Article

Automatic Evaluation of Wheat Resistance to Fusarium Head Blight Using Dual Mask-RCNN Deep Learning Frameworks in Computer Vision

Wen-Hao Su [1,†] , **Jiajing Zhang** [2,†] , **Ce Yang** [3,*] , **Rae Page** [4] , **Tamas Szinyei** [4] , **Cory D. Hirsch** [4] and **Brian J. Steffenson** [4]

1   College of Engineering, China Agricultural University, Beijing 100083, China; wenhao.su@cau.edu.cn
2   College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China; Jiajing_Zhang@cau.edu.cn
3   Department of Bioproducts and Biosystems Engineering, University of Minnesota, Saint Paul, MN 55108, USA
4   Department of Plant Pathology, University of Minnesota, Saint Paul, MN 55108, USA; pagex277@umn.edu (R.P.); tszinyei@umn.edu (T.S.); cdhirsch@umn.edu (C.D.H.); bsteffen@umn.edu (B.J.S.)
*   Correspondence: ceyang@umn.edu
†   Wen-Hao Su and Jiajing Zhang are co-first authors.

Abstract: In many regions of the world, wheat is vulnerable to severe yield and quality losses from the fungus disease of Fusarium head blight (FHB). The development of resistant cultivars is one means of ameliorating the devastating effects of this disease, but the breeding process requires the evaluation of hundreds of lines each year for reaction to the disease. These field evaluations are laborious, expensive, time-consuming, and are prone to rater error. A phenotyping cart that can quickly capture images of the spikes of wheat lines and their level of FHB infection would greatly benefit wheat breeding programs. In this study, mask region convolutional neural network (Mask-RCNN) allowed for reliable identification of the symptom location and the disease severity of wheat spikes. Within a wheat line planted in the field, color images of individual wheat spikes and their corresponding diseased areas were labeled and segmented into sub-images. Images with annotated spikes and sub-images of individual spikes with labeled diseased areas were used as ground truth data to train Mask-RCNN models for automatic image segmentation of wheat spikes and FHB diseased areas, respectively. The feature pyramid network (FPN) based on ResNet-101 network was used as the backbone of Mask-RCNN for constructing the feature pyramid and extracting features. After generating mask images of wheat spikes from full-size images, Mask-RCNN was performed to predict diseased areas on each individual spike. This protocol enabled the rapid recognition of wheat spikes and diseased areas with the detection rates of 77.76% and 98.81%, respectively. The prediction accuracy of 77.19% was achieved by calculating the ratio of the wheat FHB severity value of prediction over ground truth. This study demonstrates the feasibility of rapidly determining levels of FHB in wheat spikes, which will greatly facilitate the breeding of resistant cultivars.

Keywords: deep learning; wheat spike; Fusarium head blight; target recognition; computer vision

1. Introduction

Wheat (*Triticum aestivum* L.) is a globally significant crop for human and animal consumption. In the United States, wheat plays an important role in promoting export markets and trade balances in addition to meeting domestic food and feed production needs [1]. Many diseases affect wheat production and threaten global food security. One of the most devastating fungal diseases attacking wheat is Fusarium head blight (FHB), caused primarily by *Fusarium graminearum*. FHB attacks the spikes (ears) of wheat, causing marked reductions in both the yield and quality of the crop. Moreover, the fungus can produce an array of mycotoxins (e.g., deoxynivalenol or DON) within the grain rendering

it unsuitable for human or animal consumption [2,3]. Thus, FHB can severely impact public health [4,5] in addition to reducing the yield and quality of the crop. Spikelets infected with FHB show premature bleaching. In a susceptible wheat line, infection of just a single spikelet can eventually spread across the entire spike. The breeding of FHB resistant cultivars is one of the most important means for ameliorating the impact of this disease [6,7]. To develop resistant cultivars, hundreds of breeding lines must be evaluated for FHB severity each year, often at multiple field sites. Protocols for assessing FHB resistance have conventionally relied upon the trained eye. The severity of FHB in wheat was accurately scored by counting infected spikelets and expressing that as a percentage of total spikelets [8]. Nevertheless, this approach is laborious, costly, time-consuming, and subject to human error [9]. The spectral sensor can intelligently perceive the spectral characteristics of an object at a certain point but cannot obtain a macroscopic image of the identified object [10–12]. Thus, there is an urgent need to develop a more effective and high-throughput approach for assessing this disease in the field.

Computer vision-based phenotyping is a rapid, high-throughput, and non-destructive technique for capturing many types of traits [13–17]. Imaging techniques such as hyper-spectral imaging (HSI) [18], and red–green–blue (RGB) imaging [19], have been widely used to study the complex traits associated with plant growth, biomass, yield, and responses to biotic stresses such as disease and abiotic stresses such as cold, drought and salinity [20–22]. With respect to FHB, Whetton et al. [23] processed HSI images of infected and healthy crop canopies including wheat and barley under controlled environmental conditions. The correlations between FHB severity and HSI data were investigated for quantification of wheat resistance to FHB [24]. Based on support vector machine (SVM) and fisher linear discriminant analysis, healthy and infected spikes were classified with acceptable accuracies (79–89%) [25,26]. The performance of the diagnosis model was then improved using a particle swarm optimization support vector machine (PSO-SVM) [27]. Relevance vector machine (RVM) performed better than the logistic model for prediction of FHB severity under natural environmental conditions [28]. However, more advanced algorithms such as convolutional neural network (CNN) have not been adopted in their study. In addition, one big challenge faced by HSI technology is the difficulty for rapidly processing a large amount of high-dimensional data obtained in a continuous spectral range and to effectively execute automatic operations [29–38].

A color imaging camera with built-in RGB filters can rapidly capture RGB images through three spectral channels (red, green, and blue), and has great potential for real-time detection of wheat FHB [39]. Although K-means clustering and random forest classifier have been used for segmentations of disease areas in wheat spikes [40], the advantage of this digital imaging technique is greater when a large number of datasets are available and a more powerful machine learning algorithm is utilized. Deep learning with great merit of automatic feature learning is a core part of the larger family of machine learning based on multiple layers of artificial neural networks, which allows greater learning capabilities and higher computational performance. As a widely recognized deep neural network, CNN has become a standard algorithm in object identification [41–43]. Hasan et al. [44] tested a faster region-based convolutional neural network (Faster R-CNN) model to detect wheat spikes and output a bounding box (bbox) for each spike. Deep convolutional neural network (DCNN) models were successfully used to localize wheat spikes under greenhouse conditions and predict the FHB disease severity with high accuracy [30,45]. Zhang et al. [46] developed a pulse-coupled neural network (PCNN) with K-means clustering of the improved artificial bee colony (IABC) for the segmentation of wheat spikes infected FHB. Since only one spike in each image was considered, it would be difficult to efficiently detect disease in a high throughput way. Then, Qiu et al. [47] developed a protocol that can segment multiple spikes in an image and then used a region-growing algorithm to detect the diseased area on each wheat spike. However, conventional image processing operations including region-growing, gray-scale co-occurrence matrix, and connectivity analysis are not suitable for real-time disease detections [48]. In addition, wheat spikes located at the

image borders could not be segmented, and the accuracy of the target spike and disease area identifications was significantly reduced due to the presence of awns on the spike. In these studies, the shape or contour information was only roughly extracted; thus, it was difficult to accurately identify the targets. Such factors reduced the accurate assessment of FHB disease levels. Thus, a new strategy should be designed to reliably evaluate wheat resistance to Fusarium head blight under field conditions. Mask region convolutional neural network (Mask-RCNN) is a machine vision based deep structural learning algorithm to directly solve the problem of instance segmentation [49]. This algorithm has been successfully employed to identify fruits and plants. For instance, Ganesh et al. [50] and Jia et al. [51] developed harvesting detectors based on the Mask-RCNN for robotic detections of apple and orange in orchards with precision of 0.895 to 0.975. Yang et al. [52] revealed the potential of Mask-RCNN for the identification of leaves in plant images for rapid phenotype analysis, yielding the average accuracy value up to 91.5%. Tian et al. [53] illustrated that Mask-RCNN performed best compared to other models including CNN and SVM in automatic segmentation of apple flowers of different growth stages.

The novelty of this study is in the development of an integrated approach for FHB severity assessment based on Mask-RCNN for high-throughput wheat spike recognition and precision FHB infection segmentation under complex field conditions. The Mask-RCNN combined object detection and instance segmentation provide an efficient framework to extract object bboxes, masks, and key points [49]. The main objective of this research is to determine the performance of Mask-RCNN based dual deep learning frameworks for real-time assessments of wheat for FHB severity in field trials. The specific objectives were to: (1) develop an imaging protocol for capturing quality images of wheat spikes in the field; (2) annotate spikes and diseased spikelets in the images; (3) build a Mask-RCNN model that works well in detecting and segmenting wheat spikes under complex backgrounds; (4) develop a second Mask-RCNN model that is valuable for prediction of diseased areas in individual spikes of segmented sub-images; (5) evaluate the disease grade of wheat FHB based on the ratio of the disease area to the entire wheat spike. We believe this is the first study using dual Mask-RCNN frameworks for automatic evaluation of wheat FHB disease severity.

2. Materials and Methods

2.1. Data Collection

FHB evaluation trials were established at the Minnesota Agricultural Experiment Station on the Saint Paul campus of the University of Minnesota. Wheat samples of 55 genetic lines were sown on May 2019 and FHB inoculations were made using the conidial spray inoculation [54]. To achieve sufficient infection levels on wheat lines throughout the field nursery, three inoculations were made: the first performed one week before the heading time of the earliest maturing accessions, the second one week later, and the third coinciding with accessions having late heading dates. Daily mist irrigation (0.61 cm per day) was provided at regular intervals (10 min at the top of every hour, 0.05 cm per hour) from 6 p.m. through 5 a.m. (12 times) to promote infection and disease development. Irrigation began after the first inoculation and continued until the latest maturing accessions reached the late dough stage of development. The growth stage of wheat at the time of image acquisition is a key factor for effective FHB detection. When the wheat was in the start of flowering and the late maturing stage, distinction of diseased spikes was not possible based on the naked eye [55]. The best time to assess the disease is when the spike symptoms become visible but not yet senescence.

