# Detection of fusarium head blight using a YOLOv5s-based method improved by attention mechanism

Lei Shi[1,2], Chengkai Yang[1], Xiaoyun Sun[1], Jiayue Sun[1], Ping Dong[1], Shufeng Xiong[1], Jian Wang[1*]

(1. *College of Information and Management Science, Henan Agricultural University, Zhengzhou 450046, China*;
2. *Henan Grain Crop Collaborative Innovation Center, Zhengzhou 450046, China*)

**Abstract:** Fusarium head blight (FHB) is one of the most destructive diseases in global wheat production. In order to count the FHB-infected wheat ears under field conditions, this study proposed an algorithm for diseased wheat ear detection based on improved YOLOv5s (Tr-YOLOv5s). The Swin Transformer was used to replace the CSPDarknet backbone network to enhance the extraction of characteristic information of the population wheat ears of FHB in the field background. The convolutional block attention module (CBAM) attention mechanism was added to improve the detection effect of target wheat ears, subsequently improving the overall accuracy of the model. The original loss function complete intersection over union (CIoU) was replaced by Scylla intersection over union (SIoU) loss to accelerate the model convergence and decrease the loss value. The results showed that the mean average precision (mAP) of the Tr-YOLOv5s model reached 90.64%, making a 4.63% improvement compared to the original YOLOv5s model. The improved model could quickly detect and count wheat FHB ear in the field environment, which laid a foundation for the subsequent automatic disease identification and grading of wheat FHB under field conditions.

**Keywords:** fusarium head blight, YOLOv5s, attention mechanism, Swin Transformer, loss function
**DOI:** 10.25165/j.ijabe.20241705.8425

**Citation:** Shi L, Yang C K, Sun X Y, Sun J Y, Dong P, Xiong S F, et al. Detection of fusarium head blight using a YOLOv5s-based method improved by attention mechanism. Int J Agric & Biol Eng, 2024; 17(5): 247–254.

## 1 Introduction

Wheat is one of the major grain crops in China, and the yield and quality of wheat have always been major concerns for those involved[1,2]. Therefore, accurate and timely identification of wheat fusarium head blight (FHB) can provide important guarantees to prevent wheat FHB and improve wheat yield.

Since the mid-20th century, traditional machine-learning algorithms have gradually been applied in agriculture[3]. Early methods for intelligent disease recognition are based on machine learning, where image preprocessing and feature extraction are used to select specified features, and at the same time, machine learning algorithms are used to train the classification ability of the model for feature vectors to achieve disease recognition. Chaudhary et al.[4] proposed an ensemble particle swarm optimization algorithm that achieved 96% accuracy in identifying vegetable diseases. Abdu et al.[5] used a method to extract features from necrotic lesions, achieving precision and recall of over 99% in plant disease classification. Devi et al.[6] employed a method, H2K, to automatically identify and classify five peanut leaf diseases with an accuracy of 97.67%.

Using machine learning algorithms for disease identification requires manual selection of feature variables and has stricter requirements (image background, lighting conditions, leaf placement, etc.), which is more challenging to promote and apply. In contrast, the deep learning approach only requires the provision of labeled datasets and can extract features from datasets without human-designed features[7-9]. In the case of identifying wheat ears in a field environment, where the size of a single wheat ear varies and the background is complex and diverse, the target detection technique can directly learn features by labeling disease areas as examples, and output the location and category of each wheat ear area in the image at the same time, which is suitable for recognizing wheat ear diseases in complex scenarios. YOLO[10], as a common target detection framework, can identify diseases quickly. Agarwal et al.[11] verified the feasibility of a lightweight CNN model that achieved 98.4% recognition accuracy on the public dataset of PlantVillage. Fang et al.[12] introduced a CNN algorithm to classify 10 crop diseases in terms of disease leaf classes, and the recognition accuracy reached 95.61%. Qi et al.[13] constructed a SE-YOLOv5 algorithm to identify tomato virus disease with a mean average precision of 94.10%. Zhang et al.[14] proposed an FHB detection method based on the improved YOLOv5 and the random forest algorithm. The experimental results showed that the method could effectively detect the severity of FHB in complex field conditions. The above studies have employed different optimization strategies to improve model accuracy. However, building a real-time and accurate FHB recognition model remains a significant challenge.

In order to count the FHB-infected wheat ears accurately under field conditions, this study proposed an FHB detection model based on the improved YOLOv5s for identifying and counting healthy and diseased wheat ears in one image. The backbone network

**Received date:** 2023-07-16 **Accepted date:** 2024-03-13
**Biographies: Lei Shi**, PhD, Professor, research interest: machine learning, data mining, and their applications in agricultural fields, Email: shilei@henau.edu.cn; **Chengkai Yang**, MS, research interest: deep learning in agriculture. Email: yck03185@stu.henau.edu.cn; **Xiaoyun Sun**, PhD, research interest: signal processing and quantum computation, Email: xysun@henau.edu.cn; **Jiayue Sun**, MS, research interest: smart agriculture, Email: sjyue@stu.henau.edu.cn; **Ping Dong**, PhD, research interest: agricultural Internet of Things, Email: dongping@henau.edu.cn; **Shufeng Xiong**, PhD, Associate Professor, research interest: natural language processing and deep learing method for text mining, Email: xsf@whu.edu.cn.

