

Article

Automatic Tandem Dual BlendMask Networks for Severity Assessment of Wheat Fusarium Head Blight

Yichao Gao ^{1,†}, Hetong Wang ^{1,†}, Man Li ² and Wen-Hao Su ^{3,*} ¹ College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China² Department of Engineering and Applied Sciences, Xinhua College of Ningxia University, Yinchuan 750030, China³ College of Engineering, China Agricultural University, Beijing 100083, China

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

† These authors are co-first authors.

Abstract: Fusarium head blight (FHB) disease reduces wheat yield and quality. Breeding wheat varieties with resistance genes is an effective way to reduce the impact of this disease. This requires trained experts to assess the disease resistance of hundreds of wheat lines in the field. Manual evaluation methods are time-consuming and labor-intensive. The evaluation results are greatly affected by human factors. Traditional machine learning methods are only suitable for small-scale datasets. Intelligent and accurate assessment of FHB severity could significantly facilitate rapid screening of resistant lines. In this study, the automatic tandem dual BlendMask deep learning framework was used to simultaneously segment the wheat spikes and diseased areas to enable the rapid detection of the disease severity. The feature pyramid network (FPN), based on the ResNet-50 network, was used as the backbone of BlendMask for feature extraction. The model exhibited positive performance in the segmentation of wheat spikes with precision, recall, and MIoU (mean intersection over union) values of 85.36%, 75.58%, and 56.21%, respectively, and the segmentation of diseased areas with precision, recall, and MIoU values of 78.16%, 79.46%, and 55.34%, respectively. The final recognition accuracies of the model for wheat spikes and diseased areas were 85.56% and 99.32%, respectively. The disease severity was obtained from the ratio of the diseased area to the spike area. The average accuracy for FHB severity classification reached 91.80%, with the average *F1*-score of 92.22%. This study demonstrated the great advantage of a tandem dual BlendMask network in intelligent screening of resistant wheat lines.



Citation: 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. <https://doi.org/10.3390/agriculture12091493>

Academic Editor: Maciej Zaborowicz

Received: 21 August 2022

Accepted: 14 September 2022

Published: 18 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/>).

Keywords: deep learning; wheat spike; Fusarium head blight; object recognition; image segmentation

1. Introduction

Wheat is one of the main food sources for humans, contributing about 20% of the total dietary calories and proteins [1]. With the growing world population, the steady increase in wheat production is of great significance. Wheat production has traditionally been threatened by various diseases, pests, and abiotic stresses. According to statistics, the global wheat yield loss caused by fungal diseases is as high as 15% to 20% [2]. Among them, wheat Fusarium head blight (FHB) is one of the most harmful fungal diseases. It is mainly caused by *Fusarium graminearum*, infecting spike flowers at the flowering stage and expanding along the panicle axis during grain filling and maturation. The production and accumulation of toxins, such as deoxynivalenol (DON), *Fusarium nivalenol* (NIV), and zearalenol (ZEN) [3], can reduce the yield and quality of wheat, causing great harm to human and animal health [4]. The breeding of FHB-resistant varieties is one of the most important means to mitigate the effects of the disease. In order to develop resistant varieties, hundreds of lines must be assessed each year for FHB severity. Protocols for assessing FHB resistance often rely on manual detection. The severity of FHB in wheat can be accurately scored by counting infected spikelets and calculating its percentage in

total spikelets [5]. However, this traditional approach is time-consuming, labor-intensive, and prone to human error. Thus, there is an important need to develop a more effective, non-destructive, and high-throughput approach to assess this disease in the field.

The commonly used methods for FHB detection mainly focus on visual analysis, chromatography, polymerase chain reaction (PCR), and enzyme-linked immunosorbent assay (ELIS). Inspection by experienced experts is prone to subjective interference and human error. Biochemical methods, such as chromatography and ELIS, are very accurate, but they often require complex processing steps, which are not suitable for analyzing a large number of FHB-infected wheat spikes in reality [6]. In recent years, imaging and spectroscopic methods, including near infrared spectroscopy (NIRS) and hyperspectral imaging (HSI), have shown strong potential in agriculture and food, especially in crop disease detection [7–16]. NIRS is based on differences in the absorption, emission, or transmission of light by substances, whose fingerprint features are related to changes in the apparent color, internal composition, and structure of the sample [17]. Peiris et al. [18] proposed an automated single-kernel NIRS method for classifying healthy and FHB-infected wheat with accuracy as high as 99.9%. However, NIRS can only obtain point features of the kernel and cannot achieve large-scale and rapid classification. As a non-invasive, high-throughput, and remote sensing method for plant phenotyping, HSI can merge the spatial and spectral information into a 3D data matrix, but the data in the matrix is too large to be used in real-time analysis [4,19].

With the development of artificial intelligence, color imaging based on machine learning has made great progress. Conventional machine learning methods, such as random forest (RF), K-nearest neighbor (KNN), linear discriminant analysis (LDA), and partial least squares discriminant analysis, have been widely used for crop plant detection and disease evaluation [4,9]. However, these methods perform poorly on large-scale datasets and complex feature scenarios. As a new field of machine learning, deep learning (DL) has gradually begun to show its advantages in image classification, object detection, and natural language processing [20,21]. Deep learning methods have become a preferred approach for disease identification in agricultural fields [22]. Convolutional neural networks (CNN) have been widely used in agriculture in recent years by extracting key features that use different combinations of layers, the translation invariance of convolutional operators, and spatial relationships between adjacent data. Classical CNNs, such as LeNet-5 [23], AlexNet [24], ResNet [25], and VGG [26] were successfully employed for plant disease detection [27,28]. By combining the machine learning and deep learning methods, Hassan SM et al. [29] proposed two methods including shallow VGG with RF and shallow VGG with Xgboost to identify the diseases in corn, potato, and tomato, with the average accuracy as high as 95.70%. Hassan et al. [30] proposed a novel CNN model based on the inception and residual connection to classify the disease of four different plants. The testing accuracies on plant village, rice, and cassava datasets were 99.39%, 99.66%, and 76.59%, respectively. In practical application, the increase of convolutional layers and convolutional kernels of a CNN can enable the model to extract more abstract and refined features, thereby exhibiting excellent performance [26]. However, this procedure may cause the CNN to lose focus on the features and suffer from a vanishing gradient. Residual modules effectively create shortcuts in a sequential network. It leverages a shortcut connection method to weaken the continuous multiplication effect in gradient backpropagation to solve the vanishing gradient problem [25,31]. Girshick, the first author to apply deep learning for object detection, used the region-based convolutional neural network (R-CNN) model to increase the detection rate from 35.1% to 53.7% on the PASCAL VOC dataset [32]. Subsequently, Girshick launched Fast R-CNN and Faster R-CNN network models based on R-CNN, which greatly improved the accuracy and speed of the algorithm in wheat spike detection [33–35]. A pulse-coupled neural network (PCNN) with K-means clustering of the improved artificial bee colony (IABC) was developed by Zhang et al. [36] to segment wheat spikes infected with FHB. However, in that research, only one spike in the picture was taken into consideration, which was not practical for high throughput detection in

the field condition. Kumar et al. [37] used a deep convolution neural network (DCNN) to automatically classify four wheat rust diseases, achieving an accuracy of 97.16%.

Mask R-CNN is a deep learning algorithm that achieves instance segmentation [38]. Mask R-CNN extends Faster R-CNN by adding a branch that predicts an object mask, replacing the ROI Pooling layer with ROI Align. ROI Align solves the problem of misalignment caused by quantization in ROI Pooling operations. Kumar et al. [39] quantified the severity of loose smut in wheat using Mask R-CNN. However, the degree of disease was not graded. Additionally, the accuracy for disease degree classification was not verified. In addition, the calculation of the proportion of the diseased area of the whole leaf was not clearly indicated. Kumar et al. [40] used the Mask R-CNN to recognize the wheat yellow rust disease. However, this study did not include enough datasets, and the result of background segmentation was not adequate. Yang et al. [41] demonstrated the potential of Mask R-CNN to identify leaves in plant images for rapid phenotyping with an average accuracy of up to 91.5%. Su et al. [42] utilized a dual deep learning framework based on Mask R-CNN to detect the wheat for FHB severity in field trials. Due to the deficiencies of the previous strategies, more advanced models are needed to evaluate the resistance of wheat to FHB. In recent years, Chen et al. [43] proposed a more advanced instance segmentation model of BlendMask compared to Mask R-CNN. BlendMask is a state-of-the-art instance segmentation method based on a fully convolutional one-stage (FCOS) object detection network [43]. Xi et al. [44] used two instance segmentation networks including BlendMask and Mask R-CNN to delineate the ginkgo tree crowns. The results showed that the capability of BlendMask outperformed Mask R-CNN. Compared to Mask R-CNN, the BlendMask network model has the characteristics of less computation, higher mask quality, and stable inference time. Therefore, it is necessary to apply the BlendMask model for the image segmentation of wheat spikes and target recognition of the diseased areas.

