

Article

An Integrated Multi-Model Fusion System for Automatically Diagnosing the Severity of Wheat Fusarium Head Blight

Ya-Hong Wang ^{1,†}, Jun-Jiang Li ^{2,†}  and Wen-Hao Su ^{1,*} ¹ College of Engineering, China Agricultural University, 17 Qinghua East Road, Haidian, Beijing 100083, China² School of Mechanical Engineering, Xi'an Jiaotong University, 28 Xianning West Road, Xi'an 710048, China

* Correspondence: wenhao.su@cau.edu.cn

† These authors are co-first authors.

Abstract: Fusarium has become a major impediment to stable wheat production in many regions worldwide. Infected wheat plants not only experience reduced yield and quality but their spikes generate toxins that pose a significant threat to human and animal health. Currently, there are two primary methods for effectively controlling Fusarium head blight (FHB): spraying quantitative chemical agents and breeding disease-resistant wheat varieties. The premise of both methods is to accurately diagnosis the severity of wheat FHB in real time. In this study, a deep learning-based multi-model fusion system was developed for integrated detection of FHB severity. Combination schemes of network frameworks and backbones for wheat spike and spot segmentation were investigated. The training results demonstrated that MobileV3-Deeplabv3+ exhibits strong multi-scale feature refinement capabilities and achieved a high segmentation accuracy of 97.6% for high-throughput wheat spike images. By implementing parallel feature fusion from high- to low-resolution inputs, w48-Hrnet excelled at recognizing fine and complex FHB spots, resulting in up to 99.8% accuracy. Refinement of wheat FHB grading classification from the perspectives of epidemic control (zero to five levels) and breeding (zero to 14 levels) has been accomplished. In addition, the effectiveness of introducing HSV color feature as a weighting factor into the evaluation model for grading of wheat spikes was verified. The multi-model fusion algorithm, developed specifically for the all-in-one process, successfully accomplished the tasks of segmentation, extraction, and classification, with an overall accuracy of 92.6% for FHB severity grades. The integrated system, combining deep learning and image analysis, provides a reliable and nondestructive diagnosis of wheat FHB, enabling real-time monitoring for farmers and researchers.

Keywords: deep learning; wheat; fusarium head blight; image segmentation; all-in-one detection



Citation: Wang, Y.-H.; Li, J.-J.; Su, W.-H. An Integrated Multi-Model Fusion System for Automatically Diagnosing the Severity of Wheat Fusarium Head Blight. *Agriculture* **2023**, *13*, 1381. <https://doi.org/10.3390/agriculture13071381>

Academic Editor: Maciej Zaborowicz

Received: 12 June 2023

Revised: 8 July 2023

Accepted: 10 July 2023

Published: 12 July 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 is the third largest grain crop after maize and rice around the world, and its cob and fruit parts are rich in large amounts of starch, protein and other nutrients, which can be consumed by humans and animals [1]. However, there are many destructive diseases in nature that seriously affect wheat production and threaten global food security [2]. Among them, Fusarium head blight (FHB), caused by *Fusarium graminearum* Schw, is one of the epidemic diseases in wheat production [3]. The disease is prevalent in wheat fields of semihumid and humid regions and can reduce wheat yield losses by 10–70%, affecting more than 70,000 hectares of acreage [4]. Pathogenic fungi attack the spikes of wheat, resulting in a significant reduction in crop yield and quality. In addition, FHB can lead to a range of mycotoxins (e.g., deoxynivalenol and zearalenone) to be produced inside the grain, which can cause human and animal poisoning and pose a significant health risk to food safety [5]. In summary, research on the monitoring and warning of FHB is important for the development of scientific wheat production with high yield, high quality, and high efficiency.

For timely control of FHB in wheat, each sample plant needs to be monitored for disease. Traditionally, the assessment of FHB severity has relied on manual field sampling, where infected spikelets are counted and their percentage of total spikelets is calculated to assign a grade [6]. This method is simple to operate, but it is inefficient and costly in practice, and the accuracy of data is subjective, which seriously affects the efficiency of FHB control in wheat [7]. In recent years, imaging and spectroscopic methods including near-infrared spectroscopy (NIRS) and hyperspectral imaging (HSI) have also been rapidly developed in the field of disease surveillance [8,9]. However, such sensors can only sense the spectral characteristics of an object at a certain point and the computational data increase dramatically, which are not suitable for real-time analysis of large-scale wheat spikes in reality [10,11].

The novelty of this study is the development of an integrated multi-model fusion system for FHB severity assessment of wheat with deep learning. Taking wheat FHB as the research object, convolutional neural network (CNN) and related algorithms of image processing are used to sequentially complete the integrated operations of wheat spike segmentation, extraction of disease color features, disease spot segmentation, and disease grade assessment. Eventually, accurate images of wheat spike and disease spot segmentation and grade information are obtained in the high-throughput wheat population, thus realizing the purpose of real-time monitoring. The exploratory contributions of our method can be summarized as follows:

- Using the multi-scale feature of Deeplab for wheat spike extraction.
- Fine-grained segmentation of disease spots using multi-resolution feature of Hrnet.
- The evaluation method was optimized by the HSV color features as weighting factor.
- Mobile terminal equipped with the all-in-one system to achieve real-time diagnosis.

Overall, this paper follows an organized structure to elaborate on the processing of our research. Firstly, the introduction section provides an overview of the research background, motivation, and outlines the objectives and significance of the wheat FHB diagnosis. Secondly, a comprehensive literature review is conducted. This review examines existing research and theoretical frameworks related to segmentation, extraction, and diagnosis, identifying gaps in the current knowledge. Next, the section of materials and methods presents the systematic approach adopted to capture high-quality image datasets, train different network models and select appropriate parameters and equipment. This ensures transparency and replicability of the study process. Following that, the results and discussion section interprets the experimental findings. It analyzes the performance of 14 different models to select the optimal model and evaluate wheat FHB grading outcomes. The results are compared to previous studies, providing insights into the strengths and limitations of the system. Finally, the section concludes by summarizing highlights of the research and discussing potential applications. This is the first study of an integrated multi-model fusion system based on deep learning for diagnosing the severity of wheat FHB.

2. Literature Review

With the rapid development of computer vision technology, digital image processing based on deep learning has been widely applied to wheat crops [12–14]. The related research mainly includes four aspects: object segmentation, disease segmentation, disease feature extraction, and severity diagnosis system.

To accurately assess the severity of FHB in individual wheat spikes under field conditions, precise segmentation of each spike area within the complex background is crucial. Researchers worldwide have conducted extensive experiments and research on target segmentation methods, leveraging advancements in neural network performance and structure, resulting in promising achievements. Zhang, et al. [15] developed a pulse-coupled neural network (PCNN) based on the fully convolutional network (FCN) for segmenting wheat spikes infected with FHB. However, only one spike in the image was taken into consideration in the research, which was not practical for high throughput detection in the field environment. Su, et al. [16] and Qiu, et al. [17] developed Mask-RCNN for

independent accurate segmentation of wheat spikes with recognition rates reaching 77.76% and 92.01%, respectively. In addition, more advanced deep learning models, such as Fast R-CNN, BlendMask, and YOLOv4, have also been applied to image segmentation of wheat spikes [18–20].

Based on the segmented wheat spikelet samples, it is important to effectively distinguish healthy spikelets in the wheat disease region. This step is crucial for accurately grading the severity of FHB in wheat and achieving precise disease classification. Su, Zhang, Yang, Page, Szinyei, Hirsch and Steffenson [16] adopted Mask-RCNN to segment FHB disease spots, whose detection rate was as high as 98.61%, but the related strategies still need to be optimized. Since the color of wheat spikes changes significantly after being infected with FHB, color features are extracted as an auxiliary basis for judging the severity level of erysipelas on top of spot segmentation. For example, Sarayloo and Asemani [21] extracted texture, color and shape features of infected wheat, processed them as effective features for identifying diseases, and finally obtained 98.3% recognition accuracy.

In a recent study, HSV color threshold extraction was also used to assist the YOLO network in achieving improved accuracy and precision in wheat FHB detection [22].

For the task of estimating the severity of FHB, the deep convolutional neural network (DCNN) model built by Zhang, et al. [23] was successfully used to locate disease spots and to predict the grading with a high degree of accuracy. Furthermore, transfer learning was also used to assess the severity of FHB [24]. The approach can save time and partially address the overfitting problem, but pre-trained large models exhibited significant fluctuations in accuracy when evaluating imbalanced samples. With the growing demand for end devices, such as personal computers and smart agricultural equipment, the development of integrated intelligent diagnostic systems has gradually emerged as a current focal point [25]. Although satisfactory results are reported in the above studies based on deep learning models, there will be a problem in the practical application of disease severity diagnosing due to the high number of parameters, the large storage space, and computational consumption. Recent studies have also deployed light-weight GSEYOLOX-s models on mobile terminals to help farmers identify the severity of FHB in real time [26]. However, considering the small and subtle differences between different FHB severity levels, building a real-time accurate FHB all-in-one system is still a great challenge.

3. Materials and Methods

3.1. Data Collection

At the Minnesota Agricultural Experiment Station on the University of Minnesota St. Paul campus, wheat samples of 55 genetic lines were sown for the FHB evaluation trial in May 2019 [27]. The data used in this study were derived from high-throughput wheat images taken in the experiment station. In order to ensure that the different lines of wheat varieties could achieve adequate levels of infection with the *Fusarium* fungus, the batch was designed to be inoculated three times with appropriate amounts of wheat FHB conidia spray, and eventually different lines of wheat expressed different levels of infection [27]. In order to better assess the spot characteristics of the blast, the experiment was finally selected to collect images of wheat spikes when the symptoms became visible but before senescence.

The image acquisition equipment was based on an autofocus single lens reflex (SLR) camera (Canon EOS Rebel T7i, resolution: 6000 × 4000, a camera manufactured by Canon Inc. in Tokyo, Japan) with a fixed macro lens. The camera operates in automatic mode, allowing the appropriate acquisition parameters to be set, including white balance, ISO speed and exposure time. Eventually, the collection of wheat spike images of 55 genetic lines in the field from flowering to the late maturing stage was completed during sunny weather (10:00 to 13:00). For the method of classifying disease severity, the national standard specifies six disease levels based on the ratio of the area of FHB spots to the corresponding wheat spikes area to visualize the degree of disease occurrence. The “not occurring” level is 0: [0–1%], in which no control measures are needed at this time. The “slightly occurring” level is 1: (1–10%), in which the disease occurs sporadically at this

time, no chemical control measures are needed, and only the diseased wheat spikes need to be eradicated in time. “Light” is level 2: (10–20%]; when the disease has a tendency to spread and expand, it should be established in time with some agronomic control measures. “Moderate” is level 3: (20–30%]; when the disease is sufficient to cause significant local loss of wheat yield, there is a need to carry out the corresponding chemical control. “Slightly serious occurrence” is level 4: (30–40%], when the disease area wheat needs to focus on the general prevention; otherwise it will cause serious loss of wheat yield in the region. “Severe occurrence” is level 5: (40–100%], when all wheat plants in the field need to be extensively prevented; otherwise, a large reduction of wheat yield in the year may result. In order to better compare the disease resistance of wheat spikes, this study refined the grade intervals based on the original classification criteria with reference to the study of Su, Zhang, Yang, Page, Szinyei, Hirsch and Steffenson [16]. The images obtained contained 15 disease levels: level 0: [0–1%], level 1: (1–2.5%], level 2: (2.5–5%], level 3: (5–7.5%], level 4: (7.5–10%], level 5 level: (10–12.5%], level 6: (12.5–15%], level 7: (15–17.5%], level 8: (17.5–20%], level 9: (20–25%], level 10: (25–30%], level 11: (30–40%], level 12: (40–50%], level 13: (50–60%], and level 14. (60–100%].