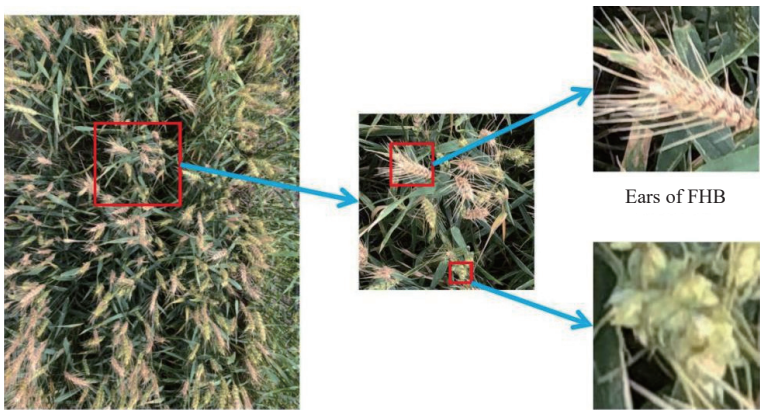
***Corresponding author: Jian Wang**, PhD, Associate Professor, research interest: remote sensing image classification and vegetation parameter inversion. College of Information and Management Science, Henan Agricultural University, Zhengzhou, China. Tel: +86-371-56990030, Email: leonw63@mail.bnu.edu.cn.

CSPDarknet was replaced by Swin Transformer, and the convolutional block attention module (CBAM) attention mechanism and the Scylla intersection over union (SIoU) loss function were incorporated into the YOLOv5s model to obtain the Tr-YOLOv5s model, improving the model detection accuracy. The proposed model aimed to improve detection accuracy effectively in the field environment and provide a reference for field detection of wheat FHB.

## 2 Materials and methods

### 2.1 Dataset construction

The field FHB trial was conducted in the Rocky Ford FHB nursery, Kansas State University in the 2021-2022 wheat growing season. Four winter wheat varieties "Clark", "Jagger", "Overley" and "Everest" were used as the plant materials. About 30 seeds per line were sown in a 1 m long single-row plot using a randomized complete block design. In this study, wheat ear images of FHB were collected in June 2022, the image acquisition equipment is a high-pixel smartphone, the shooting process is taken from above, the distance between the camera and the wheat ear is 50 cm, the image resolution is 3024×4032 pixels, a total of 500 images. The images were labeled by LabelImg graphical annotation tool, and the corresponding XML files were generated. Crop an image with 832×832 pixels resolution from each image, the train set, validation set, and test set are randomly divided according to the ratio of 8:1:1. The image dataset is shown in Figure 1.




Note: FHB: Fusarium head blight.

Figure 1 Healthy wheat ears and ears of FHB

### 2.2 Tr-YOLOv5s Model Structure

The YOLOv5 model has high detection accuracy and fast detection speed, which is one of the best choices for target detection[13]. Depending on the network depth and feature map width, they can be classified as YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. In this study, the YOLOv5s model with the smallest storage footprint is selected for improvement, which is mainly divided into the following three aspects:

1) Using Swin Transformer to replace YOLOv5s backbone network CSPDarknet[15]. The Multi-head self-attention mechanism based on the shifted windows in the Swin Transformer can effectively capture global context information and has better feature extraction capabilities.

2) The CBAM attention mechanism was added after the three feature layers were extracted by the backbone network. The CBAM attention mechanism makes the model pay more attention to the characteristics of the ears of FHB by adjusting the weights of the feature map on the channel and spatial at the same time, to improve the recognition effect.

3) The original loss function complete intersection over union (CIoU)[16] was replaced by the SIoU[17] function, which takes into account the vector angle between the desired regressions and redefines the penalty indicator to make the model converge faster and reduce the model loss.

The data input part used Mosaic data enhancement to process the data and improve the detection accuracy of small targets. Tr-YOLOv5s model is the improved model, with its structure shown in Figure 2.

#### 2.2.1 Swin Transformer

The Swin Transformer is a deep learning model based on the self-attention mechanism, which consists of two parts: Encoder and Decoder. The Multi-head self-attention mechanism based on the shifted windows in the Swin Transformer can model dependencies between features at different spatial locations, effectively capture global context information, and have better feature extraction capabilities. In this study, the Swin Transformer network was used to replace the CSPDarknet network in YOLOv5s. The powerful feature extraction capability of the Swin Transformer and the efficient inference speed of the YOLOv5s object detection algorithm were combined to improve the detection effect. The Swin Transformer structure is shown in Figure 3.

