

## Article

# GSEYOLOX-s: An Improved Lightweight Network for Identifying the Severity of Wheat Fusarium Head Blight

Rui Mao <sup>1</sup>, Zhengchao Wang <sup>1</sup>, Feilong Li <sup>1</sup>, Jia Zhou <sup>2</sup>, Yinbing Chen <sup>1</sup> and Xiaoping Hu <sup>2,3,\*</sup><sup>1</sup> College of Information Engineering, Northwest A&F University, Yangling, Xianyang 712100, China<sup>2</sup> College of Plant Protection, Northwest A&F University, Yangling, Xianyang 712100, China<sup>3</sup> Key Laboratory of Integrated Pest Management on Crops in Northwestern Loess Plateau, Ministry of Agriculture and Rural Affairs, Yangling, Xianyang 712100, China

\* Correspondence: xphu@nwafu.edu.cn

**Abstract:** *Fusarium* head blight (FHB) is one of the most detrimental wheat diseases. The accurate identification of FHB severity is significant to the sustainable management of FHB and the guarantee of food production and security. A total of 2752 images with five infection levels were collected to establish an FHB severity grading dataset (FHBSGD), and a novel lightweight GSEYOLOX-s was proposed to automatically recognize the severity of FHB. The simple, parameter-free attention module (SimAM) was fused into the CSPDarknet feature extraction network to obtain more representative disease features while avoiding additional parameters. Meanwhile, the ghost convolution of the model head (G-head) was designed to achieve lightweight and speed improvements. Furthermore, the efficient intersection over union (EIoU) loss was employed to accelerate the convergence speed and improve positioning precision. The results indicate that the GSEYOLOX-s model with only 8.06 MB parameters achieved a mean average precision (mAP) of 99.23% and a detection speed of 47 frames per second (FPS), which is the best performance compared with other lightweight models, such as EfficientDet, Mobilenet-YOLOV4, YOLOV7, YOLOX series. The proposed GSEYOLOX-s was deployed on mobile terminals to assist farmers in the real-time identification of the severity of FHB and facilitate the precise management of crop diseases.

**Keywords:** *Fusarium* head blight; wheat diseases severity identification; lightweight network architecture; YOLOX; attention mechanism



**Citation:** Mao, R.; Wang, Z.; Li, F.; Zhou, J.; Chen, Y.; Hu, X.

GSEYOLOX-s: An Improved Lightweight Network for Identifying the Severity of Wheat Fusarium Head Blight. *Agronomy* **2023**, *13*, 242.

<https://doi.org/10.3390/agronomy13010242>

Academic Editor: Haiguang Wang

Received: 29 November 2022

Revised: 10 January 2023

Accepted: 12 January 2023

Published: 13 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Wheat (*Triticum aestivum* L.) is one of the world's three major food crops. In 2021, global wheat production exceeded 776 million tons, of which China, with 136 million tons ranked first, accounting for ca. 17.6% of the world's total wheat production [1]. However, *Fusarium* head blight (FHB) is a destructive disease that restricts the safety of wheat production and food quality [2,3]. In epidemic years, FHB can reduce the yield loss of wheat by 10–70%, influencing more than 7 million hectares of wheat-planting areas [4]. The classification of FHB severity is crucial in making decisions on its control. The inaccurate identification of FHB severity will lead to ineffective control and fungicide abuse [5]. Therefore, it is essential to conduct a precise method to identify the severity of FHB.

Traditionally, the identification of the severity of FHB mostly depends on technicians' continuous field investigation. Disease severity based on scale proportions is usually estimated using the naked eye depending on the number or the area of the lesions in the whole wheat ear. This method is not only time-consuming but also biased and unreliable. Deep convolutional neural networks (DCNNs) [6] are state-of-art neural networks, which have the ability of self-learning and contribute to the automatic recognition and severity estimation of crop diseases. Esgario et al. [7] used the ResNet50 model with 25 MB parameters to estimate the severity caused by four diseases on coffee leaves, such as

cercospora leaf spot, brown leaf spot, rust, and leaf miner, which achieved an 86.51% accuracy. However, the average time required by this model to train one epoch was 21.90 s. Ji et al. [8] employed images of single crop leaves with a simple background to identify ten crop disease types and three crop disease severity levels (normal, general, and serious samples), and adopted the ResNet50 model improved by multi-label binary correlation CNN (BR-CNN) with an accuracy of 86.70% and the parameters of 23 MB. Mi et al. [9] developed a novel GradCAM++ network based on a densely connected convolutional network (DenseNet) and a convolutional block attention module (CBAM) to identify the severity of wheat stripe rust and the accuracy of the severity level improved to 97.99%. Zhang et al. [10] proposed a segmentation model of single wheat ear and the FHB lesion segmentation model based on their self-constructed IABC-K-PCNN. According to the ratio between the lesion and the whole wheat ear, the speed of the FHB classification was about 5 s per photo, with an accuracy of 92.5%. Although satisfactory results are reported in the above studies based on large CNNs, there will be a problem in the practical application of disease severity identification due to the high number of parameters, the large storage space, and computational consumption.

In order to reduce the parameters and speed up network training, some lightweight CNNs have been gradually applied in plant disease detection research. Bao et al. [11] developed SimpleNet using convolution and inverted residual blocks that achieved a 94.10% recognition accuracy of wheat ear diseases. Hong et al. [12] improved the YOLOv4 by a lightweight network, non-maximum suppression (NMS), and complete intersection over union (CIoU) to achieve model lightweight improvement and deploy to UAV to identify FHB. Liu et al. [13] improved the YOLOX-Nano model by introducing blueprint-separable convolution (BSCConv), attention mechanism, and the asymmetric ShuffleBlock. This study provided a feasible solution for the real-time detection of apple leaf disease with a 91.08% accuracy. Although the parameters of the above network models were smaller than those of the common DCNNs, these models did not directly identify the severity of FHB. In recent years, researchers have started to concentrate on extracting critical feature information and improving the performances of models by incorporating the attention mechanism into the networks. Cui et al. [14] introduced the CBAM into the autoencoder to identify maize leaf diseases from PlantVillage in laboratory scenarios, and obtained a 99.44% identification accuracy. Li et al. [15] applied a hybrid attention, the Atrous Space Pyramid Pool (ASPP), to optimize DeepLab V3+ to accurately segment lesions and automatically assess the severity of cucumber downy mildew and powdery mildew. The above studies demonstrated that using different strategies to optimize neural networks can further improve the recognition accuracy and training speed. However, considering the small and subtle differences between the different severity levels of FHB, it is still a great challenge to establish a real-time and accurate identification model for the severity of FHB on the mobile terminal.

