A LIGHTWEIGHT IMPROVED YOLOv5s MODEL-BASED RICE BLAST DETECTION METHOD AND MOBILE DEPLOYMENT

/

基于轻量化改进 YOLOv5s 模型的稻瘟病检测方法及移动部署

Fankai MENG[1)], Congkuan YAN[1)], Yuqing YANG[1)], Ruixing XING[1)], Dequan ZHU[1)], Aifang ZHANG[2)], Qixing TANG[1,3], Juan LIAO[1,3*)]

[1)] College of Engineering, Anhui Agricultural University, Hefei 230036 / China;

[2)] Institute of Plant Protection and Agricultural Product Quality and Safety, Anhui Academy of Agricultural Sciences, Hefei 230031/ China;

[3)] East Anhui Comprehensive Experimental Station, Anhui Agricultural University, Mingguang, 239400 / China

E-mail: liaojuan@ahau.edu.cn

DOI: https://doi.org/10.35633/inmateh-74-68

Keywords: *Rice blast detection; YOLOv5s; Lightweight; GhostConv; BiFPN; SE; Android Studio*

ABSTRACT

For achieving more efficient recognition results and deployment on mobile devices, a rice blast recognition model was constructed by making lightweight improvements to YOLOv5s. First, using YOLOv5s as the base, GhostConv was introduced to replace standard convolution in its backbone and neck, and LightC3 module was built to improve the C3 module in the neck. This significantly reduced the computational burden and model size. Furthermore, Concat operator was replaced with BiFPN and SE attention mechanism was integrated to maintain accuracy when reducing model complexity. These modifications enhanced the model's ability to capture fine-grained features and multi-scale information. The experimental results showed that the designed model had a 49% decrease in the number of model parameters and a 50% decrease in FLOPs without a decrease in precision on self-built rice blast dataset, compared with the YOLOv5s, achieving the good balance between detection performance and model lightweight. Then, an APP named RiceBlastDetector was built based on the model, achieving accurate detection in the scenario with the different characterization scale disease spots from experiments in the field, which can provide a reference for detecting other crop diseases.

摘要

为了在移动设备上实现更高效的识别结果与部署，我们通过对 YOLOv5s 进行轻量化改进构建了一个稻瘟病识别模型。首先，以 YOLOv5s 为基础，引入 GhostConv 替换其骨干和颈部的标准卷积，并构建 LightC3 模块，这显著减轻了计算负担和模型大小。接着，用 BiFPN 替换 Concat 算子，并集成 SE 注意力机制，从而在保持精度的同时降低复杂度，这些改进增强了模型捕捉细粒度特征和处理多尺度信息的能力。通过自建数据集验证，该模型在不损失精度的情况下，参数减少了 49%，浮点运算次数降低了 50%，这证明了我们在实现模型轻量化的同时保持高检测性能这一方法的有效性。我们还构建应用程序，用于田间的精准检测，为其他农作物病害的检测提供了参考，而且该应用程序有提升农业生产力和促进可持续农业实践的潜力。

INTRODUCTION

In terms of food production, rice ranks as one of the key crops in China, playing a crucial role in ensuring food security for the population. However, rice production is often hindered by diseases such as brown spot, rice stalk, and rice blast, which often occur and can cause considerable yield losses (*Malvade et al., 2022*). Among them, rice blast disease affects rice globally because of its rapid occurrence and easy infection. Rice blast varies significantly throughout the rice plant's development stages and can be divided into seedling blast, leaf blast and ear blast according to the time and location of rice infection, among which leaf blast and ear blast are the most harmful, which can reduce rice yield by 10%~30%, or even more than 50% (*Asibi et al., 2019; Yang et al., 2020*). Hence, it is necessary to identify and control rice blast in the early stage, which is very important to ensure a stable and efficient rice production.

The conventional process of disease recognition in agriculture relies heavily on the expertise of agronomists or plant pathologists who visually examine and manually identify lesions based on their shape, color, and other characteristics. However, this conventional method is often inefficient and subjective. As computer vision technology advances, numerous studies have concentrated on automating the detection of rice diseases using pattern recognition and machine learning techniques (*Daniya et al., 2019; Sanghavi et al., 2021; Manavalan et al., 2020*), where the essential challenge is to first extract characteristic feature information

of rice diseases from the images by using image processing and then using the classifier to judge whether the crop is infected or not according to a single image feature such as color, shape and texture or a combination of multiple features. The effectiveness of these machine learning-based methods is affected by challenges such as location of rice disease infection, variations between different diseases, manual feature extraction, and various factors present in the natural environment, leading to potentially insufficient accuracy. Therefore, there is a need for research focused on developing effective feature extraction and detection algorithms for rice blast disease.

Recently, deep learning technology has shown great prospects for application in agricultural target detection due to its ability to extract effective features, which have been extensively utilized for detecting diseases in plant leaves (*Bedi et al., 2021*), citrus fruit and leaves (*Khattak et al., 2021*), teas (*Sun et al., 2023*), maize (*Khan et al., 2023*), and other crops (*Nandhini et al., 2022*). To detect rice disease, *Zhou et al.* (*2019*) presented an algorithm that integrates the improved k-means clustering with Faster R-CNN, achieving a detection accuracy of 97.53%. *Rahman et al.* (*2020*) developed a small two-stage rice disease detection model based on CNN with 93.30% recognition accuracy. *Dogra et al. (2023*) designed a brown spot disease detection model based on CNN-VGG19 and transfer learning and achieving a detection accuracy of 93% on the developed rice leaf disease dataset. *Jia et al.* (*2023*) presented a rice disease detection model, where MobileNetV3 was used to design YOLOv7's backbone and the CA attention module was introduced in feature fusion to improve the model to integrate more key feature information, achieving an average precision of 98%.

The aforementioned studies have strongly demonstrated the feasibility of applying deep learning to recognize crop diseases (*Yan et al., 2022*). However, deep learning-based models often require significant computational resources such as high-performance workstations or servers equipped with powerful GPUs during both training and inference phases. Furthermore, most of existing deep learning-based methods prioritize higher accuracy while leading to more complex neural network models with a large number of parameters and longer prediction times. This complexity makes them difficult to deploy on embedded devices or mobile platforms, hindering their practical applications in agriculture. In this regard, some scholars have also started to focus on the lightweight CNN models for target identification. *Cheng et al.* (*2022*) present a lightweight method for crop pest recognition utilizing YOLOv3Lite, which integrated the lightweight sandglass module and the CA attention module into its residual structure. The method achieved a computational cost of 9.8 GFLOPs, which is only 8.1% of that of YOLOv3. *Li et al.* (*2024*) introduced a lightweight YOLOv8s algorithm that integrated the lightweight GhostNet to build the YOLOv8s' backbone. The approach reduced the size of model by 50.2% and FLOPs by 43.1% compared to the original YOLOv8s. In the field environment, the most portable method is to deploy neural networks in smartphones with limited computing power and use them for object detection such as crop disease detection and fruit detection (*Yu et al., 2024*). Although these approaches lay a foundation for lightweight studies in crop disease detection models, it is important to note that they are mainly designed for use on desktop computers. This focus makes it challenging to develop applications suitable for low-computation platforms. Additionally, the above studies on the model structure of general lightweight crop diseases most are based on the crops with little difference in crop disease symptoms, but similar or different symptoms of rice blast disease can develop at different stages, or on different plant parts, resulting in low accuracy of the above methods directly applied to rice blast recognition.

Aiming to better address the practical challenges of agricultural production, this study designs an improved YOLOv5 algorithm by effectively integrating various lightweight modules for rice blast detection, and develops a local Android-based application for identifying and detecting rice blast, enabling deployment on mobile devices. The algorithm replaces the standard convolution in the YOLOv5s' backbone and neck with the GhostConv module. And it substitutes the C3 modules of the YOLOv5s' backbone with GhostC3 and in the neck with LightC3 modules, thereby reducing the computational load. Furthermore, to enhance model perception of disease-related features, the lightweight channel attention mechanism squeeze-and-excitation (SE) is placed before the SPFF module and incorporated into the Bottleneck module of the YOLOv5s' neck, respectively. Additionally, the BiFPN network is introduced in the YOLOv5s' neck network to integrate more multi-scale features, and the convergence speed during model training is improved using EIoU loss. Finally, the optimized and lightweight model is implemented on mobile devices for practical monitoring of rice blast disease.

