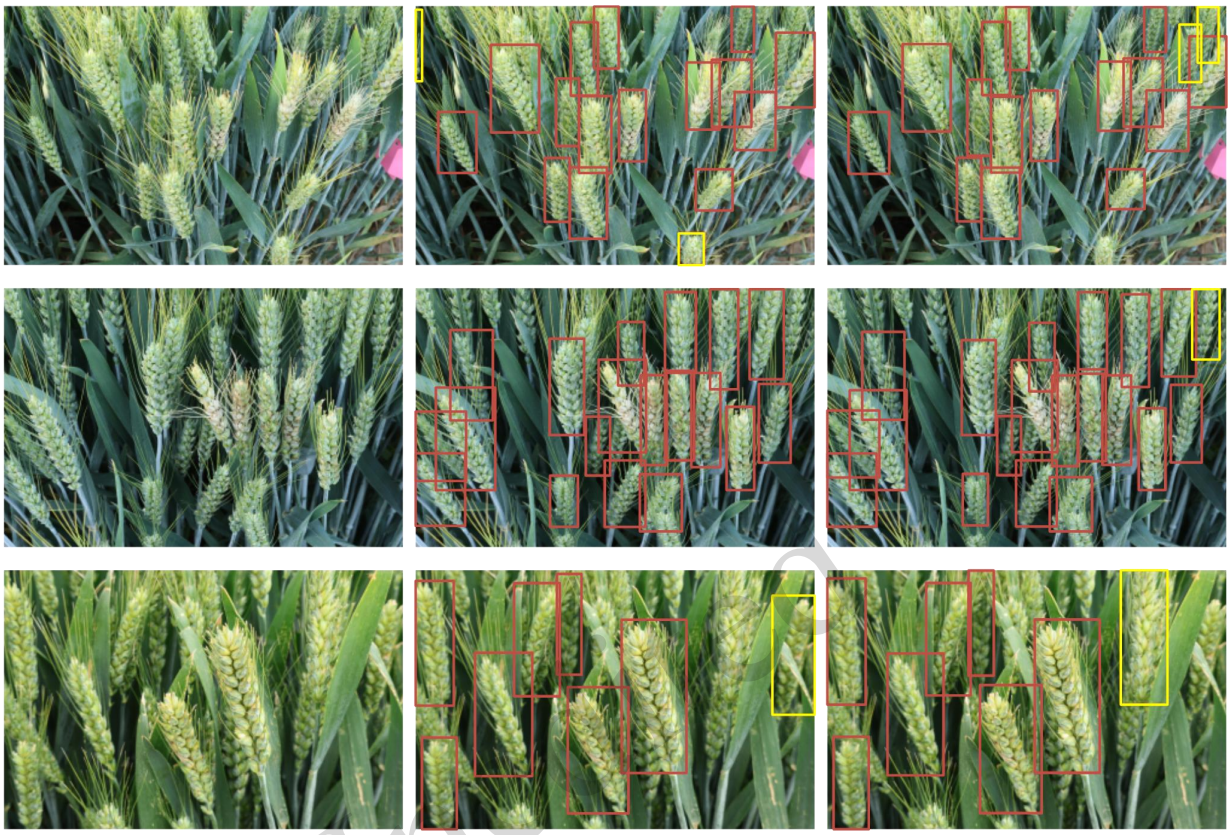


Fig. 8 Detection results of wheat spikes by the YOLOv8s and designed models. Detected spikes were marked with red and yellow rectangles. The yellow rectangles represent the detection differences between the YOLOv8s and designed models.

3.2 Results for wheat spikelet detection

To address the challenges of detecting low-resolution and small-sized wheat spikelets, a spikelet detection model was developed by integrating the MobileViT backbone network, small object detection module SPD-Conv, and the BiFormer module. Ablation experiments were conducted to evaluate the individual contributions of these modules, with the quantitative results presented in Table 4.