3.2. Data Annotation and Examination

A set of 718 images was selected as the experimental material for wheat FHB research. Through further selection and cropping, 3875 wheat spike images containing disease spots were finally obtained to serve as the dataset. In the detection task of wheat spike region, 20,488 wheat spikes were manually labeled, with each image containing about 7124 wheat spikes, and then 646 images (containing 3462 wheat spikes) and 72 images (containing 413 wheat spikes) were randomly selected as the training set and validation set of the model, respectively. In the detection task of disease areas, a total of 7684 disease spots were labeled in 3875 wheat spike images. All images were annotated using the software (Labelme, An image annotation tool developed at Massachusetts Institute of Technology’s (MIT) Computer Science and Artificial Intelligence Laboratory (CSAIL), <https://github.com/wkentaro/labelme>, accessed on 19 May 2022), which required a total of three steps. The first step was to annotate the wheat spike area of the image, the second step was to split the multiple wheat spike areas in the labeled image into separate sub-images, and the third step was to annotate the blotch areas in the separate wheat spike images. Specifically, the shape of the wheat spikes was outlined in the collected full-size image by manually drawing polygons, and then the original image was segmented into rectangular areas containing a single wheat spike. The independent images were generated into independent images by image processing in the PIL sub-module of the image library. The sub-images of the wheat spikes after filtering the background were obtained by superimposition, and the relevant features of the diseased areas of the wheat spikes were manually marked on these sub-images.

3.3. Data Enhancement and Pre-Processing

In this study, random resize, place, and distort operations were performed on image datasets to achieve data enhancement. Specifically, the image-resizing module randomly rotates, flips, pans, and scales the image; the image place module performs a cutout operation and randomly divides a certain number of sub-regions in the image for copying and swapping; and the image-distorting module changes two sets of weights to adjust the brightness, contrast, saturation, and chromaticity of the image, thus greatly enriching the background information of the image. Median filtering was used to remove the image noise generated by interference factors, such as jitter during the shooting process while maintaining the image quality of the edge contour of the wheat spike. In addition, Gaussian filtering was used to pre-process images, with the aim of giving suppression and linear smoothing to extremely complex environmental background images. A Gaussian filter with a small kernel of 3×3 was used in this study.

3.4. Network Framework

The framework selection incorporated leading advancements in semantic segmentation, selecting four neural network models (DeeplabV3+, Unet, Pspnet, and Hrnet) that have demonstrated excellence in various competitions and prominent conference papers. These models will serve as the overall architectures of the deep learning models in this study.

3.4.1. Deeplabv3+

Deeplabv3+ is a semantic segmentation network based on Deeplabv3, which introduces the encoder–decoder form to fuse multi-scale information [28,29]. Taking a two-dimensional signal as an example, the input is x , the corresponding sequence number between the output and the input is i , the convolution kernel is w , the output is y , and the expansion factor is r . The null convolution is equivalent to inserting $r - 1$ zeros into the input x in each channel dimension and convolving it between the convolution kernels generated by two consecutive convolution kernels, as shown in the following equation.

$$y[i] = \sum_k kx[i + r \times k]w[k] \quad (1)$$

3.4.2. Hrnet

The Hrnet uses a high-resolution sub-network as the underlying architecture, connects the multi-resolution sub-networks in a parallel manner, and uses repetitive multi-scale fusion to obtain a large number of high-resolution representations to predict more accurate heat maps of key points [30]. Specifically, Hrnet consists of four phases with four parallel subnetworks as the main body. There are eight switching units in the whole model, which are involved in a total of eight multi-scale fusions, and the equations of the switching units are shown as follows.

The input is the response map of $s: \{X_1, X_2, \dots, X_s\}$. The output is the response map of $s: \{y_1, y_2, \dots, y_s\}$, and the resolution of the image with the number of channels corresponds to the input. The function $a(X_i, k)$ represents the resolution of X_i from i to k by upsampling or downsampling, which is usually performed with a convolution of step size 3.

$$Y_k = \sum_{i=1}^s a(X_i, k) \quad (2)$$

The commonly used Hrnet backbone extraction networks mainly include Hrnet-w18, Hrnet-w32, and Hrnet-w48, and the network sizes used by them are in increasing order, where 18, 32, and 48 represent the number of channels C of the high-resolution subnets in the end three phases.

3.4.3. U-net

U-net is an excellent semantic segmentation model similar to the above model structure [31]. The difference is that the downsampling part of U-net is completely symmetric with the upsampling part, and it stacks the feature layers together in the dimension of channels with the overall U-shaped network structure [7].

Specifically, U-net consists of three parts. The first part is the backbone feature extraction, which is the convolution and the maximum pooling stacking to obtain multiple effective feature layers. The second part is the enhanced feature extraction, in which the network fuses the five effective feature layers output from the backbone part, by upsampling and stacking them in sequence [32]. The third part then uses the features to obtain the prediction results and consists of 1×1 convolution [33].

3.4.4. Pspnet

Pspnet is an improved network based on FCN networks, which reduces the segmentation errors of the network by introducing more contextual information [34]. In order to aggregate the contextual information of different regions, the model proposes a pyramid

pooling module, which greatly improves the model's ability to obtain global information [35]. The Pspnet structure module also serves to divide the acquired feature layers into grids with different sizes and to average pooling independently within each grid.

3.5. Backbone

To obtain a better fit of the main network framework, this study selected CNNs that have been pre-trained on high-quality datasets, including Resnet [36], Mobilenet [37], Xception [38], and Ghostnet [39]. The models were modified to become the main feature extraction network modules suitable for research tasks. In addition, the original backbones of Deeplabv3+, Unet, and Pspnet networks were replaced with these model architectures, resulting in deep learning network models with different generalization capabilities. The Hrnet model is based on the original backbone, and the parallel network was deepened and enlarged to obtain different types of architectures. The specific network model combinations are shown in Table 1.

Table 1. Fourteen types of models with different combinations of network frameworks and backbones.

Network Model Framework	Backbone Feature Network	Serial Number
Deeplabv3+	Mobilev2	1
	Mobilev3	2
	Resnet50	3
	Resnet101	4
	Resnet152	5
	Ghostnet	6
	Xceptionnet	7
Pspnet	Resnet50	8
	Mobilev2	9
U-net	Resnet50	10
	Resnet101	11
Hrnet	W18	12
	W32	13
	W48	14

Moblie: Moblienet; Resnet: Residual network; Ghostnet: More features from cheap operations.

3.6. Evaluation Metrics

In this study, the following parameters were used to evaluate the performance of the above neural network to select the optimal model for the segmentation of high-throughput wheat spikes and disease spots. Accuracy, recall, precision, *F*-score value, average precision (*AP*), and average pixel precision (*mPA*) were generated by calculating the values of true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*). *Precision*, *Recall*, *F*-score, *AP*, *IoU*, *MIoU*, and *mPA* can be calculated using the following equations.

$$Precision = \frac{TP}{TP + FP'} \quad (3)$$

$$Recall = \frac{TP}{TP + FN'} \quad (4)$$

$$F_1 = \frac{2PR}{P+R'} \quad (5)$$

$$AP = \frac{1}{11} \sum_{R_j} P(R_j), j=1,2,3,\dots, 11 \quad (6)$$

$$IoU(E, F) = \left| \frac{E \cap F}{E \cup F} \right| \quad (7)$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (8)$$

$$mPA = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij}} \quad (9)$$

TP corresponds to the number of true positives generated by the model, which is the number of correctly detected wheat spikes, FP represents the number of incorrectly identified wheat spikes, and FN represents the number of spikes that were not detected but should have been identified. E represents the manually marked true area, F represents the predicted area of the network model output, and if the model estimated IoU value is higher than the preset threshold of 0.5, it indicates that the model has a true positive TP prediction. $k+1$ represents all outputs including the background the total number of classes, P_{ii} , P_{ij} and P_{ji} represent TP , FP and FN , respectively.

3.7. Experimental Equipment and Devices

3.7.1. Hardware Equipment

The whole process of model training and validation was realized by a shared computer (processor AMD EPYC 7543, 16×4 cores, host memory 50 G, operating system Ubuntu SMP18.04.5, 64-bit) rented from the “HengYuanCloud” shared GPU platform. The GPU (NVIDIA RTX 3090Ti, 2 units, 24G of video memory) was used to optimize the training speed of the models. Table 2 provides details of the modeling parameters, such as maximum iterations, the warm-up decay rate, and the number of classes. In addition, the code for image processing was coded in Python.

Table 2. Settings of modeling parameters for identifications of wheat spikes and disease areas.

Modeling Parameters	Value
Pretrained	True
Batch_size	20/30
Max_epoch	000
Init_lr	0.001
Min_lr	0.0001
Optimizer	Adam
Weight_decay	0
Warmup_lr_ratio	0.1
No_auy_iter_ratio	0.3
Lr_decay_type	Cos
Number of classes	2

lr: Learning rate; no_auy_iter_ratio: Proportion of iterations where the learning rate remains constant after cosine annealing.

3.7.2. Optimizer Selection and Learning Rate Adjustment

The Adam optimizer was adopted in the study, which was able to adaptively adjust the corresponding learning rates for different parameters, with computationally efficient and

explanatory hyperparameters and initial learning rates. Specifically, the Adam optimizer combines the advantages of two optimization algorithms (RMSProp and AdaGrad) and provides a comprehensive consideration of the first and second order moment estimation of the gradient, and then calculates the step size to be updated [40]. Meanwhile, in order to obtain a better fitting effect, warmup+CosineAnnealingLR was used for learning rate adjustment, where the learning rate warmup method was to avoid the unstable oscillation of the model due to the higher learning rate for the untrained initialized weights. The learning rate adjustment relation after the learning rate preheating method was in accordance with the trend of cosine annealing function [41], and if the training rounds of warmup are excluded, for the cosine decay function, assuming there are still T training rounds, the learning rate η_t :

$$\eta_t = \frac{1}{2} \left(1 + \cos \left(\frac{t\pi}{T} \right) \right) \eta \quad (10)$$

Since the number of classes in this study is small (Number of classes = 2) and the number of data samples captured for each training is large (batch_size > 10), the *Dice_loss* function needs to be introduced to measure the similarity of the two samples in both wheat spikes and disease spot segmentation. For the task of wheat spike detection, the distribution of positive and negative samples was balanced, so the cross-entropy loss function (*CE_loss*) was introduced for the model training [42]. In addition, since all the spikes of wheat selected in this study contained red disease spots (positive samples), and the number of pixels belonging to the spots on each diseased spike was small, an extremely unbalanced distribution of positive and negative samples for FHB spots task was caused. Therefore, the training of the model for the task required the introduction of *Focal_loss*, which assigned weights to the losses of the model, thus enhancing the influence of positive samples on the model.

4. Results

4.1. Model Training

Fourteen different deep learning-based segmentation models were developed based on wheat spikes and disease areas. To increase confidence, all models were trained based on the same image data and devices. At least the number of training rounds for each model was between 70 and 100 rounds, since changing the network model type causes a change in the convergence rate. The only condition to determine whether the model converges is that the overall loss of the training and validation sets does not vary by more than ± 0.006 for 8 ± 2 epoch. Table 3 shows the segmentation effects of 14 models on wheat spikes and disease spot areas, including the three metrics of *IoU*, *mPA*, and accuracy.