An autofocus single-lens reflex (SLR) camera (Canon EOS Rebel T7i, Canon Inc., Tokyo, Japan) mounted with a fixed macro lens was utilized to acquire images. The camera ran in automatic mode, allowing it to set the appropriate acquisition parameters including white balance, ISO speed and exposure time. Images of wheat spikes of 55 genetic lines at the late flowering stage to the milk stage of maturity (from July 11 to August 2) were eventually collected during sunny weather (10:00 to 13:00) in the field. Different genetic

lines of wheats had different resistance to FHB. The images obtained contained FHB of 15 severity stages from grade 0 to grade 14 (Grade 0: [0–1%], Grade 1: (1–2.5%], Grade 2: (2.5–5%], Grade 3: (5–7.5%], Grade 4: (7.5–10%], Grade 5: (10–12.5%], Grade 6: (12.5–15%], Grade 7: (15–17.5%], Grade 8: (17.5–20%], Grade 9: (20–25%], Grade 10: (25–30%], Grade 11: (30–40%], Grade 12: (40–50%], Grade 13: (50–60%], Grade 14: (60–100%]). Images were captured under the ambient conditions of the field site, which included complex and variable backgrounds of blue sky, white clouds, and green wheat plants. Each image (resolution: 6000 × 4000) contained about 7–124 spikes (Figure 1a).
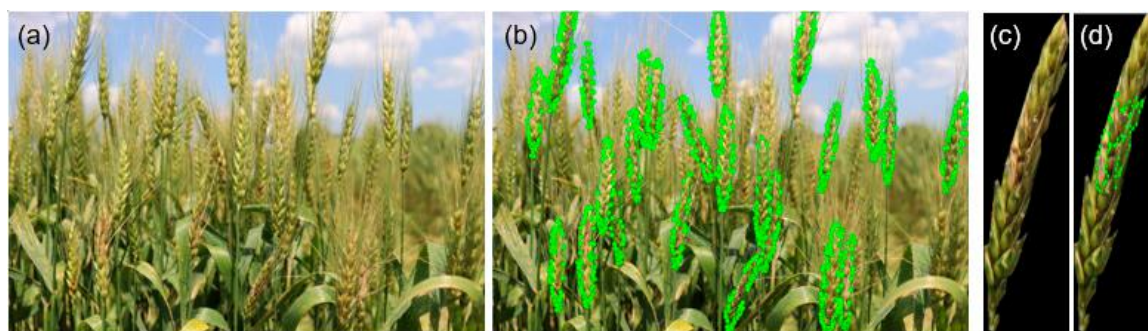

Figure 1. An example of manual annotation of target images; (**a**) a typical full-size image of spikes from a single planted row of wheat in the field; (**b**) the same image with annotated wheat spikes; (**c**) a segmented sub-image of a single wheat spike; (**d**) the individual wheat spike with annotated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

2.2. Data Annotation and Examination

Wheat spikes and diseased areas in collected images were manually annotated. A total of 690 images were captured from a large wheat germplasm collection that varied with respect to FHB reaction. Among this set, 524 images (including 12,591 spikes) and 166 images (including 4749 spikes) were randomly selected as the training set and the validation set of the model for spike identifications, respectively. For disease area detection, 2832 and 922 diseased spikes in sub-images were used, respectively, for model training and validation by random selection. All image annotations were executed by using an artificial image annotation software (Labelme, https://github.com/wkentaro/labelme). Three steps were used for image annotation. The first step was to label wheat spikes in the original images (Figure 1b); the second step was to segment annotated spikes into different sub-images, and the third step was to label the diseased areas in each individual spike. Specifically, the shapes of wheat spikes in the full-size image in the training set were marked by manually drawing polygons. Each of the labeled spikes was then automatically segmented into a sub-image containing a single spike by image processing. All areas of the sub-image were first defaulted as the background (black color) using binarization; then, only the annotated areas were allowed to be recovered (Figure 1c). The physical feature of the diseased areas on the spike of the sub-image was labeled manually (Figure 1d).

2.3. Mask-RCNN

Mask-RCNN was employed to automatically segment the diseased areas of the wheat spikes in full-size images. It is a two-stage model for object detection and segmentation. The first stage is the regional proposal network (RPN), which aims to propose candidate bboxes in the regions of interest (RoI). The second stage is based on the normalized RoIs acquired from RoI Align to output confidence, bbox, and binary mask. The Mask-RCNN is mainly composed of four parts including a backbone, feature pyramid network (FPN), RPN, and feature branches [56]. The backbone is a multilayer neural network used to extract feature maps of original images. A backbone network can be any CNN with residual network (ResNet) developed for image analysis. The ResNet was proposed to solve the vanishing gradient problem when training deep convolutional networks [57].

It relies on a series of stacked residual units as a set of building blocks to develop the network-in-network architecture [58]. The residual units consist of convolution, pooling, and layers.

A ResNet model with 101 layers (ResNet-101) mentioned by He, Zhang, Ren and Sun [57] was employed in this study. ResNet has outperformed previous networks such as visual geometry group networks (VGGNets) at many tasks including object detection and semantic image segmentation [59]. The purpose of using FPN is to completely extract multi-scale feature maps [60]. The RPN has the capacity to generate and choose a rough detection rectangle. Based on functional branches, three operations in terms of classification, detection, and segmentation can be performed. In addition, the batch normalization (BN) is added between activation functions and convolutional layers in the network to accelerate the convergence speed of network training. Ioffe and Szegedy [61] proved that BN could reduce the training steps by more than ten times without changing the model accuracy. The original full-size images with annotated wheat spikes and the sub-images with annotated diseased areas were used, respectively, as the inputs to train two Mask-RCNN models for detection of wheat spikes and diseased areas. Based on the trained dual models, the segmentation of wheat spikes and FHB diseased areas of the images in the validation set was conducted (Figure 2). The severity of FHB was examined based on the ratio of the number of pixels of diseased area to the number of pixels of entire spike area. The workflow of this study is presented in Figure 3.
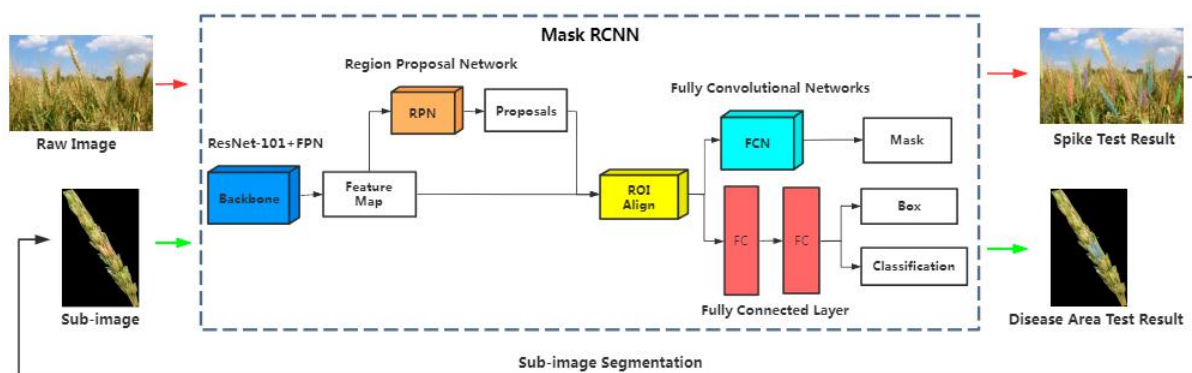

Figure 2. The architecture of the mask region convolutional neural network (Mask-RCNN) approach for wheat Fusarium head blight (FHB) disease assessment.

2.4. Evaluation Metrics

The performance of the Mask-RCNN was evaluated using several parameters. The false positive (*FP*), false negative (*FN*), and true positive (*TP*) were computed and used to generate metrics including recall, precision, *F1*-score, and average precision (*AP*). Among them, the recall (also known as sensitivity) is the proportion of the number of real positive instances in the total number of instances actually belonging to the positive category, while precision (also known as positive predictive value) is the proportion of the number of real positive instances among the total number of instances predicted as belonging to the positive category [62]. As a measure of accuracy of the test, the *F1*-score (also *F*-measure) is the harmonic mean of the recall and precision, where parameters are evenly weighted [63]. The *AP* is the area under the precision–recall (*PR*) curve [64]. The *AP* score is computed as the mean precision over 11 recall values (default values) given a preset intersection over union (*IoU*) threshold [65]. The *IoU* is defined as the degree to which the manually labeled ground truth box overlaps the bbox generated by the model. The mean intersection over union (*MIoU*) is a standard indicator for assessing the performance of image segmentation [66]. *MIoU* was computed as the number of *TP* over the sum of *TP*,