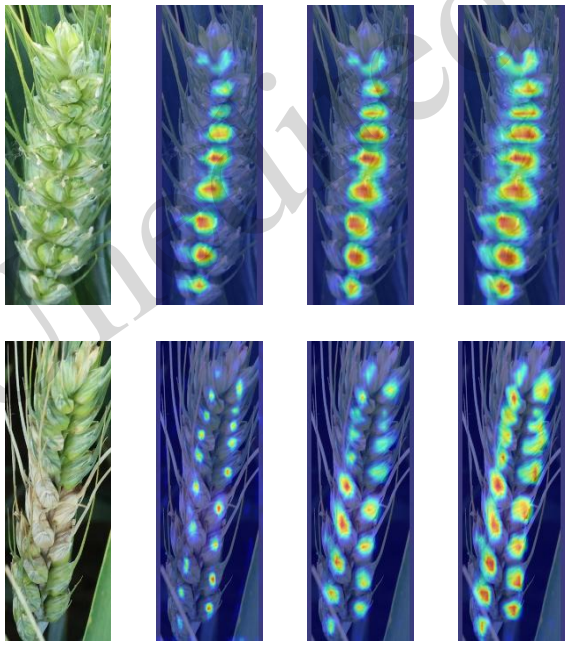
Table 4 Results of ablation experiments for wheat spikelet detection.

Model number	Models	Parameters (M)	FLOPs (G)	mAP50	mAP50:95
A	YOLOv8s	11.2	28.6	0.917	0.688
B	A+MobileViT	1.2	5.3	0.924	0.699
C	B+SPD-Conv	1.2	2.0	0.933	0.700
D	C+BiFormer	1.2	2.0	0.936	0.712

The implementation of MobileViT substantially enhanced the model's lightweightness compared to YOLOv8s, with parameters reduced from 11.2 M to 1.2 M and FLOPs from 28.6 G to 5.3 G. Notably, despite these reductions in parameters and FLOPs, model B achieved slightly higher detection precision than model A. Moreover, the small object detection module SPD-Conv was added to model B, which not only further

enhanced model's lightweight performance but also reduced FLOPs to 2.0 G. In terms of detection performance, the obtained lightweight model C outperformed model B. Finally, the detection precision of the designed model was further improved by adding the attention module BiFormer to model C. In wheat spikelet detection, both the mAP50 and mAP50:95 of model D were higher than those of the original YOLOv8s model, reaching 0.936 and 0.712, respectively. These results indicate that the designed model has high performance in wheat spikelet detection while maintaining low computational complexity.

The integration of the BiFormer attention module into both the spikelet detection and FHB detection models resulted in improved global feature extraction and enhanced detection accuracy for small spikelets. Grad-CAM was also used to assess the characteristics of models A, C and D. The feature-expression heatmaps of these three models for the frontal and lateral sides of spikelets are presented in Fig. 9. The heatmaps generated by model D exhibited broader coverage than those of other models across both the frontal and lateral sides of spikelet. Moreover, the activation regions of model D's heatmaps showed more precise concentration at spikelet centers than those of YOLOv8s. These results confirmed that the BiFormer and SPD-Conv modules can significantly improve the feature extraction capability and spikelet detection accuracy of the model.



(a) (b) (c) (d)

Fig. 9 Visualizations of spikelet features extracted by different detection models. (a) Color images of wheat spikes that contain the frontal and lateral sides of spikelets; (b) Grad-CAM visualizations for the YOLOv8s model; (c) Grad-CAM visualizations for the YOLOv8s-MobileViT-SPD-Conv model; (d) Grad-CAM visualizations for the designed model. Red regions mean that they have high scores for class, and blue means evidence for class.