For wheat spike segmentation, the results showed that there were four models with better effects, including Xception-Deeplabv3+, Mobilev3-Deeplabv3+, Resnet50-Pspnet, and w48-Hrnet. Due to the location of the operation task in the field with a complex environment and the large resolution of the captured pictures, the parameter size and the corresponding operation speed of the four models should be further compared in order to meet the real-time requirements of wheat recognition. As shown in Table 4 of the run parameters, the Mobilev3-Deeplabv3+ model had an absolute advantage in terms of the number of parameters, file size and detection speed. Considering the actual production requirements, the Mobilev3-Deeplabv3+ model was selected to complete the field wheat spike segmentation task under the condition of guaranteed accuracy.

Among the disease spot segmentation models, the Mobilev3-Deeplabv3+, Resnet50-Pspnet, and Hrnet series have achieved good results in all indexes, with w48-Hrnet performing the best. Similarly, comparing the relevant parameters during the operation of the models is shown in Table 4. The difference is that the resolution of a single wheat spike image is lower, which requires much less processing effort than the wheat spike segmentation. Therefore, the difference in running time of each model for each image was within ± 0.0405 s, and the average time of w48-Hrnet detection per image was 0.1248s, which was completely acceptable in real production. Moreover, the accuracy of identifying

the disease area is critical to the overall judgment of the severity of FHB, because the disease spots are small, complex in shape, and randomly distributed in multiple slices [43]. Based on the above considerations, the w48-Hrnet model was selected for the task of accurate spot segmentation in this study.

Table 3. Model performance parameters comparison in wheat spikes and disease spot segmentation.

Serial Number	Model Type	Wheat Spike Segmentation			Disease Spot Segmentation		
		<i>MIoU</i>	<i>mPA</i>	Accuracy	<i>MIoU</i>	<i>mPA</i>	Accuracy
1	Mobilev2-Deeplabv3+	76.63	83.74	94.76	79.97	87.90	98.16
2	Mobilev3-Deeplabv3+	76.99	85.60	94.63	83.61	90.88	98.54
3	Resnet50-Deepnabv3+	72.31	85.98	92.50	76.44	84.01	97.81
4	Resnet101-Deepnabv3+	73.61	83.60	93.53	73.61	83.60	93.53
5	Resnet152	72.84	82.82	93.33	75.84	85.21	97.64
6	Ghostnet-Deepnabv3+	73.41	82.47	93.65	79.89	91.30	97.99
7	Xceptionnet-Deeplabv3+	77.52	85.09	94.90	79.10	83.18	98.27
8	Resnet50-Pspnet	77.03	85.16	94.71	82.54	88.50	98.48
9	Mobilev2-Pspnet	74.07	83.71	93.72	83.41	89.70	98.54
10	Resnet50-Unet	73.75	80.56	94.14	77.59	86.37	97.83
11	Resnet101-Unet	73.61	83.60	93.53	74.24	84.68	97.33
12	W18-Hrnet	64.18	74.60	90.58	83.68	91.35	98.51
13	W32-Hrnet	73.86	82.40	93.85	84.60	91.02	98.64
14	W48-Hrnet	79.11	86.79	95.26	85.06	91.74	98.67

mPA: Average pixel precision; *MIoU*: Modular input/output unit.

Table 4. Parameters for the operation of the wheat spike and disease spot segmentation models.

Segmented Objects	Network Model	Network Layers	Parameters	File Size/M	Average Running Time/s
Wheat spikes	Mobilev3-Deeplabv3+	268	5,635,029	21.7	1.344
	Xception-deeplabv3+	447	54,713,557	209.7	2.382
	Resnet50-Pspnet	201	46,706,626	178.5	2.541
	W48-Hrnet	999	65,860,821	252.2	2.957
Disease spots	Mobilev3-Deeplabv3+	268	5,635,029	21.7	0.0843
	Resnet50-Pspnet	201	46,706,626	178.5	0.0852
	W32-Hrnet	999	29,547,477	113.6	0.1114
	W48-Hrnet	999	65,860,821	252.2	0.1248

4.2. Wheat Spike Segmentation

The Deeplabv3+ model was trained with Mobilev3 as the backbone based on the labeled images of wheat spikes, whose construction and loss variation trend are shown in Figure 1. It can be seen from the figure that the loss curve of the model decreases sharply in the early training phase, which means that the model generalizes well on this dataset. The decreasing trend of both loss functions slows down when the number of training rounds approaches 50 rounds. Finally, after roughly 130 rounds, the loss fluctuations stabilize and the model fitting function reaches the highest segmentation accuracy. The final losses of the training and test sets were 0.16 and 0.268, respectively. The final values of *IoU*, *Recall*, *Precision*, *F-score*, *MIoU*, *mPA*, and Accuracy obtained by the model were 59.82%, 74.08%,

75.65%, 74.86%, 76.99%, 85.6%, and 94.63%, respectively, which indicated that the model was able to segment wheat spikes in the natural environment of the field effectively.

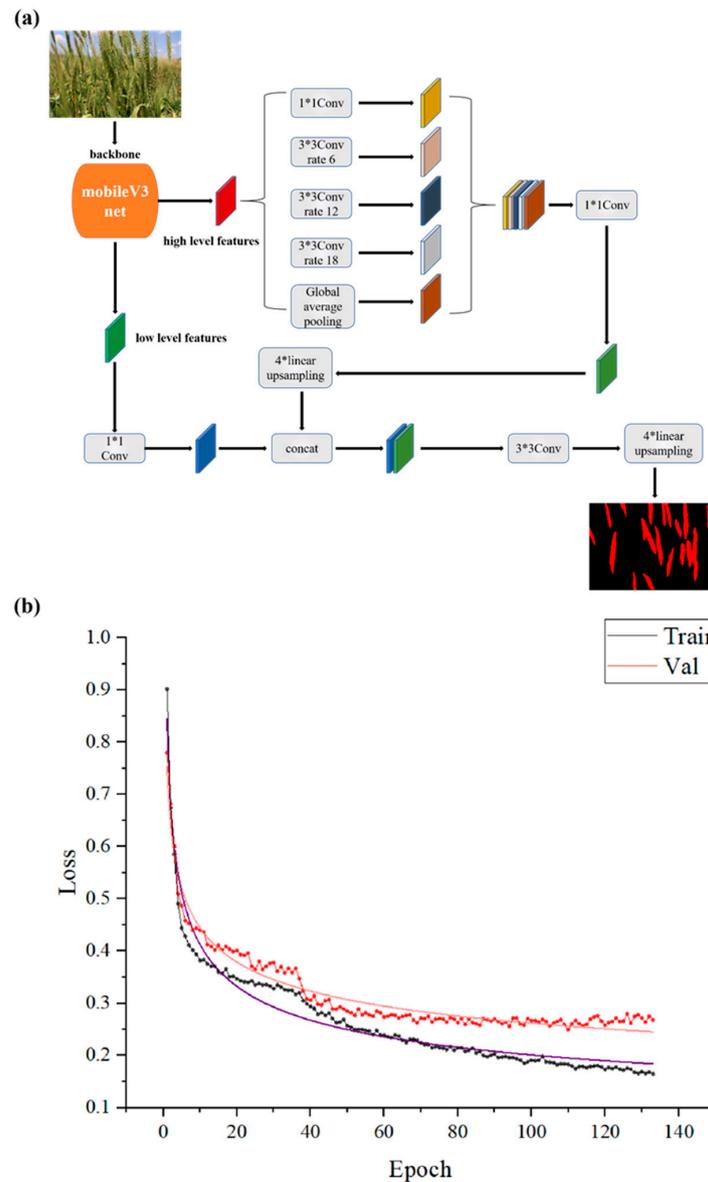


Figure 1. (a) The architecture of the MobileV3-DeeplabV3+ network for wheat spike segmentation; (b) training curves of loss against number of epochs on segmentation of wheat spikes.

The MobileV3-DeeplabV3+ model successfully identified the high density of small wheat spikes in the field (Figure 2a) and performed effective segmentation between the edges of the spikes. Due to the camera shooting angle, the images gave the appearance of wheat spikes blocking each other. For some images where the blocking phenomenon is not obvious, the network model can effectively segment the wheat spikes with some slight adhesion phenomenon of the boundary contour (red box in Figure 2b), and accurately segment what we need under the blocking of the defocused wheat spikes (yellow oval in Figure 2b). In addition, as shown in the blue box in Figure 2c, the model was able to identify the unlabeled stumpy wheat spikes that are segmented by the image boundary, which indicates the strong robustness of MobileV3-DeeplabV3+. The model was also able to accurately identify and segment the dense wheat spikes in the field of view in the images taken indoors, with dim images, and in images with low resolution, where the background is very different from the field environment (Figure 2d). It proves that MobileV3-DeeplabV3+

has good generalizability and robustness for the case where the objective conditions such as the background of the image to be detected and the brightness of the environment change significantly.

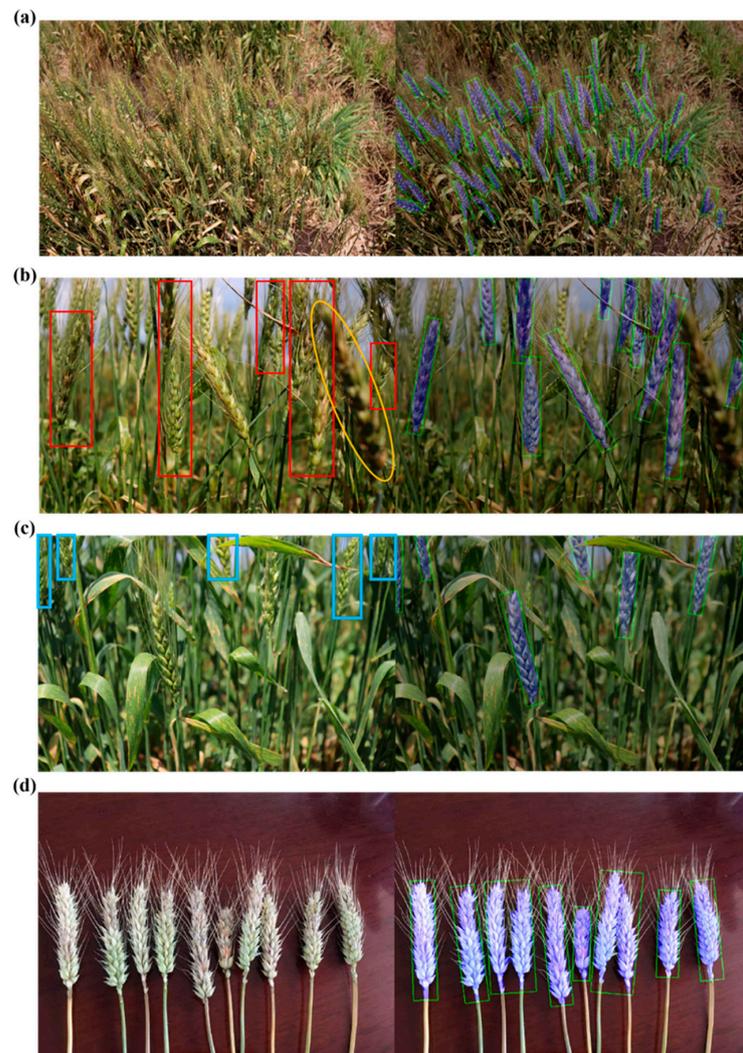


Figure 2. The prediction examples of MobileV3-Deeplabv3+ for wheat spikes under four complex conditions; (a) high density spikes; (b) sticking, shading, defocusing spikes; (c) marginal wheat spikes; (d) spikes in a dim indoor environment.