MATERIALS AND METHODS

In this study, rice blast images are captured from the experimental bases of the Rice Research Institute and the Institute of Plant Protection and Agricultural Product Quality and Safety, Anhui Academy of Agricultural

Sciences, Hefei, China. The Plant Protection Institute primarily focuses on researching the mechanisms and technologies for managing diseases and insect pests that affect crops such as rice, wheat, rapeseed, and vegetables in Anhui Province, and the Rice Research Institute is mainly engaged in basic and applied research on rice genetics and breeding, biotechnology, and development of new cultivation technologies. The images were captured from June to September 2023 using a Canon EOS 100D camera and a Huawei mobile phone. The collected image resolutions have 5184×3456 pixels and 4032×3024 pixels, respectively. To ensure the quality of the image labels, the acquired images showing rice blast disease were reviewed and validated by experts in plant protection, and then a total of 2620 images were obtained, including 1200 of ear blast and 1420 of leaf blast.

Taking into account that the high resolution of images can complicate model training and cause potential GPU memory overloads and failures, the original images were cropped to the lower resolution of images at a size of 640×640 pixels to build the dataset for this study. To label the rice blast lesions in the images, the LabelImg software was utilized to manually annotate each infected area, thereby producing the ground truth (GT) labels necessary for the model. The dataset comprises both GT and RGB images, with the specific counts for ear blast and leaf blast illustrated in Table 1. Fig. 1 gives samples of the RGB images and the corresponding GT images. To train the model and evaluate its performance, the built rice blast dataset consisted of training, validation, and test sets according to the ratio of 7:2:1.

Table 1

Rice blast disease dataset composition

	Train	Validation	Test	Total
Ear blast	840	240	120	1200
Leaf blast	994	284	142	1420

(a) Sample images of ear blast and leaf blast *(b) GT images*

Fig. 1 - RGB and ground truth (GT) images

Lightweight rice blast detection model

YOLOv5 is a widely used one-stage object detection model (*Wang et al., 2022*). Based on the model's width and depth, YOLOv5 comes in four variations: x, l, m, and s. Among these, YOLOv5s has the least depth and width in terms of network features, while the other three versions have enhancements that increase both depth and width based on this model (*Lin et al., 2022*). In this study, YOLOv5s was used as the framework for the rice blast detection model, which was composed of three networks: backbone for disease characteristics acquisition, neck for characteristics fusion and a target category prediction network *(Yonghui et al., 2024)*. In order to decrease computational and memory costs and lightweight model which can be more easily deployed on mobile phone, the YOLOv5s model has been refined through various improvements in this study, specifically:

1) using GhostConv instead of the standard convolution of YOLOv5s' backbone and neck to reduce the number of model's parameters;

2) utilizing GhostC3 module instead of C3 modules of backbone network and adding the SE attention module before SPPF module to lower computational costs and enhance the model's capability to extract features;

3) building a LightC3 module to serve as a replacement for the C3 modules of YOLOv5s' neck and introducing the BiFPN module to reduce computational burden and enhance the fusion of multi-scale features. Based on the above improvements, the architecture of the lightweight model is shown in Fig. 2.

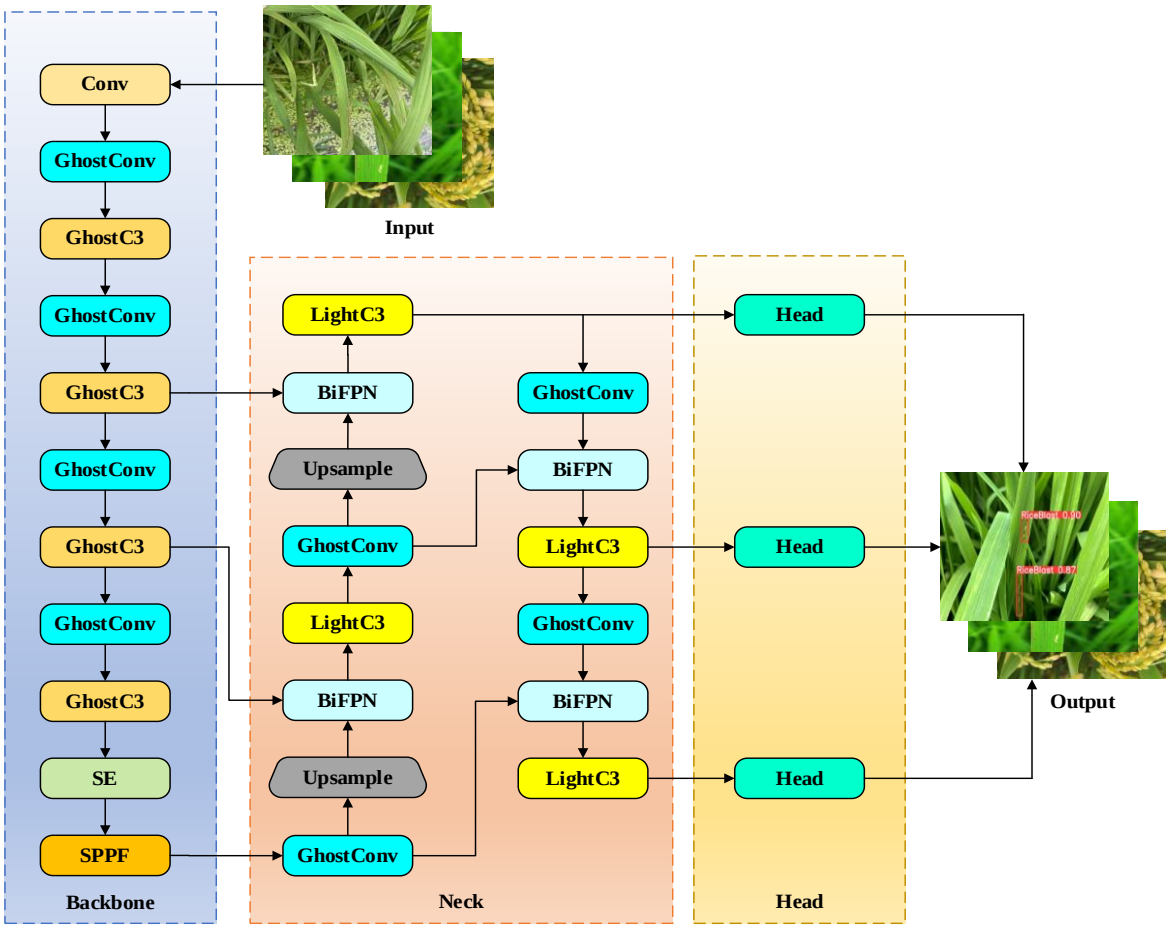


Fig. 2 - The structure of improved YOLOv5s model

Lightweight backbone network

The backbone network of YOLOv5s, with its combination of several Convs, C3, and SPP modules, is well-suited for extracting features from rice disease images, but it also presents challenges related to the number of parameters and information redundancy. To optimize the model for more efficient detection of rice diseases, the lightweight convolution GhostConv (*Han et al., 2022*) is introduced in the backbone network instead of Convs, as shown in Fig. 3. In the GhostConv, a 1×1standard convolution is utilized to decrease the channel count of input feature maps. Lightweight linear transform operations are then used to generate additional feature maps, which are then added to the previously limited number of feature maps to output a feature map with the required number of channels.

Let $X \in R^{n\times h\times w}$ be the input of the channel, height and width of n, h and w, respectively, in Fig. 3. The output with m channels is denoted as $Y \in R^{m\times h'\times w'}$. Using Ghost convolution operation, s, additional feature maps are generated, where an identity mapping operation and m(s-1) linear operations are performed. Hence, the theoretical acceleration factor of ghost convolution in relation to standard convolution with the kernel size $n\times d\times d\times m$ is:

$$\begin{aligned} r &= \frac{h\times w\times n\times k\times k\times m}{\frac{m}{s}\times h\times w\times n\times d\times d+(s-1)\times\frac{m}{s}\times h'\times w'\times d'\times d'} \\ &= \frac{s\times n\times k\times k}{n\times d\times d+(s-1)\times d'\times d'} \\ &\approx \frac{s\times n}{n+(s-1)}\approx s \end{aligned} \tag{1}$$

where $d\times d$ and $d'\times d'$ represent the convolution kernel size for the identity mapping operation and the linear operation, respectively.