FN, and *FP*. The precision, recall, *F1*-score, *AP*, *IoU*, and *MIoU* can be expressed by the following equations:

$$precision = \frac{TP}{TP + FP}, \tag{1}$$

$$recall = \frac{TP}{TP + FN}, \tag{2}$$

$$F1 = \frac{2PR}{P + R}, \tag{3}$$

$$AP = \frac{1}{11} \sum_{R_j} P(R_j), \; j = 1, 2, 3 \ldots, 11 \tag{4}$$

$$IoU(E, F) = \left| \frac{E \cap F}{E \cup F} \right|, \tag{5}$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{P_{ii}}{\sum_{j=0}^{k} P_{ij} + \sum_{j=0}^{k} P_{ji} - P_{ii}}, \tag{6}$$

where *TP* corresponds to the number of true positives generated (i.e., the number of wheat spikes correctly detected), *FP* represents the number of wheat spikes incorrectly identified, *FN* is the number of wheat spikes undetected but should have been identified. *E* represents the ground truth box labeled manually and *F* represents the bbox generated based on the Mask R-CNN model. If the estimated *IoU* value is higher than the preset threshold (0.5), the predicted result of this model is considered as a *TP*, otherwise as an *FP*. $k + 1$ is the total number of output classes including an empty class (the background), and P_{ii} represents *TP*, while P_{ij} and P_{ji} indicate *FP* and *FN*, respectively.
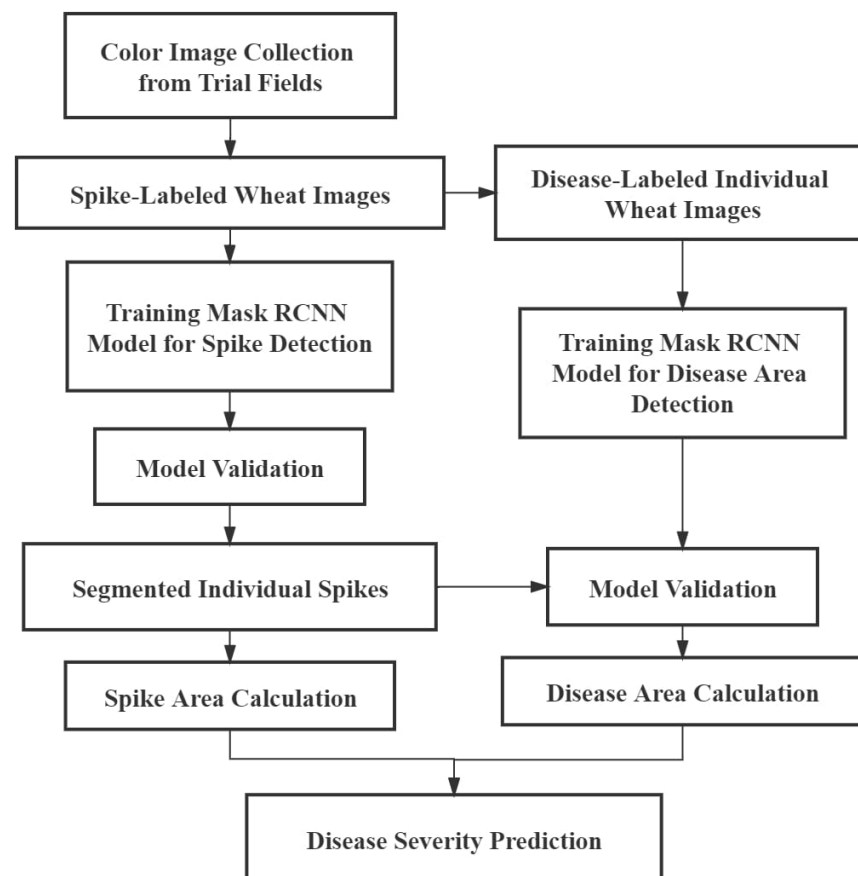

Figure 3. A flowchart of the proposed approach for evaluating FHB disease severity.

2.5. Equipment

The entire process for model training and validation was implemented by a personal computer (processor: Intel(R) Xeon(R) CPU E3-1225 v3 @ 3.20GHz; operating system: Ubuntu 18.04, 64 bits; memory: 20 Gb). The training speed was optimized in graphics processing unit (GPU) mode (NVIDIA RTX 2070 8 Gb). Table 1 presents the relevant modeling parameters (such as the base learning rate) adopted in this study. The time for model training and validation is shown in Table 2. The code for image processing was written in Python.

Table 1. Modeling parameter settings.

Modelling Parameters	Values
Base learning rate	0.02
Image input batch size	2
Gamma	0.1
Number of classes	2
Maximum iterations	2,700,000

Table 2. The total time for model training and validation.

Application	Training Time	Validation Time
Wheat spike identification	45 h 23 min 26 s	16 min 30 s
FHB disease detection	23 h 46 min 58 s	1 min 28 s

3. Results

3.1. Model Training

Dual training Mask-RCNN models of deep neural networks were established based on the annotated images of wheat spikes and FHB diseased areas. Figure 4a shows the trend of accuracy and loss during first model training for wheat spike identification. It was observed that the loss of the bbox and the mask dropped sharply from the initial iteration and tended to stabilize after 25,000 iterations. Compared with the loss, the model accuracy increased during this process. The function curves of the accuracy and the loss fluctuated during the iterations and weakened after iterating 210,000 times. Both the accuracy function and the loss functions reached convergence after 270,000 iterations. As can be seen, the loss value of the mask was always greater than that of the bbox. When the model accuracy for wheat spike increased to 1 (100%), the loss value of the bbox reached the lowest (0.001), and the loss value of the mask reduced to 0.037. Similarly, Figure 4b describes the variation of the accuracy and the loss in another model training for FHB disease assessment. Throughout the iteration process, both the loss values of the bbox and the mask gradually decreased and tended to converge, while the model accuracy maintained a trend of weak growth until convergence. Eventually, the loss values of the mask and the bbox reduced to 0.063 and 0.002, respectively, while the model accuracy for diseased areas increased to over 99.80%. These results indicate that the trained classifiers effectively learned the features of annotated wheat spikes and diseased areas.

3.2. Wheat Spike Identification

The trained Mask-RCNN model was then used to recognize wheat spikes in full-size images in the validation set. Instance segmentation of individual wheat spikes was conducted under complex conditions including occlusion and overlap. The category score (bbox) and mask of each wheat spike was generated for the test images. The algorithm successfully recognized the high-density wheat spikes in the field (Figure 5a). Due to the camera shooting angle, wheat spikes in the images inevitably obstructed each other, but the algorithm was able to segment two wheat spikes with overlapping boundaries (Figure 5b). Most FHB phenotyping was only taken from plants in the center portion of a plot and

the edge of plots were excluded due to possible edge effects, which meant wheat spikes that were incompletely segmented were usually located at the borders of full-size images. Figure 5c shows that the wheat spikes cut at the image borders are able to be successfully recognized. It is important to identify the partial spikes because such spikes can be used as a beneficial supplement to maximize the dataset and enhance the robustness of the model. The segmentation results of 166 test images showed that the *MIoU* rate for wheat spikes reached 52.49%. The algorithm presented an acceptable performance for wheat spike prediction, with the *AP* of the mask and the bbox of 57.16% and 56.69%, respectively (Table 3). Based on the results of 166 images, the overall rates of precision, recall, *IoU* and *F1*-score were 81.52%, 71.00%, 46.41% and 74.78%, respectively. The total number of spikes identified by the Mask-RCNN was compared with the actual number of spikes labeled manually. Among 4749 wheat spikes, 3693 spikes were correctly identified, yielding the recognition rate of 77.76%. This proves that the Mask-RCNN was effective for rapidly identifying wheat spikes under field conditions.
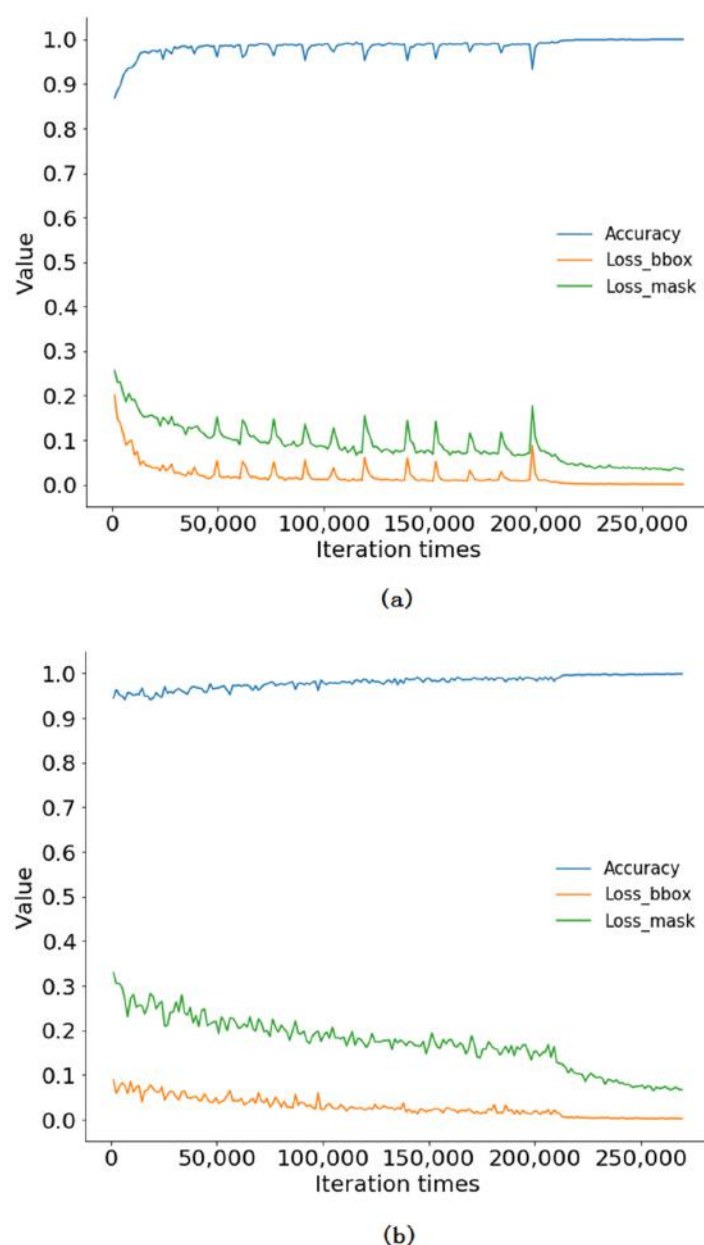