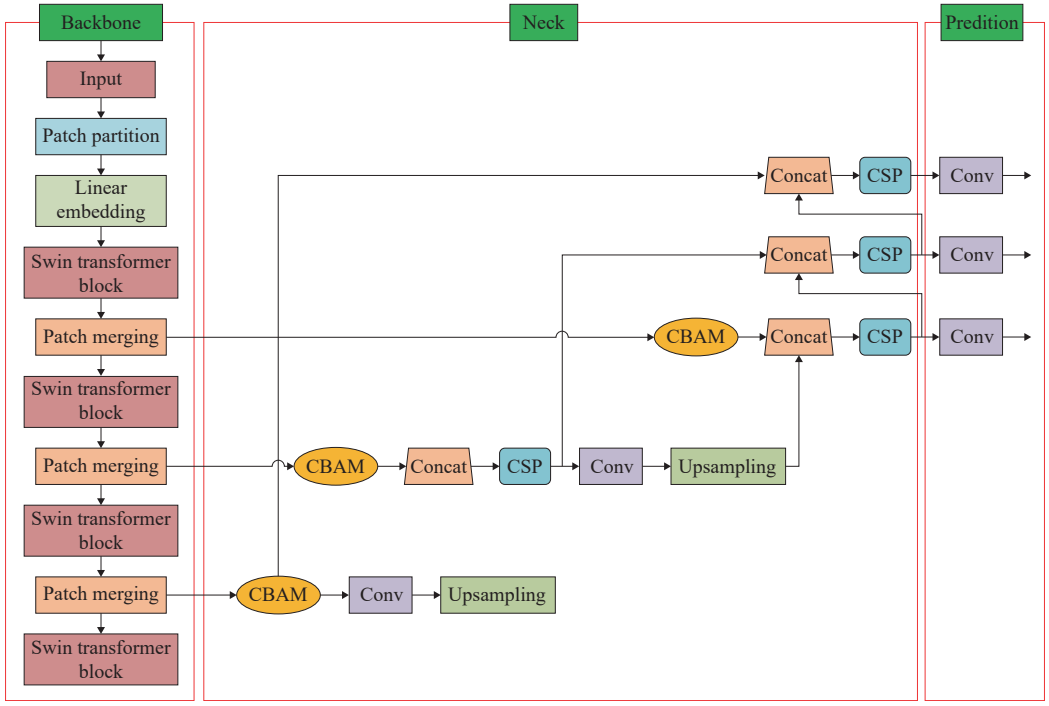
#### 2.2.2 CBAM Attention Mechanism

CBAM[18] contains two independent sub-models, channel attention module (CAM) and spatial attention module (SAM). First, the input feature layer is averaged pooling and maximum pooling to obtain two compressed feature maps, then the output is summed by a two-layer multilayer perceptron. Finally, the weight coefficients are obtained by the Sigmoid activation function and multiplied with the original feature layer to obtain the new feature map.

Due to the complex background of wheat ears growing in a field environment, occlusion between the ears of wheat is common. Therefore, this study added the CBAM attention mechanism to the YOLOv5s network to refine the features by weighting them in both channel and spatial dimensions to improve the model attention to healthy wheat ears and diseased wheat ears in both dimensions, reduce the proportion of weights occupied by irrelevant features, and thus improve the model accuracy. The CBAM attention mechanism is shown in Figure 4.

#### 2.2.3 SIoU Loss Function

The original YOLOv5s model utilizes CIoU loss to expedite




Note: CBAM: Convolutional block attention module; Conv: Convolution; CSP: Cross-stage partial network.

Figure 2 Tr-YOLOv5s model structure

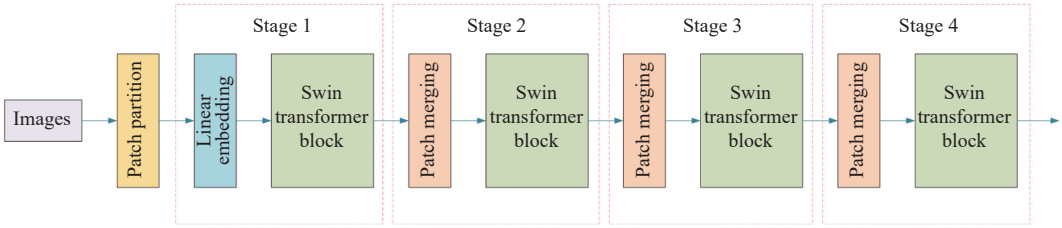



Figure 3 Swin Transformer structure

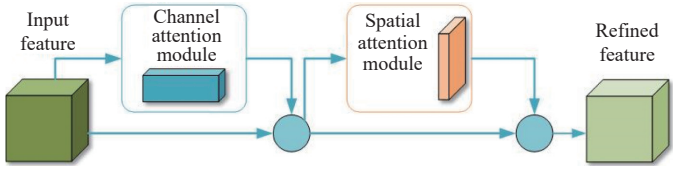



Figure 4 Implementation of CBAM attention mechanism

the regression speed of prediction frames to some extent. However, CIoU takes into account the distance between the center point of the prediction frame and the target frame, the coverage area, and the aspect ratio, but does not take into account the angle mismatch between the actual frame and the prediction frame, so this study introduced SIoU loss function to replace CIoU loss function. SIoU loss value is abbreviated as $\text{Loss}_{\text{SIoU}}$. The Equations are as follows:

$$\text{Loss}_{\text{SIoU}} = 1 - \text{IoU} + \frac{\Delta + \Omega}{2} \tag{1}$$

$$\text{IoU} = \frac{\left|B \cap B^{\text{GT}}\right|}{\left|B \cup B^{\text{GT}}\right|} \tag{2}$$

Among them,

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho_t}) = 2 - e^{-\gamma \rho_x} - e^{-\gamma \rho_y} \tag{3}$$

$$\rho_x = \left( \frac{b_{c_x}^{gt} - b_{c_x}}{c_w} \right)^2, \ \rho_y = \left( \frac{b_{c_y}^{gt} - b_{c_y}}{c_h} \right)^2, \ \gamma = 2 - \Lambda \tag{4}$$

$$\Lambda = 1 - 2\sin^2 \left( \arcsin\left(\frac{c_h}{\sigma}\right) - \frac{\pi}{4} \right) = \cos\left(2 \cdot \left(\arcsin\left(\frac{c_h}{\sigma}\right) - \frac{\pi}{4}\right)\right) \tag{5}$$

$$\Omega = \sum_{t=w,h} (1 - e^{-w_t})^{\theta} = (1 - e^{-w_w})^{\theta} + (1 - e^{-w_h})^{\theta}, \tag{6}$$

$$w_w = \frac{|w - w^{\text{gt}}|}{\max(w, w^{\text{gt}})}, \ w_h = \frac{|h - h^{\text{gt}}|}{\max(h, h^{\text{gt}})} \tag{7}$$

where, $B$ and $B^{\text{GT}}$ in Equation (2) represent the prediction box and the real box, respectively; $c_h$ and $c_w$ in Equation (4) represent the height and width of the minimum cordered rectangle of $B$ and $B^{\text{GT}}$, $c_h$ and $\sigma$ in Equation (5) represent the height difference and distance between $B$ and the center point of $B^{\text{GT}}$, respectively, $w$ and $h$ in Equation (7) represent the width and height of $B$, $w^{\text{gt}}$ and $h^{\text{gt}}$ represent the width and height of $B^{\text{GT}}$, and $\theta$ controls the degree of attention paid to shape loss.