Since the quality of the images taken in the field is easily affected by natural conditions such as light and wind speed, the prediction results of the model output grayscale map often show several defective wheat spikes that are not worthy of further study as well as some noisy images that are misjudged. In order to reduce the impact on the subsequent processes, such as disease detection on segmented wheat spikes, it is necessary to pre-process the segmented images of wheat spikes before using them as input to the disease detection model. A smaller kernel (3×3) was used to erode the binarized image to separate some wheat spikes with adhering edges. The image was then inflated with a kernel of the same size to compensate for the loss of area of the normal spikes. In order to determine the noise and missing spikes, the FindContours algorithm was introduced to complete the search for all spikes contours and finally obtained the relative area of each individual spikes for that image. The median area of the spikes in each image was calculated to achieve the standardization process.

$$Area_{out} = \frac{Area_{in}}{Area_{mid}} \quad (11)$$

The $Area_{in}$ represents the original area of the spike connected area, $Area_{mid}$ represents the median area of all spikes in the image, and $Area_{out}$ represents the area of the spike connected area after standardization. The statistical results showed that the number of samples in the corresponding interval maintained a stable trend and approached the minimum value point when the standardized area of wheat spikes fell to the range of 0.3 to 0.4, so 0.35 was taken as the threshold for area rejection. When the area of a spike-connected domain was less than 0.35 times the median area of the connected domains of the spike in the picture, the connected domain was considered a “too-defective” spike not worthy of further analysis or as a noise not eliminated by morphological processing, and was excluded from the data. The processed predicted segmentation map of wheat spike was obtained by the morphological transformation and contour operation of image processing, which is shown in Figure 3. The results showed that the processed predicted images eliminate the effects of noise and “too-defective” wheat spikes very well and have a smoothing effect on the edge contours of the identified wheat spike areas.

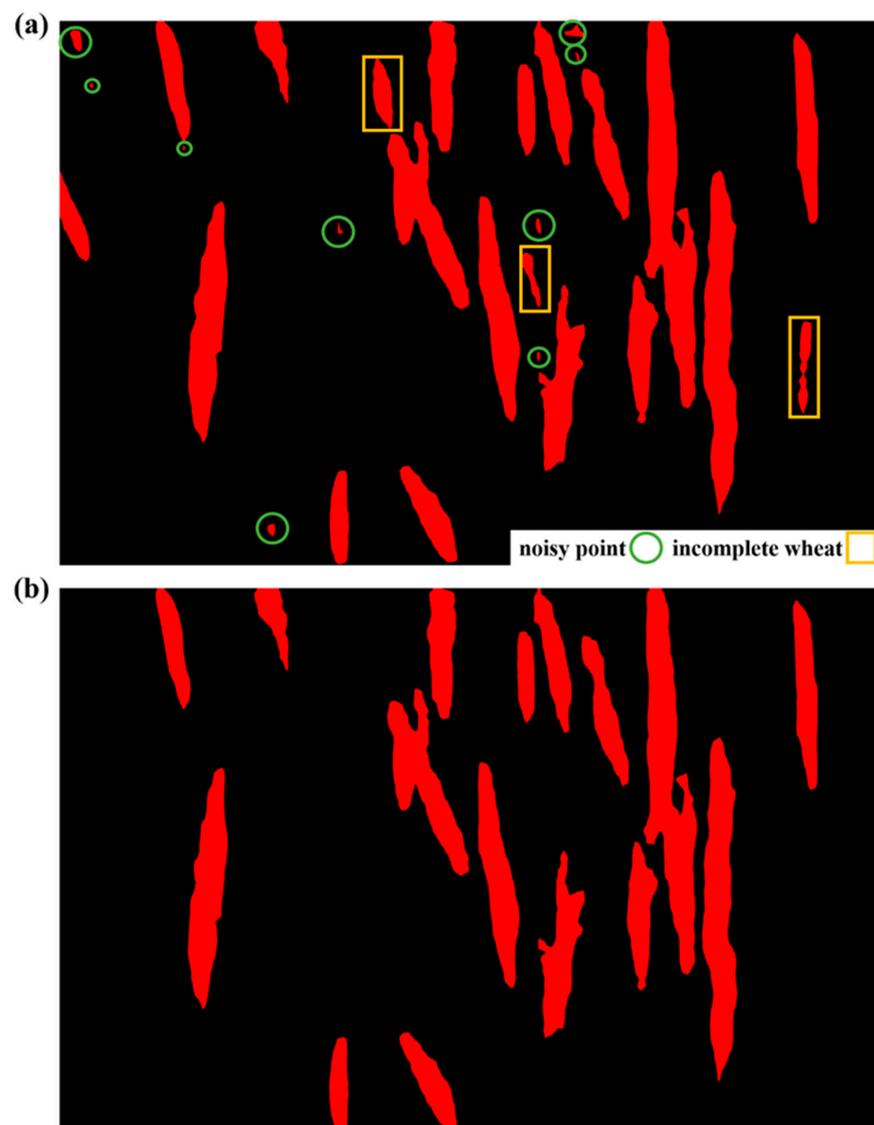


Figure 3. Comparison of the effect of wheat spikes after pre-processing; (a) the predicted wheat spike mask image before processing; (b) the predicted wheat spike mask image after processing; the green circles represent the noise in the recognition process, and the yellow boxes represent the “too-defective” wheat spikes.

To further evaluate the performance of model, 717 field real-world wheat spike images were selected in this study. The Mobilev3-Deeplabv3+ mode finally segmented 19,146 wheat ear regions among 20,488 wheat spike groups, of which 17,349 wheat spikes were correctly identified with a detection rate of 84.7%, and 1797 wheat spike regions (FN) were incorrectly identified with a false detection rate of 8.8%. The final accuracy of the wheat spike segmentation model was obtained as 97.6%. Compared to the latest wheat spike recognition study by Gao, Wang, Li and Su [19], there is a significant improvement in model accuracy (increased by 11.8%). The actual area of manually marked disease spots was specified as the independent variable x and the area of disease spots predicted by the model as the dependent variable. The 17,349 wheat crop areas were projected onto the x - y coordinate system to establish a linear regression relationship. It showed that the Pearson correlation coefficient of this fitted equation was 0.984, which indicated a perfectly positive correlation between the predicted wheat spike area and the actual wheat spike area.

4.3. Disease Spot Segmentation

After segmentation of individual wheat spikes in full-size images, the trained w48-Hrnet model was used to evaluate the lesioned areas and the structure with its loss function is shown in Figure 4. Compared with the Deeplab model used for disease spot segmentation, the convergence rate of the Hrnet model at the beginning of the model is more dramatic, which indicates that the fusion of multi-scale features in the Hrnet network is more efficient for the extraction of effective information. In addition, the model has a lower function loss at the beginning of training and the model has a stronger backbone feature extraction capability. In the training range of 25 to 50 rounds, there are significant fluctuations in the loss of the test set, which indicates that w48-Hrnet has multiple local optimal points in the direction of gradient descent, and eventually finds an optimal solution in these ranges. In the spot segmentation task, the Hrnet model showed excellent convergence performance, with the loss of the model stabilizing at 0.088 and the accuracy of the spot segmentation stabilizing at 0.099 after only 86 rounds of training. Based on the results of 3875 wheat spike images, the final values of *IOU*, *recall*, *precision*, *F-score*, *mIOU*, *mPA*, and Accuracy obtained by the model were 71.49%, 84.22%, 82.55%, 83.38%, 85.06%, 91.74%, and 98.67%, respectively, which indicated that the model was able to segment the spots on wheat spikes well.

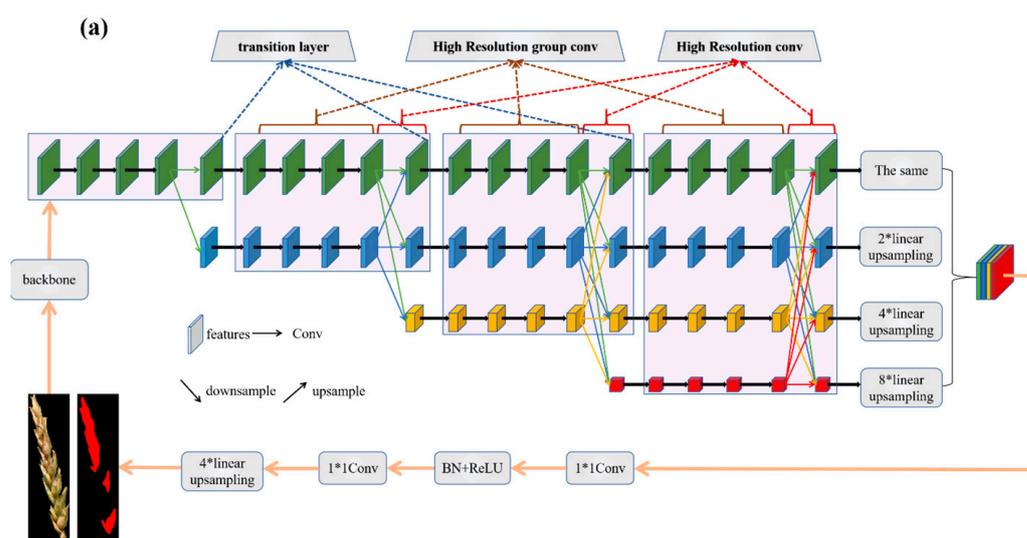


Figure 4. Cont.

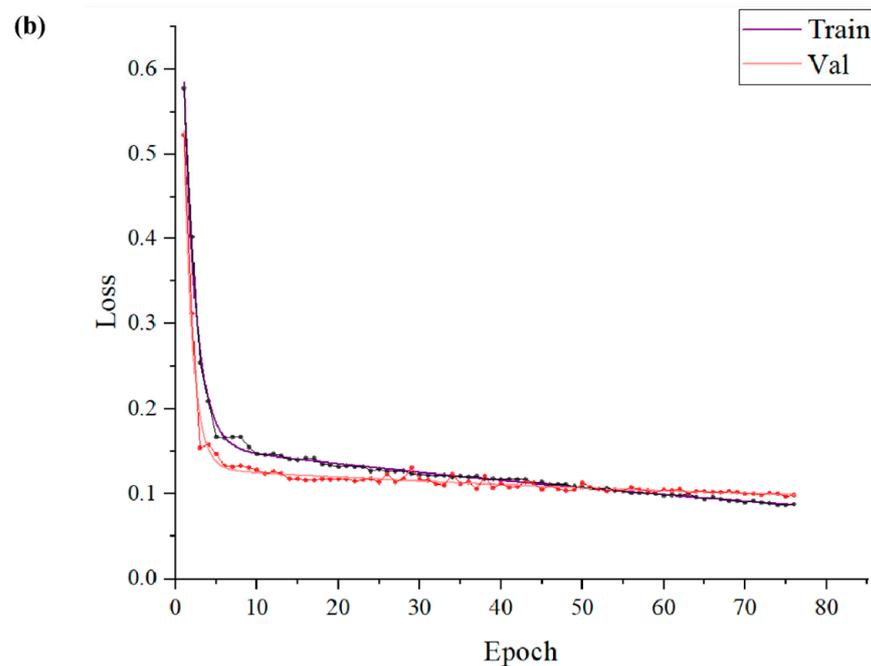


Figure 4. (a) The architecture of the w48-Hrnet network for disease spot segmentation; (b) training curves of loss against number of epochs on segmentation of disease spots; “2*” represent double, “4*” represent quadruple, “8*” represent octuple.