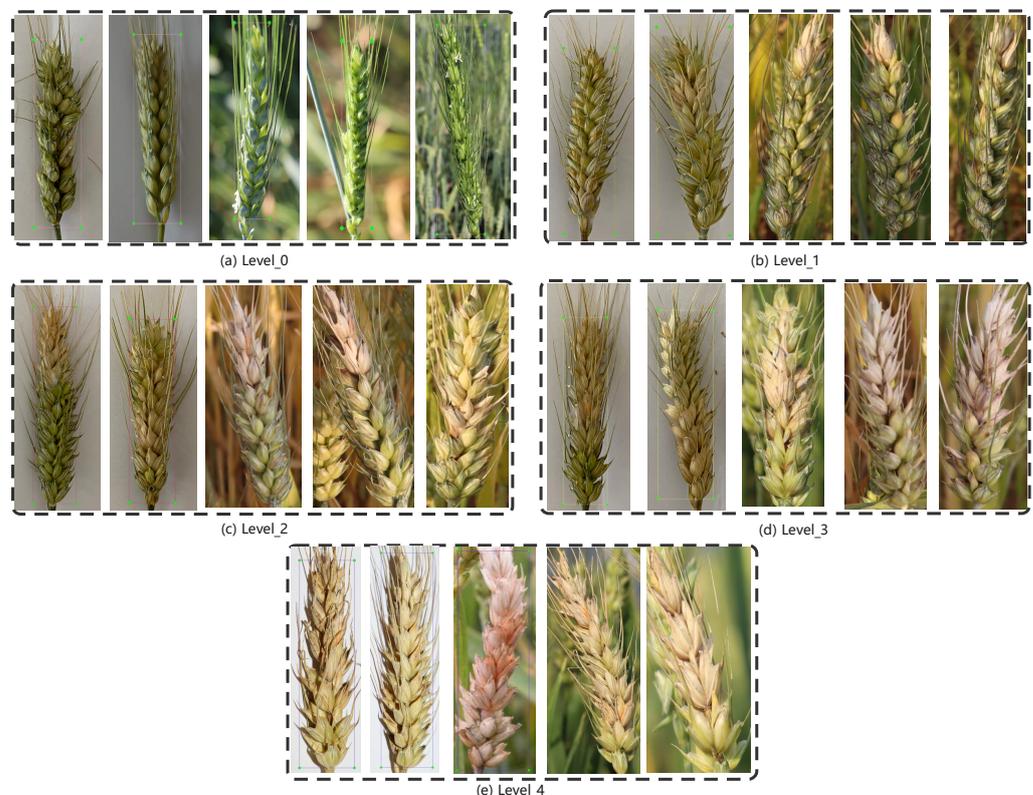
To reduce the parameters while improving the detection accuracy and speed of the classical YOLOX-s, a lightweight network named GSEYOLOX-s was proposed in this study. An FHB severity grading dataset (FHBSGD) was established, which contained a total of 2752 images with five infection levels of FHB under experimental and complex field conditions. GSEYOLOX-s model was designed by fusing a simple, parameter-free attention module (SimAM), the ghost convolution of the model head (G-head), and the efficient intersection over union (EIoU) loss function into YOLOX-s. The three improvements maximized the reduction in model parameters and facilitated the deployment of our model on the mobile terminal. The experimental results show that the proposed GSEYOLOX-s model only had 8.06 MB parameters while achieving a mean average precision (mAP) of 99.23% on the FHBSGD. The proposed model provides support for precise planting and intelligent decision-making for wheat production.

## 2. Materials and Methods

### 2.1. Data Acquisition

FHB consists in the premature fading or wilting of wheat ears that can be seen 3 weeks post *Fusarium* spp. [16]. infection, with infected spikelets showing orange or pink colors due to pathogen spores and mycelia on wheat ears [17]. According to the National Standard of the People's Republic of China for FHB (GB/T 15796-2011), the severity of FHB is divided into 5 levels ranging from Level\_0 to Level\_4. Level\_0 represents disease-free; Level\_1 indicates that the number of infected spikelets is less than 25% of all spikelets; Level\_2 indicates that the number of infected spikelets is less than 50% and more than 25% of all spikelets; Level\_3 indicates that the number of infected spikelets is less than 75% and more than 50% of all spikelets; and Level\_4 indicates that the number of infected spikelets is more than 75% of all spikelets.

The images acquisition area is Guanzhong district, Shaanxi Province, such as Wugong, Fuping, Pucheng, and Mei County. The image collection time was from 10 May 2022 to 20 May 2022 during late grouting stage of wheat. The shooting equipment include one digital camera (Canon EOS 850D) and mobile phones (Apple 11pro and Huawei Mate20), using automatic white balance and optical focusing. These devices were held 25–50 cm away from wheat ears and the time of acquiring image was from 9:00 a.m. to 18:00 p.m. The images were stored in JPEG format with a resolution of  $6000 \times 4000$  pixels and  $4032 \times 3024$  pixels. To meet the training requirements of the proposed detection model, the image size was uniformly adjusted to the size of  $640 \times 640$  pixels. According to the above standards, we classified the collected dataset into five severity levels of FHB. The dataset contained a total of 2752 images of FHB, including 556 images of Level\_0, 638 images of Level\_1, 732 images of Level\_2, 598 images of Level\_3, and 228 images of Level\_4. Specific samples of the FHBSDG are shown in Figure 1.



**Figure 1.** Wheat FHB with different severity levels under experimental and field conditions in the dataset.

## 2.2. Data Processing

Dataset annotation is essential for an identification task based on deep learning. Under the guidance of plant protection experts, we used “LabelImg” to manually label the severity grading of wheat ears in the images and generated XML-type annotation files with Pascal VOC 2007 dataset format standard. The dataset was divided into training set, validation set, and test set at 8:1:1, as shown in Table 1. To enrich the background of samples and improve the robustness against image noises, the data enhancement method in this study adopted the combination of Mixup and Mosaic [18], which was helpful to improve the detection ability of partial occlusions and subtle targets.