### 2.3 Model evaluation metrics

In order to evaluate the detection effect of the algorithm on wheat ears. The model uses precision ($P$), recall ($R$), average

precision (AP), and mean average precision (mAP) as evaluation metrics to examine the model performance. The formulas are as Equations (8)-(11).

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{8}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{9}$$

$$\mathrm{AP} = \int_0^1 P \cdot R \mathrm{d}R \tag{10}$$

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{11}$$

where, TP represents the number of positive samples detected correctly, FP represents the number of positive samples detected incorrectly, FN represents the number of negative samples detected incorrectly, and N represents the number of categories of data.

While detecting wheat ears, it is necessary to analyze the counting performance of the final counting results to measure the accuracy and applicability of the algorithm in wheat ear counting. In this study, the determination coefficient $R^2$, root mean squared error (RMSE), and mean absolute error (MAE) are selected as algorithm counting evaluation metrics. The formulas are as Equations (12)-(14).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(t_i - p_i)^2}{\sum_{i=1}^{n}\left(t_i - \overline{t_i}\right)^2} \tag{12}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(t_i - p_i)^2}{n}} \tag{13}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|t_i - p_i| \tag{14}$$

where, $n$ is the total number of images participating in the calculation of the evaluation index, and $t_i$ and $p_i$ represent the number of measured ears of wheat and the number of ears of wheat counted by the algorithm, respectively, $\overline{t_i}$ represents the average number of ears of wheat in each image.

### 2.4 Test Environment and Parameters

This study used the PyTorch deep learning framework, on a computer with an InteI(R) Core(TM) i5-7300HQ CPU@2.50 GHz. The software tools included Python 3.6, CUDA10.0, Cudnn7.4.1.5. The batch size was set to 4, the epoch was set to 400, the learning rate was 0.001, and the optimizer was the Adam optimizer.

## 3 Results and analysis

### 3.1 Model training results

In this study, the YOLOv5s, YOLOv4[19], Faster Region-based Convolutional Neural Networks (R-CNN)[20], and Tr-YOLOv5s models are trained while ensuring that the initial parameters such as training batch, number of steps, optimizer, and learning rate are consistent with the device. Table 1 lists the model training results. Compared with the Faster R-CNN model, the improved model exhibits increased precision by 20.58%, recall by 2.28%, and mAP by 11.23%. In comparison to the YOLOv4 model, the enhanced precision is 7.74%, recall improves by 6.30%, and mAP value increases by 8.73%. Furthermore, when contrasted with the original YOLOv5s model, the improved model demonstrates a precision rise of 4.15%, achieves a recall rate of 86.01%, and attains an elevated mAP value of 4.63%. From the training results, it can be seen that the improved model achieves the highest mAP value after replacing CSPDarknet with Swin Transformer and adding the CBAM attention mechanism, which can improve the model's ability to extract features from both healthy and diseased wheat ears in the field environment and improve the recognition effectiveness while ensuring the mAP value.

**Table 1 Comparison of different models**

| Model | Recall | Precision | mAP |
|---|---|---|---|
| Faster R-CNN | 84.12% | 63.06% | 79.41% |
| YOLOv4 | 80.10% | 75.90% | 81.91% |
| YOLOv5s | 86.42% | 79.49% | 86.01% |
| Tr-YOLOv5s | 86.40% | 83.64% | 90.64% |

Note: mAP: mean average precision. Same below.

From the comparison of model parameters results listed in Table 2, it can be obtained that the improved model in this study exhibits minimal differences from the YOLOv7 model in terms of mAP. Moreover, the number of parameters and floating point operations has been reduced compared with the original YOLOv7, and the model operation speed has been improved. Therefore, the improved Tr-YOLOv5s model in this study achieves better results in model size and complexity compared with the original YOLOv7 model in wheat ear FHB target identification, and the model performance has been improved and can be used for wheat ear FHB target identification.

**Table 2 Comparison of model parameters**

| Model | mAP | Parameters | Gflops |
|---|---|---|---|
| YOLOv7 | 91.61% | 37.2 MB | 101.2 G |
| Tr-YOLOv5s | 90.64% | 31.3 MB | 50.7 G |

### 3.2 Results of model ablation experiments

In this study, the Tr-YOLOv5s model replaces the CSPDarknet network with the Swin Transformer network based on the YOLOv5s model and adds the CBAM attention module and the SIoU loss function. In order to analyze the improvement of the Tr-YOLOv5s model compared with the original model more intuitively, five groups of experiments were conducted and the mAP of the experiment were compared, and the results are listed in Table 3.

**Table 3 Model ablation experiments**

| Experiment number | Swin Transformer | CBAM | SIoU | mAP |
|---|---|---|---|---|
| 1 | | | | 86.01% |
| 2 | √ | | | 87.19% |
| 3 | | √ | | 86.85% |
| 4 | √ | √ | | 88.55% |
| 5 | √ | √ | √ | 90.64% |

Note: √ indicates that the baseline model uses this module.

Experiment 1 is the original YOLOv5s model with an mAP of 86.01%, without the addition of Swin Transformer, CBAM, and SIoU; Experiment 2 and Experiment 3 improve the mAP by 1.18% and 0.84%, respectively, compared to the original YOLOv5s model with only the addition of Swin Transformer and CBAM modules. Experiment 4 is the Tr-YOLOv5s model with both Swin Transformer and CBAM modules, and the mAP is improved by 2.54%; Experiment 5 is the improved model of this study, and the mAP is 90.64%, which is improved by 4.63%.

Figure 5 shows the loss functions of the original YOLOv5s model and the Tr-YOLOv5s model after training. It can be seen that the training stabilizes after 300 rounds and the model converges, and the improved model has a lower loss value compared with the original model, and the training effect is improved.