The w48-Hrnet model successfully performed accurate spot segmentation of single diseased wheat spikes from full-size field images and could effectively identify diseased areas of wheat spikes in strong light, low light, different types of spikes (different spikes sizes, different spike lengths, different spike tilting postures, different degrees of integrity), different degrees of disease, and under conditions of shadow interference or awn shading (Figure 5a–e). Additionally, Figure 6 shows the results of spot segmentation when the wheat spike was partially obscured by the awn or stalk of other wheat (where red boxes represented stalk obscuration and blue boxes represented awn obscuration). The image results demonstrated that the w48-Hrnet model was highly adaptable to the masking condition, and eventually transformed the original field wheat image into an output highlighting the diseased regions. The above segmentation results illustrated that the w48-Hrnet model was capable of identifying and segmenting wheat FHB spots in complex environments.

To evaluate the segmentation accuracy of the w48-Hrnet model more comprehensively, a total of 3876 sub-images containing individual diseased wheat spikes were selected as a large sample test set, and the final detection rate of 99.8% was obtained. The accuracy of the w48-Hrnet model is higher than that of FHB disease spot recognition models developed so far, including PCNN, Mask-RCNN, BlendMask, etc., with the highest surpassing by as much as 21.6% [15,19,20]. The total area of all the spots detected by the model in each sub-image was taken as the dependent variable y , and the total area of its corresponding true labeled spots was taken as the independent variable x . Then, the spot identification results of the above 3868 sub-images were labeled in the x – y coordinate system, and a linear regression relationship ($y = kx$) between x and y was established. The slope of this linear regression equation was 0.998, which proved that the trained w48-Hrnet model can be sufficiently sensitive to FHB spots and was suitable for feature generalization for most types of spots.

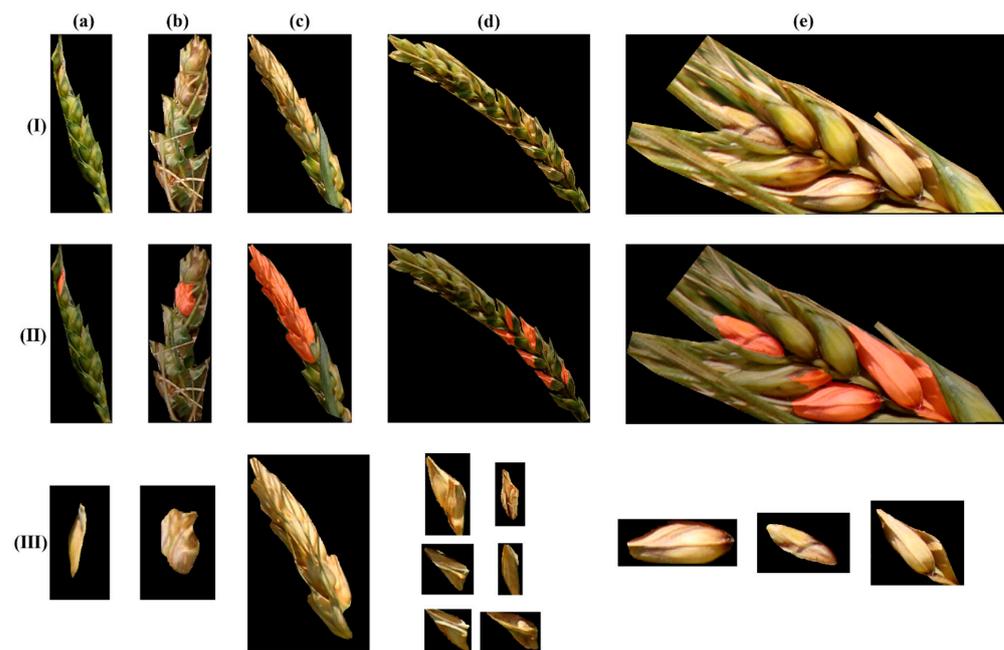


Figure 5. The prediction examples of w48-Hrnet for disease spots; (a) small-shaped spots; (b) wheat awn masking; (c) large-shaped spots; (d) multiple spots; (e) tilted wheat spikes. (I) (II) (III) represent the original image, disease spot prediction mask, and disease spot segmentation image, respectively.

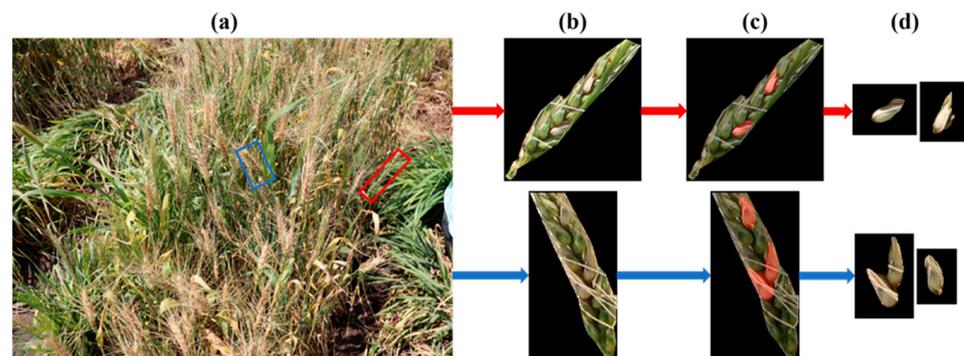


Figure 6. The prediction examples of w48-Hrnet for disease spots under masking conditions; (a) high-throughput wheat spike images in the field; (b) segmented single wheat spike images; (c) masking of disease spot predictions by w48-Hrnet; (d) segmented images of disease spots.

4.4. Classification of Wheat FHB Severity Grades

The FHB fungus is particularly damaging to the spikes of wheat. At the beginning of infection, a small amount of light brown water-soaked patches appear on the spikelet and glumes and then gradually expand to the whole wheat spike and eventually become yellow [44]. Compared with RGB, HSV color space can reflect the lightness, hue, and vividness of colors more visually. Therefore, based on the large difference in color features between diseased wheat spikes and healthy wheat spikes, 400 images of wheat spikes were selected for this study, and the information of hue, saturation, and brightness of wheat spikes was extracted using HSV color space. Firstly, the images were transformed from RGB color space to HSV color space, and then the distribution of the values of H, S and V was calculated and normalized, and finally the corresponding color bars were set to facilitate the visual observation of the difference between H, S, and V values. In this paper, wheat spikes with different disease levels (mild, medium, and severe) were selected for feature extraction of HSV color channel, and the results are shown in Figure 7. By comparing the distribution of the values of the three diseased wheat spikes on the H,

S, and V channels to better find the significant distinguishing factors between different disease levels, hue H was finally determined as the main characteristic parameter, and its value range was $[0.51, 0.80]$, i.e., the value of hue H was $[183.6^\circ, 288^\circ]$. The saturation S and the luminance V were auxiliary characteristic parameters, where the value range of S is $[0, 216^\circ]$ and the value range of V is $[144^\circ, 360^\circ]$. The higher the values of HSV, the more intense the gradient of color change of the wheat spikes caused by the FHB and the higher the corresponding disease level. For the wheat FHB grade evaluation task, the simple w48-Hrnet model had omissions in the prediction of heavily diseased areas. Therefore, the color extraction amount of disease spot features based on the HSV color channel was introduced as a weighting factor into the grade evaluation model to improve the prediction ability for higher grade diseased wheat spikes.

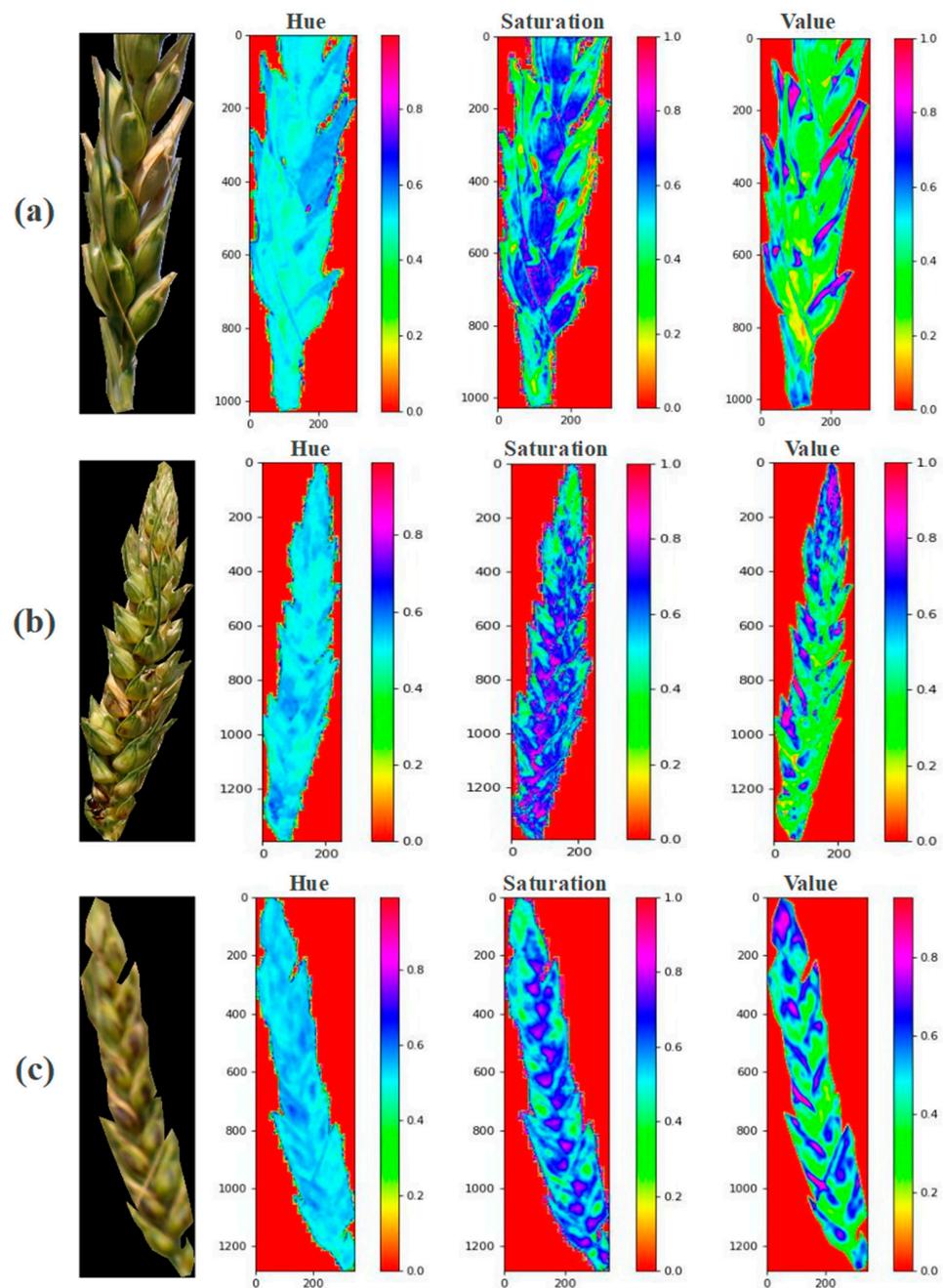


Figure 7. Distribution of wheat spike images in the H, S, and V channels; (a) slightly diseased; (b) moderately diseased; (c) severely diseased.