Figure 4. Training curves of accuracy and loss against number of iterations on identifications of (**a**) wheat spikes and (**b**) diseased areas.
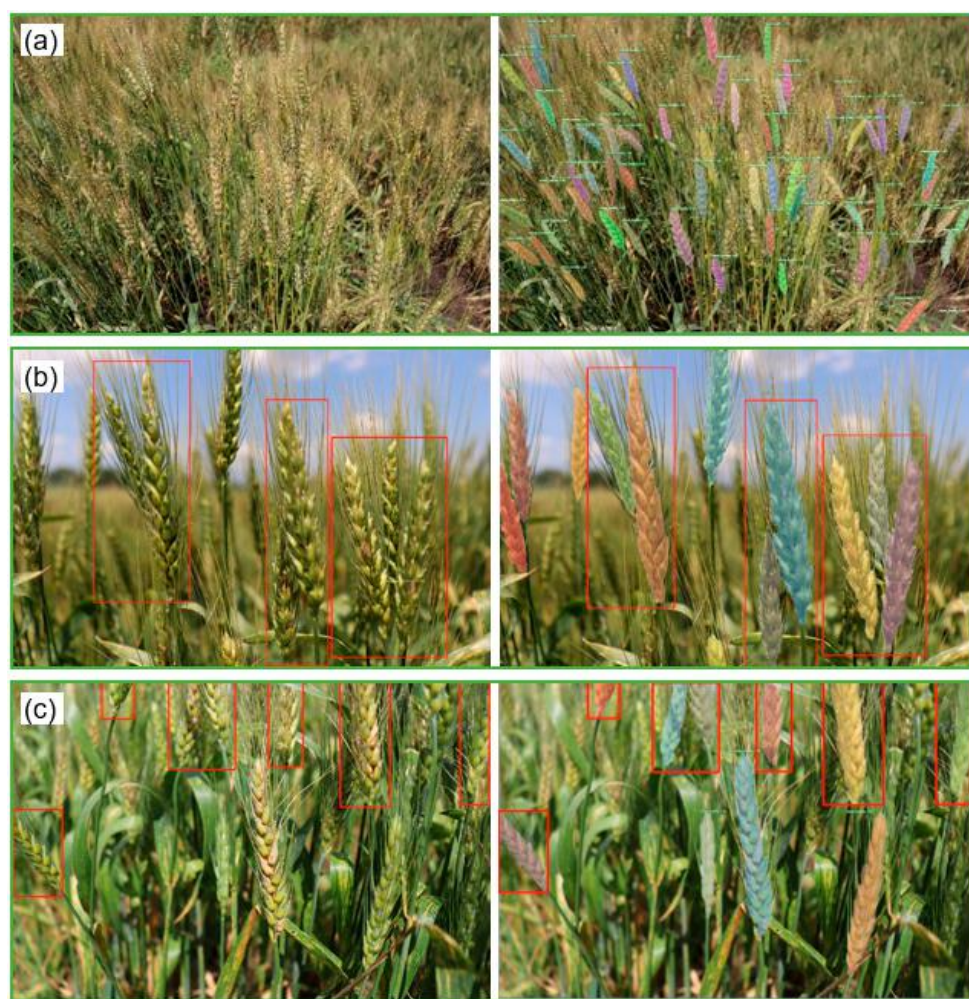
Figure 5. Selected examples of instance segmentation on correct detections of wheat spikes under three complex conditions in the test dataset, (**a**) high-density wheat spikes, (**b**) wheat spikes with overlapping boundaries, and (**c**) wheat spikes cut at the image borders. For each example, the original image (left) and the corresponding recognition result (right) are displayed.

Table 3. Results of Mask-RCNN for wheat spike and Fusarium head blight (FHB) disease detection.

Type	P (%)	R (%)	*F1*-Score (%)	IoU (%)	AP of Bbox (%)	AP of Mask (%)	MIoU (%)
Wheat spike	81.52	71.00	74.78	46.41	56.69	57.16	52.49
FHB disease	72.10	76.16	74.04	51.24	63.38	65.14	51.18

The failures of the proposed methodology for wheat spike identification were also evaluated. Figure 6 depicts the predictions of selected wheat spike images from the validation dataset. As can be observed, most spikes were successfully detected. For spikes not detected, the *FN* were detection failures calculated in the cases of wheat spikes (blue rectangles) highly occluded (Figure 6a,c), cut at the image border (Figure 6b) or missing annotation on slightly blurred spikes (Figure 6d) due to human error. The *FP* were detection failures (red rectangles) caused by various factors, such as the presence of awns obscuring background spikes (Figure 6a,b), overlapping spikes (Figure 6d), and out-of-focus spikes (Figure 6c). These factors should not be considered by the model. Specifically, Figure 6a shows the misdetection (a blue rectangle with a red rectangle inside) of a wheat spike (blue rectangle) due to the occlusion of awns. In contrast, the detection (a red rectangle with a blue rectangle inside) observed in Figure 6b is a true spike (blue rectangle) covered by

a red mask (red rectangle) that also has been obscured by the awns (presenting a similar pattern to spikes). This spike was undetectable due to the model error. Figure 7a,b show that there was one wheat spike misclassified as two spikes. The multi-detections were *FP*. As the number of long spikes in the dataset was not large, the specific features of these spikes were not fully learned by the algorithm, generating the same result as in the case of overlapped wheat spikes (Figure 7c,d).
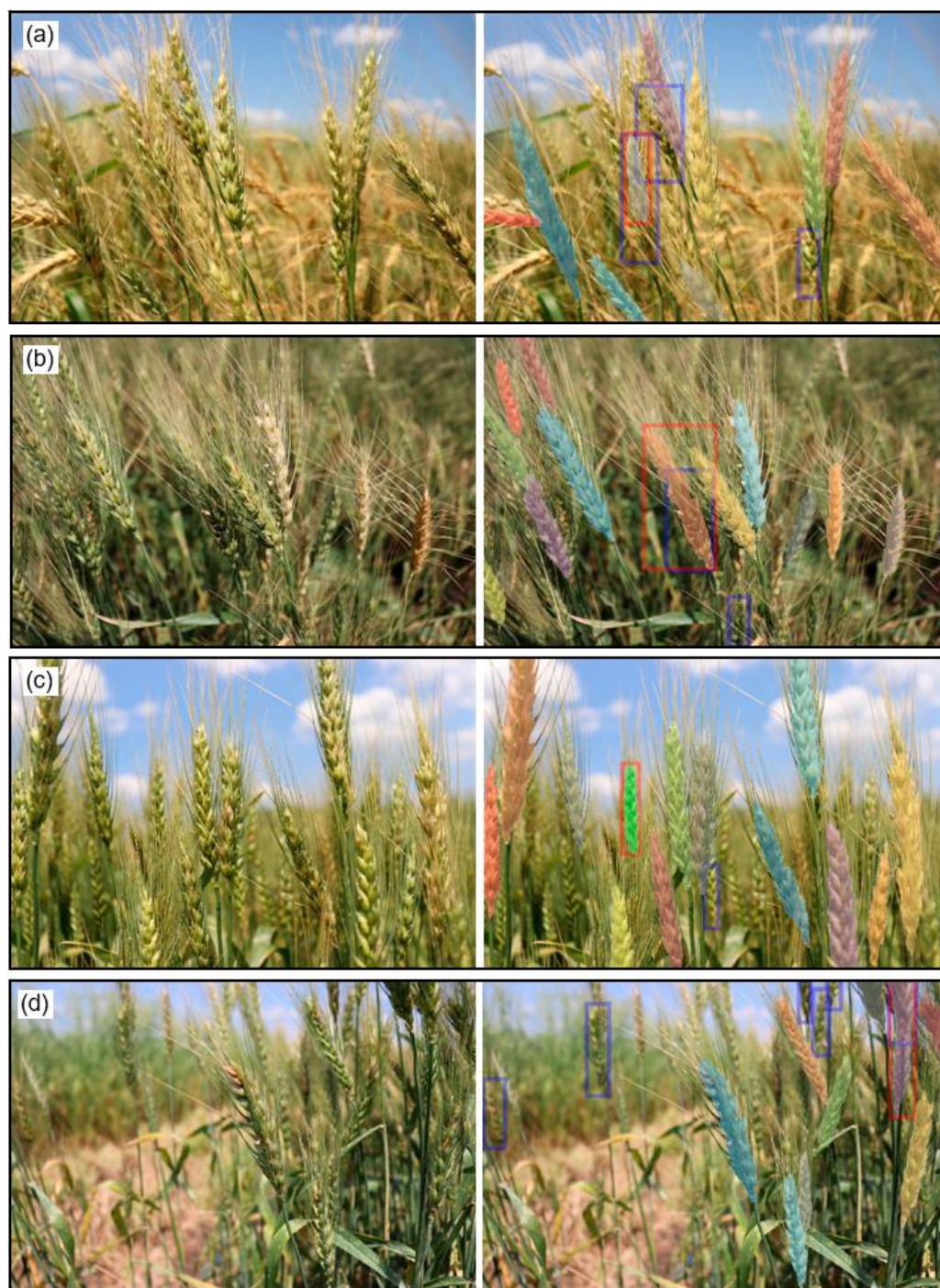