The comparisons between YOLOv8s and the designed model in wheat spikelet and FHB detections are shown in Fig. 10. In the frontal view of a wheat spike, spikelets and their florets are visible. However, the mutual overlap of florets poses challenges for the YOLOv8s model in spikelet detection, as displayed in Fig. 10b. In the lateral view of a wheat spike, spikelets can be detected by recognizing their glumes. Typically, FHB progression commences at the point of infection and spreads bidirectionally along the spike. FHB pathogens damage spike tissue, inducing the loss of chlorophyll and water, as well as triggering color changes, which often obscure the boundaries between the infected and healthy areas. As shown in Figs. 10c and 10f, the designed model exhibits better performance than YOLOv8s in feature extraction and spikelet detection accuracy.

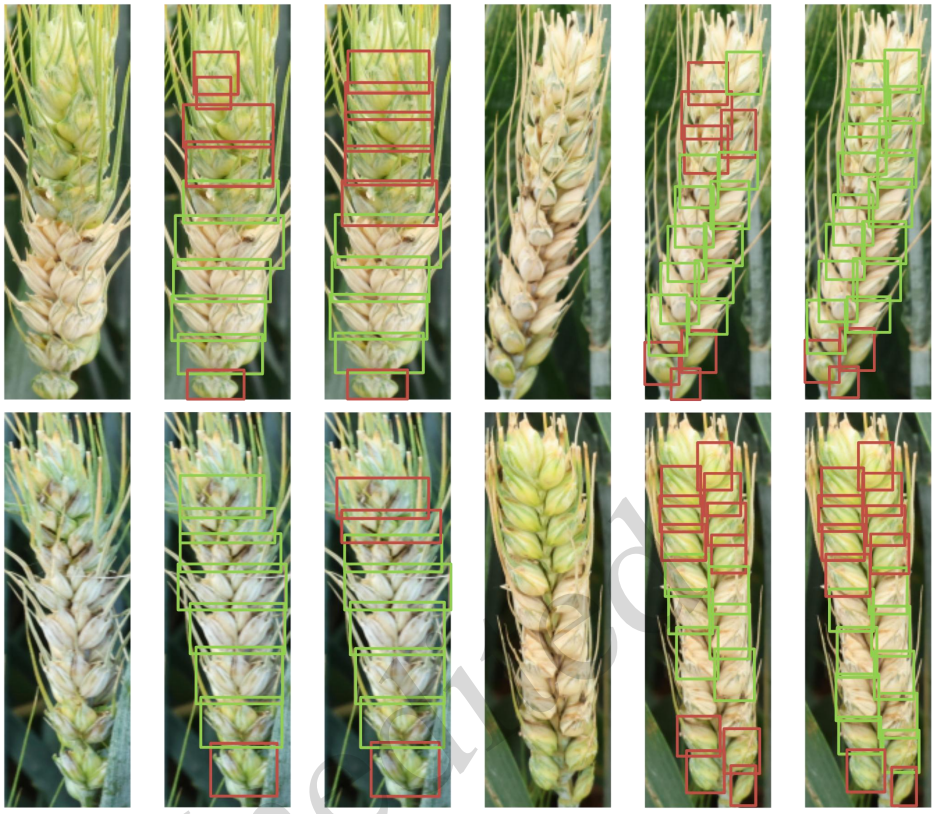


Fig. 10 Detection results of wheat spikelets by the YOLOv8s and designed models. (a) Color images of wheat spikes that contain the frontal side of spikelets; (d) Color images of wheat spikes that contain the lateral side of spikelets; (b, e) Detection results by the YOLOv8s model; (c, f) Detection results by the designed model. Healthy and infected spikelets were labeled by red and green rectangles, respectively.