The assessment of plant disease severity is another important and challenging task in agriculture. Efficient evaluation methods should be of great help to growers and breeders. Table 1 summarizes some studies using advanced deep learning methods to assess crop disease severity. Esgario et al. [45] established five models to successfully classify the severity of coffee disease into five grades. ResNet50 achieved the best accuracy of 84.13%. Pan et al. [46] proposed Faster R-CNN (VGG16) and Siamese networks for strawberry leaf scorch severity estimation. An accuracy of 88.3% was achieved on a new dataset, but the manual labeling method is time-consuming and prone to subjective errors. Joshi et al. [47] used VirLeafNet to classify Vigna mungo disease into three grades and the accuracy reached 91.5%. Although the aforementioned studies performed well in determining the severity of plant diseases, the accuracy was lower than the protocol proposed in the current study. In another study, Zhang et al. [48] used the ratio of the number of diseased wheats to the total number of wheats as a method for evaluating disease severity. However, this method ignored the overlapping wheat spikes in the image. In the studies of Ji et al. [49] and Wu et al. [50], improved YoLo V5 and deeplabV3+ achieved the accuracies of 97.75% and 95.34%, respectively, in evaluating disease severities of grape and pepper, respectively. Different from the single-stage segmentation methods mentioned above, Liu et al. [51] developed a two-stage framework to automatically estimate the severity of apple leaf disease in the field, yielding an accuracy of 96.41%. However, the applicability of the framework used in their study lacked validation on multi-leaf images.

With the advancements in machine learning and deep learning techniques, the methods for plant disease detection have shown promising performance. However, the original data for model training were mostly acquired in a lab, which would limit the performance of the method in real field conditions [52]. The main objective of this study was to investigate the feasibility of automatic tandem dual BlendMask networks for assessments of wheat for FHB severity in field trials. The specific steps of this study were to: (1) capture high-quality images of wheat spikes in the field, (2) annotate wheat spikes and diseased areas in the raw images, (3) train a BlendMask model to detect and segment wheat spikes in full-size images, (4) train a second BlendMask model to predict diseased areas in individual

spikes, (5) write a program that combines dual BlendMask networks to simultaneously display the results of wheat spike detection and diseased area segmentation in full-size images, and (6) evaluate the disease grade of wheat FHB based on the ratio of the diseased area to the overall wheat spike. To our knowledge, this is the first study to assess the severity of wheat FHB based on automatic tandem dual BlendMask networks.

Table 1. Summary of the studies on the evaluation crop disease severities based on deep learning methods.

References	Model	Crop	Severity Levels	Accuracy (%)
Esgario et al. [45]	AlexNet, GoogleNet, VGGNet, ResNet, MobileNet	Coffee	Healthy, low, very low, high, very high	84.13
Pan et al. [46]	Faster R-CNN (VGG16)	Strawberry	Healthy, general, serious	88.3
Joshi et al. [47]	VirLeafNet	Vigna mungo	Healthy, mild, severe	91.5
Zhang et al. [48]	Improved YoLo V5	Wheat	Minor, light, medium, heavy, major	91.0
Ji et al. [49]	DeepLabV3+	Grape	Healthy, mild, medium, severe	97.75
Wu et al. [50]	MultiModel_VGR	Pepper	Healthy, general, serious	95.34
Liu et al. [51]	DeepLabV3+, PSPNet, UNet	Apple	Healthy, early, mild, moderate, severe	96.41

2. Materials and Methods

2.1. Data Collection and Annotation

Images of wheat spikes from the late flowering stage to maturity were collected in different weather conditions (sunny, cloudy, rainy, etc.) and different backgrounds (other spikes, sky, clouds, and soil). A total of 690 images were captured using an autofocus single-lens reflex (SLR) camera (Canon EOS Rebel T7i, resolution: 6000×4000). Each image consists of several wheat spikes. For spike identifications, 524 images (including 12,591 spikes) were selected for the training set and the other 166 images (including 4749 spikes) were selected for the validation set by random. For diseased area detection, a total of 2832 and 922 sub-images of diseased spikes were selected for the training set and validation set by random, respectively. An image annotation software (Labelme, <https://github.com/wkentaro/labelme> (accessed on 30 October 2021)) was used to label the ground truth for spikes in full-size images and FHB-diseased areas in sub-images. The first step in image annotation was to label wheat spikes in the raw images (Figure 1a,b). After segmenting the labeled individual spikes into sub-images, the diseased regions in the sub-images were then annotated in the second step (Figure 1c,d).

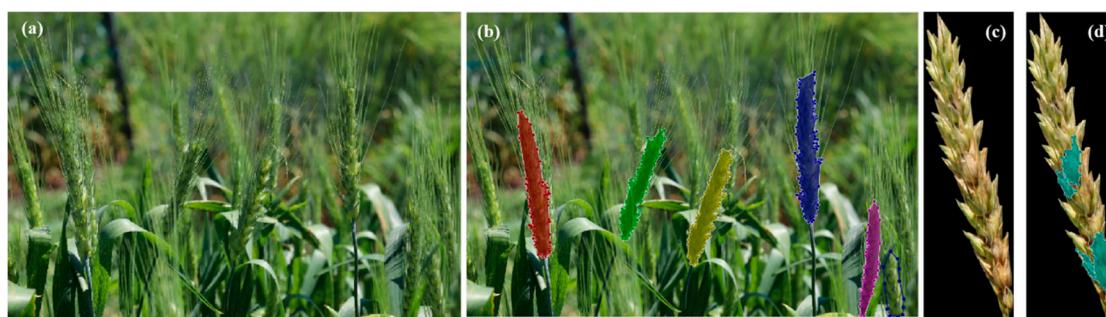


Figure 1. An example of manual annotation of images: (a) a typical full-size image of spikes from a single planted row of wheat in the field; (b) the same image with annotated wheat spikes; (c) a segmented sub-image of a single wheat spike; (d) the individual wheat spike with annotated diseased areas.

The FHB severity was calculated as the proportion of the diseased area to the total spike area. When the diseased area does not exceed 20%, the wheat can be considered to have the potential to resist FHB. If the infection ratio is above 50%, it means the wheat

is very susceptible to the disease. According to these principles, the FHB severity can be divided into four grades: healthy [0–5%), mild [5–10%), moderate [10–20%), and severe [20–50%]. The result of the Levene variance equivalence test and quality *t*-test of the means are both zeros, which means there are significant differences between any two classes.

2.2. BlendMask

BlendMask is a one-stage dense instance segmentation method model. The main contribution of BlendMask is to propose the Blender module, which can merge the feature both from top-down and bottom-up methods. Specifically, BlendMask is composed of a detector network and a mask branch. In the detector module, FCOS is used to generate the feature from top-down, including the bounding boxes and attention maps. FCOS is composed of a fully convolutional network, which is similar to the pixel-by-pixel prediction method of semantic segmentation to solve target detection. This detection not only becomes anchor free and proposal free, but also reduces the amount of design parameters significantly. The mask branch consists of three parts, including bottom module, top layer, and blender module. Bottom module is used to process the underlying features, generating a score map called Base. The top layer is concatenated on the box head of the detector, which predicts top-level attentions through a convolution layer on each of the detection towers. The blender module is the most important module of BlendMask, which can merge bounding boxes, attention maps and masks to generate the final prediction of instance segmentation.