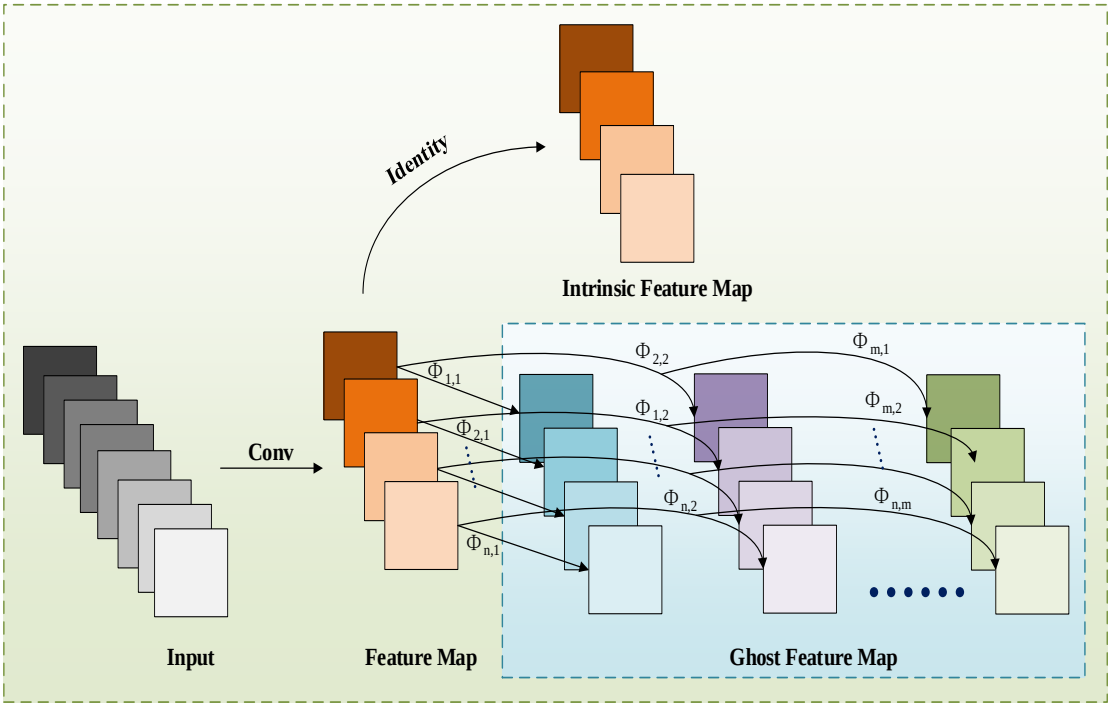


Fig. 3 - The structure of GhostConv module

From Equation (1), it can be seen that GhostConv presents a compelling alternative to standard convolutional layers, achieving significant reductions in both memory usage and FLOPs. Based on this, the improved backbone therefore uses the GhostC3 module (*Han et al., 2022*) instead of the original C3 module.

As shown in Fig. 4, the GhostConv and GhostBottleneck structures have been used to replace the traditional standard convolution and bottleneck structures, creating a lightweight C3 module in the backbone. The GhostBottleneck module, which refers to the residual structure in ResNet, replaces standard convolution with GhostConv to reduce the computational burden and memory usage (*Yang et al., 2022*).

The GhostBottleneck module's structure is given in Fig. 4, which is suitable for stride = 1 and stride = 2. In Fig. 4, the GhostBottleneck module contains two GhostConv modules and one shortcut when stride =1. The first GhostConv module expands the input feature map, while the second GhostConv module has the same number of channels as the input by mapping the output of the first module. Subsequently, the output of the second GhostConv module is added to the input of the first GhostConv module to achieve a shortcut connection. Additionally, batch normalization (BN) is applied after the two GhostConv modules to prevent gradient vanishing and to speed up training and convergence. Besides, a depth-separable convolution (DWConv) (*Wen et al., 2023*) is inserted between the two GhostConv modules when stride =2, as shown in Fig. 4. The depth convolution of DWConv is used to down-sample the output from the first GhostConv, while the point-wise convolution of DWConv is utilized to generate the intrinsic features for efficiency.

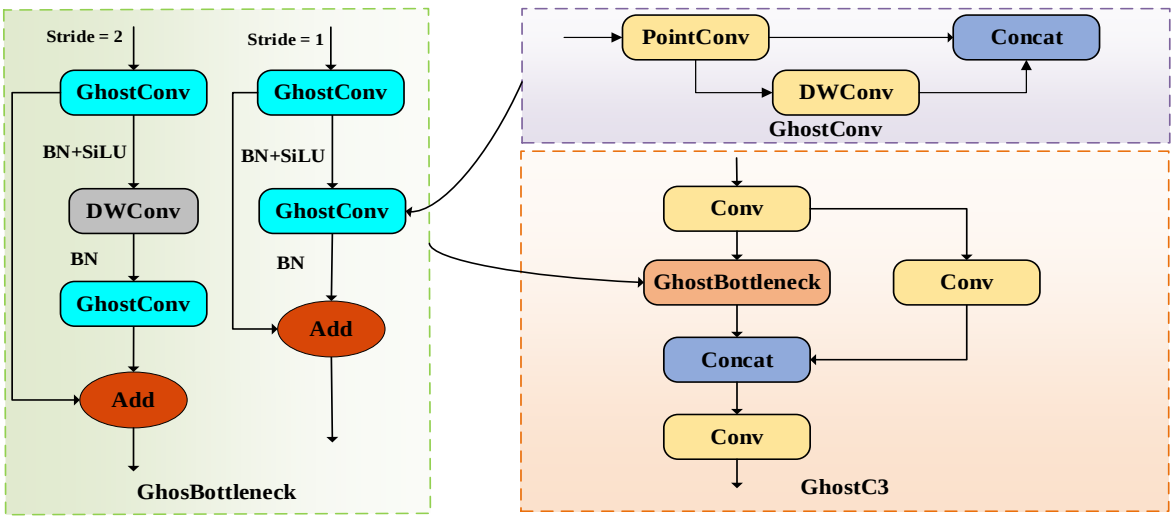


Fig. 4 - The structures of GhostC3 and GhostBottleneck modules

Added attention module

As mentioned above, using GhostConv and GhostBottleneck modules can effectively lightweight the YOLOv5s' backbone. However, in the field natural environment, the background is complex and variable, such as multiple rice plants, changing light conditions, large healthy rice leaves and other complex background information, while the disease characteristics of early leaf blast and ear blast are relatively small. During the model training process, these complex background elements will also be transmitted. As the model deepens, the weight of these background elements in the feature map may grow, potentially overwhelming the target information and negatively impacting the model's performance *(Ye et al., 2023)*. Additionally, in the YOLOv5s backbone network, the convolutional layer primarily extracts feature information from adjacent positions within each feature map. However, it does not consider the correlation among channel information, even though each channel contains distinct feature data. Therefore, the squeeze-and-excitation (SE) attention mechanism (*Xin et al., 2020*) is positioned before the SPFF module to consider the relationships between channels and dynamically adjust their weights, so as to enhance the model's perception of diseases while suppressing the influence of irrelevant information, thereby enhancing the model's perception of disease feature and improving overall accuracy.

Figure 5 illustrates the structure of SE module, which consists of squeeze, excitation and scale steps. The squeeze operation uses global average pooling to compress each 2D feature channel into a single real number with a global acceptance field. Let $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\bullet\bullet\bullet,\mathbf{x}_C]\in\mathbf{R}^{C\times H\times W}$ be the input. The output $\mathbf{S}\in\mathbf{R}^{C}$ from the squeeze operation is given in Equation (2). Following this, an excitation process is performed to assess the relationships of the different channels and output weights that represent the importance of the different channels. The excitation operation consists of a sequence of layers with 1×1 convolution, ReLU and sigmoid activation, which is defined by Equation (3). Following the excitation operation, the scale operation employs multiplication to weight the previously selected features for each channel, effectively rescaling the original features along the channel dimension. The output $\hat{\mathbf{X}}\in\mathbf{R}^{C\times H\times W}$ of the scale operation is expressed in Equation (4). According to Equation (4), the output generated by the SE module is obtained by rescaling the input with the activation of each channel. Therefore, the SE module assigns weight values to each channel of the feature map, enhancing the contribution of important features while reducing the impact of less relevant ones. This process enhances the network's capacity to learn important features related to rice diseases.

$$s_c=\mathrm{F}_{se}(\mathbf{x}_c)=\frac{\mathbf{1}}{W\times H}\sum_{i=1}^{W}\sum_{\mathrm{j}=1}^{H}\mathbf{x}_c(i,j) \tag{2}$$

$$\mathbf{z}=\mathrm{Sigmoid}(\mathbf{W}_2\times\mathrm{Re}\,LU(\mathbf{W}_1\mathbf{S})) \tag{3}$$

$$\hat{\mathbf{x}}_c=\mathrm{z}_c\times\mathbf{x}_c \tag{4}$$

where s_c is the c-th element of $\mathbf{S}$, and $\mathbf{x}_c\in\mathbf{R}^{H\times W}$ is a feature map at the c-th channel. $\mathbf{W}_2\in\mathbf{R}^{\frac{C}{r}\times C}$ and $\mathbf{W}_1\in\mathbf{R}^{C\times\frac{C}{r}}$ are weights generated for each feature channel, and r is reduction ratio. $\mathrm{Sigmoid}(\bullet)$ refers to the sigmoid function. $\hat{\mathbf{x}}_c$ is the c-th element of $\hat{\mathbf{X}}=[\hat{\mathbf{x}}_1,\hat{\mathbf{x}}_2,\bullet\bullet\bullet,\hat{\mathbf{x}}_C]$.