**Table 1.** The sample distribution in FHBSGD.

|                | Level_0 | Level_1 | Level_2 | Level_3 | Level_4 |
|----------------|---------|---------|---------|---------|---------|
| training set   | 444     | 510     | 586     | 478     | 182     |
| validation set | 56      | 64      | 73      | 60      | 23      |
| test set       | 56      | 64      | 73      | 60      | 23      |
| total          | 556     | 638     | 732     | 598     | 228     |

## 2.3. GSEYOLOX-s for FHB Severity Identification

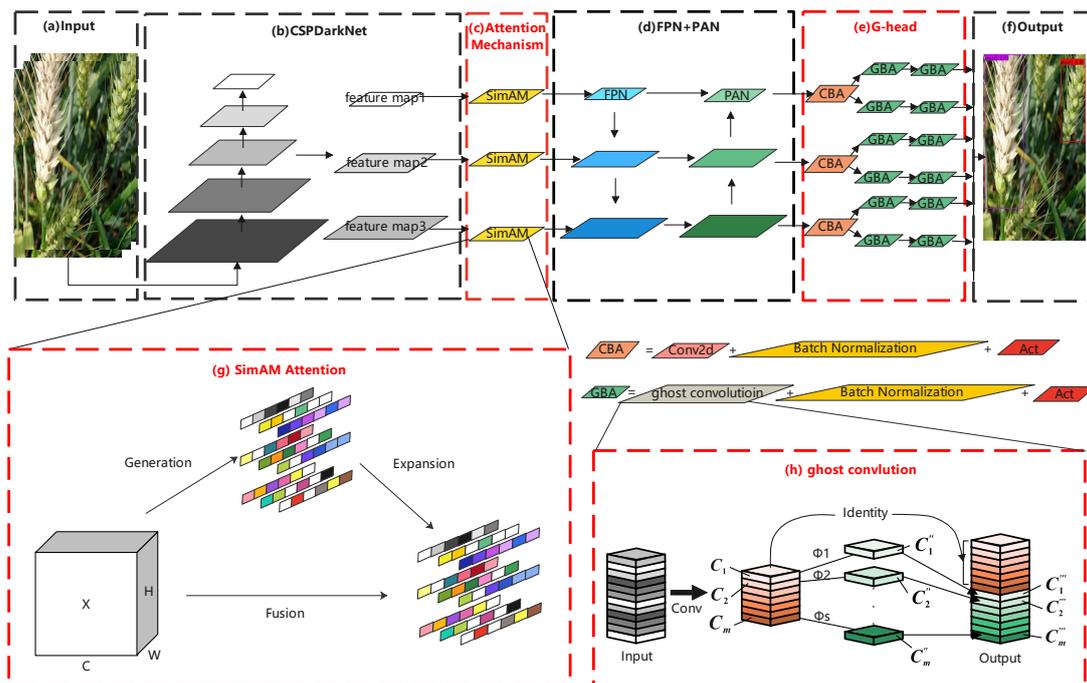
### 2.3.1. The Network Architecture

YOLOX is a new high-performance target-detection model presented by Ge and Liu et al. [19]. Its network architecture baseline includes CSPDarknet [20], feature pyramid networks (FPNs) [21] structure combined with the path aggregation networks (PANs) [22] structure to extract features, the decoupled head with the anchor-free method, and IoU loss [23] for bounding box regression. YOLOX-s, a lightweight version of YOLOX series, was selected as the base network architecture in this study. Although the parameter number of YOLOX-s had been significantly reduced compared to other versions of the YOLOX series, the amount of calculation for the mobile terminal was still greater. Moreover, there is still a challenge to accurately detect the subtle differences between different severity levels of FHB.

A lightweight network based on improved YOLOX-s, named GSEYOLOX-s (Figure 2), was proposed in this study. To fully minimize the parameters while enhancing the detection accuracy and speed of the classical YOLOX-s, three main strategies were designed in our model. Firstly, the attention mechanism, SimAM, was introduced into the CSPDarknet backbone network of YOLOX-s (Figure 2c) to focus on essential features without additional parameters, which can effectively improve the expressive ability of feature extraction by optimizing an energy function to calculate the contribution of each neuron. Secondly, G-head (Figure 2e) was applied to take full advantage of redundant feature information among similar images, thus greatly reducing parameters and ensuring the performance of the model. In addition, EIoU function was employed for rapid convergence and more accurate localization of wheat ear lesions.

### 2.3.2. A Simple, Parameter-Free Attention Module (SimAM)

Referring to the attention module of the human brain, Yang and Zhang presented a novel SimAM [24] based on the theory of neuroscience, which can directly calculate the full three-dimensional weights on the feature graph. The SimAM improved the performance of convolutional neural networks without referring to any additional parameters to calculate the attention weight on the feature map. In this study, we introduced the SimAM to the three feature maps extracted by CSPDarknet. This operator was used to boost the representation ability of feature extraction and a favorable effect for FPN and PAN structure feature fusion. The calculation procedure of the SimAM is shown in Figure 2g.



**Figure 2.** The structure of the GSEYOLO-s.

In the theory of neuroscience, the attention of the human brain evaluates the importance of each neuron, in which info-rich neurons usually display different firing modes from surrounding neurons. Furthermore, the activated neurons usually inhibit the surrounding neurons, which need to be given more attention. The study defines the energy of neurons by measuring the linear separability between neurons, which is simplified to produce the following energy function (Equation (1)) [24] to assess the importance of each neuron.

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{u})^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{1}$$

where  $\hat{u} = \frac{1}{M} \sum_{i=1}^M x_i$ ,  $\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - u_t)^2$ ,  $x_i$  is the target neuron, and  $M$  is the number of neurons on this channel. The significance of neurons can be gained by  $1/e_t^*$ , where  $e_t^*$  represents the energy value of neurons at the current position. Combined with the neuroscience theory, the lower the current neuronal energy, the greater the discrepancy between the current neuron and the surrounding neuron, the greater the importance. Then, a scaling operator refines the whole feature.

$$\tilde{X} = sig\left(\frac{1}{E}\right) \odot X \tag{2}$$

In Equation (2),  $sig$  is the sigmoid activation function,  $E$  is the integration of all  $e_t^*$  across spatial and channel dimensions, and  $X \in R^{C \times H \times W}$  is the input features. The sigmoid activation function is applied to limit the E-value range without affecting the relative importance of each neuron.

### 2.3.3. G-head Module

Traditional convolution processing has the redundancy problem of feature maps, resulting in a large amount of computational costs. The ghost convolution designed simple operations to simplify the redundant calculation process of traditional convolution [25]. By introducing the ghost convolution into the G-head, we achieved the lightweight and speed improvement of the model.

The ghost convolution (Figure 2h) first uses a primary convolution to generate  $m$  intrinsic feature maps according to the input feature information where  $m$  is less than the desired  $n$  feature maps; then, each intrinsic feature is subjected to a series of cheap linear operations to generate  $s$  ghost features according to the following function:

$$C_i''' = \Phi_i(C_j'') \quad \forall i = 1, \dots, s, j = 1, \dots, m \quad (3)$$

where  $C_j''$  is the  $j$ -th intrinsic feature map generated by primary convolution,  $\Phi_i$  is the  $i$ -th linear operation that generates the  $i$ -th ghost feature map  $C_i'''$ , and  $C_j''$  generates total  $s$  corresponding ghost feature maps through different mappings. Finally, these feature maps are linked together to form the desired  $n$  feature maps, which reduces the learning cost of non-critical features without reducing the accuracy of the model.