3.3 Results for wheat FHB evaluation

Traditionally, wheat FHB severity levels are evaluated based on the conditions of wheat spikes and spikelets. The designed models facilitate quantitative FHB evaluation via automated spike and spikelet analysis. Eighty wheat images, collected from various plots and not used in model training, were used to assess the performance of the proposed models and methods. The wheat FHB levels in these images were firstly determined through visual observation to establish ground-truth RD_S and DI_W indices. These wheat images were then processed using the designed models to evaluate the FHB severity levels and output the corresponding diseased indices. For comparison purposes, the YOLOv8s model was also applied to these wheat images. The results obtained from both the designed and YOLOv8s models are summarized in Table 5. For the RD_S index, the R^2 values, RMSEs and MAEs between the manual and automatic evaluations were 0.55, 9.82% and 6.29% for the YOLOv8s model, and 0.71, 7.08% and 4.86% for the designed model, respectively. Regarding the DI_W index, the R^2 values, RMSEs and MAEs between the manual and automatic evaluations were 0.85, 3.44% and 2.51% for the YOLOv8s model, and 0.93, 2.28% and 1.74% for the designed model, respectively. For the evaluation of both the RD_S and DI_W indices in wheat FHB, the designed models generally outperformed the original model.

Table 5 Performance of the YOLOv8s and designed models on FHB evaluation.

Disease indices	Models	R^2	RMSE (%)	MAE (%)
RD_S	YOLOv8s	0.55	9.82	6.29
	Designed models	0.71	7.08	4.86
DI_W	YOLOv8s	0.85	3.44	2.51
	Designed models	0.93	2.28	1.74

Fig. 11 shows the improvements in the evaluation performance of the designed models. The estimated values of the designed models were all closer to the 1:1 lines than those of the original model, indicating that the proposed FHB evaluation method can more accurately assess the disease severity levels.

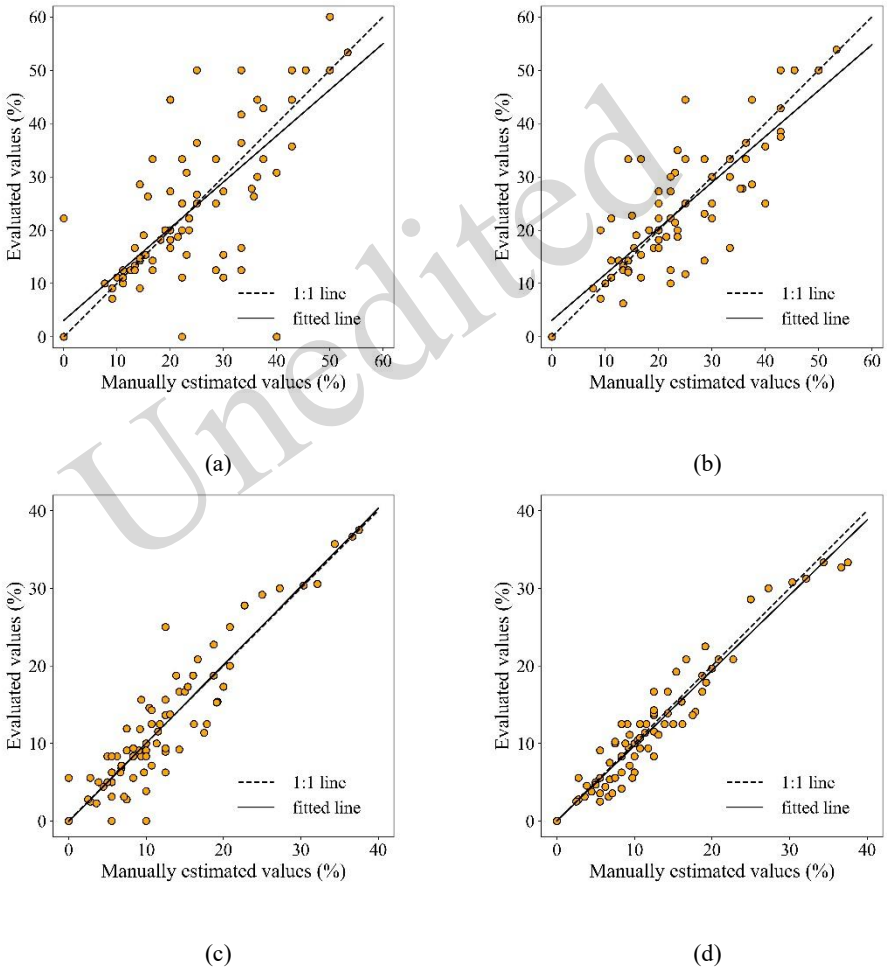


Fig. 11 Regression between manually estimated and evaluated FHB levels. (a) Regression for RD_S evaluated by the YOLOv8s model. (b) Regression for RD_S evaluated by the designed models. (c) Regression for DI_W evaluated by the YOLOv8s model. (d) Regression for DI_W evaluated by the designed models. Solid lines are regression lines using the least squares linear method, while dotted lines are 1:1 lines.