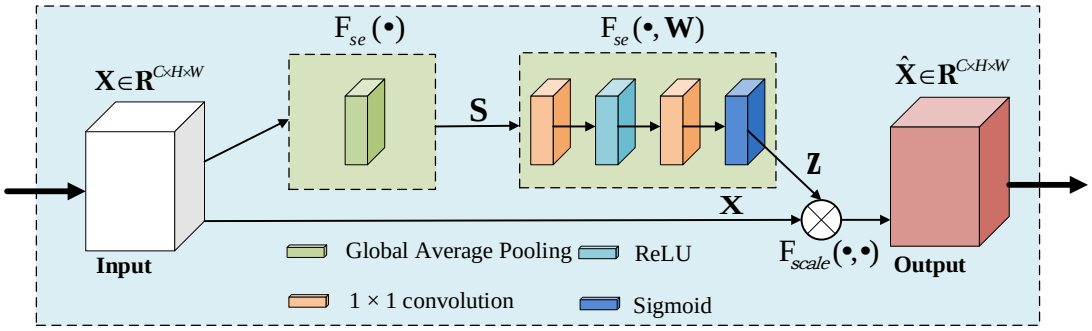
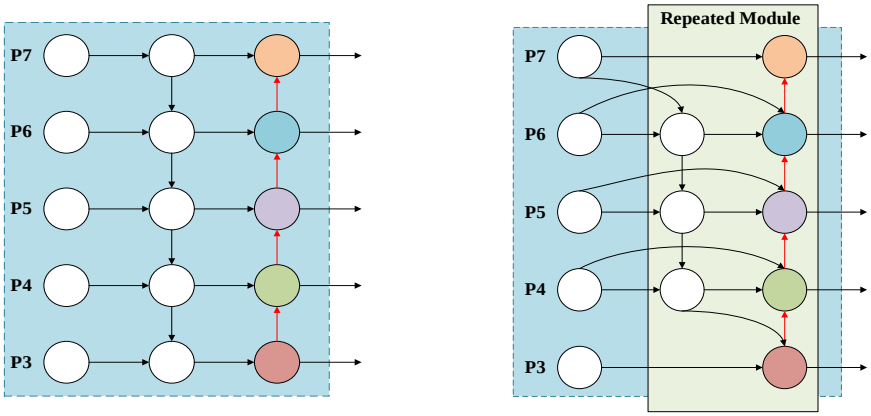


Fig. 5 - The structure of SE module

Enhanced feature fusion



(a) PANet *(b) BiFPN*

Fig. 6 - Structures of PANet and BiFPN modules

In the original YOLOv5s, the neck network performs convolution, up-sampling, and concatenation operations on features output from the YOLOv5s' backbone before being sent to the detection layer. In this process, the PANet structure in the feature pyramid is utilized to establish connections between different level feature information, as seen in Fig. 6(a). In PANet, the feature fusion process is achieved with a direct summation operation, which can potentially cause some loss of critical feature information originally extracted from the YOLOv5 backbone, particularly important features of small objects in shallow layers. However, early disease spots on rice leaves and panicles often manifest as small targets, and the characterization scale for diseases like leaf blast can vary significantly. Moreover, these small target lesions frequently coexist with larger targets such as healthy leaves. During the training process, the model may prioritize learning from the larger targets, hereby diminishing attention to or neglecting the small target lesions. This can adversely affect the accuracy in distinguishing between similar representations of diseases and small target lesions in images. Therefore, it becomes crucial to improve the ability to fuse features at different scales of the model through weighted designs.

To effectively fuse multi-scale features, a bidirectional feature pyramid network (BiFPN) (*Li et al., 2022*) is introduced and integrated in the YOLOv5 neck. As illustrated in Fig. 6(b), BiFPN does not utilize a single top-down and bottom-up path like PANet. Instead, BiFPN employs a pair of paths as a feature layer, fusing different level features from top to bottom and bottom to top. It assigns additional weights to each input, enabling the model to gradually learn the importance of each feature during the training process of feature aggregation. In this study, BiFPN is integrated to replace the Concat operation in the neck network according to the fact that different input features possess varying resolutions and contribute differently to feature fusion. This improvement based on BiFPN module can enhance the transmission of feature information crossed different network layers, and allow the model to better understand and learn the importance of features by integrating features from multiple levels both top-down and bottom-up. Hence, it can make all levels of the feature pyramid better balanced to provide a more global and semantic feature representation, which is helpful to accurately detect rice blast disease in complex environments, and improve the detection performance of different scale targets.

Lightweight neck network

Despite the lightweight in the YOLOv5s' backbone due to the aforementioned improvements, the feature fusion process still makes a significant contribution to the total number of parameters, affecting the speed of fusion. To achieve a lightweight neck network, standard convolution is replaced with the efficient GhostConv. Additionally, LightBottleneck module is designed based on the DWConv module, since DWConv extracts local detail features while utilizing fewer parameters compared to traditional convolution, thereby reducing the model's computational burden (*Liu et al., 2023*). The structure of LightBottleneck is given in Fig. 7. The LightBottleneck replaces the Bottleneck design in the original C3, resulting in the creation of a new module called LightC3. This reduces the parameter count associated with standard convolution, leading to a smaller model size. Furthermore, to reduce the redundant feature information introduced by the LightBottleneck and

to enhance feature discrimination and representation between channels during the fusion of multiple scale features without significantly increasing the computational load, the SE module is incorporated into the LightBottleneck module. As shown in Fig. 7, the SE module calculates the weights for each channel by leveraging global average pooling in conjunction with fully connected layers, allowing for adaptive adjustment of each channel's weight in the feature maps. It enables the network to better leverage multi-scale inputs by emphasizing significant features while diminishing irrelevant ones. Thus, it enhances the network's capacity to percept the key information about the disease characteristics, especially crucial features of small objects. In summary, the LightBottleneck module is designed based on DWConv and SE modules, which can enhance efficiency and accuracy in obtaining information related to disease features, thus lightweight model while maintaining accuracy in detecting rice blast disease.

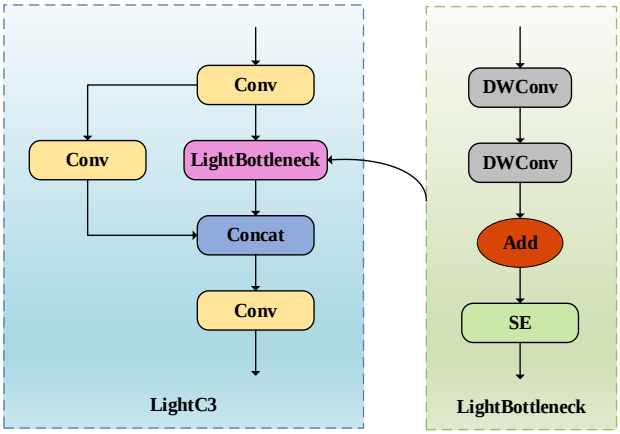


Fig. 7 - The structure of LightC3 module

Loss function

The loss function used in YOLOv5 is $CIoU$, which is expressed as follows:

$$L_{CIou} = 1 - IoU + \frac{\rho^2(b^p, b^{gt})}{c^2} + \alpha v \tag{5}$$

where IoU is the intersection ratio of a predicted box b^p and a ground-truth box b^{gt}; c denotes the length of the shortest diagonal of the smallest bounding box that encompasses both b^p and b^{gt}; $\rho^2(b^p, b^{gt})$ is the Euclidean distance between the center points of b^p and b^{gt}; α and v are used to calculate the discrepancy of the width-to-height ratio of b^p and b^{gt}, which are calculated as follows:

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{6}$$

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{7}$$

where w and h, h^{gt} and w^{gt} are the height and width of b^p, and height and width of b^{gt}, respectively.