Figure 6. (**a**–**d**) Various examples of wheat spike detections in the validation dataset; the spike with the blue box rectangle only has the ground truth (a.k.a. is labeled) with no prediction results (a.k.a. is missed), which is a false negative, while the spike with the red rectangle only has the prediction result (a.k.a. is detected) with no ground truth (a.k.a. is not labeled), which is false positive.
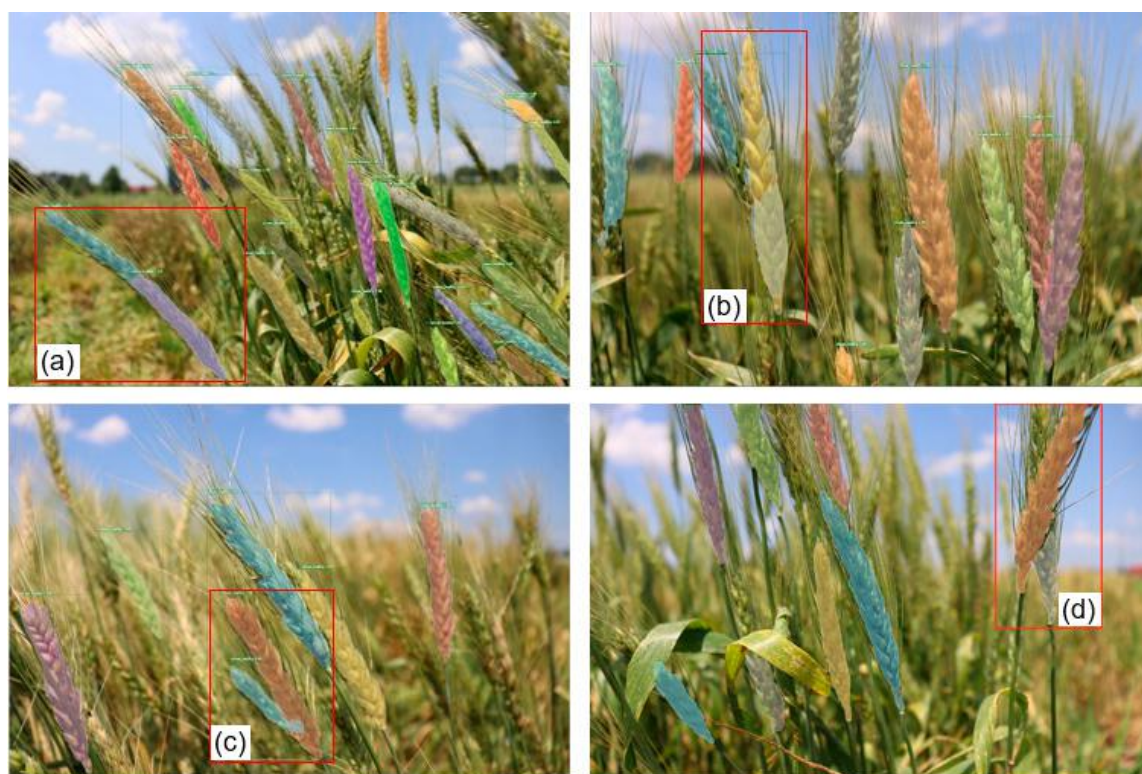
Figure 7. Results of instance segmentation of selected wheat spikes (red rectangles); (**a**,**b**) examples of multi-detections; (**c**,**d**) examples of wheat spike occlusion.

3.3. FHB Disease Evaluation

After segmenting the individual wheat spikes in full-size images, a second trained Mask-RCNN model was employed to evaluate the diseased areas on these infected spikes. A dataset of 922 sub-images of diseased wheat spikes was used as the validation set in Mask-RCNN. As shown in Figure 8I, each sub-image contained one spike. The diseased spikelets in each spike were successfully recognized and marked using the category scores, bboxes, and masks (Figure 8II). The instance area of FHB disease can be segmented and extracted. Figure 8III shows the segmentation images of infected spikelets from the spikes. Results showed that the *MIoU* rate for disease area instance segmentation reached 51.18%. The *AP* rates of the mask and the bbox for FHB disease detection were 65.14% and 63.38%, respectively. The diseased areas with shadow, strong light, low light, or awn occlusion were effectively recognized (Figure 8). Figure 9 shows the results for disease detection when the entire wheat spike is occluded by a straw. Mask-RCNN achieved the accurate identification of diseased areas. Eventually, a total of 911 diseased spikes were recognized from 922 samples with the detection rate of 98.81%. Moreover, Mask-RCNN generated acceptable results for detecting disease, yielding the overall rates of precision, recall, *F1*-score, and *IoU* of 72.10%, 76.16%, 74.04% and 51.24%, respectively (Table 3).

Although the trained model showed excellent performance in disease area recognition, there were difficulties in some cases. Figure 10 shows selected examples of incorrect segmentation of FHB diseased areas. As can be seen, the diseased areas of two wheat spikes were not successfully detected, including the one with the presence of multi-detections (Figure 10a1–a4) and a highly occluded one (Figure 10b1–b4). Such errors were associated with *FP*. As shown in Figure 11a, this detection failure (red rectangle) was reported as *FP* caused by the occlusion of awns. Other *FP* detections included were spikelets that were misannotated when labeled due to human error (red rectangles in Figure 11c,d). The undetected wheat spikelets (blue rectangles) in Figure 11a,b,d were the *FN* due to the model error. In addition, environment factors including sunlight reflection or appearance

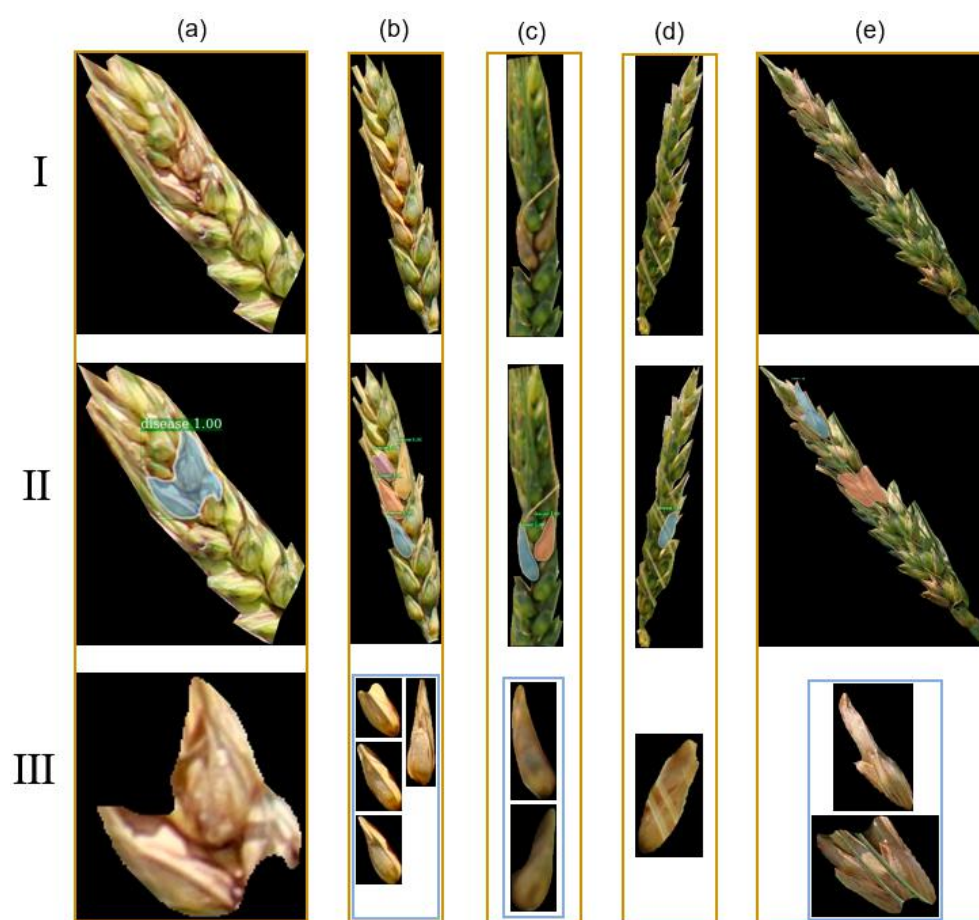variations of the diseased areas in view angle, shape, or occlusion may result in the failure identifications.

Figure 8. Illustration of (**II**) disease area detection and (**III**) segmentation results from (**I**) 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**,**e**) diseased areas under awn occlusion.

Figure 9. The process of disease detection for an entire wheat spike partially occluded by the peduncle (neck) of another plant; (**a**) a selected example of an original image with multiple wheat spikes; (**b**) a zoom-in view of a spike occluded by a peduncle; (**c**) Mask-RCNN for wheat spike detection; (**d**) Mask-RCNN for disease area identification; (**e**) the segmentation result of diseased spikelets.

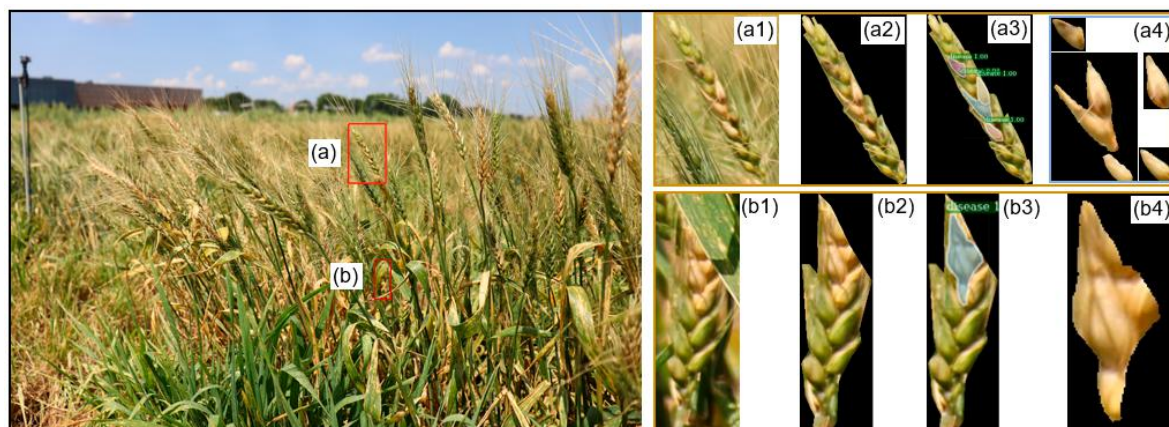**Figure 10.** Examples of misjudgment of the diseased areas on (**a**,**b**) two wheat spikes; (**a1**,**b1**) zoom-in views of the spike; (**a2**,**b2**) Mask-RCNN for wheat spike detection; (**a3**,**b3**) Mask-RCNN for disease area identification; (**a4**,**b4**) the segmentation result of diseased spikelets.
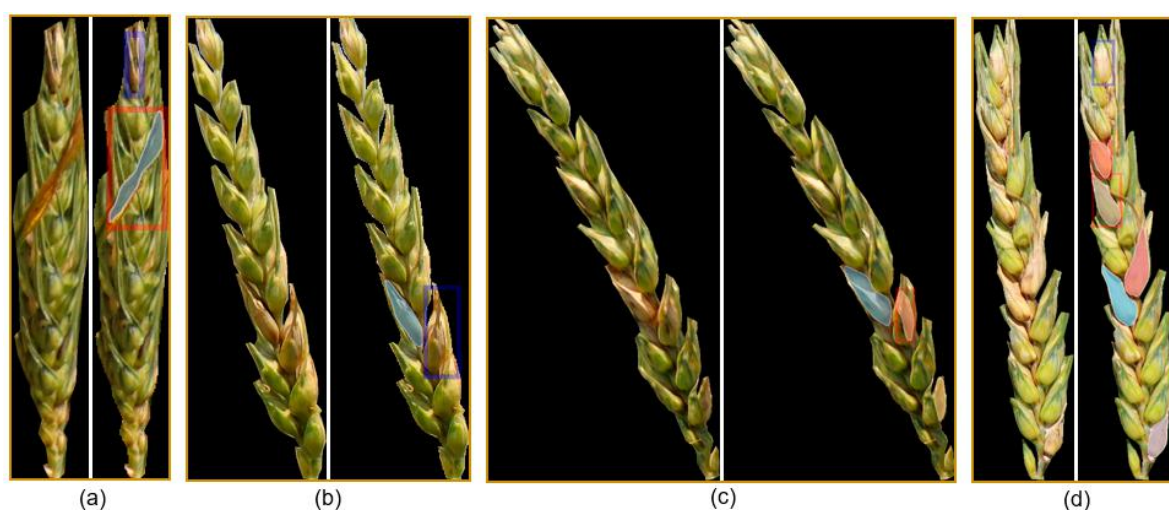