3.4 Discussion

Color imaging and deep learning techniques were used to automatically detect wheat FHB and quantitatively evaluate its severity. Leveraging currently popular object detection models, lightweight models

for spike, spikelet and FHB detections were designed by integrating lightweight backbones, attention modules and convolution modules. This approach reduced the number of model parameters and FLOPs while maintaining comparable performance.

For wheat FHB detection and evaluation, the proposed method consisted of three steps. First, complete spikes in the color images were detected and extracted. Second, spikelets within the extracted spikes were detected and classified as healthy or diseased. Finally, the RD_S and DI_W indices were calculated to evaluate the severity levels of wheat FHB. In contrast to the instance segmentation models employed by Su et al. (2021), our study designed object detection models to reduce computational complexity. The features of wheat spikelets between the frontal and lateral sides of spikes are different, particularly when observed from the frontal side. Additionally, the diseased areas of spikelets are typically located on their frontal sides. Our model can detect both the healthy and diseased spikelets from the frontal and lateral views of spikes, an aspect not explored by other studies.

Many studies detect the spike diseased areas based on their color features and evaluated the FHB severity levels based on the percentage of diseased pixels. However, this approach fails to consider the gaps between the spikelets in images. In addition, not all florets within a spikelet may be infected, and some disease areas are too small to be detected. The shapes and postures of wheat spikes can also cause errors in pixel-based evaluation methods. All these factors lead to an underestimation of FHB infection. In this study, quantitative evaluation methods for wheat FHB were developed based on the number of diseased spikelets per spike, in line with the evaluation criteria of crop protection and plant sciences. A model was designed to first detect spikes from wheat images, and another model was developed to further process spike images. Then, spikelets were detected and classified to evaluate FHB severity levels. The design of these two models can clearly identify the detection range of wheat spikes, reduce the computational complexity, and enhance spikelet detection. The sizes and FLOPs of these two models were significantly smaller than those of the original YOLOv8s model: the spikelet detection model has a size of merely 1.2 M. The performance of our model in spike detection is better than that of Zhang's model (Zhang et al., 2023); the mAP50 achieved 0.964. Therefore, the detection errors of wheat spike detection have a negligible impact on the overall evaluation of FHB severity. The test results indicate that the designed models outperform the original YOLOv8s model in FHB severity evaluation. Furthermore, the proposed evaluation methods achieved an accuracy of 0.93, which is higher than the 0.91 obtained by Zhang's method using an improved YOLOv5 model (Zhang et al., 2022). Our method has a similar performance compared to Wang's method, while our model's parameter sizes are much smaller (Wang et al., 2023).

To ensure the accuracy of FHB evaluation, only complete spikes in the images were detected, while some infected spikes remained undetected due to the occlusions caused by other spikes or leaves. This caused inaccuracies in evaluating FHB severity, especially in calculating the RD_i, with the R^2 value achieving 0.71. Although the designed models can detect the spikelets on both the frontal and lateral sides of spikes, some spikes in images display both frontal and lateral sides owing to their angles, leaving certain spikelets undetectable by the models. In future work, more attention needs to be paid to handling the error problems caused by infected but incomplete spikes, thereby further improving the precision of FHB evaluation. Moreover, the designed lightweight models for wheat FHB detection and evaluation can be migrated and deployed on embedded systems or mobile device platforms, in order to promote the development and application of portable detectors and monitors in the field. The timely detection and evaluation of wheat FHB can significantly contribute to assessing the resistance of wheat varieties and advancing the precision management of diseases in the field.