The backbone of this study is a ResNet model with 50 layers (ResNet-50) mentioned by He et al. [25]. With the increase of the CNN depth, the problems of vanishing gradients [53] and the degradation [25] have been exposed. ResNet applies the shortcut connection of the residual module, which alleviates the problems of gradient disappearance and network degradation to a certain extent [54]. FPN is a top-down hierarchy with lateral connections for constructing high-level semantic features at various scales [55]. The raw images with annotated wheat spikes and the sub-images with annotated diseased areas were used as the inputs to train two BlendMask models for the detection of wheat spikes and diseased areas, respectively. After training the dual BlendMask models, the images in the validation set can be used for the segmentation of the wheat spikes and FHB-diseased areas. The recognition and segmentation results are displayed on the image below as output (Figure 2). The workflow of training dual BlendMask models is presented in Figure 3.

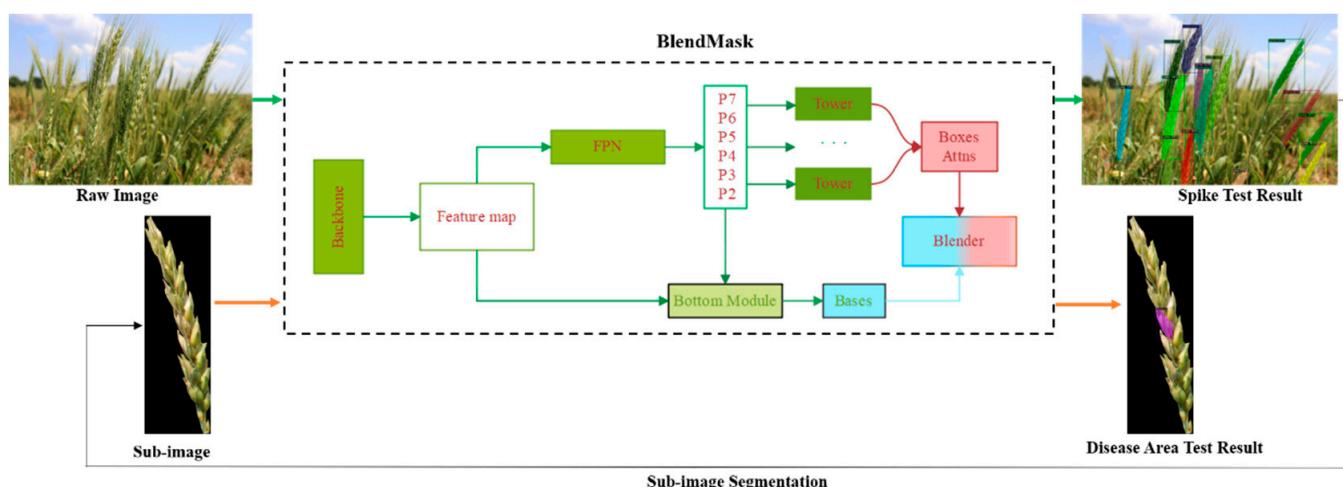


Figure 2. The architecture of the BlendMask approach for wheat Fusarium head blight (FHB) disease detection.

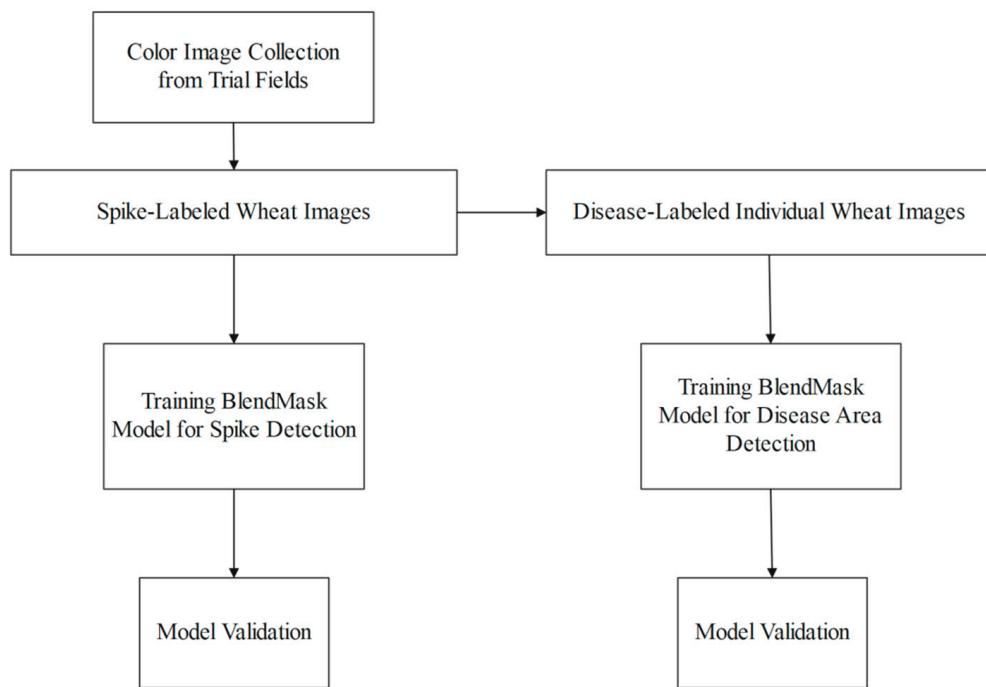


Figure 3. A flowchart of the proposed approach for training dual models.

2.3. Evaluation Metrics

In order to evaluate the segmentation results of BlendMask networks on wheat spikes and FHB-diseased areas, several evaluation metrics were computed. It includes precision, recall, *F*1-score, and average precision (*AP*). Precision is defined as the proportion of the number of real positive instances among the total number of instances predicted as belonging to the positive category. Recall is defined as the proportion of the number of real positive instances in the total number of instances actually belonging to the positive category [56]. *F*1-score denotes the harmonic weighted average of precision and recall values, which takes both false positives and false negatives into account [57]. Precision and recall are used to calculate *AP*, which represents the area under the precision–recall (*P-R*) curve. The intersection over union (*IoU*) is defined as the degree to which the manually annotated ground truth box overlaps the bbox (category score) generated by the model. The *IoU* threshold was used to determine whether a wheat spike was correctly segmented. If the predicted result exactly matches the ground truth, the *IoU* reaches 100%. The mean intersection over union (*MIoU*) is another standard metric for image segmentation, which can be calculated as the number of true positive (*TP*) over the sum of true positive (*TP*), false negative (*FN*), and false positive (*FP*) [58]. The equations of precision, recall, *F*1-score, *AP*, *IoU*, and *MIoU* are shown below:

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$AP = \frac{1}{10} \sum_{IoU=0.5}^{0.95} \int_0^1 P_{IoU}(r) dr$$

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

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

where TP (true positive) is the number of correctly segmented wheat spikes or FHB-diseased areas, FP (false positive) is the number of wrong segmented wheat spikes or FHB-diseased areas, and FN (false negative) is the number of ground truth that were not segmented. $P_{IoU}(r)$ is the precision–recall curve with different IoU values. E represents the manually labeled ground truth box. F represents the bbox generated based on the BlendMask model. The intersection over union (IoU) computes the overlapping area between the predicted wheat spike and the ground truth wheat spike divided by the area of union between them. In some studies, the IoU threshold value is set to 50%, which means that a wheat spike is considered to be correctly segmented if $IoU \geq 0.5$. It is considered to be mis-segmented if $IoU < 0.5$. $k+1$ is the total number of output classes, including an empty class (the background). P_{ii} denotes TP , while P_{ij} and P_{ji} represent FP and FN , respectively.

2.4. Equipment

A computer with a graphics processing unit (GPU) mode (NVIDIA RTX 3090 24 Gb, Shanghai, China) was used to complete the process for model training and validation. It is important to emphasize that these eight GPUs are leased from a GPU sharing platform, and the personal computer can control these leased GPUs on the GPU rental website. Table 2 shows the BlendMask modeling parameters. Table 3 depicts the time for model training and validation.

Table 2. Modeling hyperparameter settings.

Hyperparameter	Values
Backbone	Resnet-50
Batch size	16
Base learning rate	0.01
Attention size	14
Max iteration	1,700,000
Bottom resolution	56
Number of classes	2

Table 3. The total time for model training and validation.