Figure 11. The detection of diseased areas in selected wheat sub-images in the validation dataset; (**a**,**b**) examples of detection failures caused by model errors; (**c**) an example of detection failures caused by human error due to mis-annotation; (**d**) an example of detection failures caused by model error or human error (the spike with the blue box rectangle was only captured in the ground truth set with no prediction results, while the spike with the red rectangle was only captured in the prediction results not in the ground truth results).

3.4. Examination of Wheat FHB Severity

The FHB disease severity was evaluated according to the ratio of the disease area to the entire spike area. As shown in Figure 12, disease levels of each spike were calculated and divided into 14 FHB severity grades. Spikes with lower disease levels were separated into more numerous grades with narrower severity intervals, because selecting among lines with lower disease levels is more critical to the breeding process, as lines with high disease levels are undesirable. Figure 12a depicts the ground truth (ground truth is the visual rating of spikes by an expert from the acquired images) of wheat spikes at different disease grades in the training set. As seen in Figure 12a, 83.51% of samples in the training set were categorized with disease grades of 2–9, while 87.74% of the ground truth in the validation set was assigned to this group, which was a little bit lower than that (92.10%) in the prediction set as described in Figure 12b. The statistical results of the FHB severity of wheat spikes in the training and validation sets are shown in Table 4. By inspecting

the distribution of samples, it was observed that the ground truth of FHB severity in the validation set was close to that of the training set. The overall predicted ratio of the disease area over the entire spike area was 9.27% (grade 4) based on data in validation set. For 92.10% of wheat spikes, the infected area of an individual spike ranged from 2.5% to 25% (grades 2–9). Samples with infection areas of 2.5% to 10% (grades 2–4) accounted for 60.59%, followed by the 27.55% samples with the infection area between 10% and 20% (grades 5–9). When the disease level was over grade 4, it was observed that the predicted number (e.g., 95 for grade 5) of the diseased wheat spikes in each grade was lower than the ground truth (e.g., 105 for grade 5). In disease grades 5–12, the differences between blue bars and orange bars are the false negatives. When the disease level was no more than grade 4, the predicted number (e.g., 133 for grade 4) of spikes in each grade was higher than its actual number (e.g., 124 for grade 4). These differences in categories (1–4) are the false positives. Nevertheless, the average disease severity (9.27%) from the prediction was comparable to that of the ground truth (12.01%). Eventually, the prediction accuracy (77.19%) for diseased wheat spikes was calculated by the severity value of prediction (9.27%) over ground truth (12.01%).
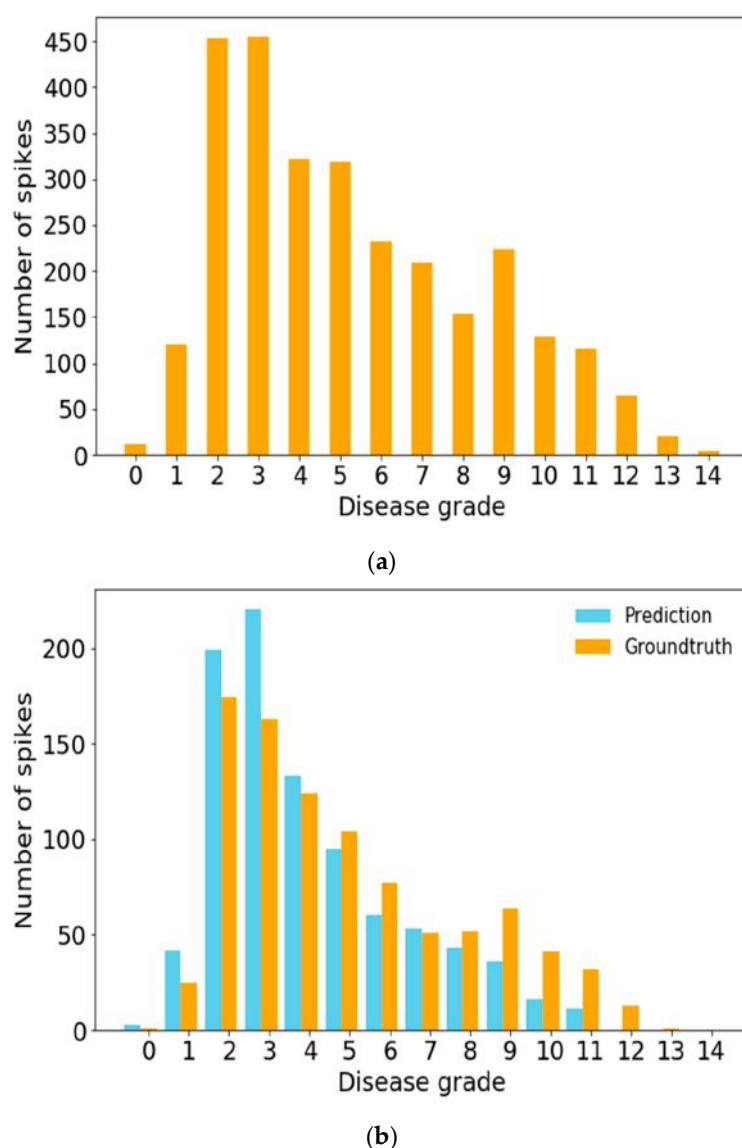

(**a**)

(**b**)

Figure 12. Frequency of wheat spikes of each disease grade; (**a**) the number of wheat spikes at different disease grades in the training set; (**b**) the number of predicted and ground truth spikes in the validation set.

Table 4. Statistical results of wheat spike FHB disease severity.

Dataset	Type	No. of Spikes	Severity (%)		
			Mean ± SD	Max	Min
Training	Ground truth	2382	13.23 ± 10.44	85.51	0.50
Validation	Ground truth	922	12.01 ± 8.81	50.16	0.89
	Prediction	911	9.27 ± 6.15	34.68	0.86

4. Discussion

This research proposed a new approach using two Mask-RCNN models in a row for automatic determination of the symptom location and the disease severity of wheat spikes in the field. The artificial inoculation adopted in this study is able to ensure sufficient infection to decrease high environmental variations. Natural infections tend to be highly dependent on weather conditions, but genotypes with high resistance cannot be reliably identified from natural infections because of low infection pressure [67]. Nevertheless, the symptoms based on the inoculation may be slightly different from a natural infection. In future studies, images from naturally infected wheat samples will be used for modeling. The extent and dynamics of FHB development under the same environmental conditions depend on the resistance of host plants. This requires a lot exploration to find new sources of resistance through wild and domesticated wheat germplasm diversity. In this study, the images of wheat spikes from 55 genetic lines were acquired from the fields. Each genetic line of wheat had specific resistance to FHB, which means that this sample complexity can ensure the generalization of deep models. The results of field trials can provide a more comprehensive estimate of FHB resistance than greenhouse screenings, which is of great practical significance for breeders to identify resistant varieties.

Data annotation task is a very laborious and time consuming and very complex progress. There is a need for more senior experts to annotate input images, as existing experts may be susceptible to errors when facing a challenging task in disease annotation. In the future, it should also be possible to make the annotation automatically by the more advanced software. This kind of automation will be the process desired for the future application of our algorithm. The differences between blue bars and orange bars in Figure 12 were due to false positives (disease grades 1–4) or false negatives (disease grades 5–12). These false rates could affect the decision making when selecting resistant lines. This would underestimate the disease grade of a wheat line, which means that the disease grade of this wheat originally belongs to a high grade, but it may be classified into a lower disease category by the model. Nevertheless, this methodology yielding an accuracy of 77.19% is good for breeders as manually screening hundreds of wheat lines will also be inaccurate and this process is very time-consuming and inefficient. Although it is very difficult to overcome human errors, the annotation of all spikes for modeling can be achieved by carefully labeling. A big barrier in the use of Mask-RCNN is the need for more representative and larger datasets including wheat spikes with awns and overlapping in model training. Samples under different climatic conditions should be collected in the future. In addition, the number of samples should be increased by data augmentation to establish a more robust model and avoid the potential overfitting problem. A high-throughput method utilizing streamlined image analysis for real-time FHB assessment in the field could be developed, thus helping to reduce the amount of time, cost (labor), and subjectivity error from conventional manual phenotyping methods. Recently, Zhang et al. [68] developed an FHB diagnostic system for detection of individual wheat spikes with a black cloth background. Only 79 samples were used for training and a very small dataset of 41 wheat spikes were tested in the study. In contrast, the sample number was much larger in our study. In total, 524 images with 12,591 wheat spikes and 166 images with 4749 wheat spikes were utilized. These images were taken under complex backgrounds (including blue sky, white clouds, and green wheat plants) in the field.