4 Conclusions

In this study, lightweight CNN models were designed to explore the potential of detecting and evaluating FHB using color images. Two lightweight models were designed for wheat spike and spikelet detection,

respectively, and FHB evaluation. These models achieved mAP50 values of 0.964 and 0.936, while their parameters were significantly reduced by 4.4 M and 10.0 M, respectively. A total of 80 images were tested, and the RD_S and DI_W indices were computed. The R^2 values between the manually evaluated and estimated wheat FHB levels were 0.71 for RD_S and 0.93 for DI_W, respectively. The results demonstrate that wheat FHB can be detected and evaluated using color images. The method proposed in this study can efficiently detect spikes and quantitatively evaluate FHB levels, thereby contributing to the monitoring of wheat FHB. In future research, the models will be transferred and deployed on edge and mobile devices to facilitate the practical application of the research findings.

Data availability statement

All data generated or analyzed during this study can be made available by the corresponding author upon reasonable request.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (Grant No. 2022YFD200150100-2022YFD200150101) and the 2115 Talent Development Program of China Agricultural University.

Author contributions

Wang ZHANG, Yi REN, Zidi GUO performed the experimental research and data analysis, wrote the manuscript. Han LI, Man ZHANG contributed to the funding acquisition. Jie LIU, Ruicheng QIU performed the study design, reviewed the manuscript. All authors read and approved the final manuscript and, therefore, had full access to all the data in the study and take responsibility for the integrity and security of the data.

Compliance with ethics guidelines

Wang ZHANG, Yi REN, Zidi Guo, Han LI, Man ZHANG, Jie LIU and Ruicheng QIU declare that they have no conflict of interest.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Abdu AM, Mokji MM, Sheikh UU, 2020. Automatic vegetable disease identification approach using individual lesion features. *Comput Electron Agr*, 176:105660. http://doi.org/10.1016/j.compag.2020.105660

Agarwal M, Gupta SK, Biswas KK, 2020. Development of efficient CNN model for Tomato crop disease identification. *Sustain Comput Infor*, 28:100407. http://doi.org/10.1016/j.suscom.2020.100407

Ahlgren P, Kekalainen J, 2007. Indexing strategies for Swedish full text retrieval under different user scenarios. *Inform Process Manag*, 43(1):81-102. http://doi.org/10.1016/j.ipm.2006.03.003

Almoujahed MB, Rangarajan AK, Whetton RL, et al., 2022. Detection of fusarium head blight in wheat under field conditions using a hyperspectral camera and machine learning. *Comput Electron Agr*, 203:107456. https://doi.org/10.1016/j.compag.2022.107456

Basavaiah J, Anthony AA, 2020. Tomato leaf disease classification using multiple feature extraction techniques. *Wireless Pers Commun*, 115:633-651. http://doi.org/10.1007/s11277-020-07590-x

Chen J, Kao S, He H, et al., 2023. Run, Don't Walk: Chasing higher FLOPS for faster neural networks. In: 2023 IEEE Conference On Computer Vision and Pattern Recognition, Vancouver, p.12021-12031. http://doi.org/10.1109/CVPR52729.2023.01157

Chen J, Zhang D, Zeb A, et al., 2021. Identification of rice plant diseases using lightweight attention networks. *Expert Syst Appl*, 169(1):114514. http://doi.org/10.1016/j.eswa.2020.114514

Chen Y, Wang X, Zhang X, et al., 2023. Spectral quantitative analysis and research of fusarium head blight infection degree in wheat canopy visible areas. *Agronomy*, 13(3):933. http://doi.org/10.3390/agronomy13030933

Chen Z, Zhang Y, Wang Y, et al., 2023. Robust image inpainting forensics by using an attention-based feature pyramid network. *Appl Sci*, 13(16):9196.