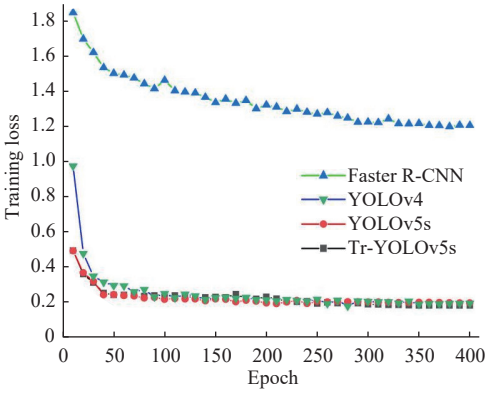



Figure 5 Training loss of the models.

### 3.3 Wheat ear FHB counting

#### 3.3.1 Comparison of wheat count results on the test set

In order to measure the difference between the proposed Tr-YOLOv5s model and the original YOLOv5s model in wheat count results, the algorithmic predicted values of wheat counts and manually obtained true values were counted on the test set images separately. Subsequently, a linear fit was performed on the predicted and true values of wheat ear counting. From the figure, it can be seen that the points in the count results of the Tr-YOLOv5s model are close to the fitted line, while the points in the count results of the original YOLOv5s model are more scattered, which indicates that the significance of the Tr-YOLOv5s model is stronger. The determination coefficient $R^2$ of the Tr-YOLOv5s model was 0.92, which was 0.02 higher than that of the original YOLOv5s model, indicating that the Tr-YOLOv5s model had a significant linear correlation between the predicted values of wheat ear counting on a single wheat ear image and the true values of the manual statistics. The fitting results of wheat ear counting before and after YOLOv5s improvement are shown in Figure 6.

Table 4 lists the results of wheat ear counting before and after the improvement of the YOLOv5s model. From the table, it can be seen that the RMSE of the original YOLOv5s model is 0.13 higher than that of the Tr-YOLOv5s model, and the MAE is 0.16 higher, indicating that the Tr-YOLOv5s model in this study outperforms the original YOLOv5s model in terms of wheat ear counting results.

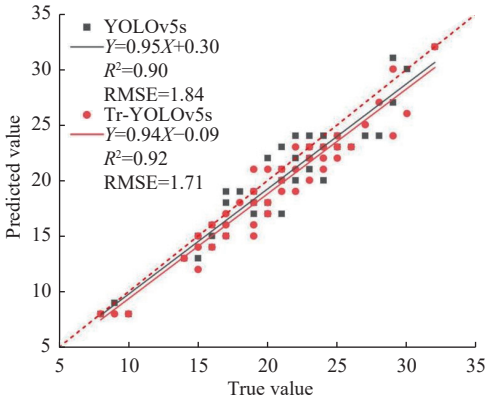


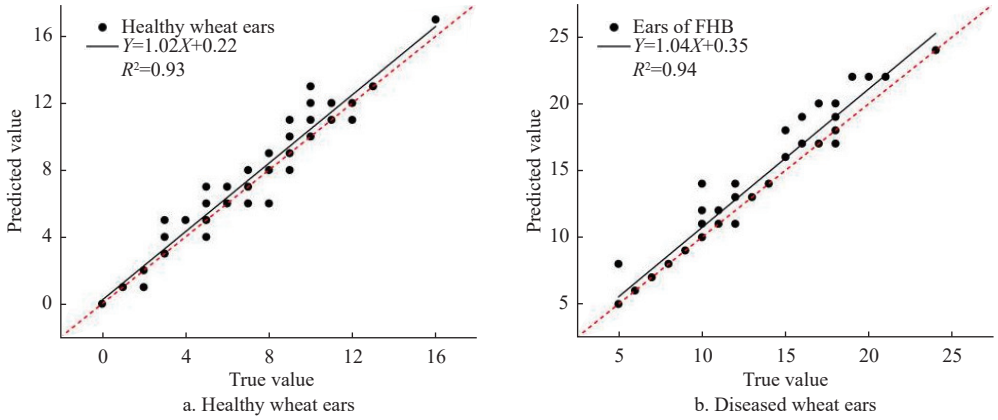

Figure 6 Fitting results of wheat ear counting before and after YOLOv5s improvement

**Table 4 Comparison of wheat ear counting before and after model improvement**

| Model | $R^2$ | RMSE | MAE |
|---|---|---|---|
| YOLOv5s | 0.90 | 1.84 | 1.40 |
| Tr-YOLOv5s | 0.92 | 1.71 | 1.24 |

#### 3.3.2 Counting Results of Healthy Wheat ears and Diseased Wheat ears

In order to further analyze the effect of improved YOLOv5s algorithm on wheat ear counting, this study was tested by 50 images. The measured values of healthy wheat ears and ears of FHB in a single image were fitted with the predicted values of the algorithm by linear regression method, and the fitting results are shown in Figure 7.




a. Healthy wheat ears b. Diseased wheat ears

Figure 7 Fitting results of healthy wheat ears and diseased wheat ears

From the regression plot, it can be seen that the fitting results for FHB wheat ears counting in the improved YOLOv5s algorithm surpass those of healthy wheat ears, with an $R^2$ of 0.94, which is 0.1 higher than that of healthy wheat ears. This could be attributed to the initial dataset containing fewer healthy wheat ears than diseased wheat ears, resulting in less training for healthy wheat ears. Consequently, the model recognition effect on healthy wheat ears is weaker than that on diseased wheat ears.