In order to facilitate farmers spraying the corresponding concentrations of control agents according to the different grades of FHB in a timely manner, this study divided the FHB degrees into 6 levels according to the national standard. The wheat spikes FHB severity grades in the whole dataset were predicted and statistically analyzed by Hrnet, as shown in Figure 8a. The disease distribution in this dataset was mainly in the range of “slightly occurring” to “light”, which requires timely and appropriate agronomic control measures. In addition, 57.3% of the wheat spikes were “light” and 28.9% were “moderate”, with an overall normal distribution trend. By studying the prediction of each sample level in the dataset, it can be observed that the final disease level predicted by the model is quite close to the actual disease level of the wheat spikes, which ensures that all cases of FHB can be detected and the spread of the disease can be avoided. For the classification of disease grades from the epidemic prevention and control perspective, the overall accuracy rate of 86.9% for the entire dataset of 3875 wheat spikes indicates that the model is effective in accurately determining the level of FHB in each wheat spike, ensuring that growers target the appropriate spray doses. As shown in Figure 8b, the model also achieved an overall correct rate of 83.2% for 386 wheat spikes in the validation set, indicating that the model has a relatively good robustness and can be applied to identify wheat FHB severity in different environments.

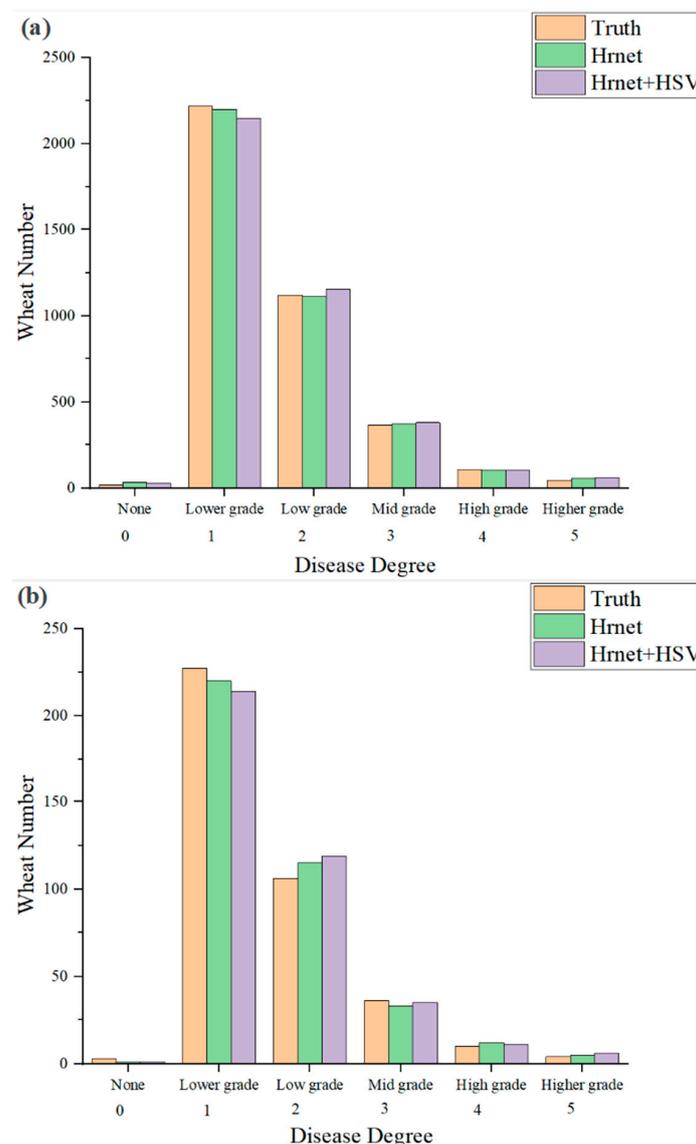


Figure 8. Distribution of wheat Fusarium head blight grades for each wheat spikelet from the epidemic prevention and control perspective (0 to 5 levels); (a) on the entire dataset; (b) on the validation set.

As the results in Figure 8 show, the neural network model predicted the severity of FHB with the majority of the samples having a regional disease level in the range of ‘slight occurrence’ to ‘light occurrence’ (0–20%). This may be due to the fact that only the spot condition of the grain was taken into account in the labelling of the dataset, but not the color change of the awn during the lesion. Therefore, in order to take a more holistic view of the disease grades, the aforementioned spot color feature extraction was introduced, and the spot area calculated by the Hrnet was weighted with the characteristic color area extracted from the HSV channel. The results of the experimental tests showed that the highest prediction accuracy was achieved when the weights accounted for by the neural network calculation and the HSV feature extraction were 0.9 and 0.1, respectively. Mao, Wang, Li, Zhou, Chen and Hu [26] proposed using an improved lightweight network to identify the severity of FHB in wheat and directly classify the dataset into zero to five levels by the model. However, because human labeling cannot account for the overall color change of wheat FHB lesions, the neural network model alone may fail to predict the moderate or severe disease areas, which may delay the control period. Although the overall prediction accuracy of the model is reduced with the introduction of spot color feature extraction, it compensates for the disease omission of the Hrnet model, so that effective control of the wheat FHB epidemic, which may continue to grow, can be achieved without overspraying.

To prioritize disease-resistant wheat lines, this study has further classified FHB severity into 15 different levels, with a greater emphasis on lines with less severity. Since higher-level diseased wheat varieties lack significance in breeding efforts, there is no need to incorporate the extraction of disease spot color features in this task. Breeders visually rated 3875 wheat spikes with FHB. The results showed that 75.7% of the overall wheat spikes in the entire dataset had good resistance (disease levels zero to six with high FHB resistance), with the highest number of spikes with level two, about 20.7% of the overall. Wheat spikes with poor disease resistance accounted for 7.0% of the total (disease levels 10 to 14). The statistics of wheat FHB severity on the entire training and validation sets are shown in Table 5.

Table 5. Severity of wheat Fusarium head blight overall verification.

Date	Type	Number of Wheat Spikes	Severity (%)		
			Mean ± Standard Deviation	Maximum	Minimum
Training set	Actual value	3490	10.8 ± 8.6	58.2	3.7
	Predicted value	3490	10.9 ± 8.7	80.3	0.0089
Test set	Actual value	386	9.7 ± 8.7	57.4	57.8
	Predicted value	386	9.9 ± 8.7	0.72	0.048

Figure 9a shows the distribution of samples on the entire dataset. The infected area of individual spikes ranged from 2.5% to 25% (disease levels two to nine) for 86.2% of the wheat spikes, with 50.9% of the samples having an infected area of 2.5% to 10% (disease levels two to four) and 35.1% having an infected area of 10% to 20% (disease level 5 to 9). Finally, using the entire dataset as the research subject, the predictive accuracy of wheat FHB severity was calculated to be 98.6% by comparing the predicted disease severity value (10.8%) with the actual disease severity value (10.7%). For the validation set (Figure 9b), the final distribution of disease grades predicted by the model was very close to the interval distribution of true grades, with no false negative samples, and its disease check rate was 100%. The test results indicated that the model’s best accuracy for classifying the severity levels of wheat FHB was of 98.1%, which was determined by comparing the predicted disease severity value (9.9%) with the actual value (9.7%). Compared to the previous Dual Mask-RCNN model [16], w48-Hrnet showed a closer approximation to the ground truth

values in predicting the severity levels of 15 disease levels. The classification accuracy improved by 20.1%.

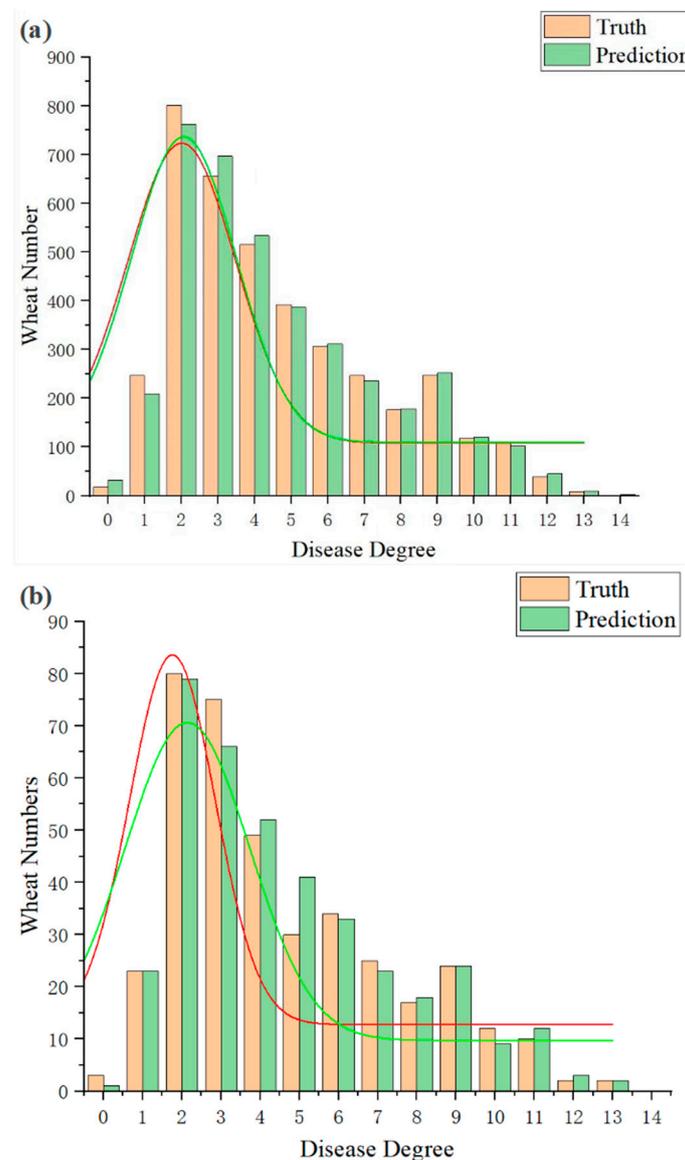


Figure 9. Distribution of wheat Fusarium head blight grades for each wheat spikelet from the breeding perspective (0 to 14 levels); (a) on the entire dataset; (b) on the validation set; the red line represents the fitted curve of Gaussian distribution for true data; the green line represents the fitted curve of Gaussian distribution for predicted data.

To further validate the accuracy of the classification results obtained from the model, the confusion matrix was applied to analyze the similarities and differences between the predicted and true results. Figure 10a depicts the confusion matrix predicted by the model over the entire dataset, where the average accuracy of the FHB severity classification was 60.3%, with the correct predictions for both lower disease levels (zero to four) and higher disease levels (nine to 14) exceeding 60%. To test the robustness of the model, Figure 10b depicts the confusion matrix predicted by the model on the validation set, and the average correct rate of disease grade classification to distinguish differences in disease resistance among different wheat spikes was 53.9%, and the probability that the grade classification error was within one level was 87.0%. This further demonstrates the outstanding detection performance of the method.