The Mask-RCNN used in our study performed very well yielding the detection rate as high as 98.81% compared to the study of Williams et al. [69], in which they reported a detection rate of 89.6% using CNN for kiwifruit. There are two main factors that led to the success of the current study. The first reason is the superiority of Mask-RCNN over CNN. Another main reason is probably because each wheat spike used to develop the training model is labeled with high precision, so that the performance and robustness of the trained Mask-RCNN model have been significantly improved. This algorithm for wheat spike detection showed a similar performance to the PCNN in eliminating background interference [46]. Mask-RCNN has the capability of providing segmentation masks for rapid detection of multiple wheat spikes (7–124) in one image, which is more feasible for a high-throughput and real-time assay in the field. Zhang et al. [46] reported a high accuracy (0.981) for segmentation of individual wheat spikes in an image, but the metrics such as precision, recall, *F1*-score, *MIoU*, and *AP* were not considered in their study. Although Li et al. [70] reported an accuracy (*F1*-score = 71.8–76.4%) for recognition of rice sheath blight disease that was similar to that found for FHB disease detection in our study (*F1*-score = 74.04%), their *IoU* threshold (0.2) for detection was set very low. In this study, only the network framework of ResNet-101 was considered for target detection because residual networks including ResNet-44, 47, 50, 71 and 101 have been used, respectively, as a backbone network of Mask R-CNN in a previous study for strawberry identification. It was found that the ResNet-101 achieved the highest detection accuracy [71]. In the study of Kiratiratanapruk et al. [72], Mask R-CNN provided higher performance than other models including Faster R-CNN and RetinaNet in detection of rice leaf diseases, but YOLOv3 achieved the highest accuracy. Although the YOLOv3 detector had higher computation efficiency, the Faster-RCNN outperformed YOLOv3 in apple detection [73]. This indicated that the more advanced YOLO algorithms (such as YOLOv4 and YOLOv5) should be utilized in future studies for assessing wheat resistance to FHB.

A ground-based motorized phenocart could be designed in the future to provide real-time and accurate FHB assessment data to assist in disease phenotyping, which will hasten the time required to develop new breeding lines with enhanced FHB resistance. The ideal vehicle should be low-cost, lightweight, and easy to maneuver across variable field surfaces. The image capture equipment should be designed to collect images under the variable environmental conditions (e.g., wind and sunlight) that occur in the field. The human error caused by manual labeling should be corrected, and more samples should be analyzed for training a more reliable model to reduce the error. The severity of FHB on individual wheat lines can increase over the course of the season. In practice, conventional assessments of FHB severity are usually made just once when the disease has reached its maximum. Given that different wheat lines mature at different times, conducting multiple disease assessments of an entire breeding nursery would be prohibitive. The results from this study demonstrate that the developed framework has great potential for real-time assessment of FHB severity in wheat spikes in the field. This development that will greatly enhance the efficiency of reliably selecting wheat lines with the desired level of FHB resistance. The use of a uniform background panel fixed to a phenocart would produce a clearer and more uniform background contrast for wheat spikes, allowing easier identification of the targets and improving the detection accuracy of the algorithm. In addition, new classifiers are expected to be developed to assist rapid labeling of training data in the future.

5. Conclusions

A high-throughput framework of deep-learning based disease detection algorithms was established to automatically assess wheat resistance to FHB under field conditions. The protocols involved image collection, processing and deep learning modeling. Dual Mask-RCNN models were developed for rapid segmentations of wheat spikes and FHB diseased areas. Based on the methodology, mask images of individual wheat spikes and diseased areas were outputted, with detection rates of 77.76% and 98.81%, respectively. The Mask-RCNN model demonstrated strong capacity for recognition of the targets occluded by

wheat awns or cut at the image borders. By calculating the wheat FHB severity value of prediction over ground truth, acceptable prediction accuracy was achieved. The knowledge generated by this study will greatly aid in the efficient selection of FHB resistant wheat lines in breeding nurseries. This, in turn, will contribute to the development of resistant wheat cultivars that will ameliorate the losses due to FHB, thereby contributing to global food security and sustainable agricultural development.

References

1. McMullen, M.; Bergstrom, G.; De Wolf, E.; Dill-Macky, R.; Hershman, D.; Shaner, G.; Van Sanford, D. A unified effort to fight an enemy of wheat and barley: Fusarium head blight. *Plant Dis.* **2012**, *96*, 1712–1728. [CrossRef] [PubMed]
2. 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**, 128507. [CrossRef]
3. Su, W.-H.; Arvanitoyannis, I.S.; Sun, D.-W. Trends in food authentication. In *Modern Techniques for Food Authentication*; Elsevier: Amsterdam, The Netherlands, 2018; pp. 731–758.
4. Pedersen, J. Distribution of deoxynivalenol and zearalenone in milled fractions of wheat. *Cereal Chem.* **1996**, *73*, 388–391.
5. Stenglein, S. Fusarium poae: A pathogen that needs more attention. *J. Plant Pathol.* **2009**, *91*, 25–36.
6. Buerstmayr, H.; Ban, T.; Anderson, J.A. QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: A review. *Plant Breed.* **2009**, *128*, 1–26. [CrossRef]
7. Horsley, R.D.; Schmierer, D.; Maier, C.; Kudrna, D.; Urrea, C.A.; Steffenson, B.J.; Schwarz, P.; Franckowiak, J.; Green, M.; Zhang, B. Identification of QTLs associated with Fusarium head blight resistance in barley accession CIho 4196. *Crop Sci.* **2006**, *46*, 145–156. [CrossRef]
8. Stack, R.W.; McMullen, M.P. *A Visual Scale to Estimate Severity of Fusarium Head Blight in Wheat*; North Dakota State University: Dakota, NE, USA, 1998.
9. Fetch, T.G., Jr.; Steffenson, B.J. Rating scales for assessing infection responses of barley infected with Cochliobolus sativus. *Plant Dis.* **1999**, *83*, 213–217. [CrossRef]
10. Su, W.-H.; Bakalis, S.; Sun, D.-W. Potato hierarchical clustering and doneness degree determination by near-infrared (NIR) and attenuated total reflectance mid-infrared (ATR-MIR) spectroscopy. *J. Food Meas. Charact.* **2019**, *13*, 1218–1231. [CrossRef]
11. Su, W.-H.; Bakalis, S.; Sun, D.-W. Fingerprinting study of tuber ultimate compressive strength at different microwave drying times using mid-infrared imaging spectroscopy. *Dry. Technol.* **2019**, *37*, 1113–1130. [CrossRef]
12. Su, W.-H.; Bakalis, S.; Sun, D.-W. Fourier transform mid-infrared-attenuated total reflectance (FTMIR-ATR) microspectroscopy for determining textural property of microwave baked tuber. *J. Food Eng.* **2018**, *218*, 1–13. [CrossRef]
13. Su, W.-H.; Zhang, J.; Yang, C.; Page, R.; Szinyei, T.; Hirsch, C.D.; Steffenson, B.J. Evaluation of mask RCNN for learning to detect fusarium head blight in wheat images. In *Proceedings of the 2020 ASABE Annual International Meeting, Omaha, NE, USA, 12–15 July 2020*; American Society of Agricultural and Biological Engineers: Saint Joseph, MI, USA, 2020; p. 1.