## 4 Discussion

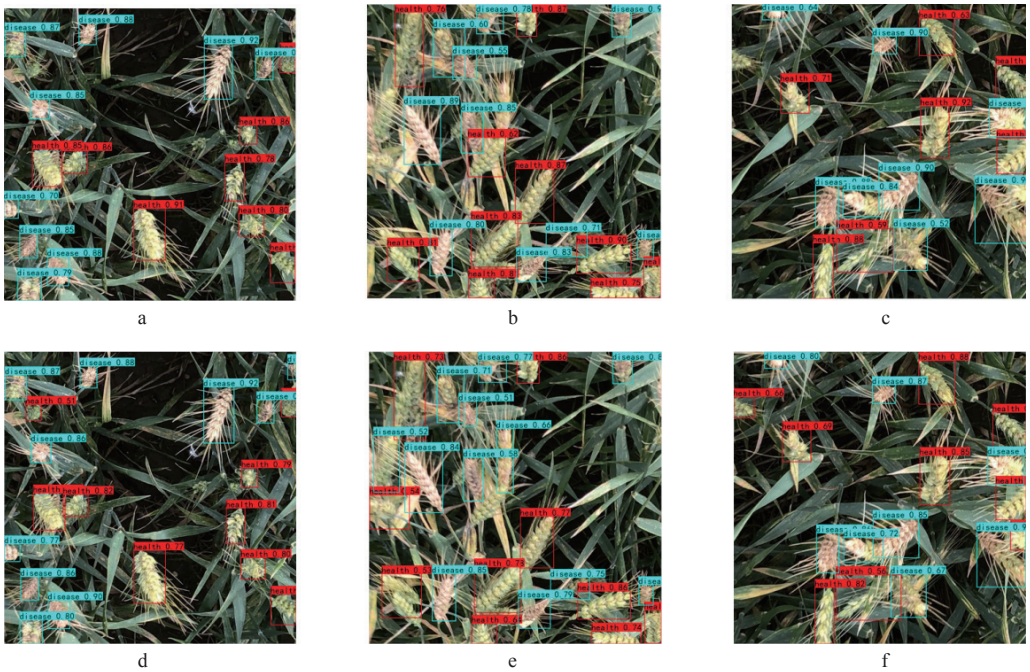
Studies have shown that traditional machine learning algorithms have been widely used in many fields and have achieved certain research results. These algorithms are also gradually being applied to the field of agricultural production[21,22]. The main

processes related to crop disease identification using traditional machine learning algorithms include image pre-processing[23,24], feature extraction, and classification recognition. In order to improve the accuracy of crop disease recognition, more and more researchers have been improving and innovating traditional machine learning algorithms. Basavaiah et al.[25] proposed a method for identifying tomato leaf diseases that incorporates multiple features, achieving a classification accuracy of over 90%.

Although crop disease identification methods based on traditional machine learning algorithms can achieve relatively good results, the identification process is tedious and requires manual extraction of the best features. To avoid the inconvenience of manually searching for features, researchers have gradually turned to using deep convolutional neural networks(CNN) to classify and identify crop diseases and achieved better results than traditional machine learning algorithms in crop classification[26,27], weed detection[28,29], disease identification[30,31], and pest monitoring[32,33]. Liu et al.[34] used an improved convolutional neural network-based algorithm to identify 6 grape leaf diseases with an accuracy of 97.22%. Chen et al.[35] improved the YOLOv4 algorithm to detect weeds with an average accuracy of 98.52%. However, the impact of different deep learning models on different datasets varies; therefore, to enhance the model performance on the dataset in this study, improvements are made to increase recognition accuracy.

In this study, three models, Faster R-CNN, YOLOv4, and YOLOv5s, were used to train the collected wheat dataset. The results showed that the YOLOv5s model achieved the highest accuracy of 86.01% with the same parameters as the three models. Therefore, the YOLOv5s model was chosen for improvement. Firstly, the CSPDarknet network is replaced by Transformer to improve the model's recognition ability for different scales of wheat ears. Secondly, The CBAM attention mechanism is added after the three feature layers extracted from the backbone network to further enhance the recognition ability for small wheat ears. Finally, we replace the CIoU loss function in the original YOLOv5s model with the SIoU loss function to make the model convergence. The results show that the Tr-YOLOv5s model has an mAP value of 90.64%, which is 4.63% higher than the original model, with an accuracy of 92.55% for the identification of FHB wheat ears and 88.73% for the identification of healthy wheat ears. The possible reason is that in this study, the occurrence of FHB is artificially induced by blast spores, and there are more diseased ears than healthy ears in the image dataset, resulting in a lower training amount of the model for healthy ears compared to diseased ears. Hence, the recognition accuracy of the model for healthy ears is lower than that for diseased ears. The convergence speed of the Tr-YOLOv5s model is faster because SIoU introduces the vector angle between the real frame and the predicted frame, redefining the correlation loss function. In contrast, CIoU does not consider the direction between the real frame and the predicted frame, resulting in a slower convergence speed than SIoU. The experimental results show that the Tr-YOLOv5s model is feasible for identifying healthy and diseased wheat ears under field conditions, and can achieve better results. This serves as a reference for the subsequent automatic identification of FHB in wheat under field conditions.

In addition to the above work, this study also achieved an estimated counting of wheat ears. The research on recognition counting of wheat using CNN mainly includes three methods: image segmentation[36], object detection[37,38], and regression counting[39]. Wang et al.[40] constructed an improved EfficientDet-D0 algorithm based on counting wheat ears, and its counting accuracy reached 94%. Yang et al.[41] employed an improved YOLOv4-based model with a mAP of 94% on the local wheat dataset. Figure 8 shows the visualization image of part of the original YOLOv5s model and the Tr-YOLOv5s model, it can be seen that the recognition effect and confidence of the Tr-YOLOv5s model are better than the original YOLOv5s model. It can be seen that the YOLOv5s improvement model proposed in this study counts the wheat ears in the close-up image, and directly counts the number of ears in each image, which can provide data information for subsequent accurate estimation of wheat yield.



Note: a-c: Visualization results of the original YOLOv5 model with different samples; d-f: Visualization results of the original Tr-YOLOv5 model with different samples.