From Equation (5), it is important to note that CIoU involves the overlap area, aspect ratio, and the distance between the central points when calculating the loss between b^p and b^{gt}. However, it does not account for the differences in width and height between b^p and b^{gt}, resulting in a slower convergence speed. Thus, the EIoU Loss (*Zhang et al., 2022*) is used in this study, which is calculated as follows:

$$\begin{aligned} L_{EIoU} &= L_{IoU} + L_{dis} + L_{asp} \\ &= 1 - IoU + \frac{\rho^2(b^p, b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \end{aligned} \tag{8}$$

where w^c and h^c represent the width and height of the smallest box that encompasses b^p and b^{gt}. In Equation (8), the loss function is composed of L_{IoU}, L_{dis} and loss L_{asp}. These components effectively quantify the differences in overlap area, central point position, and the variations in width and height between b^p and b^{gt}, leading to faster convergence and accurate localization.

RESULTS AND ANALYSIS

Experimental setup and evaluation metrics

The experimental setup used a GeForce RTX 4080 16G GPU, an Intel I7-11700K CPU running at 2.5 GHz, and operating on Windows 11. The software involved Python 3.9, CUDA 12.1, and the PyTorch 1.13.0 framework. The initial learning rate was 0.01. The batch size was 64. The training epochs were 300. The stochastic gradient descent (SGD) optimizer was employed, weight decay at 0.0005. The momentum was 0.937. To assess model complexity and detection performance, precision (*P*), recall (*R*), the number of parameters, FLOPs, and mean average precision (m*AP*) were considered, which were calculated as follows:

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$\mathrm{m}AP = \frac{1}{N}\sum_{i=1}^{N}\int_{0}^{1} P dR \tag{11}$$

where TP denote positive samples that were correctly detected, while FP refer to negative samples incorrectly classified as positive, and FN represent positive samples that were not detected. N indicates the total number of categories. Higher $\mathrm{m}AP$ values indicate better detection performance.

Ablation experiment results

The YOLOv5s model is the basic framework for this study. The use of GhostConv and LightC3 modules contributed to lightweight model. Additionally, BiFPN was introduced for feature fusion, while SE attention mechanism was added to strengthen the focus on the disease features. To assess the effect of each introduced module on the performance of the rice blast disease detection, ablation experiments were conducted on eight models using the same training parameters. Table 2 displays comparisons of evaluation metrics for various enhancement methods, with "√" marking the inclusion of each improvement strategy in the network. The detection results for each model are show in Fig. 8.

From Table 2 and Fig. 8, it is observed that the integration of the LightC3 module reduced model parameters as well as FLOPs, the recall and mAP by 14.39%, 13.29%, respectively, and achieved rice blast detection results that were on par with YOLOv5s. The addition of GhostConv alone led to a reduction of 33.9% in FLOPs and a 0.6% decrease in mAP and increased the precision of leaf blast and ear blast by 0.9% and 0.2%, respectively. These show that adding the LightC3 or GhostConv module notably decreased model parameters and FLOPs, with only a minor reduction of mAP, indicating that integrating the LightC3 and GhostConv modules into the YOLOv5s model is a practical solution for deployment on resource-constrained mobile devices. Precision, recall and mAP metrics are improved with the introduction of BiFPN only compared to YOLOv5s. As depicted in Fig. 8, compared with YOLOv5s, the YOLOv5-B obtained better detection results for ear blast and leaf blast, indicating that BiFPN can improve detection of targets at different scales. Furthermore, compared with the baseline model, the model that incorporated the BiFPN module along with LightC3 or GhostConv achieved reductions in model parameters by 15.67% and 35.19%, respectively, and increased the precision. With the simultaneous introduction of LightC3, GhostConv and BiFPN modules, model parameters and FLOPs were reduced by 49.93% and 50.63% respectively. However, it causes a decrease in recall and mAP by 2.5% and 1.4%, respectively, since the detection results of YOLOv5-BGL on small-scale targets were worse than those of other models, as shown in Fig. 8. This proposed model incorporates the SE module within the YOLOv5s' backbone with incorporation of the LightC3, GhostConv and BiFPN modules achieved 3.58M parameters and 7.9 G FLOPs, which reduced by 49% and 50% compared with YOLOv5s. In terms of accuracy, the detection performance was comparable to YOLOv5s, with only 0.6% and 1.0% decrease in recall and mAP, respectively. Additionally, the proposed model integrated the SE module in the neck network, which can enhance feature discrimination and representation between channels in the multi-scale feature fusion process without adding too much calculation amount, thus yielding superior performance over YOLOv5-BGL with respect to detecting different scale disease targets. Thus, this demonstrates that the proposed model obtains an optimal balance between disease identification performance and lightweight.

Table 2

Results of ablation experiments

Light C3	Ghost Conv	BiFPN	SE	Model	P (leaf blast)	P (ear blast)	R	mAP	Params (M)	FLOPs (G)
				YOLOv5s	0.977	0.963	0.936	0.976	7.02	15.8
√				YOLOv5-L	0.977	0.958	0.931	0.967	6.01	13.7
	√			YOLOv5-G	0.986	0.965	0.938	0.970	4.64	10.0
		√		YOLOv5-B	0.992	0.985	0.938	0.978	6.90	15.7
√		√		YOLOv5-BL	0.982	0.969	0.934	0.967	5.92	13.6
	√	√		YOLOv5-BG	0.987	0.972	0.938	0.976	4.55	9.90
√	√	√		YOLOv5-BGL	0.966	0.932	0.911	0.962	3.55	7.80
√	√	√	√	Ours	0.986	0.963	0.930	0.966	3.58	7.90