http://doi.org/10.3390/app13169196

Fang T, Chen P, Zhang J, et al., 2020. Crop leaf disease grade identification based on an improved convolutional neural network. *J Electron Imaging*, 29(1):013004.
http://doi.org/10.1117/1.JEI.29.1.013004

Ferentinos KP, 2018. Deep learning models for plant disease detection and diagnosis. *Comput Electron Agr*, 145:311-318.
http://doi.org/10.1016/j.compag.2018.01.009

Figueroa M, Hammond-Kosack KE, Solomon PS, 2018. A review of wheat diseases - a field perspective. *Mol Plant Pathol*, 19(6):1523-1536.
http://doi.org/10.1111/mpp.12618

Gao C, Guo W, Yang C, et al., 2024. A fast and lightweight detection model for wheat fusarium head blight spikes in natural environments. *Comput Electron Agr*, 216:108484.
http://doi.org/10.1016/j.compag.2023.108484

Hellemans T, Landschoot S, Dewitte K, et al., 2018. Impact of crop husbandry practices and environmental conditions on wheat composition and quality: A review. *J Agr Food Chem*, 66(11):2491-2509.
http://doi.org/10.1021/acs.jafc.7b05450

Hong Q, Jiang L, Zhang Z, et al., 2022. A lightweight model for wheat ear fusarium head blight detection based on RGB images. *Remote Sens*, 14(14):3481.
http://doi.org/10.3390/rs14143481

Howard A, Sandler M, Chu G, et al., 2019. Searching for MobileNetV3. In: 2019 IEEE/CVF International Conference on Computer Vision, Seoul, p.1314-1324.

Hu J, Xie L, Gu X, et al., 2022. Information-interaction feature pyramid networks for object detection. In: 2022 IEEE International Conference On Tools with Artificial Intelligence, Aomen, p.1301-1306.
http://doi.org/10.1109/ICTAI56018.2022.00197

Khan MS, Uandai SB, Srinivasan H, 2019. Anthracnose disease diagnosis by image processing, support vector machine and correlation with pigments. *J Plant Pathol*, 101:749-751.
http://doi.org/10.1007/s42161-019-00268-9

Liu G, Reda FA, Shih KJ, et al., 2018. Image inpainting for irregular holes using partial convolutions. In: 2018 European Conference on Computer Vision, Munich, p.89-105.
http://doi.org/10.1007/978-3-030-01252-6_6

Liu S, Huang D, Wang Y, 2019. Learning spatial fusion for single-shot object detection. arXiv:1911.09516.
https://doi.org/10.48550/arXiv.1911.09516

Liu T, Zhao Y, Sun Y, et al., 2024. High-throughput identification of fusarium head blight resistance in wheat varieties using field robot-assisted imaging and deep learning techniques. *J Clean Prod*, 480:144024.
https://doi.org/10.1016/j.jclepro.2024.144024

Mao R, Wang Z, Li F, et al., 2023. GSEYOLOX-s: An improved lightweight network for identifying the severity of wheat fusarium head blight. *Agronomy*, 13(1):242.
http://doi.org/10.3390/agronomy13010242

Mehta S, Rastegari M, 2021. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. arXiv:2110.02178.
https://doi.org/10.48550/arXiv.2110.02178

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

Ouyang D, He S, Zhang G, et al., 2023. Efficient multi-scale attention module with cross-spatial learning. In: 2023 IEEE International Conference on Acoustics, Speech and Signal Processing, Rhodes Island, p.1-5.
https://doi.org/10.1109/ICASSP49357.2023.10096516

Pantazi XE, Moshou D, Tamouridou AA, 2019. Automated leaf disease detection in different crop species through image features analysis and one class classifiers. *Comput Electron Agr*, 156:96-104.
http://doi.org/10.1016/j.compag.2018.11.005

Park J, Kim J, Kim C, 2023. Biformer : learning bilateral motion estimation via bilateral transformer for 4K video frame interpolation. In: 2023 IEEE Conference on Computer Vision and Pattern Recognition, Vancouver, p.1568-1577.
http://doi.org/10.1109/CVPR52729.2023.00157