Figure 8 YOLOv5s and Tr-YOLOv5s detection results

In future work, researchers can acquire more images of wheat ears in real environments to expand the dataset to train a more robust model and improve the accuracy of wheat ear counting. In addition, this study used the YOLOv5s algorithm to identify healthy and diseased wheat ears under field conditions. Although this study can achieve good results, it may not always get ideal results when identifying wheat ears in a highly complex farmland environment. In the future, it is hoped to further study this algorithm and enhance its accuracy in recognizing wheat ears in complex farmland environments.

## 5 Conclusions

In order to achieve the rapid identification and counting of healthy and diseased wheat ears in a field environment, this study introduced the Swin Transformer structure to fuse global and local information to improve the generalization performance of the model based on the YOLOv5 model, added the Convolutional Block Attention Module (CBAM) attention module to enhance the extraction of the network for the ears of wheat feature information. Additionally, replaced the original Complete Intersection over Union (CIoU) loss function with the Scylla Intersection over Union (SIoU) loss function; and counted the healthy wheat ears and diseased wheat ears.

1) Compared with the YOLOv5s model, the Tr-YOLOv5s model improved the precision of wheat ears by 4.15% and the mean average precision (mAP) by 4.63%. In comparison to the YOLOv4 model, the Tr-YOLOv5s model enhanced precision by 7.74%, recall by 6.30%, and mAP by 8.73%. Furthermore, when compared with the Faster Region-based Convolutional Neural Networks (R-CNN) model, the Tr-YOLOv5s model showed improvements in precision by 20.58%, recall by 2.28%, and mAP by 11.23%. The Tr-YOLOv5s model has improved in both mAP and precision and has better recognition results.

2) In the ear counting experiment, the improved YOLOv5s model has root mean squared error (RMSE) and mean absolute error (MAE) values of 1.71 and 1.24, respectively, indicating a decrease of 0.13 and 0.16 compared to the original YOLOv5s model. The improved model exhibits the lowest error in the ear counting experiment, further reducing the counting error of wheat ears, and can provide data information for subsequent estimation of wheat yield.

The results show that the Tr-YOLOv5s model had a better effect on the detection and counting of fusarium head blight (FHB) in wheat under field environments.

## Acknowledgements

The authors are thankful to Guihong Yin and Guihua Bai for their strong support for this work. This study was supported by the Natural Science Foundation of Henan Province (NO. 222301420113, 232102520006); Major Science and Technology Special Project of Henan Province (NO. 221100210600); Henan Province key research and development project (NO. 231111110100); Key Scientific and Technological Project of Henan Province (NO. 242102111193), and the Natural Science Foundation of China(NO. 31501225, 42101362). The authors would like to express their appreciation for the valuable discussions and perceptive inspirations provided by the editors and anonymous reviewers, who signiffcantly enhanced the quality of this work.

## [References]

[1] Han J C, Zhang Z, Cao J, Luo Y C, Zhang L L, Li Z Y, Zhang J. Prediction of winter wheat yield based on multi-source data and machine learning in China. Remote Sensing, 2020; 12(2): 236.

[2] Huang J, Chen J H, Zhang F M, Hu Z H. Spatial-temporal changes in risk of climate related yield reduction of winter wheat during 1973-2014 in Anhui province, southeast China. Theoretical and Applied Climatology, 2022; 148(1-2): 49–63.

[3] Zhao S Q, Gu J B, Zhao Y Y, Hassan M, Li Y N, Ding W M. A method for estimating spikelet number per panicle: Integrating image analysis and a 5-point calibration model. Scientific Reports, 2015; 5: 16241.

[4] Chaudhary A, Thakur R, Kolhe S, Kamal R. A particle swarm optimization based ensemble for vegetable crop disease recognition. Computers and Electronics in Agriculture, 2020; 178: 105747.

[5] Abdu A M, Mokji M M, Sheikh U U. Automatic vegetable disease identification approach using individual lesion features. Computers and Electronics in Agriculture, 2020; 176: 105660.

[6] Suganya Devi K, Srinivasan P, Bandhopadhyay S. H2K - A robust and optimum approach for detection and classification of groundnut leaf diseases. Computers and Electronics in Agriculture, 2020; 178: 105749.

[7] Rahman C R, Arko P S, Ali M E, Khan M A I, Apon S H, Nowrin F, et al. Identification and recognition of rice diseases and pests using convolutional neural networks. Biosystems Engineering, 2020; 194: 112–120.

[8] Li C L, Li H Y, Gao G S, Liu Z F, Liu P C. An accelerating convolutional neural networks via a 2D entropy based-adaptive filter search method for image recognition. Applied Soft Computing, 2023; 142: 110326.

[9] Zhao Z Q, Zheng P, Xu S T, Wu X D. Object detection with deep learning: A review. IEEE Transactions on Neural Networks and Learning Systems, 2019; 30(11): 3212–3232.

[10] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas: IEEE, 2016; pp.779–788. doi: 10.1109/CVPR.2016.91.

[11] Agarwal M, Gupta S K, Biswas K K. Development of efficient CNN model for tomato crop disease identification. Sustainable Computing: Informatics & Systems, 2020; 28: 100407.

[12] Fang T, Chen P, Zhang J, Wang B. Crop leaf disease grade identification based on an improved convolutional neural network. Journal of Electronic Imaging, 2020; 29(1): 013004.

[13] Qi J T, Liu X N, Liu K, Xu F R, Guo H, Tian X L, et al. An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease. Computers and Electronics in Agriculture, 2022; 194: 106780.

[14] Zhang D Y, Luo H S, Wang D Y, Zhou X G, Li W F, Gu C Y, et al. Assessment of the levels of damage caused by fusarium head blight in wheat using an improved YOLOv5 method. Computers and Electronics in Agriculture, 198: 107086.