Equations (4) and (5) represent the calculation of parameter numbers of ordinary convolution ( $P_{bsc}$ ) and ghost convolution ( $P_{ghc}$ ), respectively.

$$P_{bsc} = c_1 c_2 k \times k \quad (4)$$

$$P_{ghc} = \frac{1}{2} c_1 c_2 k \times k + \frac{1}{2} c_2 k \times k = \frac{1}{2} c_1 c_2 k \times k \quad (5)$$

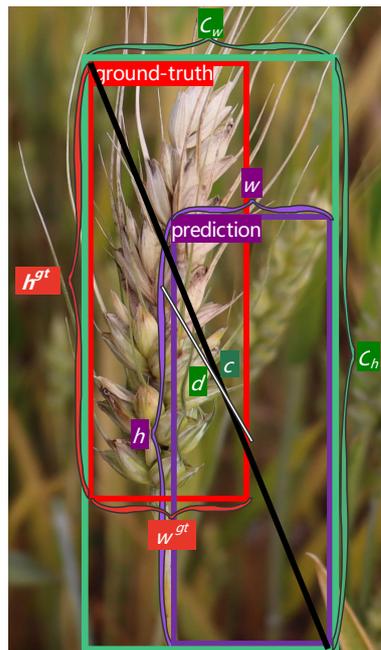
where  $c_1$  is the channel of the input feature map,  $c_2$  is the channel of the output feature map, and  $k \times k$  is the size of the convolution kernel. The ghost convolution model has only half as many parameters as the conventional convolution for the same input–output feature maps.

#### 2.3.4. Efficient Intersection over Union (EIoU) Loss

To obtain faster convergence and more accurate localization of the prediction target, EIoU loss [26] was used for bounding box regression. YOLOX-s adopts IoU loss to calculate the difference of the prediction box and the real box. However, when the prediction box and the real box do not intersect, the IoU loss [27] will exhibit gradient disappearance. The CIoU [28] considers three geometric factors, such as the overlap area, the central point distance, and the aspect ratio, to ameliorate the above problem. Due to the fact that the aspect ratio difference of CIoU cannot reflect the true discrepancy in width and height as well as their confidence, EIoU breaks down the aspect ratio factor into width and height loss separately. EIoU is defined in Equation (6). The two previous parts follow the approach in CIoU, while the width–height loss module directly minimizes the width–height difference between the target frame and the anchor frame. It effectively helps to speed up the convergence and improve positioning accuracy.

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} = 1 - IoU + \frac{d^2}{c^2} + \frac{|w - w^{gt}|^2}{C_w^2} + \frac{|h - h^{gt}|^2}{C_h^2} \quad (6)$$

where  $L_{IoU}$  is the overlap area loss,  $L_{dis}$  is the center point distance loss,  $L_{asp}$  is the aspect ratio loss,  $d$  represents the center distance between the groundtruth box and the prediction box, and  $c$  represents the minimum border diagonal distance that can include both the prediction box and the real box,  $C_w$  and  $C_h$  are the width and height of the minimum bounding box covering the two boxes, respectively,  $w^{gt}$  and  $h^{gt}$  represent the width and height of the groundtruth box, respectively,  $w$  and  $h$  are the width and height of the prediction box, respectively. In summary, the illustration of EIoU loss for box regression is shown in Figure 3.



**Figure 3.** The EIoU loss for box regression. The  $h^{gt}$ ,  $w^{gt}$ ,  $w$ ,  $h$ ,  $C_w$ ,  $C_h$ ,  $d$ , and  $c$  are the same definition as Equation (6).

#### 2.4. Experimental Configurations and Hyperparameters Setting

The experiment employed Python as the programming language and Pytorch as the deep learning framework. The configurations of hardware and software were shown in Table 2. The hyperparameters of the GSEYOLOX-s were set as follows. The initial learning rate was 0.001, and when the loss value of the verification set did not decrease after 10 epochs, the learning rate dropped to 1/10 of the initial value. The batch size was 8, and the number of epochs (iterations) was 200. Adam optimizer was applied to optimize the model.

**Table 2.** Hardware and software configurations.

| Configuration Item      | Value                               |
|-------------------------|-------------------------------------|
| CPU                     | Intel Xeon CPU E5-2683 v3 @ 2.00GHz |
| GPU                     | NVIDIA Geforce RTX 3090             |
| CUDA                    | 11.7                                |
| Memory                  | 64 G                                |
| Operating system        | Ubuntu 18.04.6 LTS (64-bit)         |
| Deep learning framework | Pytorch                             |

#### 2.5. Evaluation Indices

In this study, precision, recall, F1-score, mean average precision (mAP), frames per second (FPS), and parameters are selected as evaluation indices to comprehensively evaluate the performance of deep learning modules. These evaluation indices are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \times 100 \quad (8)$$

$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

$$FPS = \frac{1}{t} \quad (10)$$

$$mAP = \frac{\sum_{i=1}^S AP(S)}{S} \quad (11)$$

where  $TP$ ,  $FP$ , and  $FN$  are the numbers of true positive cases, false positive cases, and false negative cases, respectively. Moreover,  $t$  is the average time the model takes to recognize an image and  $S$  is numbers of the FHB severity level. Precision is the ratio of the quantity of positive samples correctly predicted to the quantity of all positive samples predicted. Recall is the ratio of the quantity of positive samples correctly predicted to the total quantity of true samples. F1-score can be used to balance accuracy and recall. FPS is used to evaluate the real-time processing speed of a model and parameters are used to estimate the model size. The mAP is the mean of the average precision (AP) for each severity level of FHB, which is used to reflect the performance of the model. The parameters (MB) represent the spatial complexity of the model, which is calculated by the size of the convolution kernels and the number of input and output channels.

### 2.6. WeChat Mini Program Development

The GSEYOLOX-s model was deployed on the cloud server. The front-end user interface of WeChat Mini Program was written in WXML and WXSS, and its logic layer was developed in JavaScript. The application consisted of disease detection home page, encyclopedia, and detection results page. The open-source Flask framework was used to communicate between the front-end applet and the cloud server.