14. Su, W.-H.; Slaughter, D.C.; Fennimore, S.A. Non-destructive evaluation of photostability of crop signaling compounds and dose effects on celery vigor for precision plant identification using computer vision. *Comput. Electron. Agric.* **2020**, *168*, 105155. [CrossRef]
15. Su, W.-H.; Fennimore, S.A.; Slaughter, D.C. Development of a systemic crop signaling system for automated real-time plant care in vegetable crops. *Biosyst. Eng.* **2020**, *193*, 62–74. [CrossRef]
16. Su, W.-H.; Fennimore, S.A.; Slaughter, D.C. Fluorescence imaging for rapid monitoring of translocation behavior of systemic markers in snap beans for automated crop/weed discrimination. *Biosyst. Eng.* **2019**, *186*, 156–167. [CrossRef]
17. Su, W.-H. Systemic crop signaling for automatic recognition of transplanted lettuce and tomato under different levels of sunlight for early season weed control. *Challenges* **2020**, *11*, 23. [CrossRef]
18. Su, W.-H.; Sun, D.-W. Facilitated wavelength selection and model development for rapid determination of the purity of organic spelt (*Triticum spelta* L.) flour using spectral imaging. *Talanta* **2016**, *155*, 347–357. [CrossRef]
19. 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*, 767–792. [CrossRef]
20. Moghimi, A.; Yang, C.; Miller, M.E.; Kianian, S.F.; Marchetto, P.M. A novel approach to assess salt stress tolerance in wheat using hyperspectral imaging. *Front. Plant Sci.* **2018**, *9*, 1182. [CrossRef]
21. Li, L.; Zhang, Q.; Huang, D. A review of imaging techniques for plant phenotyping. *Sensors* **2014**, *14*, 20078–20111. [CrossRef]
22. Enders, T.A.; St. Dennis, S.; Oakland, J.; Callen, S.T.; Gehan, M.A.; Miller, N.D.; Spalding, E.P.; Springer, N.M.; Hirsch, C.D. Classifying cold-stress responses of inbred maize seedlings using RGB imaging. *Plant Direct* **2019**, *3*, e00104. [CrossRef]
23. Whetton, R.L.; Hassall, K.L.; Waine, T.W.; Mouazen, A.M. Hyperspectral measurements of yellow rust and fusarium head blight in cereal crops: Part 1: Laboratory study. *Biosyst. Eng.* **2018**, *166*, 101–115. [CrossRef]
24. Alisaac, E.; Behmann, J.; Kuska, M.; Dehne, H.-W.; Mahlein, A.-K. Hyperspectral quantification of wheat resistance to Fusarium head blight: Comparison of two Fusarium species. *Eur. J. Plant Pathol.* **2018**, *152*, 869–884. [CrossRef]
25. Mahlein, A.-K.; Alisaac, E.; Al Masri, A.; Behmann, J.; Dehne, H.-W.; Oerke, E.-C. Comparison and combination of thermal, fluorescence, and hyperspectral imaging for monitoring fusarium head blight of wheat on spikelet scale. *Sensors* **2019**, *19*, 2281. [CrossRef] [PubMed]
26. Ma, H.; Huang, W.; Jing, Y.; Pignatti, S.; Laneve, G.; Dong, Y.; Ye, H.; Liu, L.; Guo, A.; Jiang, J. Identification of Fusarium head blight in winter wheat ears using continuous wavelet analysis. *Sensors* **2020**, *20*, 20. [CrossRef] [PubMed]
27. 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] [PubMed]
28. Xiao, Y.; Dong, Y.; Huang, W.; Liu, L.; Ma, H.; Ye, H.; Wang, K. dynamic remote sensing prediction for wheat fusarium head blight by combining host and habitat conditions. *Remote Sens.* **2020**, *12*, 3046. [CrossRef]
29. 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]
30. 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]
31. Su, W.-H.; Sun, D.-W. Advanced analysis of roots and tubers by hyperspectral techniques. *Adv. Food Nutr. Res.* **2019**, *87*, 255–303.
32. Su, W.-H.; Bakalis, S.; Sun, D.-W. Chemometric determination of time series moisture in both potato and sweet potato tubers during hot air and microwave drying using near/mid-infrared (NIR/MIR) hyperspectral techniques. *Dry. Technol.* **2020**, *38*, 806–823. [CrossRef]
33. Su, W.H.; Sun, D.W. Multispectral imaging for plant food quality analysis and visualization. *Compr. Rev. Food Sci. Food Saf.* **2018**, *17*, 220–239.
34. Su, W.-H.; Sun, D.-W. Evaluation of spectral imaging for inspection of adulterants in terms of common wheat flour, cassava flour and corn flour in organic Avatar wheat (Triticum spp.) flour. *J. Food Eng.* **2017**, *200*, 59–69. [CrossRef]
35. Su, W.-H.; Sun, D.-W. Chemical imaging for measuring the time series variations of tuber dry matter and starch concentration. *Comput. Electron. Agric.* **2017**, *140*, 361–373. [CrossRef]
36. Su, W.-H.; He, H.-J.; Sun, D.-W. Non-destructive and rapid evaluation of staple foods quality by using spectroscopic techniques: A review. *Critical Rev. Food Sci. Nutr.* **2017**, *57*, 1039–1051. [CrossRef] [PubMed]
37. Su, W.-H.; Sun, D.-W. Multivariate analysis of hyper/multi-spectra for determining volatile compounds and visualizing cooking degree during low-temperature baking of tubers. *Comput. Electron. Agric.* **2016**, *127*, 561–571. [CrossRef]
38. Su, W.-H.; Sun, D.-W. Comparative assessment of feature-wavelength eligibility for measurement of water binding capacity and specific gravity of tuber using diverse spectral indices stemmed from hyperspectral images. *Comput. Electron. Agric.* **2016**, *130*, 69–82. [CrossRef]
39. Dammer, K.-H.; Möller, B.; Rodemann, B.; Heppner, D. Detection of head blight (*Fusarium* ssp.) in winter wheat by color and multispectral image analyses. *Crop Prot.* **2011**, *30*, 420–428. [CrossRef]
40. Zhang, D.; Wang, Z.; Jin, N.; Gu, C.; Chen, Y.; Huang, Y. Evaluation of efficacy of fungicides for control of wheat fusarium head blight based on digital imaging. *IEEE Access* **2020**, *8*, 109876–109890. [CrossRef]

41. Fu, J.; Zheng, H.; Mei, T. Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4438–4446.

42. Lawrence, S.; Giles, C.L.; Tsoi, A.C.; Back, A.D. Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Netw.* **1997**, *8*, 98–113. [CrossRef]

43. Zhou, L.; Gu, X. Embedding topological features into convolutional neural network salient object detection. *Neural Netw.* **2020**, *121*, 308–318. [CrossRef]

44. 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*, 100. [CrossRef]

45. Pound, M.P.; Atkinson, J.A.; Wells, D.M.; Pridmore, T.P.; French, A.P. Deep learning for multi-task plant phenotyping. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 2055–2063.

46. 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]

47. 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]

48. Prakash, R.M.; Saraswathy, G.; Ramalakshmi, G.; Mangaleswari, K.; Kaviya, T. Detection of leaf diseases and classification using digital image processing. In Proceedings of the 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 17–18 March 2017; pp. 1–4.

49. 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.

50. Ganesh, P.; Volle, K.; Burks, T.; Mehta, S. Deep orange: Mask R-CNN based orange detection and segmentation. *IFAC PapersOnLine* **2019**, *52*, 70–75. [CrossRef]

51. Jia, W.; Tian, Y.; Luo, R.; Zhang, Z.; Lian, J.; Zheng, Y. Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot. *Comput. Electron. Agric.* **2020**, *172*, 105380. [CrossRef]

52. Yang, K.; Zhong, W.; Li, F. Leaf segmentation and classification with a complicated background using deep learning. *Agronomy* **2020**, *10*, 1721. [CrossRef]

53. Tian, Y.; Yang, G.; Wang, Z.; Li, E.; Liang, Z. Instance segmentation of apple flowers using the improved mask R–CNN model. *Biosyst. Eng.* **2020**, *193*, 264–278. [CrossRef]

54. Steffenson, B. Fusarium head blight of barley: Impact, epidemics, management, and strategies for identifying and utilizing genetic resistance. In *Fusarium Head Blight of Wheat and Barley*; American Pytopathology Press: St. Paul, MN, USA, 2003.

55. Bauriegel, E.; Giebel, A.; Geyer, M.; Schmidt, U.; Herppich, W. Early detection of Fusarium infection in wheat using hyper-spectral imaging. *Comput. Electron. Agric.* **2011**, *75*, 304–312. [CrossRef]

56. Cai, L.; Long, T.; Dai, Y.; Huang, Y. Mask R-CNN based detection and segmentation for pulmonary nodule 3D visualization diagnosis. *IEEE Access* **2020**, *8*, 44400–44409. [CrossRef]

57. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

58. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In *Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016*; Springer: Cham, Switzerland, 2016; pp. 630–645.

59. Dai, J.; He, K.; Sun, J. Instance-aware semantic segmentation via multi-task network cascades. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 3150–3158.

60. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.

61. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv* **2015**, arXiv:1502.03167. Available online: https://arxiv.org/abs/1502.03167 (accessed on 21 December 2020).

62. Powers, D.M. *Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation*; School of Informatics and Engineering Flinders University: Adelaide, AUS, Australia, 2011.

63. 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, Sardinia, Italy, 13–15 May 2010; pp. 956–963.

64. Hsu, K.-J.; Tsai, C.-C.; Lin, Y.-Y.; Qian, X.; Chuang, Y.-Y. Unsupervised CNN-based co-saliency detection with graphical optimization. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 485–501.

65. Shi, J.; Chang, Y.; Xu, C.; Khan, F.; Chen, G.; Li, C. Real-time leak detection using an infrared camera and Faster R-CNN technique. *Comput. Chem. Eng.* **2020**, *135*, 106780. [CrossRef]

66. 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* **2017**, arXiv:1704.06857. Available online: https://arxiv.org/abs/1704.06857 (accessed on 21 December 2020).

67. Leplat, J.; Mangin, P.; Falchetto, L.; Heraud, C.; Gautheron, E.; Steinberg, C. Visual assessment and computer–assisted image analysis of Fusarium head blight in the field to predict mycotoxin accumulation in wheat grains. *Eur. J. Plant Pathol.* **2018**, *150*, 1065–1081. [CrossRef]

68. Zhang, D.; Wang, D.; Du, S.; Huang, L.; Zhao, H.; Liang, D.; Gu, C.; Yang, X. A rapidly diagnosis and application system of fusarium head blight based on smartphone. In Proceedings of the 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Istanbul, Turkey, 16–19 July 2019; pp. 1–5.

69. Williams, H.A.; Jones, M.H.; Nejati, M.; Seabright, M.J.; Bell, J.; Penhall, N.D.; Barnett, J.J.; Duke, M.D.; Scarfe, A.J.; Ahn, H.S. Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms. *Biosyst. Eng.* **2019**, *181*, 140–156. [CrossRef]

70. Li, D.; Wang, R.; Xie, C.; Liu, L.; Zhang, J.; Li, R.; Wang, F.; Zhou, M.; Liu, W. A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network. *Sensors* **2020**, *20*, 578. [CrossRef]

71. Yu, Y.; Zhang, K.; Yang, L.; Zhang, D. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Comput. Electron. Agric.* **2019**, *163*, 104846. [CrossRef]

72. Kiratiratanapruk, K.; Temniranrat, P.; Kitvimonrat, A.; Sinthupinyo, W.; Patarapuwadol, S. Using deep learning techniques to detect rice diseases from images of rice fields. In *Proceedings of the 33rd International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Kitakyushu, Japan, 21–24 July 2020*; Springer: Cham, Switzerland, 2020; pp. 225–237.

73. Kang, H.; Chen, C. Fast implementation of real-time fruit detection in apple orchards using deep learning. *Comput. Electron. Agric.* **2020**, *168*, 105108. [CrossRef]