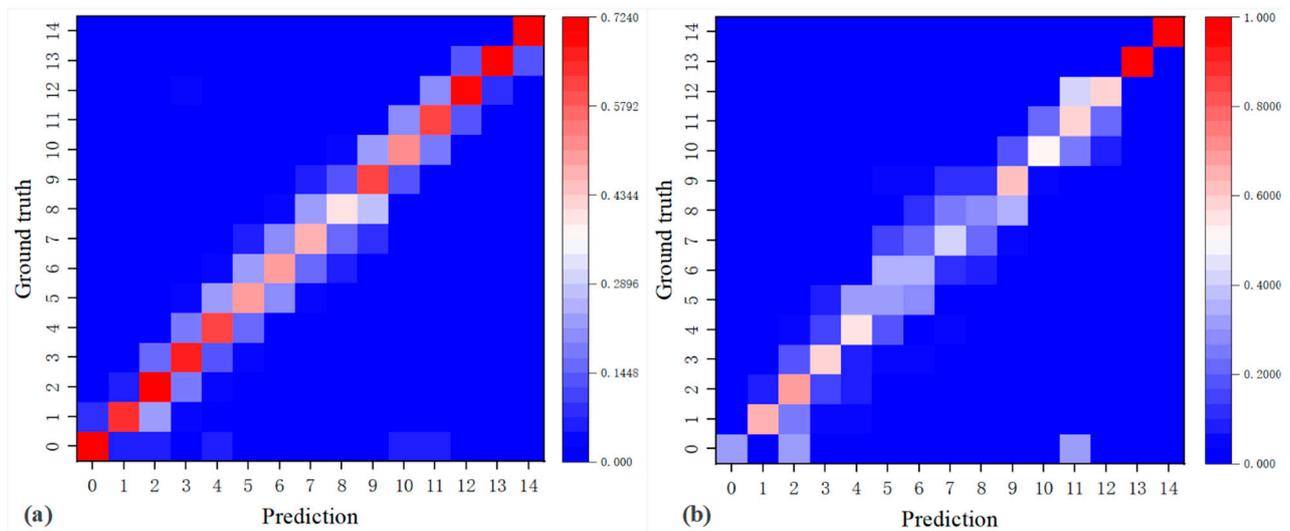


Figure 10. Confusion matrix for 15 severity levels of wheat FHB; (a) on the entire dataset; (b) on the validation set.

4.5. Wheat FHB Grades Integrated Detection System

The various detection and computational modules of the study were matched and combined to build an integrated detection model for wheat FHB levels. The image of the wheat spikes to be detected was input into the model, and the detection system preprocessed the image with some filtering operations then passed it to the trained MobileV3-Deeplabv3+ model for segmentation prediction of wheat spikes, and finally output the mask map of the spike segmentation. The system then performed some morphological operations on the mask map and calculated the connectivity domain to remove image noise. The individual mask images were obtained by segmenting each spike of wheat in the whole image. Finally, the mask image was binarized and overlaid with the original image to obtain a single RGB color image of the segmented wheat spike. After the above operation, the system would pass the images of wheat spikes to the disease color feature extraction model and the trained w48-Hrnet network, and the former extracted the disease color features in HSV color space in terms of hue, saturation, and luminance according to the appropriate thresholds and obtain the corresponding color factor values of the disease spots. The latter segmented the spots in the wheat spikes according to other geometric features such as texture and shape of the spots learned from the model and generated the corresponding mask map. The number of pixels in the spots was calculated, and the color image of the spots was generated by the same binarization and superposition operations.

Then, the system performed the wheat FHB severity ranking according to the operation purpose. For the grade evaluation based on disease control, the model weighted the number of pixel points extracted from the color feature of disease spots with the number of pixel points obtained from the neural network segmentation, and then classified the disease into six levels from zero to five. For the grade evaluation based on the breeding of resistant wheat lines, the disease was classified into 15 levels from zero to 14 based directly on the percentage of the total number of diseased pixels obtained from the model segmentation to the total pixels number of the whole wheat spike. Finally, the system would mark and annotate the obtained wheat spikes, the disease spot segmentation, and the disease grade prediction in the corresponding position of the original wheat spike image and saved and displayed them as the output of the whole detection synthesis process. The specific procedure working steps are shown in Figure 11.

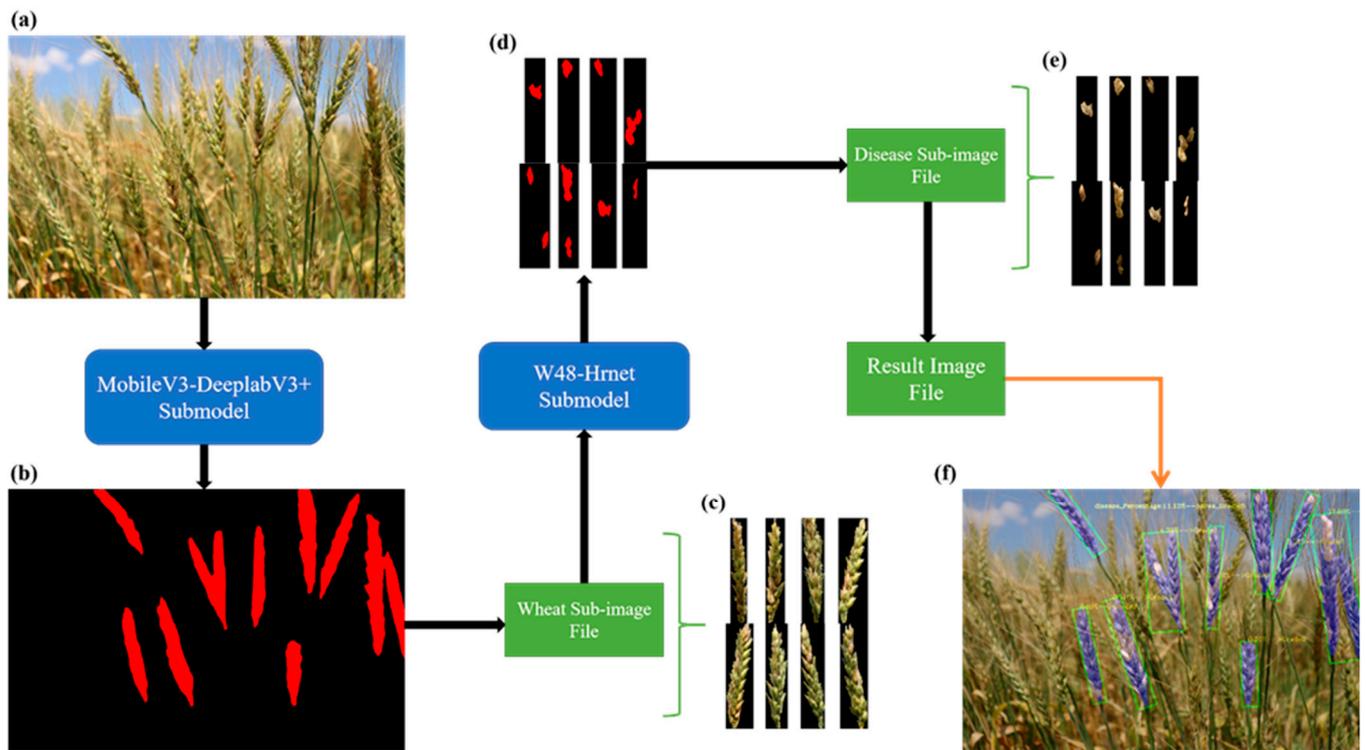


Figure 11. Working steps of the integrated multi-model fusion system for wheat Fusarium head blight severity diagnosis; (a) image to be detected; (b) wheat spike segmentation mask map; (c) wheat spike sub image; (d) disease spot segmentation mask map; (e) disease spot sub-image; (f) result image.

The output images are shown in Figure 12, where Figure 12a,b represent typical images of different wheat spike distributions and different disease levels in the dataset. The results showed that the model can complete the processes of spike segmentation, disease spot segmentation and grade evaluation in a consistent and integrated manner. The accuracy of segmentation and grade prediction has achieved a good result, with 92.6% correct diagnosis of the regional wheat population for the grade of FHB disease.

accuracy of wheat spikes was effectively improved by 19.8% compared to the study by Su, Zhang, Yang, Page, Szinyei, Hirsch and Steffenson [16]. Additionally, FHB spots detection rate was also improved by 6% over the latest MobileNetv2-YOLOv4 model [20]. Although the segmentation model used in this study was effective in identifying all wheat spike regions in the image, there was a problem of single spike segmentation for multiple wheat with adhering edges. In the future, more advanced classifiers (such as YOLOv5 and YOLOv6) [45,46] should be added to the model to fully learn the morphological features of independent wheat spikes. In addition, the annotation of FHB spot data was time-consuming and determined by experience. Meanwhile, more advanced algorithms should be developed to accomplish automated annotation for improving the capacity and precision of the dataset.

Feature color extraction of visible symptoms combined with segmentation model to evaluate wheat disease severity was innovatively applied in this study. The HSV channel color analysis based on Hrnet was introduced, both of which were used to make a comprehensive diagnosis of wheat disease from the perspectives of epidemic control with 9:1 weighting factor, making up for the weakness of deep learning color perception and allowing for more timely and effective spraying of agents. In a recent study, Gao, Wang, Li and Su [19] proposed a Dual BlendMask network architecture to classify wheat FHB severity with 91.8% accuracy. The spectral vegetation index was used for FHB evaluation with an accuracy of 89.8% [47]. None of these methods using only network models or spectral features to determine disease severity were as accurate as the scheme proposed in the current study (98.6%). There was a study that fused image and spectral features to build diagnostic models [48]. The particle swarm optimization support vector machine (PSO-SVM) algorithm was used to analyze wheat FHB features, but it was far less predictive than deep learning architectures because of its simple structure. Although it is superior in terms of accuracy, the method in this study needs to extract a large amount of image feature information. This process demands a large, labeled dataset for training, which can be time-consuming and labor-intensive. The generative adversarial networks (GAN) [49] should be considered to augment high-throughput wheat images in the next work. Experimentally collected RGB images by remote sensing equipment are used instead of existing datasets to ensure the generalizability and stability of the model.

The results of this study showed that the established architecture has great potential for real-time assessment of FHB severity in wheat. For field operations, wheat spike detection, disease spot segmentation, feature color extraction, and disease grade classification were integrated, resulting in an all-in-one system for wheat FHB diagnosis. A WeChat applet based on the integrated process and web application is also being developed. According to the current progress, the platform can detect a high-throughput wheat image in 70~80s time to obtain disease data information and treatment measures. It is expected that the system will be built on a higher configuration server and the running time will be greatly reduced soon. The proximal sensing equipment developed in the future would be deployed on UAV phenotyping platforms and a ground-based motorized vehicle used in the field. It is expected that drones and mobile phenotyping devices will enable real-time photography and use cloud servers for RGB image processing, ultimately making the FHB assessment reports available to end users. To address the interference of high-throughput environmental factors during real-time diagnosis, it is proposed that a mechanical arm or similar motion scheme be utilized to place a whiteboard behind the target wheat spikes. This strategy aims to provide a clearer and more consistent background contrast, ultimately improving detection accuracy. An intelligent monitoring platform is planned to integrate the device with agricultural IoT and field weather stations to expand the potential of deep learning for agricultural applications. The implementation of related research will significantly contribute to the large-scale precision control of FHB in wheat and guarantee national food security.

6. Conclusions

An all-in-one diagnostic system for FHB severity assessment in wheat based on multi-model fusion was established. Fourteen different network models were developed and trained. Compared with other network models, the Deeplabv3+ model with Mobilenetv3 backbone showed the best comprehensive performance with mPA, accuracy and running time values of 83.74%, 94.76% and 1.344s in wheat segmentation, respectively. The w48-Hrnet model exhibited the highest training accuracy of 98.67% in disease area detection. Wheat FHB was precisely classified into six and 15 evaluation levels, respectively, and the severity identification effect met the requirements of the target tasks. The method integrating HSV color feature extraction and CNN demonstrated more rational grading results to provide valid information on the efficacy of disease. Furthermore, a monitoring process combining segmentation, extraction and grading was proposed, which is being systematically deployed on mobile terminals. Further work is needed to enhance the development of the multi-modal wheat image assessment system with new classifiers. Thus, the smart monitoring platform of FHB will be accomplished. Obviously, this study will be of great help in determining the appropriate amount of agents to spray and breed-resistant wheat varieties, which provides technology for the development of precision agriculture.