## 3. Results

### 3.1. Ablation Experiments

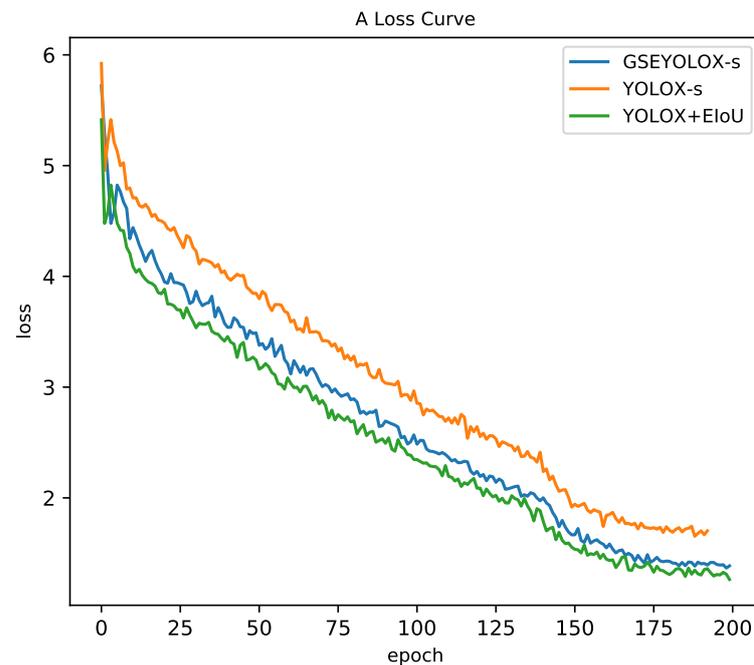
To verify the effectiveness of the G-head, SimAM, and EIou loss function for the YOLOX-s model, ablation experiments were conducted on the FHBSGD dataset under the same experimental setting.

As shown in Table 3, the CBAM, SimAM, and EIou all provided positive feedback to the model in terms of precision or light weight. Firstly, compared with the attention mechanism of CBAM, the SimAM achieved better results with 97.92% of mAP and without increasing the number of parameters. Secondly, although the accuracy of the model was slightly reduced by introducing G-head to YOLOX-s, the parameters of YOLOX-s + G-head were reduced by 9.84%. Additionally, when the G-head was combined with SimAM attention and EIou loss, our proposed GSEYOLOX-s could not only reduce the parameters to 8.06 MB, but also increase the average precision of the model to 99.23%. Thirdly, as shown in Figure 4, the EIou loss function can obtain the smaller loss value and accelerate convergence compared with the IoU loss function used in YOLOX-s. In summary, by uniformly applying the three above methods to the proposed model, the simultaneous tuning of model accuracy maximization and lightweight parameters was achieved.

**Table 3.** Ablation experiments of the GSEYOLOX-s on FHBSGD.

| Models                   | CBAM | SimAM | G-head | EIoU | mAP (%)      | Parameters (MB) |
|--------------------------|------|-------|--------|------|--------------|-----------------|
| YOLOX-s                  |      |       |        |      | 96.71        | 8.94            |
| YOLOX-s + CBAM           | ✓    |       |        |      | 97.81        | 9.03            |
| YOLOX-s + SimAM          |      | ✓     |        |      | 97.92        | 8.94            |
| YOLOX-s + G-head         |      |       | ✓      |      | 96.48        | 8.06            |
| YOLOX-s + EIou           |      |       |        | ✓    | 98.54        | 8.94            |
| YOLOX-s + SimAM + G-head |      | ✓     | ✓      |      | 96.95        | 8.06            |
| YOLOX-s + SimAM + EIou   |      | ✓     |        | ✓    | 98.56        | 8.94            |
| <b>GSEYOLOX-s</b>        |      | ✓     | ✓      | ✓    | <b>99.23</b> | <b>8.06</b>     |

A '✓' indicates that the corresponding module is used.



**Figure 4.** The loss curves during training process.

### 3.2. Comparison of Different Target Detection Models

The GSEYOLOX-s was compared with five typical target detection models including EfficientDet [29], Mobilenet-YOLOV4 [30], YOLOV7 [31], YOLOX-s, and YOLOX-m. The detection results on the FHBSGD are shown in Table 4. Among the five target detection models, YOLOX-s employed relatively small parameters of 8.93 MB to achieve the relatively high mAP of 96.71% and the fastest detection speed of 50 FPS, while YOLOX-m used nearly three times as many parameters (25.28 MB) to achieve the highest mAP of 98.73%, and Mobilenet-YOLOV4 achieved the highest F1-score of 94%. However, compared with the above models, the mAP, F1-score, and parameter numbers of our proposed GSEYOLOX-s were all optimized. In addition, the detection speed with 47 FPS of our model was also comparable to that of the fastest YOLOX-s, which can meet the requirements of real-time detection. Above all, the proposed model is more suitable for mobile devices.

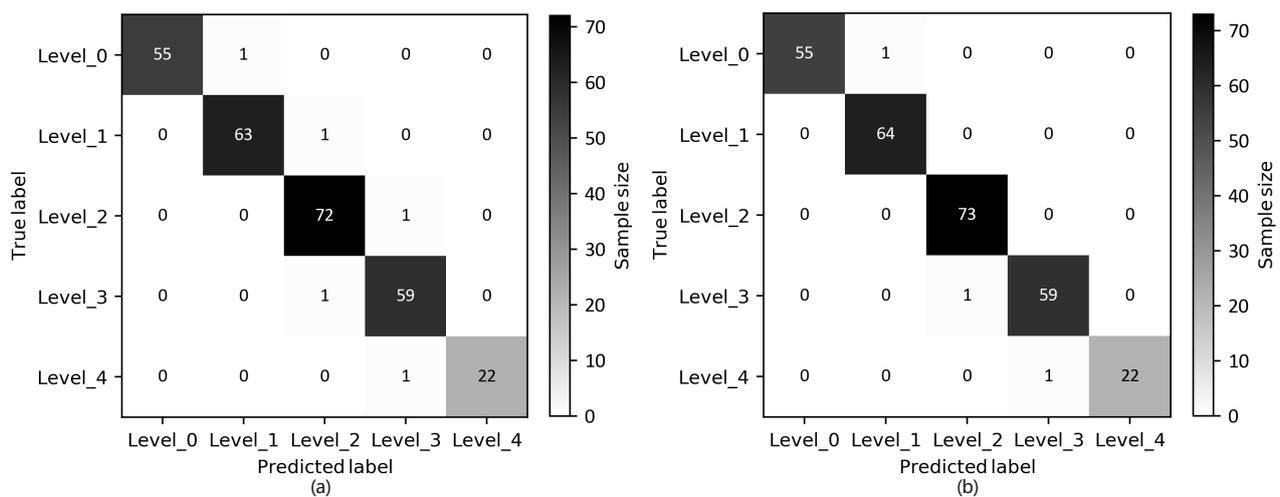
**Table 4.** Detection results of different target detection models on FHBSGD.