[15] Wang C-Y, Liao H-Y M, Wu Y-H, Chen P-Y, Hsieh J-W, Yeh I-H. CSPNet: A new backbone that can enhance learning capability of CNN. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle: IEEE, 2020; pp.1571–1580. doi: 10.1109/CVPRW50498.2020.00203.

[16] Zheng Z H, Wang P, Liu W, Li J Z, Ye R G, Ren D W. Distance-IoU loss: Faster and better learning for bounding box regression. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2020; 34(7): 12993–13000.

[17] Du S J, Zhang B F, Zhang P. Scale-sensitive IOU loss: An improved regression loss function in remote sensing object detection. IEEE Access, 2021; 9: 141258–141272.

[18] Woo S, Park J, Lee J Y, Kweon I S. CBAM: Convolutional Block Attention Module. In: Proceedings of the European Conference on Computer Vision (ECCV), 2018; pp.3–19. doi: 10.1007/978-3-030-01234-2_1.

[19] Wu D H, Lv, S C, Jiang M, Song H B. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. Computers and Electronics in Agriculture, 2020; 178: 105742.

[20] Ren S, He K M, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017; 39(6): 1137–1149.

[21] Xiao D Q, Feng J Z, Lin T Y, Pang C H, Ye Y W. Classification and

recognition scheme for vegetable pests based on the BOF-SVM model. Int J Agric & Biol Eng, 2018; 11(3): 190–196.

[22] Wang L A, Zhou X D, Zhu X K, Dong Z D, Guo W S. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. The Crop Journal, 2016; 4(3): 212–219.

[23] Sampathkumar S, Rajeswari R. An automated crop and plant disease identification scheme using cognitive fuzzy C-means algorithm. IETE Journal of Research, 2022; 68(5): 3786–3797.

[24] Bao W X, Zhao J, Hu G S, Zhang D Y, Huang L S, Liang D. Identification of wheat leaf diseases and their severity based on elliptical-maximum margin criterion metric learning. Sustainable Computing: Informatics & Systems, 2021; 30: 100526.

[25] Basavaiah J, Anthony A A. Tomato leaf disease classification using multiple feature extraction techniques. Wireless Personal Communications, 2020; 115(1): 633–651.

[26] Chew R, Rineer J, Beach R, O'Neil M, Ujeneza N, Lapidus D, et al. Deep neural networks and transfer learning for food crop identification in UAV images. Drones, 2020; 4(1): 7.

[27] Zhang F, Chen Z J, Ali S, Yang N, Fu S L, Zhang Y K. Multi-class detection of cherry tomatoes using improved YOLOv4-Tiny. Int J Agric & Biol Eng, 2023; 16(2): 225–231.

[28] Tripathi S P, Yadav R K, Rai H. Chapter 9 - Weednet: A deep neural net for weed identification. In: Deep Learning for Sustainable Agriculture, Academic Press, 2022; pp.223–236. doi: 10.1016/B978-0-323-85214-2.00010-0.

[29] Wang Z, Guo J X, Zhang S W. Lightweight convolution neural network based on multi-scale parallel fusion for weed identification. International Journal of Pattern Recognition and Artificial Intelligence, 2022; 36(7): 2250028.

[30] Tian J Y, Zhang Y, Wang Y L, Wang C, Zhang S Y, Ren T H. A method of corn disease identification based on convolutional neural network. In: 2019 12th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou: IEEE, 2019; pp.245–248. doi: 10.1109/ISCID.2019.00063.

[31] Wu Y H. Identification of maize leaf diseases based on convolutional neural network. Journal of Physics: Conference Series, 2021; 1748(3): 032004.

[32] Cheng X, Zhang Y H, Chen Y Q, Wu Y Z, Yue Y. Pest identification via deep residual learning in complex background. Computers and Electronics in Agriculture, 2017; 141: 351–356.

[33] Wang D W, Deng L M, Ni J G, Gao J Y, Zhu H F, Han Z Z. Recognition pest by image-based transfer learning. Journal of the Science of Food and Agriculture, 2019; 99(10): 4524–4531.

[34] Liu B, Ding Z F, Tian L L, He D J, Li S Q, Wang H Y. Grape leaf disease identification using improved deep convolutional neural networks. Frontiers in Plant Science, 2020; 11: 1082.

[35] Chen J D, Zhang D F, Zeb A, Nanehkaran Y A. Identification of rice plant diseases using lightweight attention networks. Expert Systems with Applications, 2021; 169: 114514.

[36] Marszalek M, Körner Ms, Schmidhalter U. Prediction of multi-year winter wheat yields at the field level with satellite and climatological data. Computers and Electronics in Agriculture, 2022; 194: 106777.

[37] Ma J C, Li Y X, Du K M, Zheng F X, Zhang L X, Gong Z H, et al. Segmenting ears of winter wheat at flowering stage using digital images and deep learning. Computers and Electronics in Agriculture, 2020; 168: 105159.

[38] Madec S, Jin X L, Lu H, De Solan B, Liu S Y, Duyme F, et al. Ear density estimation from high resolution RGB imagery using deep learning technique. Agricultural and Forest Meteorology, 2019; 264: 225–234.

[39] Ma J C, Li Y X, Liu H J, Wu Y F, Zhang L X. Towards improved accuracy of UAV based wheat ears counting: A transfer learning method of the ground-based fully convolutional network. Expert Systems With Applications, 2022; 191: 116226.

[40] Wang Y D, Qin Y X, Cui J L. Occlusion robust wheat ear counting algorithm based on deep learning. Frontiers in Plant Science, 2021; 12: 645899.

[41] Yang B H, Gao Z W, Gao Y, Zhu Y. Rapid detection and counting of wheat ears in the field using yolov4 with attention module. Agronomy, 2021; 11(6): 1202.