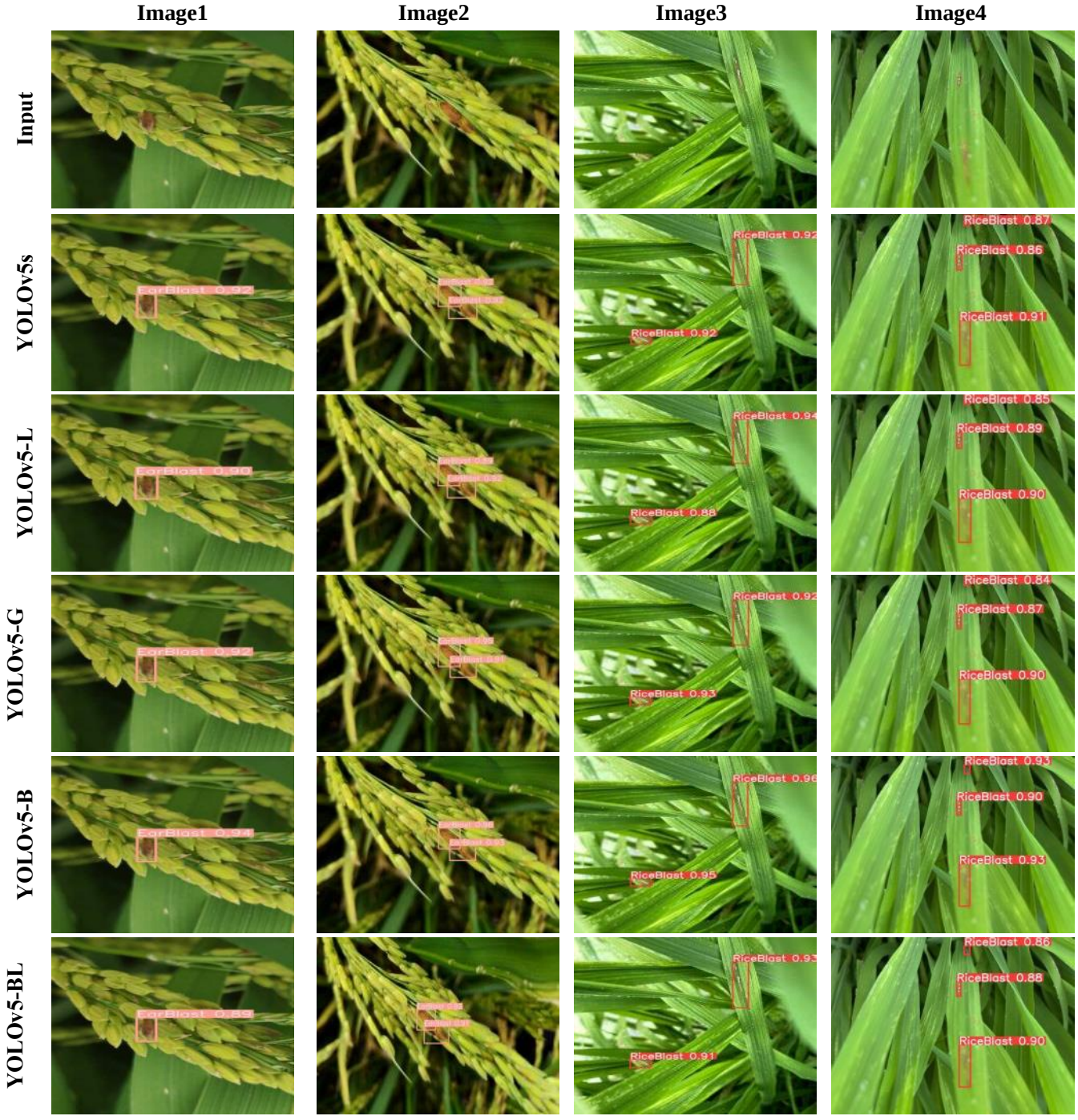


Fig. 8 - Rice blast detection results of each model based on YOLOv5s

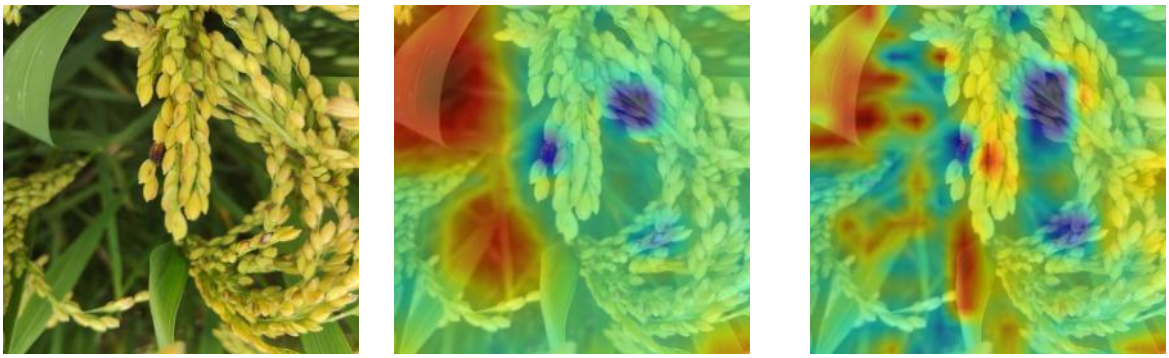
Performance comparison using different attention mechanisms

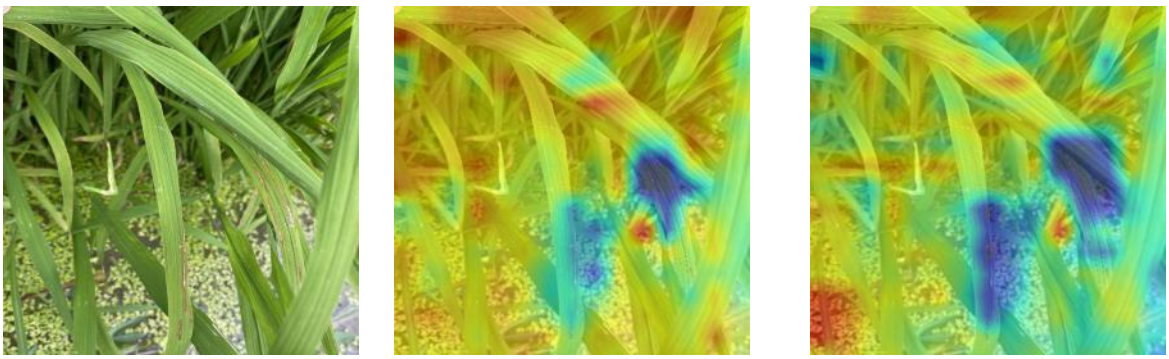
To assess the effectiveness of the incorporated SE module, the additional attention mechanisms such as SimAM (*Yang et al., 2021*), CA (*Hou et al., 2021*) and CBAM (*Woo et al., 2018*) were compared. Table 3 shows comparison results with the use of various attention mechanisms on our dataset. Our improved model demonstrated optimal performance including precision, recall, mAP, and the number of parameters, as well as FLOPs, compared to the other two attention modules. Moreover, Fig. 9 shows the heat map of the YOLOv5-BGL without SE and after adding SE, which can intuitively display the degree of regional attention and the higher the feature is, the higher the heat map value at the corresponding position. It can be noted that the heat map shown in Fig. 9(c) shows that the extracted features of small rice blast are more focused and accurate compared with Fig. 9(b). As a result, incorporating the SE module led to improved accuracy in perception of disease features, especially for smaller target objects.

Table 3

Results of the comparison of different attention mechanisms

Model	P	R	mAP	Params(M)	FLOPs(G)
YOLOv5-BGL	0.949	0.911	0.962	3.55	7.80
YOLOv5-BGL +SimAM	0.960	0.900	0.949	3.55	7.80
YOLOv5-BGL +CA	0.957	0.913	0.955	3.57	7.90
YOLOv5-BGL +CBAM	0.949	0.905	0.955	4.01	8.70
YOLOv5-BGL +SE(Ours)	0.975	0.930	0.966	3.58	7.90





(a) Input *(b) Heat map without SE module* *(c) Heat map with SE module*

Fig. 9 - The heat map of the final output result of the model

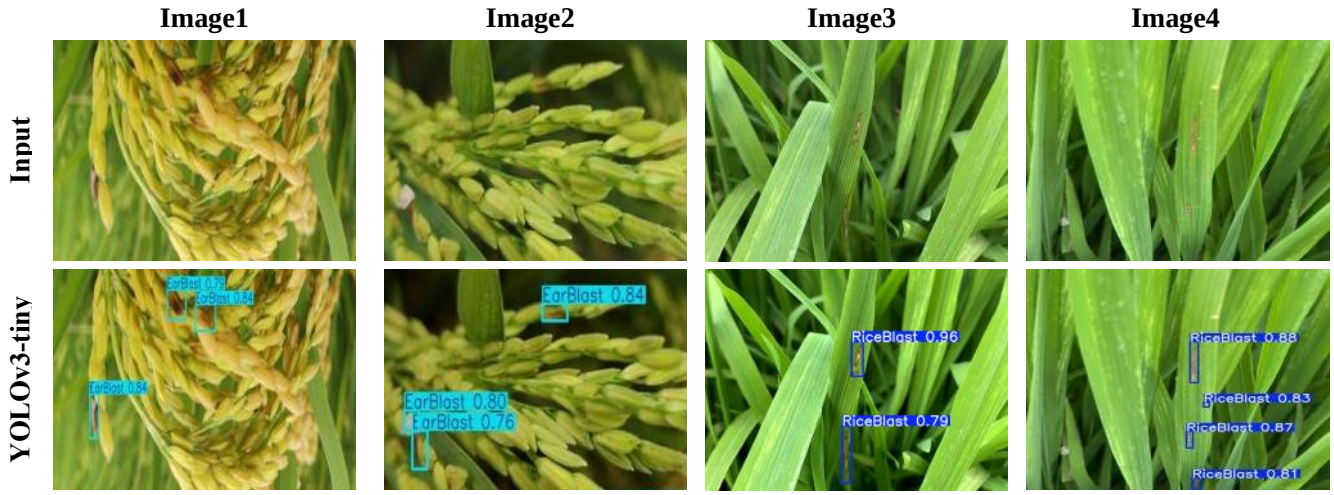
Performance comparison using different detection model

To evaluate the efficiency and lightweight of the model for rice blast disease recognition, our model was compared to the YOLOv3-tiny, YOLOv4-tiny, YOLOv5s, YOLOv7-tiny, and YOLOv8s. The performance comparisons are given in Table 4. As shown in Table 4, while the YOLOv8s achieves high scores in detection accuracy of rice blast disease, it does not compare favorably to the proposed model in terms of parameters, FLOPs as well as model size. Comparison with the other lightweight models including YOLOv3-tiny, YOLOv4-tiny, and YOLOv7-tiny, the proposed model outperforms them across all metrics. Compared with the YOLOv5s model, the reductions in precision, recall, and mAP do not exceed 2%. However, there is a 49% decrease in model parameters, a 50% decrease in FLOPs, and a 74.61% decrease in model size, respectively. Additionally, as shown in Fig. 10, the detection performance of the YOLOv5s, YOLOv8s and the proposed model does not differ much in the case of rice blast detection. However, the target confidence of the YOLOv7-tiny and YOLOv4-tiny models were lower than that of the other three models, while in early disease spots on rice leaves and panicles, the YOLOv7-tiny and YOLOv4-tiny models, on the other hand, demonstrated some missing detections. Hence, it can be concluded that the proposed lightweight model outperforms comparable algorithms for rice blast detection, featuring fewer FLOPs and a more lightweight architecture that is suitable for mobile deployment.