| Methods           | mAP(%)       | F1-Score (%) | Parameters (MB) | FPS       | Recall (%)   | Precision (%) |
|-------------------|--------------|--------------|-----------------|-----------|--------------|---------------|
| EfficientDet      | 88.71        | 86.20        | 3.87            | 11        | 81.23        | 92.21         |
| Mobilenet-YOLOV4  | 96.28        | 93.81        | 12.29           | 20        | 95.93        | 92.15         |
| YOLOV7            | 88.52        | 81.18        | 37.22           | 27        | 78.57        | 85.51         |
| YOLOX-s           | 96.71        | 79.67        | 8.93            | 50        | 96.71        | 70.08         |
| YOLOX-m           | 98.73        | 95.88        | 25.28           | 27        | 96.62        | 95.15         |
| <b>GSEYOLOX-s</b> | <b>99.23</b> | <b>96.02</b> | <b>8.06</b>     | <b>47</b> | <b>96.70</b> | <b>95.33</b>  |

In detail, the mAP of FHB severity grading detection using the six models is shown in Table 5. It can be seen that the mAP of the GSEYOLOX-s was the best among the six models and all above 99% for five FHB severity levels, while the confusion matrixes of YOLOX-s (a) and GSEYOLOX-s (b) on the test set are shown in Figure 5. Among the images of the test set, the YOLOX-s incorrectly predicted one sample for each severity level of FHB. However, the GSEYOLOX-s mis-predicted 1 Level\_0 sample, 1 Level\_3 sample, and 1 Level\_4 sample. All samples in Level\_1 and Level\_2 were correctly identified. This further demonstrates the outstanding detection performance of our model.

**Table 5.** The mAP of FHB severity grading detection using six target detection models.

| Methods           | mAP(%)       |              |              |              |              |
|-------------------|--------------|--------------|--------------|--------------|--------------|
|                   | Level_0      | Level_1      | Level_2      | Level_3      | Level_4      |
| EfficientDet      | 70.57        | 88.56        | 91.36        | 96.28        | 96.77        |
| Mobilenet-YOLOV4  | 96.44        | 95.69        | 96.74        | 96.76        | 95.78        |
| YOLOV7            | 90.84        | 82.20        | 89.57        | 91.32        | 88.68        |
| YOLOX-s           | 96.62        | 97.85        | 97.73        | 97.54        | 93.83        |
| YOLOX-m           | 98.70        | 98.68        | 98.79        | 99.32        | 98.16        |
| <b>GSEYOLOX-s</b> | <b>99.28</b> | <b>99.23</b> | <b>99.09</b> | <b>99.33</b> | <b>99.23</b> |

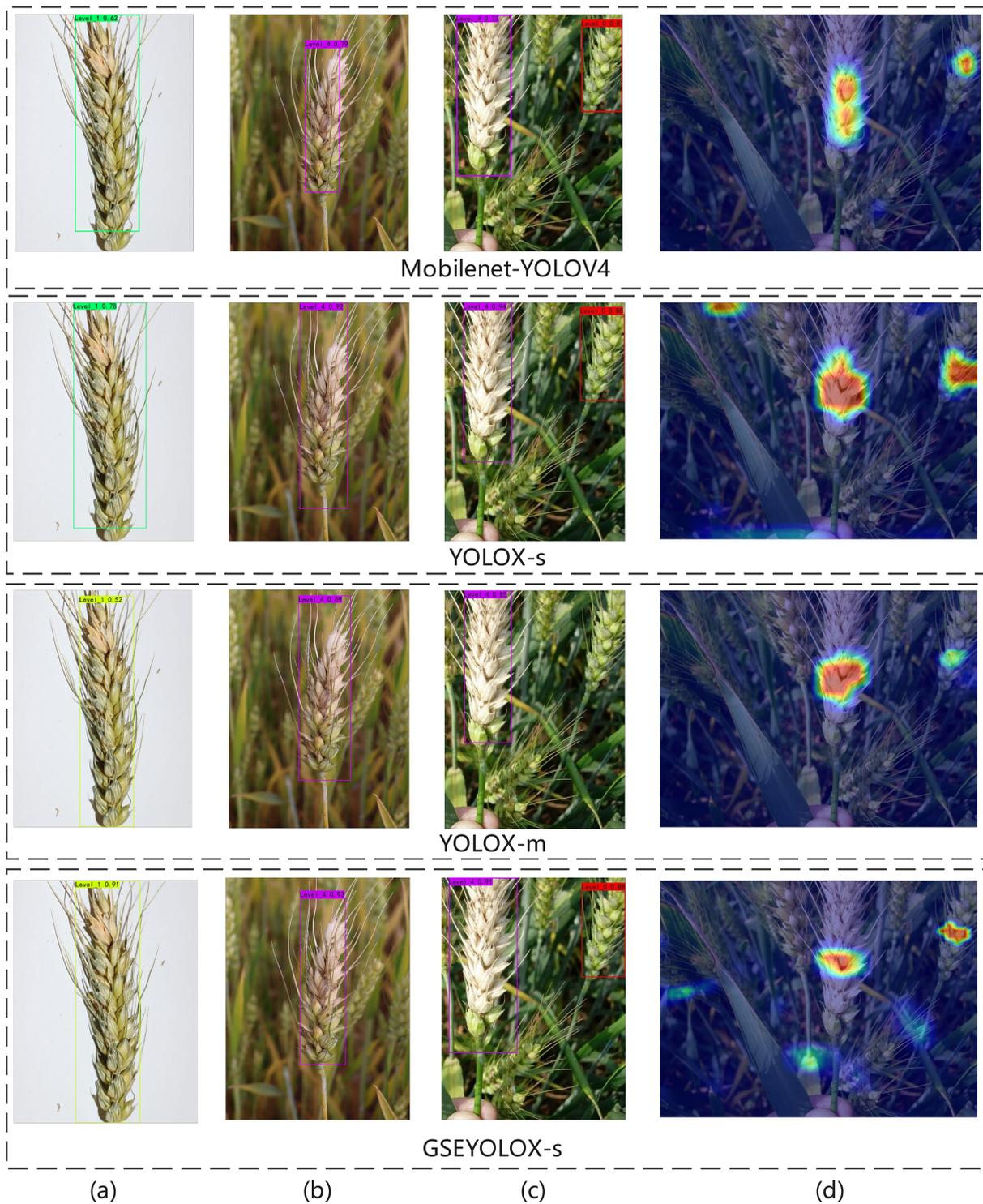
**Figure 5.** Confusion matrixes of YOLOX-s (a) and GSEYOLOX-s (b) on test set.