Application	Training Time	Validation Time
Wheat spike identification	72 h 45 min 28 s	14 min 36 s
FHB disease detection	46 h 25 min 43 s	2 min 30 s

3. Results

3.1. Model Training

Dual BlendMask networks were trained based on the labeled images of wheat spikes and FHB-diseased areas. Figure 4a shows the changing trend of loss during the model training for wheat spike identification. It can be seen from the figure that the curves of the loss of the bbox and the mask decrease sharply in the initial iteration, but the downtrend slows down after 30,000 iterations. The loss curve then fluctuates until the curve trend stabilizes after 160,000 iterations. It can be seen that the loss value of the mask is always greater than that of the bbox. Finally, the loss value of the bbox and mask reached the lowest levels: 0.0012 and 0.024, respectively. Similarly, Figure 4b describes the variation of the loss in the model training for FHB disease. The overall change trend of the loss curve in the image is similar to the previous model for segmenting spikes. The loss values of the mask and the bbox reduced to 0.032 and 0.0016, eventually.

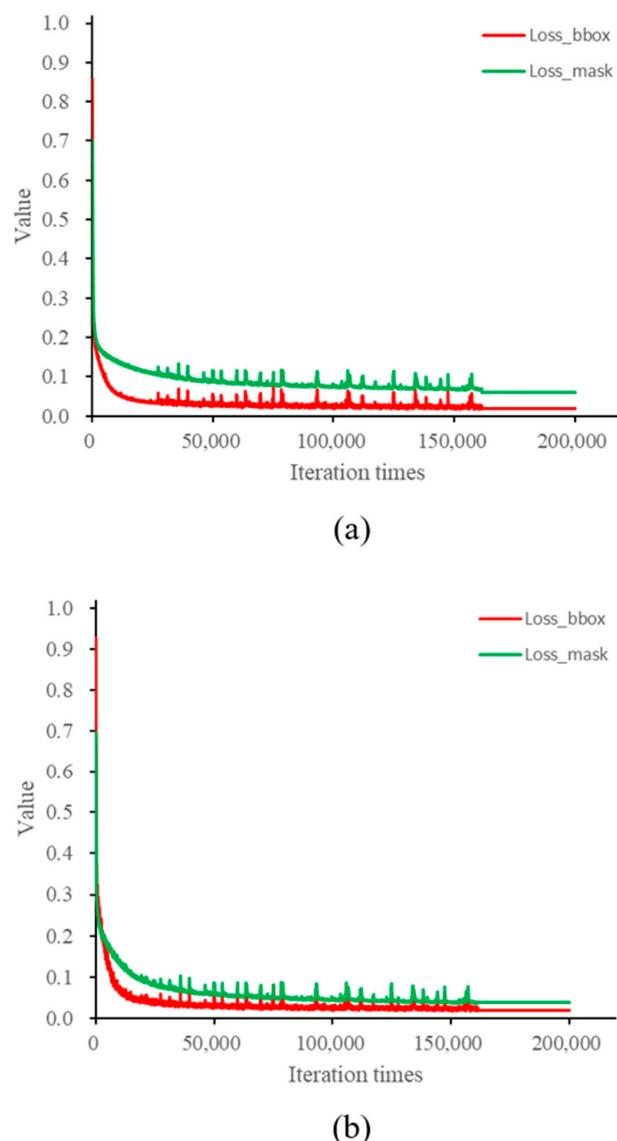


Figure 4. Training curves of loss values against iteration times on segmentation of (a) wheat spike and (b) diseased area.

3.2. Wheat Spike Identification

The segmentation results of the wheat spikes on the BlendMask network are shown in Table 3. As can be seen from the data in Table 4, it is apparent that the algorithm showed an outstanding performance for wheat spike segmentation, with the *AP* of the mask and the bbox at 57.16% and 56.69%, respectively. In the 166 test images, the overall values of *IoU*, *F1*-score, precision, and recall were 48.23%, 80.17%, 85.36%, and 75.58%, respectively. According to statistics, 4053 were correctly identified from the 4749 wheat spikes, yielding a recognition accuracy of 85.56%. All the above results indicate that the BlendMask network has a strong capability to identify wheat spikes under field conditions.

Table 4. Results of BlendMask for wheat spikes and FHB disease detection.

Type	Precision (%)	Recall (%)	F1-score (%)	IoU (%)	Ap of Mask (%)	MIoU (%)
Wheat spike	85.36	75.58	80.17	48.23	59.28	56.21
FHB disease	78.16	79.46	78.89	52.41	66.74	55.34

Figure 5 depicts the prediction of selected wheat spike images from the validation dataset. It was observed that most of the spikes were successfully detected. The network model successfully detected the high-density wheat spikes in the field (Figure 5a). Due to the shooting position of the camera and the natural growth state of the wheat spikes, the wheat spikes in many images are partially occluded. There are mainly three types of occlusions, including occlusion by the leaves of the wheat (Figure 5b), occlusion by the wheat awns or straw (Figure 5c), and occlusion by the other wheat spikes (Figure 5d). The training results of the model showed that BlendMask successfully segmented the occluded spikes of wheat, demonstrating the strong performance of the model. Figure 5e shows that wheat spikes cut at the image borders were successfully identified. It is important to identify local spikes, as this allows full utilization of wheat spikes in the dataset to enhance the robustness of the model.

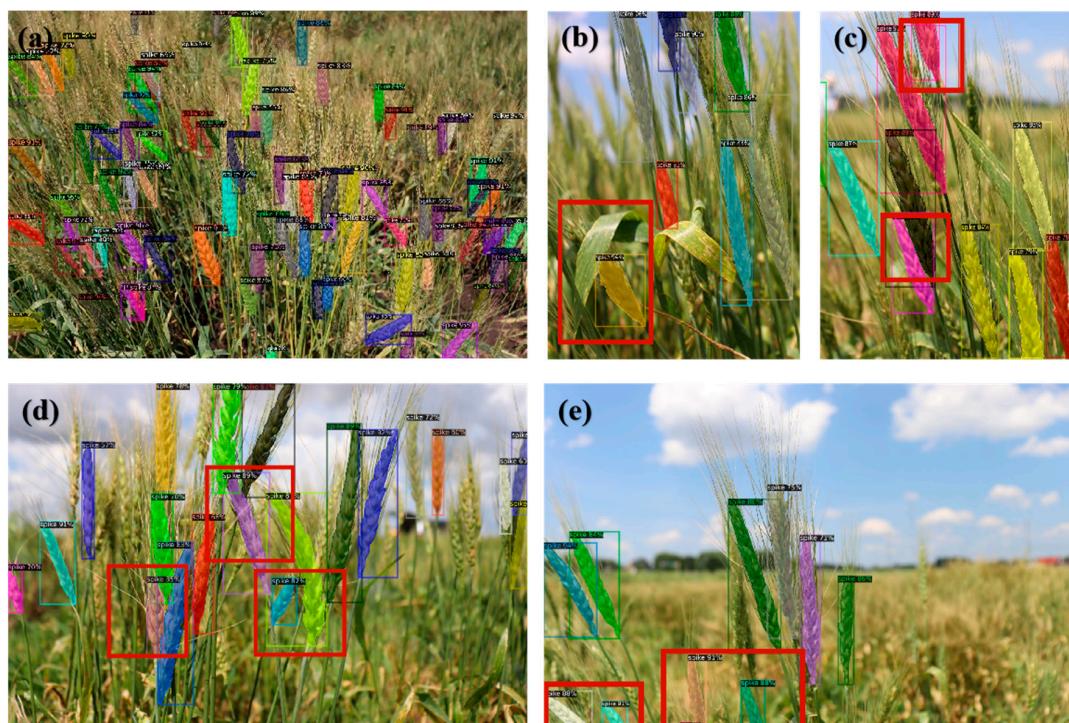


Figure 5. Successful segmentation results for wheat in five complex environments: (a) high-density wheat spikes in the field; (b) wheat spikes occluded by the leaves of other wheat; (c) wheat spikes occluded by other wheat awns or stems; (d) wheat spikes occluded by the other spikes; (e) wheat spikes cut at the image borders. (The red boxes are to emphasize different situations with complex backgrounds.)