Table 4

Results of the comparison of different detection model

model	P	R	mAP	Params(M)	FLOPs(G)	Size(M)
YOLOv3-tiny	0.891	0.805	0.871	8.67	12.9	17.5
YOLOv4-tiny	0.759	0.387	0.337	6.01	16.2	23.57
YOLOv5s	0.977	0.936	0.976	7.01	15.8	14.1
YOLOv7-tiny	0.761	0.714	0.774	6.01	13.2	12.28
YOLOv8s	0.986	0.942	0.977	11.1	28.4	22.54
Ours	0.975	0.930	0.966	3.58	7.90	7.61



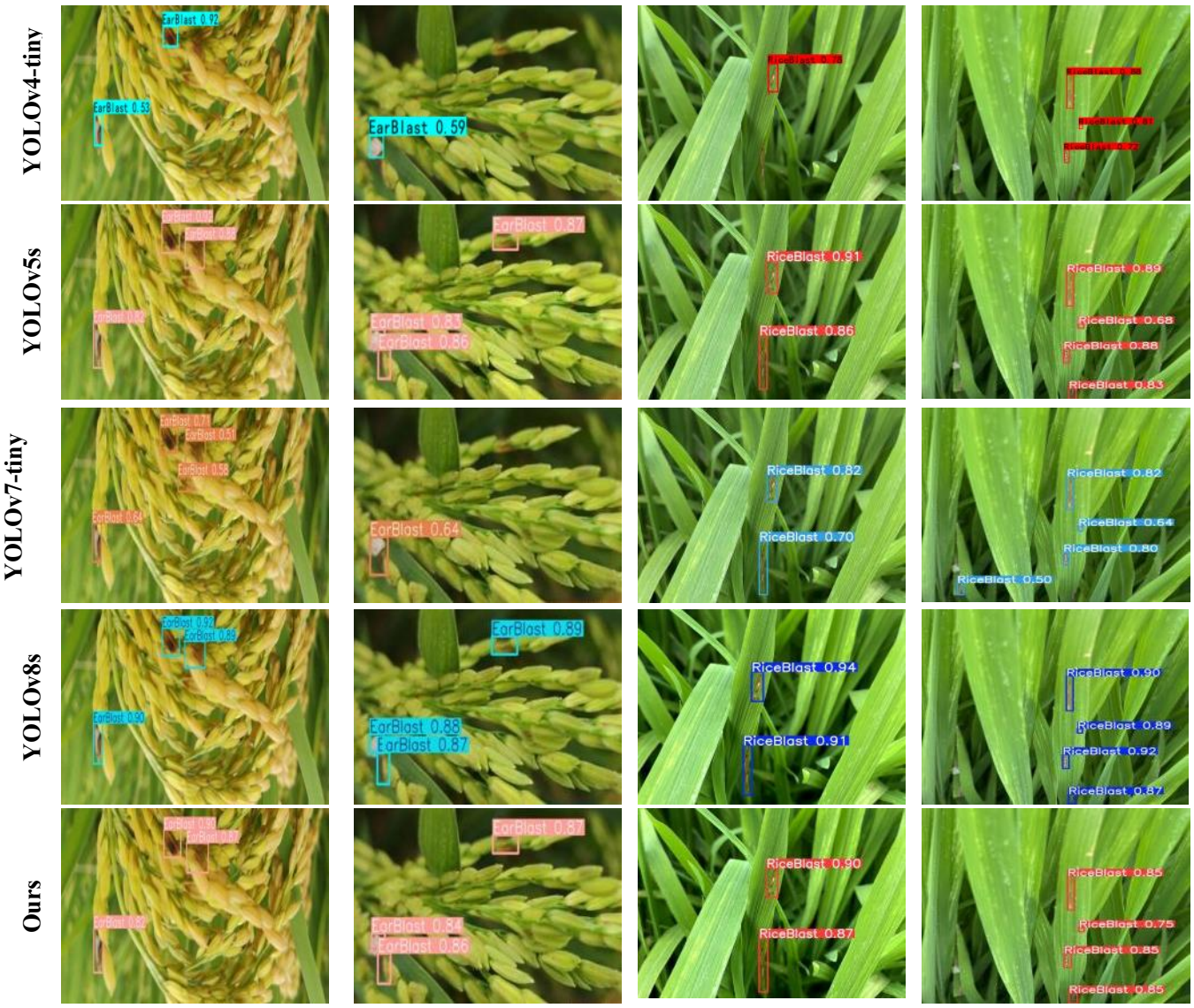


Fig. 10 - Comparison of rice blast detection results

Android deployment experimentation

The NCNN framework was used to quantify the trained rice blast model and then created an application RiceBlastDetector based on the model, as illustrated in Fig.10. The app was packaged as an APK file, generated using Android Studio, and installed on a Redmi K30 Pro smartphone running Android 12. As shown in Fig.11, the app's main interface features a unit for displaying image framing, photo button and video button. The user can use images saved on the phone or call the smartphone camera through “photo” or “video” button. The image framing display unit is capable of showing labels for disease categories, confidence scores, and predicted boxes. The effect of rice blast detection using the APP in the field is shown in Fig. 12. It can be seen from Fig.12 that the APP can achieve accurate detection in the scenario with the different characterization scale disease spots.

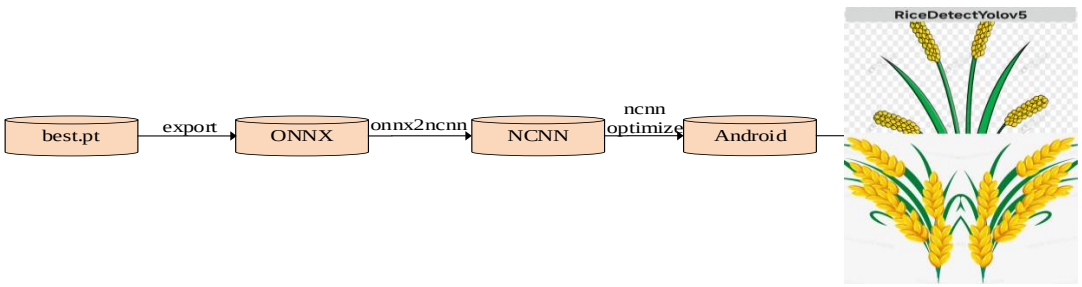


Fig. 11 - Model quantification and APP interface

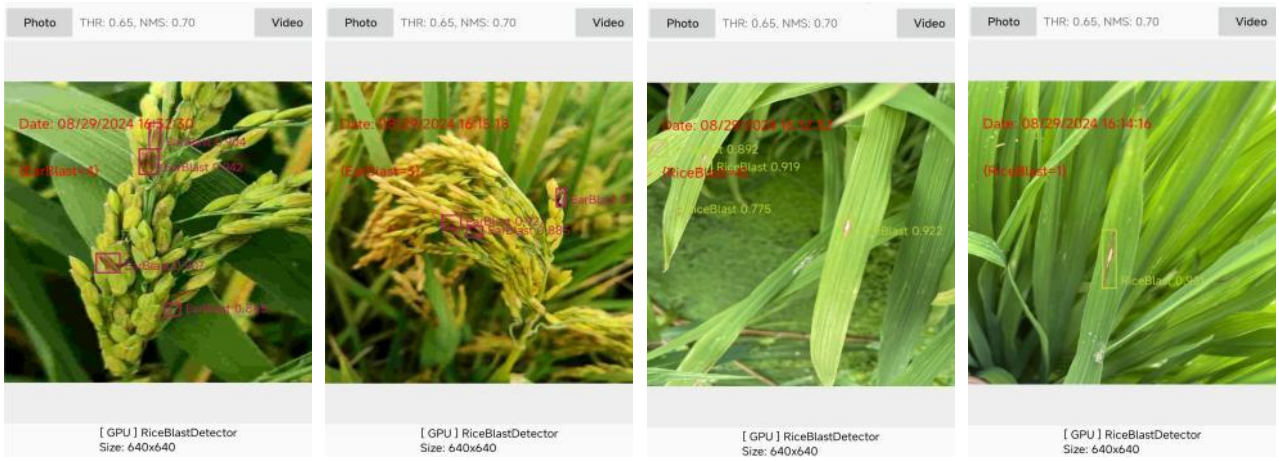


Fig. 12 - Detection rice blast based on APP

CONCLUSIONS

To effectively lightweight the rice blast detection model under the premise of ensuring the recognition performance, a lightweight YOLOv5s model was designed, verified and deployed on Android devices, facilitating detection of rice blast. The model replaced the standard convolution in its backbone and neck with GhostConv and implemented the LightC3 module, reducing complexity. The BiFPN module was used for multi-scale feature fusion, and the SE module was integrated into the backbone to enhance detection precision. The RiceDiseaseDetector app was developed and used for field detection. The results indicate that the designed model has a 49% decrease in model parameters, a 50% decrease in FLOPs, and a 74.61% decrease in model size, respectively, compared with the YOLOv5s model, which provides technical support for future intelligent plant protection initiatives. Furthermore, this approach can serve as a reference for detecting other minor rice disease-related issues.