Qiu R, Yang C, Moghimi A, et al., 2019. Detection of Fusarium head blight in wheat using a deep neural network and color imaging. *Remote Sens*, 11:2658.
http://doi.org/10.3390/rs222658

Redmon J, Divvala S, Girshick R, et al., 2016. You Only Look Once: Unified, real-time object detection. In: 2016 IEEE Conference On Computer Vision and Pattern Recognition, Las Vegas, p.779-788.

http://doi.org/10.1109/CVPR.2016.91

Rößle D, Prey L, Ramgraber L, et al., 2023. Efficient noninvasive FHB estimation using RGB images from a novel multiyear, multirater dataset. *Plant Phenomics*, 5:0068.
https://doi.org/10.34133/plantphenomics.0068

Saccon FM, Parcey D, Paliwal J, et al., 2017. Assessment of fusarium and deoxynivalenol using optical methods. *Food Bioprocess Tech*, 10:34-50.
http://doi.org/10.1007/s11947-016-1788-9.

Sampathkumar S, Rajeswari R, 2022. An automated crop and plant disease detection scheme using cognitive fuzzy C-Means algorithm. *IETE J Res*, 68(5):3786-3797.
http://doi.org/10.1080/03772063.2020.1780163.

Su W, Zhang J, Yang C, et al., 2021. Automatic evaluation of wheat resistance to fusarium head blight using dual mask-rcnn deep learning frameworks in computer vision. *Remote Sens*, 13:26.
http://doi.org/10.3390/rs13010026.

Sunkara R, Luo T, 2022. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects. arXiv: 2208.03641.
https://doi.org/10.48550/arXiv.2208.03641.

Tan M, Le QV, 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. In: 2019 International Conference on Machine Learning, Long Beach, p.6105-6114.

Vincke D, Eylenbosch D, Jacquemin G, et al., 2023. Near infrared hyperspectral imaging method to assess Fusarium Head Blight infection on winter wheat ears. *Microchemical J*, 191:108812.
https://doi.org/10.1016/j.microc.2023.108812.

Wang G, Chen M, Lin YC, et al., 2024. Efficient multi-branch dynamic fusion network for super-resolution of industrial component image. *Displays*, 82:102633.
http://doi.org/10.1016/j.displa.2023.102633

Wang Y, Li J, Su W, 2023. An integrated multi-model fusion system for automatically diagnosing the severity of wheat Fusarium Head Blight. *Agriculture*, 13:1381.
https://doi.org/10.3390/agriculture13071381

Zhang D, Luo H, Wang D, et al., 2022. Assessment of the levels of damage caused by Fusarium head blight in wheat using an improved YoloV5 method. *Comput Electron Agr*, 198:107086.
https://doi.org/10.1016/j.compag.2022.107086

Zhang D, Luo H, Cheng T, et al., 2023. Enhancing wheat Fusarium head blight detection using rotation Yolo wheat detection network and simple spatial attention network. *Comput Electron Agr*, 211:107968.
https://doi.org/10.1016/j.compag.2023.107968

Zhao Y, Liu L, Xie C, et al., 2020. An effective automatic system deployed in agricultural Internet of Things using multi-context fusion network towards crop disease recognition in the wild. *App Soft Comput*, 89:106128.
http://doi.org/10.1016/j.asoc.2020.106128

Zheng L, Shen L, Tian L, et al., 2015. Scalable person re-identification: a benchmark. In: 2015 IEEE International Conference On Computer Vision, Santiago, p.1116-1124.
http://doi.org/10.1109/ICCV.2015.133

Zhu L, Wang X, Ke Z, et al., 2023. Biformer: vision transformer with bi-level routing attention. In: 2023 IEEE Conference On Computer Vision and Pattern Recognition, Vancouver, p.10323-10333.
http://doi.org/10.1109/CVPR52729.2023.00995