3.3. FHB Disease Evaluation

The trained BlendMask model was used to detect the diseased areas on wheat spikes. Figure 6 shows the diseased area detection and segmentation results from the individual wheat spikes under different conditions including diseased areas with shadow, diseased areas under strong light, diseased areas under low light, and diseased areas under awn occlusion. The bboxes, masks, and category scores were shown in individual spikes. The $MIoU$ value reached 55.34%. The AP values of the mask and the bbox for FHB disease detection were 66.74% and 65.38%, respectively. As seen in Table 3, the overall values of precision, recall, $F1$ -score, and IoU were 78.16%, 79.46%, 78.8%, and 52.41%, respectively. Due to the complexity of crop growth in natural fields and the interference of light intensity or weeds, the individual wheat spikelet that is segmented from full-sized images generate a certain amount of noise, which is mainly manifested in the partial shadow area of the wheat spike, weak light, strong light, and occlusion by wheat awns or straw. All of these

bring challenges to the accurate recognition of diseased areas. Nevertheless, the BlendMask model was effective in identifying the lesion areas of these conditions (Figure 6). Eventually, a total of 916 diseased spikes were recognized from 922 samples with an accuracy of 99.32%. The results prove that BlendMask achieved the accurate identification of FHB-diseased areas. Figure 7 shows the flowchart of the combination of dual BlendMask models for evaluating FHB disease severity. Figure 8 shows the visualization of the recognition accuracy and disease severity of different wheat spikes from a full-sized image based on the automatic tandem dual BlendMask framework. It is worth noting that the dual BlendMask framework runs very fast. The recognition results can be output in about 2 s.



Figure 6. Illustration of diseased area detection and segmentation results from the individual wheat spikes in test set: (a) diseased areas with shadow; (b) diseased areas under strong light; (c) diseased areas under low light; (d) diseased areas under awn occlusion.

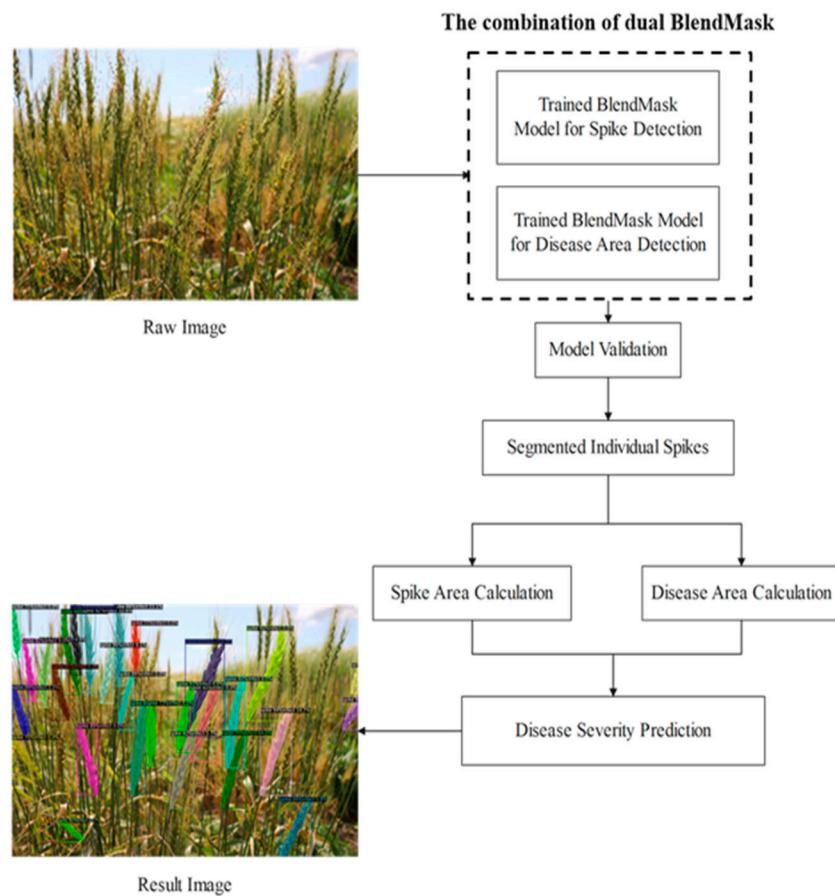


Figure 7. A flowchart of the combination of dual BlendMask models for evaluating FHB disease severity.

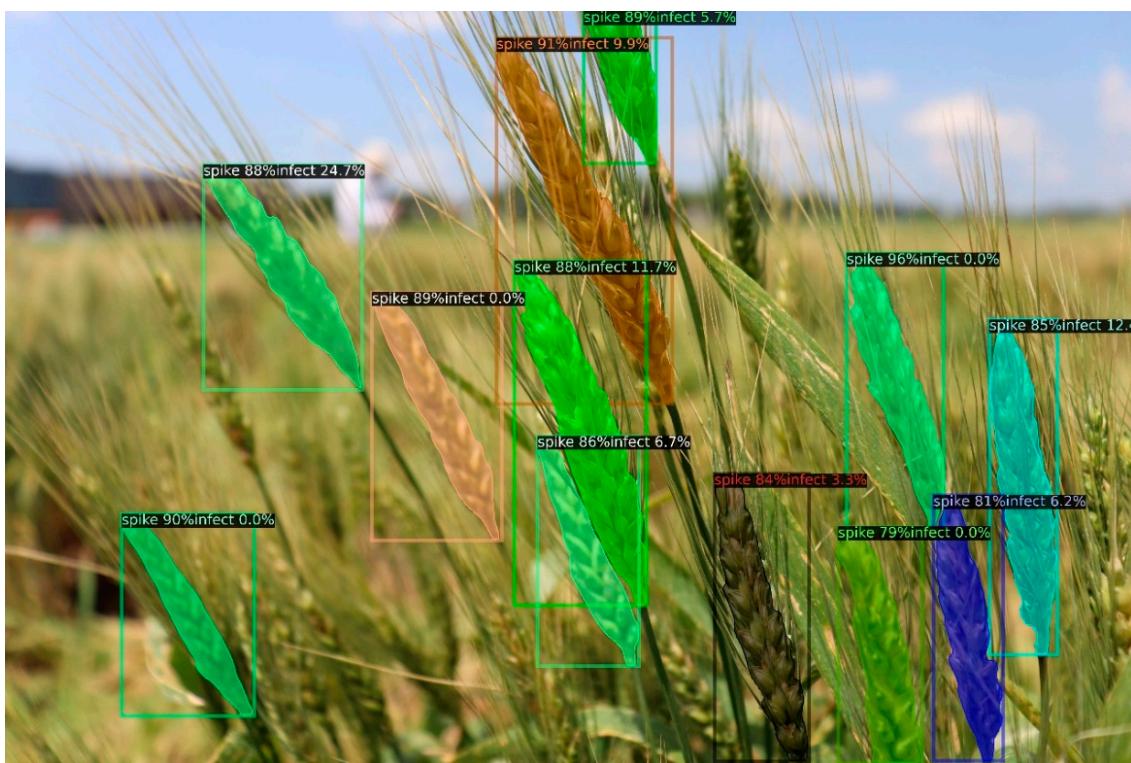


Figure 8. Visualization of the recognition accuracy and disease severity of different wheat spikes from a full-sized image based on automatic tandem dual BlendMask framework.

Although the trained model performed very well in diseased area detection, some mistakes occurred during validation. For the case of identification failure, the main reasons were model error and annotation error. As shown in Figure 9a, the part adjacent to the lesion area is identified as the lesion area (red rectangle). This type of error belongs to *FP*. The same error is also shown in Figure 9d in the blue box. The color of the wheat spikes near the diseased area is similar to the unique color of the diseased area due to natural light. Figure 9b shows that the model incorrectly identifies several adjacent lesion areas as one. Figure 9c shows that some lesions (blue rectangle) are not identified due to model errors, which belongs to *FN*. Figure 9d shows a failure due to annotation errors (red rectangle).