### 3.3. Visualization

The visual detection results of the top four precision models on FHBSGD are shown in Figure 6, and it can be seen that GSEYOLOX-s outperforms the other comparison models in terms of the accuracy of the grading identification and the prediction of the bounding box for single ear wheat images in column (a) and column (b). For the multi-ear wheat image in column 3, GSEYOLOX-s and Mobilenet-YOLOV4 can accurately detect the severity level of the two wheat ears at the same time, but for the wheat ear obscured partially in column (c), the GSEYOLOX-s detected the severity level of FHB more accurately than Mobilenet-YOLOV4. From the heatmaps in column (d), due to the application of SimAM, the GSEYOLOX-s more accurately focused on the center of the wheat than the other models, making it easier to extract the typical features of FHB. Overall, the comprehensive performance of GSEYOLOX-s was superior to other models in the identification of severity levels of FHB.

### 3.4. Model Deployment

We developed a WeChat Mini Program (Figure 7) to deploy the GSEYOLOX-s model. The service allowed croppers to upload wheat pictures taken in the field or in mobile phone albums (Figure 7a). By using our proposed model on mobile terminals, croppers can easily obtain the real-time feedback to identify the severity levels of FHB (Figure 7b). In addition, some popular science knowledge about FHB was graphically and literally provided in the platform (Figure 7c).



**Figure 6.** Visualization of detection results on FHBSGD. (a) Single wheat ear in laboratory scenes; (b) Single wheat ear in field scenes; (c) Multiple wheat ears in field scenes; (d) Heatmaps of detection results.

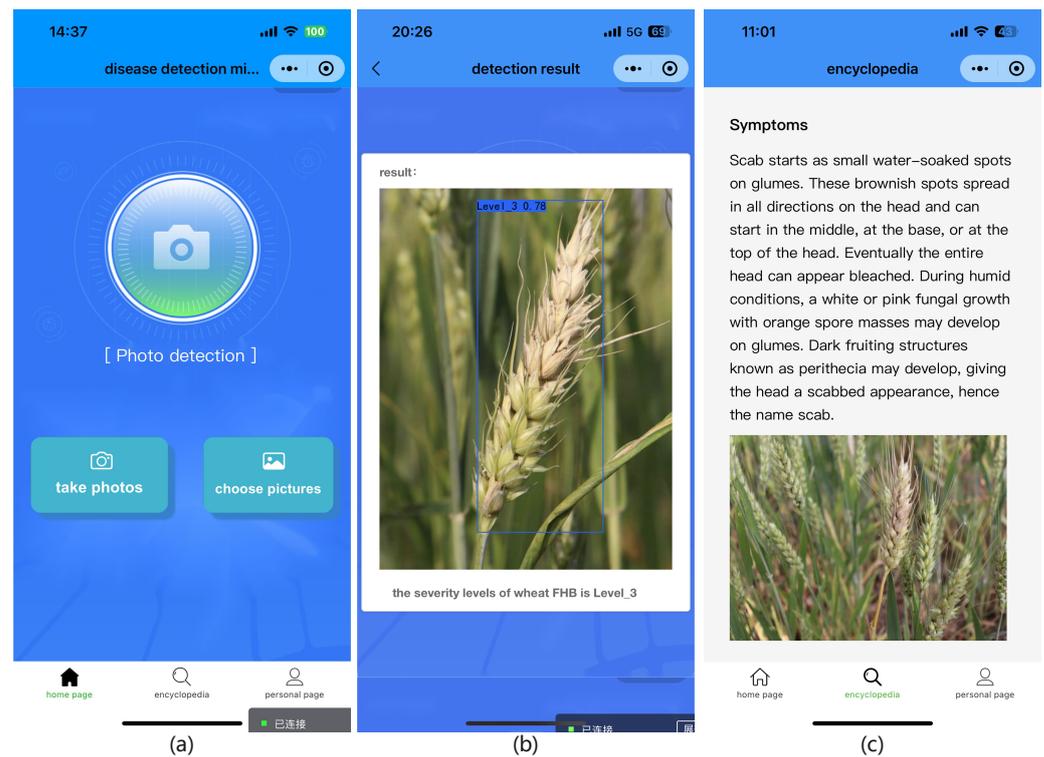


Figure 7. User interfaces of Wechat Mini program.

#### 4. Discussion

The light weight and accuracy of the model determine whether it can be easily applied to the actual planting environment. Gao et al. [32] employed the ResNet50 model to estimate the severity of FHB. However, the model with about 25 MB parameters was constructed under simple experimental conditions, which faced the problem of decreasing accuracy in complex field conditions. The FHBSGD dataset established in this study collected disease images under experimental and complex field conditions. Training on this dataset enabled the proposed GSEYOLOX-s to learn differential features of FHB severity while avoiding the interference of complex backgrounds. Moreover, the combined optimization of SimAM, G-head, and EIou improved the identification speed and accuracy of FHB severity, reduced the parameters of the model, and promoted the deployment of our model on the WeChat Mini Program to achieve real-time detection in the field.

FHB is a typical head disease of wheat, and directly induces the yield decline of wheat crops. Given the cost of disease control and the yield loss, Level\_1 is an appropriate threshold to intervene on the disease with treatments of chemical fungicide. While the proportion of diseased spikelets is less than 25%, the intervention could retrieve a loss which far outweighs the cost of disease control. When the severity of FHB exceeds Level\_1, there are currently no better control measures to save the affected crops. GSEYOLOX-s aims to accurately predict the severity of FHB as early as possible, which can help farmers in their decision toward better crop management.

#### 5. Conclusions

To develop a real-time automatic identification of FHB grading, a lightweight GSEYOLOX-s model was proposed in support of the sustainable management of FHB. The SimAM attention mechanism was introduced into CSPDarkNet to enhance the ability to represent essential disease features without additional parameters. The G-head module was designed to decrease the number of parameters and ensure the performance of the model. Additionally, the use of EIou loss further accelerated the convergence and increased the positioning accuracy of the prediction box. The results indicate that the proposed GSEYOLOX-s can

efficiently and accurately identify severity levels of the FHB with a mAP of 99.23% and 47 FPS of detection speed. Meanwhile, the proposed model was deployed on the mobile terminal to enable farmers and technicians to automatically identify the severity levels of FHB. Further work is required to enrich the construction of the canopy scale multi-severity detection model. Thus, the automatic intelligent monitoring platform of FHB will be achieved.