ACKNOWLEDGEMENT

Thank you to all the participating authors. The research was funded by the National Key Research and Development Program (2022YFD2001801-3) and the National Natural Science Foundation of China (No.32201665).

REFERENCES

[1] Asibi, A. E., Chai, Q., & Coulter, J. A. (2019). Rice blast: A disease with implications for global food security [J]. *Agronomy*, 9(8), 451.

[2] Bedi, P., & Gole, P. (2021). Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network [J]. *Artificial Intelligence in Agriculture*, 5, 90-101.

[3] Cheng, Z., Huang, R., Qian, R., Dong, W., Zhu, J., & Liu, M. (2022). A lightweight crop pest detection method based on convolutional neural networks. *Applied Sciences*, 12(15), 7378.

[4] Daniya, T., & Vigneshwari, S. (2019). A review on machine learning techniques for rice plant disease detection in agricultural research [J]. *System*, 28(13), 49-62.

[5] Dogra, R., Rani, S., Singh, A., Albahar, M. A., Barrera, A. E., & Alkhayyat, A. (2023). Deep learning model for detection of brown spot rice leaf disease with smart agriculture. *Computers and Electrical Engineering*, 109, 108659.

[6] Han, K., Wang, Y., Xu, C., Guo, J., Xu, C., Wu, E., & Tian, Q. (2022). GhostNets on heterogeneous devices via cheap operations. *International Journal of Computer Vision*, 130(4), 1050-1069.

[7] Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 13713-13722).

[8] Jia, L., Wang, T., Chen, Y., Zang, Y., Li, X., Shi, H., & Gao, L. (2023). MobileNet-CA-YOLO: An improved YOLOv7 based on the MobileNetV3 and attention mechanism for Rice pests and diseases detection. *Agriculture*, 13(7), 1285.

[9] Khan, F., Zafar, N., Tahir, M. N., Aqib, M., Waheed, H., & Haroon, Z. (2023). A mobile-based system for maize plant leaf disease detection and classification using deep learning [J]. *Frontiers in Plant Science*, 14, 1079366.

[10] Khattak, A., Asghar, M. U., Batool, U., Asghar, M. Z., Ullah, H., Al-Rakhami, M., & Gumaei, A. (2021). Automatic detection of citrus fruit and leaves diseases using deep neural network model [J]. *IEEE* access, 9, 112942-112954.

[11] Li, R., Li, Y., Qin, W., Abbas, A., Li, S., Ji, R., Wu Y., & Yang, J. (2024). Lightweight Network for Corn Leaf Disease Identification Based on Improved YOLO v8s. *Agriculture*, 14(2), 220.

[12] Li, T., Zhang, Y., Li, Q., & Zhang, T. (2022). AB-DLM: an improved deep learning model based on attention mechanism and BiFPN for driver distraction behavior detection. *IEEE* Access, 10, 83138-83151.

[13] Lin, Y., Chen, T., Liu, S., Cai, Y., Shi, H., Zheng, D., Lan, Y., & Zhang, L. (2022). Quick and accurate monitoring peanut seedlings emergence rate through UAV video and deep learning. *Computers and Electronics in Agriculture*, 197, 106938.

[14] Liu, K., Wang, J., Zhang, K., Chen, M., Zhao, H., & Liao, J. (2023). A lightweight recognition method for rice growth period based on improved YOLOv5s. *Sensors*, 23(15), 6738.

[15] Malvade, N. N., Yakkundimath, R., Saunshi, G., Elemmi, M. C., & Baraki, P. (2022). A comparative analysis of paddy crop biotic stress classification using pre-trained deep neural networks [J]. *Artificial Intelligence in Agriculture*, 6, 167-175.

[16] Manavalan, R. (2020). Automatic identification of diseases in grains crops through computational approaches: A review [J]. *Computers and Electronics in Agriculture*, 178, 105802.

[17] Nandhini, M., Kala, K. U., Thangadarshini, M., & Verma, S. M. (2022). Deep learning model of sequential image classifier for crop disease detection in plantain tree cultivation. *Computers and Electronics in Agriculture*, 197, 106915.

[18] Rahman, C. R., Arko, P. S., Ali, M. E., Khan, M. A. I., Apon, S. H., Nowrin, F., & Wasif, A. (2020). Identification and recognition of rice diseases and pests using convolutional neural networks. *Biosystems Engineering*, 194, 112-120.

[19] Sanghavi, K., Sanghavi, M., & Rajurkar, A. M. (2021). Early stage detection of Downey and Powdery Mildew grape disease using atmospheric parameters through sensor nodes [J]. *Artificial Intelligence in Agriculture*, 5, 223-232.

[20] Sun, Y., Wu, F., Guo, H., Li, R., Yao, J., & Shen, J. (2023). TeaDiseaseNet: multi-scale self-attentive tea disease detection [J]. *Frontiers in Plant Science*, 14, 1257212.

[21] Wang, H., Xu, Y., He, Y., Cai, Y., Chen, L., Li, Y., Sotelo, M. A., & Li, Z. (2022). YOLOv5-Fog: A multiobjective visual detection algorithm for fog driving scenes based on improved YOLOv5. *IEEE Transactions on Instrumentation and Measurement*, 71, 1-12.

[22] Wen, Y., Xue, J., Sun, H., Song, Y., Lv, P., Liu, S., Chu, Y., & Zhang, T. (2023). High-precision target ranging in complex orchard scenes by utilizing semantic segmentation results and binocular vision. *Computers and Electronics in Agriculture*, 215, 108440.

[23] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 3-19).

[24] Xin, D., Chen, Y. W., & Li, J. (2020). Fine-grained butterfly classification in ecological images using squeeze-and-excitation and spatial attention modules. *Applied Sciences*, 10(5), 1681.

[25] Li, Y., Xiao, L., Li, W., Li, H., Liu, J. (2022). Research on recognition of occluded orange fruit on trees based on YOLOv4. *INMATEH-Agricultural Engineering*, 67(2), 137-146. DOI: https://doi.org/10.35633/inmateh-67-13

[26] Yang, N., Yu, J., Wang, A., Tang, J., Zhang, R., Xie, L., Shu F., & Kwabena, O. P., (2020). A rapid rice blast detection and identification method based on crop disease spores' diffraction fingerprint texture [J]. *Journal of the Science of Food and Agriculture*, 100(9), 3608-3621.

[27] Yang, L., Zhang, R. Y., Li, L., & Xie, X. (2021, July). Simam: A simple, parameter-free attention module for convolutional neural networks [C]. In *International conference on machine learning* (pp. 11863-11874). PMLR.

[28] Yang, Y., Wang, L., Huang, M., Zhu, Q., & Wang, R. (2022). Polarization imaging based bruise detection of nectarine by using ResNet-18 and ghost bottleneck. Postharvest Biology and Technology, 189, 111916.

[29] Xia, Y., Lei, X., Herbst, A., & Lyu, X. (2023). Research on pear inflorescence recognition based on fusion attention mechanism with YOLOv5. *INMATEH-Agricultural Engineering*, 69(1), 11-20. DOI: https://doi.org/10.35633/inmateh-69-01

[30] Du, Y., Gao, A., Song, Y., Guo, J., Ma, W., Ren, L. (2024). Young apple fruits detection method based on improved YOLOv5. *INMATEH-Agricultural Engineering*, 73(2), 84-93.

[31] Yu, C., Feng, J., Zheng, Z., Guo, J., & Hu, Y. (2024). A lightweight SOD-YOLOv5n model-based winter jujube detection and counting method deployed on Android. *Computers and Electronics in Agriculture*, 218, 108701.

[32] Zhang, Y. F., Ren, W., Zhang, Z., Jia, Z., Wang, L., & Tan, T. (2022). Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing*, 506, 146-157.

[33] Zhou, G., Zhang, W., Chen, A., He, M., & Ma, X. (2019). Rapid detection of rice disease based on FCM-KM and faster R-CNN fusion. *IEEE* access, 7, 143190-143206.