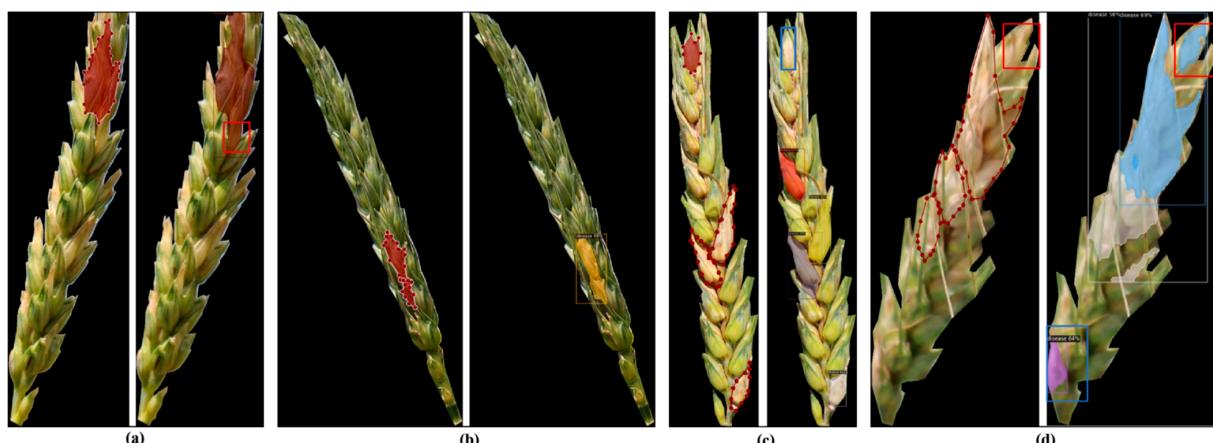


Figure 9. The detection of diseased areas in selected wheat sub-images in the validation dataset: (a–c) examples of detection failures caused by model errors; (d) an example of detection failure caused by model error or annotation error.

4. Classification of Wheat FHB Severity Grades

Figure 10a depicts the ground truth (the visual rating of spikes in the acquired images by an expert) of wheat spikes with different disease grades in the training sets. As seen in Figure 10a, 21.2%, 28.2%, 32.5%, and 18.1% of the samples in the training set are categorized as healthy, mild, moderate, and severe, respectively. Figure 10b shows the ground truth and prediction of wheat spikes with different disease grades in the validation sets. It was noticed that 20%, 29.6%, 30.3%, and 20.1% of the samples in the validation set were categorized as healthy, mild, moderate, and severe in the ground truth, which is generally similar to those from the training set. The distribution of the predicted results on the four grades is almost identical to that of the ground truth. The maximum error between the predicted quantity and the actual quantity occurs for the severe grade and the number is 10, while the minimum bias is only 1 when it comes to the mild grade.

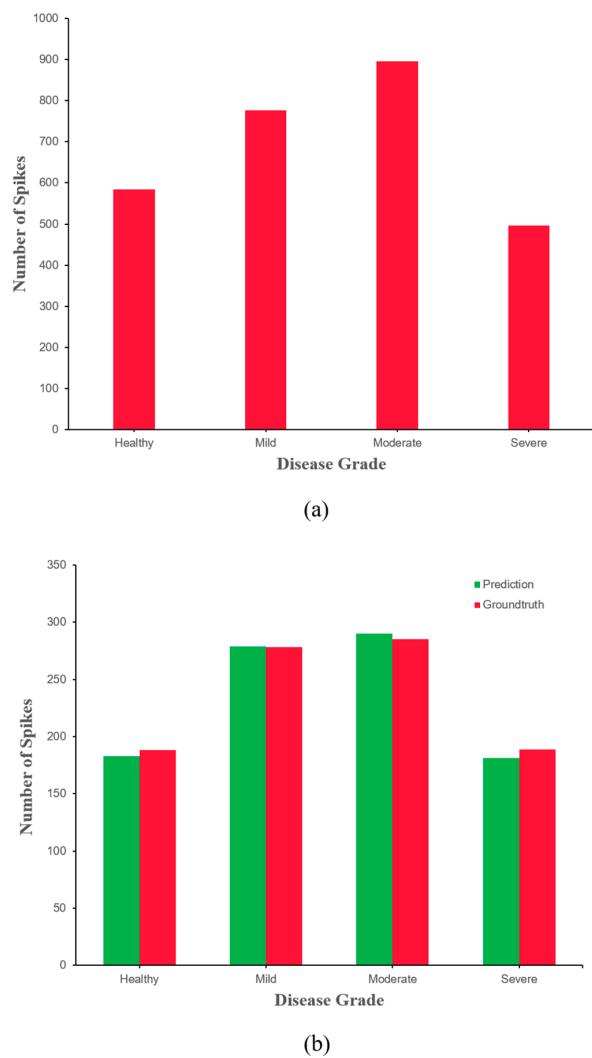


Figure 10. Frequency of wheat spikes for each disease grade: (a) the number of wheat spikes for different disease grades in the training set; (b) the number of predicted and ground truth spikes for different disease grades in the validation set.

Through the statistical analysis of the actual value and predicted value of the disease severity in the four disease grades shown in Table 5, it can be seen that the predicted value is always less than or equal to the actual value whether it is the average value or the maximum and minimum value of the disease severity. Therefore, the model may have a tendency to underestimate the severity of the disease in practical applications.

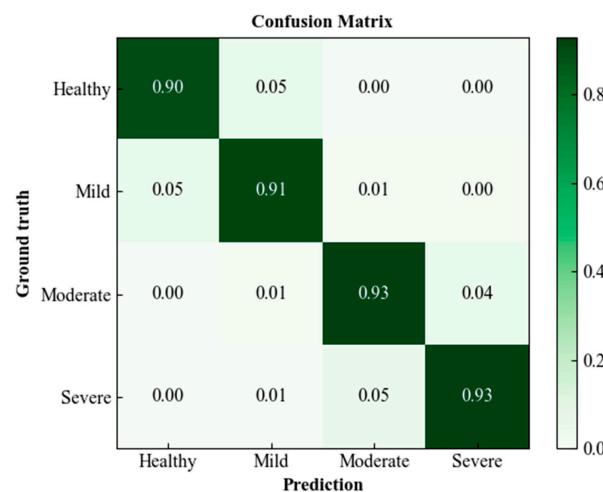
Table 5. Statistical results of wheat spike FHB disease severity for four grades.

Grade	Type	Severity (%)		
		Means ± SD	Max	Min
Healthy	Ground truth	2.8 ± 1.1	5.0	0.0
	Prediction	2.9 ± 0.8	4.9	0.0
Mild	Ground truth	7.4 ± 2.1	10.0	5.1
	Prediction	6.9 ± 1.4	9.8	5.0
Moderate	Ground truth	15.5 ± 2.3	20.0	10.3
	Prediction	14.4 ± 2.1	18.9	10.1
Severe	Ground truth	38.6 ± 4.6	49.8	21.2
	Prediction	36.4 ± 4.1	48.7	20.8

To further verify the accuracy of the classification results obtained from the model, the confusion matrix was applied to analyze the similarities and differences between the predicted results and the ground truth. Precision and sensitivity were first obtained. Based on the precision and sensitivity, the *F1*-score for each of the four grades was calculated. As shown in Table 6, it is obvious that all the values are around 90%. The lowest *F1*-score score was 91.1% while the highest score was 93.2%. The average *F1*-score was 92.22%. Figure 11 depicts a confusion matrix for wheat FHB grades. The average accuracy for FHB severity classification was 91.8%. Accuracies of 90%, 91%, 93%, and 93% were obtained for the four grades, respectively. As shown in Figure 11, mild grade samples are most easily misclassified to the healthy grade. This may be because the healthy and mild grades have similar area ratios. Since the diseased area for these two grades is relatively small, the probability of misidentification is likely to increase. Additionally, moderate grade samples are more likely misclassified to the severe grade. This may be due to the smaller difference in the ratio of diseased areas to spike areas for the moderate and severe grades.

Table 6. Results of precision, sensitivity, and *F1*-score for four grades.

Grade	Precision (%)	Sensitivity (%)	<i>F1</i> -score (%)
Healthy	93.6	88.9	91.1
Mild	93.7	90.9	91.7
Moderate	92.6	92.8	92.9
Severe	94.8	93.1	93.2

**Figure 11.** Confusion matrix for four wheat FHB resistance grades (For each grade, the total number of the ground truth is not equal to that of the prediction since the model fails to detect every spike of the ground truth).

5. Discussion

The research proposed a new approach using tandem dual BlendMask networks for automatic severity estimation of wheat FHB in the field. Three main parts were involved in this study, including wheat spike segmentation, FHB disease segmentation, and disease severity classification. RGB images were used to train the dual BlendMask framework to evaluate the severity of wheat FHB disease. Although positive detection results were obtained, there were still some errors in the prediction of the disease severity. These errors may come from the algorithm model or the image annotation. Data annotation is a bottleneck for segmentation tasks [59]. The annotation process is laborious and time-consuming. The annotation results are greatly affected by human factors. In the future, instance segmentation based on semi- and weakly-supervised methods can be considered. In order to maximize the utilization of the dataset, data augmentation is also an essential step. Fang et al., used data augmentation to solve the issue of insufficient training datasets in instance segmentation [60]. The application of random transformations, such as flipping, cropping, and changing saturation to generate new images, can generate new images that effectively augment the training set.