**Author Contributions:** R.M.: Conceptualization, Writing—Review and Editing, and Funding Acquisition; Z.W.: Writing—Original Draft, Methodology, and Investigation; F.L.: Data collation and annotation, Methodology; J.Z.: Resources and Data Verification; Y.C.: Software; X.H.: Paper Revising and Funding Acquisition. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was financially supported by the China Agriculture Research System of Wheat (CARS-03-37); Regional Innovation Capability Guidance Plan of Science and Technology Department of Shaanxi Province (2022QFY11-03); Major Science and Technology Project of Shaanxi Agricultural Collaborative Innovation and Promotion Alliance in 2022 (LMZD202203); Innovation Training Program for College Students “Identification Research for the Severity of Wheat *Fusarium* Head Blight Based on Lightweight Deep Learning Method”.

**Data Availability Statement:** Data are available upon request from the authors.

**Conflicts of Interest:** The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Chen, Y.; Zhang, Z.; Tao, F.; Wang, P.; Wei, X. Spatio-temporal patterns of winter wheat yield potential and yield gap during the past three decades in North China. *Field Crop. Res.* **2017**, *206*, 11–20. [[CrossRef](#)]
2. Pirgozliev, S.R.; Edwards, S.G.; Hare, M.C.; Jenkinson, P. Strategies for the control of Fusarium head blight in cereals. *Eur. J. Plant Pathol.* **2003**, *109*, 731–742. [[CrossRef](#)]
3. Ochodzki, P.; Twardawska, A.; Wiśniewska, H.; Góral, T. Resistance to Fusarium Head Blight, Kernel Damage, and Concentrations of Fusarium Mycotoxins in the Grain of Winter Wheat Lines. *Agronomy* **2021**, *11*, 1690. [[CrossRef](#)]
4. Miao, J.; Zhang, G.P.; Zhang, S.J.; Ma, J.Q.; Wu, Y.Q. The orange wheat blossom midge promotes fusarium head blight disease, posing a risk to wheat production in northern China. *Acta Ecol. Sin.* **2023**, *43*, 112–116. [[CrossRef](#)]
5. Chen, B.; Shen, X.; Li, Z.; Wang, J.; Li, X.; Xu, Z.; Shen, Y.; Lei, Y.; Huang, X.; Wang, X.; et al. Antibody generation and rapid immunochromatography using time-resolved fluorescence microspheres for propiconazole: Fungicide abused as growth regulator in vegetable. *Foods* **2022**, *11*, 324. [[CrossRef](#)] [[PubMed](#)]
6. Ferentinos, K.P. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **2018**, *145*, 311–318. [[CrossRef](#)]
7. Esgario, J.G.; Krohling, R.A.; Ventura, J.A. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Comput. Electron. Agric.* **2020**, *169*, 105162. [[CrossRef](#)]
8. Ji, M.; Zhang, K.; Wu, Q.; Deng, Z. Multi-label learning for crop leaf diseases recognition and severity estimation based on convolutional neural networks. *Soft Comput.* **2020**, *24*, 15327–15340. [[CrossRef](#)]
9. Mi, Z.; Zhang, X.; Su, J.; Han, D.; Su, B. Wheat stripe rust grading by deep learning with attention mechanism and images from mobile devices. *Front. Plant Sci.* **2020**, *11*, 558126. [[CrossRef](#)]
10. Zhang, D.; Wang, D.; Gu, C.; Jin, N.; Zhao, H.; Chen, G.; Liang, H.; Liang, D. Using neural network to identify the severity of wheat Fusarium head blight in the field environment. *Remote Sens.* **2019**, *11*, 2375. [[CrossRef](#)]
11. Bao, W.; Yang, X.; Liang, D.; Hu, G.; Yang, X. Lightweight convolutional neural network model for field wheat ear disease identification. *Comput. Electron. Agric.* **2021**, *189*, 106367. [[CrossRef](#)]
12. Hong, Q.; Jiang, L.; Zhang, Z.; Ji, S.; Gu, C.; Mao, W.; Li, W.; Liu, T.; Li, B.; Tan, C. A Lightweight Model for Wheat Ear Fusarium Head Blight Detection Based on RGB Images. *Remote Sens.* **2022**, *14*, 3481. [[CrossRef](#)]
13. Liu, S.; Qiao, Y.; Li, J.; Zhang, H.; Zhang, M.; Wang, M. An Improved Lightweight Network for Real-Time Detection of Apple Leaf Diseases in Natural Scenes. *Agronomy* **2022**, *12*, 2363. [[CrossRef](#)]
14. Cui, S.; Su, Y.L.; Duan, K.; Liu, Y. Maize leaf disease classification using CBAM and lightweight Autoencoder network. *J. Ambient. Intell. Humaniz. Comput.* **2022**, 1–11. [[CrossRef](#)]
15. Li, K.; Zhang, L.; Li, B.; Li, S.; Ma, J. Attention-optimized DeepLab V3+ for automatic estimation of cucumber disease severity. *Plant Methods* **2022**, *18*, 1–16. [[CrossRef](#)]
16. Chrpová, J.; Orsák, M.; Martinek, P.; Lachman, J.; Trávníčková, M. Potential role and involvement of antioxidants and other secondary metabolites of wheat in the infection process and resistance to *Fusarium* spp. *Agronomy* **2021**, *11*, 2235. [[CrossRef](#)]

17. Osborne, L.E.; Stein, J.M. Epidemiology of Fusarium head blight on small-grain cereals. *Int. J. Food Microbiol.* **2007**, *119*, 103–108. [[CrossRef](#)]
18. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
19. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
20. 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 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Seattle, WA, USA, 14–19 June 2020.
21. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
22. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
23. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T. Unitbox: An advanced object detection network. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; pp. 516–520.
24. Yang, L.; Zhang, R.Y.; Li, L.; Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the International Conference on Machine Learning PMLR, Virtual, 18–24 July 2021; pp. 11863–11874.
25. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1580–1589.
26. Zhang, Y.F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146–157. [[CrossRef](#)]
27. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
28. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12993–13000.
29. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
30. Fu, J.; Chen, X.; Lv, Z. Rail Fastener Status Detection Based on MobileNet-YOLOv4. *Electronics* **2022**, *11*, 3677. [[CrossRef](#)]
31. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.
32. Gao, C.; Gong, Z.; Ji, X.; Dang, M.; He, Q.; Sun, H.; Guo, W. Estimation of Fusarium Head Blight Severity Based on Transfer Learning. *Agronomy* **2022**, *12*, 1876. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.