Author Contributions: Conceptualization, W.-H.S.; methodology, J.-J.L. and W.-H.S.; investigation, J.-J.L.; resources, W.-H.S.; writing—original draft preparation, Y.-H.W. and J.-J.L.; writing—review and editing, W.-H.S.; visualization, J.-J.L. and Y.-H.W.; supervision, W.-H.S.; project administration, W.-H.S.; funding acquisition, W.-H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 32101610.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Goyal, L.; Sharma, C.M.; Singh, A.; Singh, P.K. Leaf and spike wheat disease detection & classification using an improved deep convolutional architecture. *Inform. Med. Unlocked* **2021**, *25*, 100642.
- Wang, Y.-H.; Su, W.-H. Convolutional neural networks in computer vision for grain crop phenotyping: A review. *Agronomy* **2022**, *12*, 2659. [[CrossRef](#)]
- Saccon, F.A.; Parcey, D.; Paliwal, J.; Sherif, S.S. Assessment of Fusarium and deoxynivalenol using optical methods. *Food Bioprocess Technol.* **2017**, *10*, 34–50. [[CrossRef](#)]
- 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](#)]
- Femenias, A.; Gatiús, F.; Ramos, A.J.; Sanchis, V.; Marín, S. Use of hyperspectral imaging as a tool for Fusarium and deoxynivalenol risk management in cereals: A review. *Food Control* **2020**, *108*, 106819. [[CrossRef](#)]
- Sood, S.; Singh, H. An implementation and analysis of deep learning models for the detection of wheat rust disease. In Proceedings of the 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 3–5 December 2020; pp. 341–347.
- Liu, B.-Y.; Fan, K.-J.; Su, W.-H.; Peng, Y. Two-stage convolutional neural networks for diagnosing the severity of alternaria leaf blotch disease of the apple tree. *Remote Sens.* **2022**, *14*, 2519. [[CrossRef](#)]
- Su, W.-H.; Yang, C.; Dong, Y.; Johnson, R.; Page, R.; Szinyei, T.; Hirsch, C.D.; Steffenson, B.J. Hyperspectral imaging and improved feature variable selection for automated determination of deoxynivalenol in various genetic lines of barley kernels for resistance screening. *Food Chem.* **2021**, *343*, 128507. [[CrossRef](#)]
- Peiris, K.; Pumphrey, M.; Dong, Y.; Maghirang, E.; Berzonsky, W.; Dowell, F. Near-infrared spectroscopic method for identification of fusarium head blight damage and prediction of deoxynivalenol in single wheat kernels. *Cereal Chem.* **2010**, *87*, 511–517. [[CrossRef](#)]
- Jin, X.; Jie, L.; Wang, S.; Qi, H.J.; Li, S.W. Classifying wheat hyperspectral pixels of healthy heads and Fusarium head blight disease using a deep neural network in the wild field. *Remote Sens.* **2018**, *10*, 395. [[CrossRef](#)]
- Zhao, J.; Yan, J.; Xue, T.; Wang, S.; Qiu, X.; Yao, X.; Tian, Y.; Zhu, Y.; Cao, W.; Zhang, X. A deep learning method for oriented and small wheat spike detection (OSWSDet) in UAV images. *Comput. Electron. Agric.* **2022**, *198*, 107087. [[CrossRef](#)]
- Kumar, D.; Kukreja, V. Deep learning in wheat diseases classification: A systematic review. *Multimed. Tools Appl.* **2022**, *81*, 10143–10187. [[CrossRef](#)]

13. Genaev, M.A.; Skolotneva, E.S.; Gulyaeva, E.I.; Orlova, E.A.; Bechtold, N.P.; Afonnikov, D.A. Image-based wheat fungi diseases identification by deep learning. *Plants* **2021**, *10*, 1500. [[CrossRef](#)] [[PubMed](#)]
14. 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](#)]
15. 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](#)]
16. Su, W.-H.; Zhang, J.; Yang, C.; Page, R.; Szinyei, T.; Hirsch, C.D.; Steffenson, B.J. Automatic evaluation of wheat resistance to fusarium head blight using dual mask-RCNN deep learning frameworks in computer vision. *Remote Sens.* **2020**, *13*, 26. [[CrossRef](#)]
17. Qiu, R.; Yang, C.; Moghimi, A.; Zhang, M.; Steffenson, B.J.; Hirsch, C.D. Detection of fusarium head blight in wheat using a deep neural network and color imaging. *Remote Sens.* **2019**, *11*, 2658. [[CrossRef](#)]
18. Hasan, M.M.; Chopin, J.P.; Laga, H.; Miklavcic, S.J. Detection and analysis of wheat spikes using convolutional neural networks. *Plant Methods* **2018**, *14*, 1–13. [[CrossRef](#)]
19. Gao, Y.; Wang, H.; Li, M.; Su, W.-H. Automatic Tandem Dual BlendMask Networks for Severity Assessment of Wheat Fusarium Head Blight. *Agriculture* **2022**, *12*, 1493. [[CrossRef](#)]
20. 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](#)]
21. Sarayloo, Z.; Asemani, D. Designing a classifier for automatic detection of fungal diseases in wheat plant: By pattern recognition techniques. In Proceedings of the 2015 23rd Iranian Conference on Electrical Engineering, Tehran, Iran, 10–14 May 2015; pp. 1193–1197.
22. Zhang, D.-Y.; Luo, H.-S.; Cheng, T.; Li, W.-F.; Zhou, X.-G.; Gu, C.-Y.; Diao, Z. Enhancing wheat Fusarium head blight detection using rotation Yolo wheat detection network and simple spatial attention network. *Comput. Electron. Agric.* **2023**, *211*, 107968. [[CrossRef](#)]
23. Zhang, D.-Y.; Chen, G.; Yin, X.; Hu, R.-J.; Gu, C.-Y.; Pan, Z.-G.; Zhou, X.-G.; Chen, Y. Integrating spectral and image data to detect Fusarium head blight of wheat. *Comput. Electron. Agric.* **2020**, *175*, 105588. [[CrossRef](#)]
24. 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](#)]
25. Navale, P.R.; Basapur, S.B. Deep Learning based Automated Wheat Disease Diagnosis System. In Proceedings of the 2023 International Conference for Advancement in Technology (ICONAT), Goa, India, 21–22 January 2023; pp. 1–5.
26. 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. [[CrossRef](#)]
27. Steffenson, B. Fusarium head blight of barley: Impact, epidemics, management, and strategies for identifying and utilizing genetic resistance. In *Fusarium Head Blight Wheat Barley*; ASP Press: St. Paul, MN, USA, 2003; pp. 241–295.
28. Yurtkulu, S.C.; Şahin, Y.H.; Unal, G. Semantic segmentation with extended DeepLabv3 architecture. In Proceedings of the 2019 27th Signal Processing and Communications Applications Conference (SIU), Sivas, Turkey, 24–26 April 2019; pp. 1–4.
29. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
30. Wu, H.; Liang, C.; Liu, M.; Wen, Z. Optimized HRNet for image semantic segmentation. *Expert Syst. Appl.* **2021**, *174*, 114532. [[CrossRef](#)]
31. Zhao, X.; Yuan, Y.; Song, M.; Ding, Y.; Lin, F.; Liang, D.; Zhang, D. Use of unmanned aerial vehicle imagery and deep learning unet to extract rice lodging. *Sensors* **2019**, *19*, 3859. [[CrossRef](#)] [[PubMed](#)]
32. Zou, K.; Liao, Q.; Zhang, F.; Che, X.; Zhang, C. A segmentation network for smart weed management in wheat fields. *Comput. Electron. Agric.* **2022**, *202*, 107303. [[CrossRef](#)]
33. Arinichev, I.; Polyanskikh, S. Arinicheva IV. Semantic segmentation of rusts and spots of wheat. *Comput. Opt.* **2023**, *47*, 118–125.
34. Pan, Q.; Gao, M.; Wu, P.; Yan, J.; Li, S. A deep-learning-based approach for wheat yellow rust disease recognition from unmanned aerial vehicle images. *Sensors* **2021**, *21*, 6540. [[CrossRef](#)]
35. Deng, J.; Zhou, H.; Lv, X.; Yang, L.; Shang, J.; Sun, Q.; Zheng, X.; Zhou, C.; Zhao, B.; Wu, J. Applying convolutional neural networks for detecting wheat stripe rust transmission centers under complex field conditions using RGB-based high spatial resolution images from UAVs. *Comput. Electron. Agric.* **2022**, *200*, 107211. [[CrossRef](#)]
36. Wu, Z.; Shen, C.; Van Den Hengel, A. Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recognit.* **2019**, *90*, 119–133. [[CrossRef](#)]
37. Chen, H.-Y.; Su, C.-Y. An enhanced hybrid MobileNet. In Proceedings of the 2018 9th International Conference on Awareness Science and Technology (iCAST), Fukuoka, Japan, 19–21 September 2018; pp. 308–312.
38. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
39. 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.
40. Zhang, Z. Improved adam optimizer for deep neural networks. In Proceedings of the 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 4–6 June 2018; pp. 1–2.

41. Lu, Z.; Wang, J.; Song, J. Multi-resolution CSI feedback with deep learning in massive MIMO system. In Proceedings of the ICC 2020–2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–6.
42. Zhang, Z.; Sabuncu, M. Generalized cross entropy loss for training deep neural networks with noisy labels. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 8792–8802.
43. Aboukhaddour, R.; Fetch, T.; McCallum, B.D.; Harding, M.W.; Beres, B.L.; Graf, R.J. Wheat diseases on the prairies: A Canadian story. *Plant Pathol.* **2020**, *69*, 418–432. [[CrossRef](#)]
44. Bai, G.; Shaner, G. Management and resistance in wheat and barley to Fusarium head blight. *Annu. Rev. Phytopathol.* **2004**, *42*, 135–161. [[CrossRef](#)] [[PubMed](#)]
45. Yung, N.D.T.; Wong, W.; Juwono, F.H.; Sim, Z.A. Safety Helmet Detection Using Deep Learning: Implementation and Comparative Study Using YOLOv5, YOLOv6, and YOLOv7. In Proceedings of the 2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST), Virtual, 26–28 October 2022; pp. 164–170.
46. Liu, M.; Su, W.-H.; Wang, X.-Q. Quantitative Evaluation of Maize Emergence Using UAV Imagery and Deep Learning. *Remote Sens.* **2023**, *15*, 1979. [[CrossRef](#)]
47. Zhang, N.; Pan, Y.; Feng, H.; Zhao, X.; Yang, X.; Ding, C.; Yang, G. Development of Fusarium head blight classification index using hyperspectral microscopy images of winter wheat spikelets. *Biosyst. Eng.* **2019**, *186*, 83–99. [[CrossRef](#)]
48. Huang, L.; Li, T.; Ding, C.; Zhao, J.; Zhang, D.; Yang, G. Diagnosis of the severity of Fusarium head blight of wheat ears on the basis of image and spectral feature fusion. *Sensors* **2020**, *20*, 2887. [[CrossRef](#)]
49. Gui, J.; Sun, Z.; Wen, Y.; Tao, D.; Ye, J. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Trans. Knowl. Data Eng.* **2021**, *4*, 3313–3332. [[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.