It is a challenging task to realize the recognition of wheat spikes in the field. In this study, FCOS in the BlendMask deep learning framework was used for wheat FHB detection. The model was capable of detecting wheat spikes within a complex environment. Comparing the neural network method with Laws texture energy [61], the BlendMask model achieved a higher accuracy of 85.56% for identification of high-density wheat spikes. The main reason for the success of the proposed model is the FCOS used in the detector module, which can generate the feature from top-down. Although the SpikeSegNet model achieved an accuracy of 99.91% in the study of Misra et al. [62], the wheat spikes were low-density in collected images.

Two trained BlendMask models are connected in series to directly display the recognition results of wheat spikes and diseased areas in the original images. The BlendMask performed very well, yielding a detection rate as high as 99.32% for FHB detection compared to 98.81% for the study of Su et al. [42]. The main reasons for the success of our model are as follows:(1) compared to Mask R-CNN, the Blender module of BlendMask provides higher-quality masks and (2) the wheat in the dataset has been marked with high precision, which helps to improve the performance and robustness of the BlendMask model. The proportion of diseased areas of the whole wheat spike displayed directly on the original input image, thus high-throughput and real-time analysis can be constructed in the field. Many deformation models of the transformer, such as vision transformer (ViT) [63] and swin transformer [64], perform well in image classification, target recognition, and image segmentation. In the future, more advanced semantic segmentation algorithms based on transformers should be considered.

It was noticed that the automatic tandem dual BlendMask networks successfully segmented individual wheat spikes and FHB-diseased areas simultaneously from images of multiple spikes with complex backgrounds. The proposed method showed great potential for non-destructive and high-throughput evaluation of the severity of wheat FHB disease. Su et al. [42] used the same dataset as this study and the Mask R-CNN model to segment wheat individuals and the disease spots on wheat individuals, and the accuracy of segmentation detection reached 77.76% and 98.81%, respectively. In this study, the segmentation detection accuracy of these two parts has been improved to a certain extent, reaching 85.56% and 99.32%. Mask R-CNN takes 250,000 iterations to achieve the accuracy while BlendMask reaches the accuracy mentioned above after 170,000 iterations. Therefore, the BlendMask model is more concise and efficient. By linking the two models, the average time to identify the severity of a wheat plant was 0.09 s.

The objective of this study is to apply the constructed model to a car taking photos in the field, so the model is biased towards proximal perceptual detection. Nowadays, unmanned aerial vehicles (UAVs) are widely used for efficient detection of crop diseases because they can sense a larger range and consume less manpower, but such detection

generally relies on the expensive remote sensing equipment on the UAVs. In future studies, we will try to improve the model and make it adapt to the high density of wheat images or videos taken by UAVs. When the UAVs fly at a low altitude over the same kind of wheat in the field, an ordinary camera with a relatively low cost can be used to take photos or video and evaluate the severity of the wheat FHB disease. This is helpful to screen out the wheat with better disease resistance.

6. Conclusions

This study proposed an integrated method for the evaluation of wheat for FHB severity based on the automatic tandem dual BlendMask networks for simultaneous segmentations of wheat spikes and diseased areas under complex field conditions. The BlendMask model demonstrated outstanding performance in the detection of the wheat spikes occluded by awns or cut at the image borders. The recognition accuracies of the model for wheat spikes and diseased areas were 85.56% and 99.32%, respectively. The inference time was within 2 s on average, which would be helpful for real-time monitoring of high-throughput wheat spikes in the field. By calculating the ratio of the lesion area to the overall area, the disease degree of wheat FHB was divided into four grades (healthy, mild, moderate, and severe) and the accuracy of disease level prediction reached 91.8%. This study demonstrates the feasibility of dual BlendMask networks for severity evaluation of wheat FHB in the field. This study will be of great help to the breeding of FHB resistant lines in breeding nurseries.

Author Contributions: Conceptualization, W.-H.S.; methodology, Y.G., H.W., M.L., and W.-H.S.; software, Y.G. and H.W.; validation, Y.G., H.W., and M.L.; formal analysis, Y.G., H.W., and M.L.; investigation, Y.G., H.W., and M.L.; resources, W.-H.S.; writing—original draft preparation, Y.G. and H.W.; writing—review and editing, M.L. and W.-H.S.; 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: Data available on request due to privacy.

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

References

- Shiferaw, B.; Smale, M.; Braun, H.-J.; Duveiller, E.; Reynolds, M.; Muricho, G. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Secur.* **2013**, *5*, 291–317. [[CrossRef](#)]
- Figueroa, M.; Hammond-Kosack, K.E.; Solomon, P.S. A review of wheat diseases—A field perspective. *Mol. Plant Pathol.* **2018**, *19*, 1523–1536. [[CrossRef](#)]
- O'Donnell, K.; Ward, T.J.; Geiser, D.M.; Kistler, H.; Aoki, T. Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the Fusarium graminearum clade. *Fungal Genet. Biol.* **2004**, *41*, 600–623. [[CrossRef](#)] [[PubMed](#)]
- Femenias, A.; Gatius, 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* **2019**, *108*, 106819. [[CrossRef](#)]
- Stack, R.W.; McMullen, M.P. A visual scale to estimate severity of Fusarium head blight in wheat. 1998. Available online: https://library.ndsu.edu/ir/bitstream/handle/10365/9187/PP1095_1998.pdf?sequence=1 (accessed on 13 September 2022).
- Barbedo, J.; Tibola, C.S.; Lima, M.I.P. Deoxynivalenol screening in wheat kernels using hyperspectral imaging. *Biosyst. Eng.* **2017**, *155*, 24–32. [[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.* **2020**, *343*, 128507. [[CrossRef](#)]
- Su, W.-H.; Xue, H. Imaging Spectroscopy and Machine Learning for Intelligent Determination of Potato and Sweet Potato Quality. *Foods* **2021**, *10*, 2146. [[CrossRef](#)]
- Su, W.-H.; Sun, D.-W.; He, J.-G.; Zhang, L.-B. Variation analysis in spectral indices of volatile chlorpyrifos and non-volatile imidacloprid in jujube (*Ziziphus jujuba* Mill.) using near-infrared hyperspectral imaging (NIR-HSI) and gas chromatograph-mass spectrometry (GC-MS). *Comput. Electron. Agric.* **2017**, *139*, 41–55. [[CrossRef](#)]
- Su, W.-H.; Sun, D.-W. Rapid determination of starch content of potato and sweet potato by using NIR hyperspectral imaging. *Hortscience* **2019**, *54*, S38.

11. Su, W.-H.; Sun, D.-W. Advanced Analysis of Roots and Tubers by Hyperspectral Techniques. In *Advances in Food and Nutrition Research*; Elsevier BV: Amsterdam, The Netherlands, 2019; Volume 87, pp. 255–303.
12. Su, W.-H.; Sun, D.-W. Potential of hyperspectral imaging for visual authentication of sliced organic potatoes from potato and sweet potato tubers and rapid grading of the tubers according to moisture proportion. *Comput. Electron. Agric.* **2016**, *125*, 113–124. [[CrossRef](#)]
13. Su, W.-H.; He, H.-J.; Sun, D.-W. Non-Destructive and rapid evaluation of staple foods quality by using spectroscopic techniques: A review. *Crit. Rev. Food Sci. Nutr.* **2016**, *57*, 1039–1051. [[CrossRef](#)]
14. Su, W.-H.; Sun, D.-W. Fourier Transform Infrared and Raman and Hyperspectral Imaging Techniques for Quality Determinations of Powdery Foods: A Review. *Compr. Rev. Food Sci. Food Saf.* **2017**, *17*, 104–122. [[CrossRef](#)] [[PubMed](#)]
15. Su, W.-H. Advanced Machine Learning in Point Spectroscopy, RGB-and hyperspectral-imaging for automatic discriminations of crops and weeds: A review. *Smart Cities* **2020**, *3*, 39. [[CrossRef](#)]
16. Nagasubramanian, K.; Jones, S.; Singh, A.K.; Sarkar, S.; Singh, A.; Ganapathysubramanian, B. Plant disease identification using explainable 3D deep learning on hyperspectral images. *Plant Methods* **2019**, *15*, 98. [[CrossRef](#)] [[PubMed](#)]
17. Raghavendra, A.; Guru, D.; Rao, M.K. Mango internal defect detection based on optimal wavelength selection method using NIR spectroscopy. *Artif. Intell. Agric.* **2021**, *5*, 43–51. [[CrossRef](#)]
18. Peiris, K.H.S.; Pumphrey, M.O.; Dong, Y.; Maghirang, E.B.; Berzonsky, W.; Dowell, F.E. 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](#)]
19. Femenias, A.; Gatius, F.; Ramos, A.J.; Teixido-Orries, I.; Marín, S. Hyperspectral imaging for the classification of individual cereal kernels according to fungal and mycotoxins contamination: A review. *Food Res. Int.* **2022**, *155*, 111102. [[CrossRef](#)] [[PubMed](#)]
20. Lu, J.; Hu, J.; Zhao, G.; Mei, F.; Zhang, C. An in-field automatic wheat disease diagnosis system. *Comput. Electron. Agric.* **2017**, *142*, 369–379. [[CrossRef](#)]
21. Weng, S.; Tang, P.; Yuan, H.; Guo, B.; Yu, S.; Huang, L.; Xu, C. Hyperspectral imaging for accurate determination of rice variety using a deep learning network with multi-feature fusion. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *234*, 118237. [[CrossRef](#)]
22. Ahmad, A.; Saraswat, D.; El Gamal, A. A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. *Smart Agric. Technol.* **2023**, *3*, 100083. [[CrossRef](#)]
23. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
24. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *NIPS* **2012**, *60*, 84–90. [[CrossRef](#)]
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
26. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint* **2014**, arXiv:1409.1556.
27. Dong, M.; Mu, S.; Shi, A.; Mu, W.; Sun, W. Novel method for identifying wheat leaf disease images based on differential amplification convolutional neural network. *Int. J. Agric. Biol. Eng.* **2020**, *13*, 205–210. [[CrossRef](#)]
28. Wagle, S.A.; R, H. Comparison of Plant Leaf Classification Using Modified AlexNet and Support Vector Machine. *Trait. Signal* **2021**, *38*, 79–87. [[CrossRef](#)]
29. Hassan, S.M.; Jasinski, M.; Leonowicz, Z.; Jasinska, E.; Maji, A.K. Plant Disease Identification Using Shallow Convolutional Neural Network. *Agronomy* **2021**, *11*, 2388. [[CrossRef](#)]
30. Hassan, S.M.; Maji, A.K. Plant Disease Identification Using a Novel Convolutional Neural Network. *IEEE Access* **2022**, *10*, 5390–5401. [[CrossRef](#)]
31. Fenu, G.; Mallochi, F.M. Using Multioutput Learning to Diagnose Plant Disease and Stress Severity. *Complexity* **2021**, *2021*, 6663442. [[CrossRef](#)]
32. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [[CrossRef](#)]
33. Girshick, R. Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
34. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [[CrossRef](#)]
35. 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](#)]
36. 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](#)]
37. Kukreja, V.; Kumar, D. Automatic Classification of Wheat Rust Diseases Using Deep Convolutional Neural Networks. In Proceedings of the 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 3–4 September 2021; pp. 1–6. [[CrossRef](#)]

38. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
39. Kumar, D.; Kukreja, V. Quantifying the Severity of Loose Smut in Wheat Using MRCNN. In Proceedings of the 2022 International Conference on Decision Aid Sciences and Applications (DASA), Chiangrai, Thailand, 23–25 March 2022; pp. 630–634.
40. Kumar, D.; Kukreja, V. An Instance Segmentation Approach for Wheat Yellow Rust Disease Recognition. In Proceedings of the 2021 International Conference on Decision Aid Sciences and Application (DASA), Online, 7–8 December 2021; pp. 926–931.
41. Yang, K.; Zhong, W.; Li, F. Leaf Segmentation and Classification with a Complicated Background Using Deep Learning. *Agronomy* **2020**, *10*, 1721. [[CrossRef](#)]
42. Su, W.-H.; Zhang, J.; Yang, C.; Page, R.; Szinyei, T.; Hirsch, C.; Steffenson, B. 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](#)]
43. Chen, H.; Sun, K.; Tian, Z.; Shen, C.; Huang, Y.; Yan, Y. Blendmask: Top-down meets bottom-up for instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8573–8581.
44. Xi, X.; Xia, K.; Yang, Y.; Du, X.; Feng, H. Evaluation of dimensionality reduction methods for individual tree crown delineation using instance segmentation network and UAV multispectral imagery in urban forest. *Comput. Electron. Agric.* **2021**, *191*, 106506. [[CrossRef](#)]
45. 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](#)]
46. Pan, J.; Xia, L.; Wu, Q.; Guo, Y.; Chen, Y.; Tian, X. Automatic strawberry leaf scorch severity estimation via faster R-CNN and few-shot learning. *Ecol. Informatics* **2022**, *70*, 101706. [[CrossRef](#)]
47. Joshi, R.C.; Kaushik, M.; Dutta, M.K.; Srivastava, A.; Choudhary, N. VirLeafNet: Automatic analysis and viral disease diagnosis using deep-learning in Vigna mungo plant. *Ecol. Informatics* **2020**, *61*, 101197. [[CrossRef](#)]
48. Zhang, D.-Y.; Luo, H.-S.; Wang, D.-Y.; Zhou, X.-G.; Li, W.-F.; Gu, C.-Y.; Zhang, G.; He, F.-M. Assessment of the levels of damage caused by Fusarium head blight in wheat using an improved YoloV5 method. *Comput. Electron. Agric.* **2022**, *198*, 107086. [[CrossRef](#)]
49. Ji, M.; Wu, Z. Automatic detection and severity analysis of grape black measles disease based on deep learning and fuzzy logic. *Comput. Electron. Agric.* **2022**, *193*, 106718. [[CrossRef](#)]
50. Wu, Q.; Ji, M.; Deng, Z. Automatic Detection and Severity Assessment of Pepper Bacterial Spot Disease via MultiModels Based on Convolutional Neural Networks. *Int. J. Agric. Environ. Inf. Syst.* **2020**, *11*, 29–43. [[CrossRef](#)]
51. 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](#)]
52. Thakur, P.S.; Khanna, P.; Sheorey, T.; Ojha, A. Trends in vision-based machine learning techniques for plant disease identification: A systematic review. *Expert Syst. Appl.* **2022**, *208*, 118117. [[CrossRef](#)]
53. Hochreiter, S. The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **1998**, *6*, 107–116. [[CrossRef](#)]
54. Wu, Z.; Shen, C.; Van Den Hengel, A. Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recogn.* **2019**, *90*, 119–133. [[CrossRef](#)]
55. 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.
56. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint* **2020**, arXiv:2010.16061.
57. Zhang, X.; Graepel, T.; Herbrich, R. Bayesian online learning for multi-label and multi-variate performance measures. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010; pp. 956–963.
58. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint* **2017**, arXiv:1704.06857.
59. Li, Q.; Arnab, A.; Torr, P.H. Weakly-and semi-supervised panoptic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Online, 8–14 September 2018; pp. 102–118.
60. Fang, H.-S.; Sun, J.; Wang, R.; Gou, M.; Li, Y.-L.; Lu, C. Instaboost: Boosting instance segmentation via probability map guided copy-pasting. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 682–691. [[CrossRef](#)]
61. Qiongyan, L.; Cai, J.; Berger, B.; Okamoto, M.; Miklavcic, S.J. Detecting spikes of wheat plants using neural networks with Laws texture energy. *Plant Methods* **2017**, *13*, 83. [[CrossRef](#)]
62. Misra, T.; Arora, A.; Marwaha, S.; Chinnusamy, V.; Rao, A.R.; Jain, R.; Sahoo, R.N.; Ray, M.; Kumar, S.; Raju, D.; et al. SpikeSegNet—a deep learning approach utilizing encoder-decoder network with hourglass for spike segmentation and counting in wheat plant from visual imaging. *Plant Methods* **2020**, *16*, 40. [[CrossRef](#)]

63. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint 2020*, arXiv:2010.11929.
64. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Online, 11–17 October 2021; pp. 9992–10002. [[CrossRef](#)]