

Article

A Weakly Supervised Approach for Disease Segmentation of Maize Northern Leaf Blight from UAV Images

Shuo Chen ¹ , Kefei Zhang ^{1,2,*} , Suqin Wu ¹ , Ziqian Tang ¹ , Yindi Zhao ¹ , Yaqin Sun ¹
and Zhongchao Shi ³

¹ School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China

² Satellite Positioning for Atmosphere, Climate and Environment (SPACE) Research Center, School of Science (SSCI), RMIT University, Melbourne, VIC 3001, Australia

³ Department of Restoration Ecology and Built Environment, Faculty of Environmental Studies, Tokyo City University, Kanagawa 224-8551, Japan

* Correspondence: profkzhang@cumt.edu.cn

Abstract: The segmentation of crop disease zones is an important task of image processing since the knowledge of the growth status of crops is critical for agricultural management. Nowadays, images taken by unmanned aerial vehicles (UAVs) have been widely used in the segmentation of crop diseases, and almost all current studies use the study paradigm of full supervision, which needs a large amount of manually labelled data. In this study, a weakly supervised method for disease segmentation of UAV images is proposed. In this method, auxiliary branch block (ABB) and feature reuse module (FRM) were developed. The method was tested using UAV images of maize northern leaf blight (NLB) based on image-level labels only, i.e., only the information as to whether NBL occurs is given. The quality (intersection over union (IoU) values) of the pseudo-labels in the validation dataset achieved 43% and the F1 score reached 58%. In addition, the new method took 0.08 s to generate one pseudo-label, which is highly efficient in generating pseudo-labels. When pseudo-labels from the train dataset were used in the training of segmentation models, the IoU values of disease in the test dataset reached 50%. These accuracies outperformed the benchmarks of the ACoL (45.5%), RCA (36.5%), and MDC (34.0%) models. The segmented NLB zones from the proposed method were more complete and the boundaries were more clear. The effectiveness of ABB and FRM was also explored. This study is the first time supervised segmentation of UAV images of maize NLB using only image-level data was applied, and the above test results confirm the effectiveness of the proposed method.

Keywords: weakly supervised segmentation; maize northern leaf blight; disease segmentation; unmanned aerial vehicle; semantic segmentation



Citation: Chen, S.; Zhang, K.; Wu, S.; Tang, Z.; Zhao, Y.; Sun, Y.; Shi, Z. A Weakly Supervised Approach for Disease Segmentation of Maize Northern Leaf Blight from UAV Images. *Drones* **2023**, *7*, 173.

<https://doi.org/10.3390/drones7030173>

Academic Editors: Görres Grenzdörffer and Emanuel Peres

Received: 31 January 2023

Revised: 16 February 2023

Accepted: 1 March 2023

Published: 3 March 2023



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

1. Introduction

The segmentation of plant disease plays an important role in obtaining the information regarding as well as further control of the disease. Images taken by unmanned aerial vehicles (UAVs) could serve as an important data source for non-destructive and efficient disease monitoring due to their large-scale imaging capability, and the costs of purchasing and using them has gradually reduced in the recent years [1].

Maize Northern Leaf Blight (NLB) is a serious crop disease which could cause a large reduction in the yield of maize crops, as this disease is one of the most economically damaging maize diseases [2]. Diseased areas caused by maize NLB are usually large; therefore, they can be seen in the images taken by UAVs. In other words, using images from UAVs compared with traditional manual inspection in the field is an efficient way to monitor maize NLB. Typical plant disease extraction methods are fully supervised segmentation methods, which require a lot of manually labeled pixel-wise data. Specifically, pixel-wise

labeled data are from manually depicting the edges of the disease zone in the label software. Moreover, it should be noted that labeling for crop disease in one image needs a large amount of time due to the dispersion and uneven sizes of crop-diseased zones.

To solve the problem of requiring a large amount of labelled sample data for fully supervised disease segmentation, a weakly supervised segmentation experiment was carried out in this study using only image-label data, and a weakly supervised segmentation method was proposed for NLB segmentation. Specifically, only the image-level labels, i.e., whether the NLB occurred, were used in this study. Therefore, the potential of the segmentation of plant disease using only image-level labels was explored in this study. To the best of our knowledge, this is also the first time a weakly supervised paradigm for plant disease segmentation using UAV images was applied.

In this study, the whole process was divided into two stages: pseudo-labels extraction and the training of the disease segmentation model using pseudo-labels. An overall description is given in Figure 1. In the end, the accuracies of the generated pseudo-labels and segmented diseased areas from the proposed method were better than those of the compared benchmark of weakly supervised segmentation models.

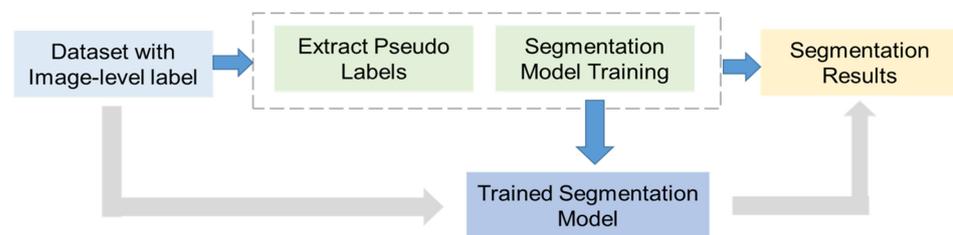


Figure 1. Overall description of the proposed method.

The main contributions of this study include (1) the first time the potential of applying weakly supervised segmentation to crop disease segmentation using UAV images was investigated. (2) The proposal of a weakly supervised method suitable for maize NLB using UAV remote sensing images, which method only relies on image-level label. High quality pseudo-labels could be generated through classification activation maps (CAMs), and the IoU of the pseudo-labels on the validation dataset achieved 43%. (3) The validated effectiveness of the proposed method in terms of the accuracies of generated pseudo-labels and segmentation results. The time for generating a pseudo-label was 0.08 s, and the segmentation accuracy of the proposed method achieved 50% without using any pixel-level ground truths.

Subsequently, related works on plant disease segmentation and weakly semantic segmentation are given in Section 1.1. Section 2 illustrates the details of the proposed method and specific algorithms, which include the proposed auxiliary branch block (ABB) and feature reuse module (FRM) module. The experimental results and analysis of the NLB dataset are listed in Section 3. Section 4 discusses the improvements brought about by the proposed modules and the limitations of the proposed method. Finally, a conclusion is given in Section 5.

1.1. Related Work

1.1.1. Plant Disease Segmentation

During the early stage of plant disease segmentation studies, researchers used manually designed features templates to extract useful information from original images; subsequently, the extracted features were classified by various classifiers. Tetila et al. [3] used the simple linear iterative cluster (SLIC) method to segment the soybean disease in UAV images. Yue et al. [4] proposed a fast disease monitoring method using UAV images through the SIFT algorithm and object-oriented information processing to extract the stress of disease and pest, and the map of plant-stress levels as such were obtained. Yang et al. [5] classified the tea disease from thermal UAV images, and the correlation score reached

0.97 when compared with manual observations. However, it takes a large amount of time and expert knowledge to manually design feature templates, and they are limited for application scenarios.

Tractor-based systems have also played an increasingly important role in plant stress detection. Some of them are already in use, for example, the See & Spray Ultimate tractor-based system detected weeds in a field to reduce weed stress [6]. To detect specific symptoms of crop diseases with sub-millimeter ground resolution, Dammer et al. [7] adopted near-surface images from a tractor combined with UAV images to detect yellow stripe rust in winter wheat. Gillis et al. [8] used a tractor equipped with global positioning system (GPS) to control *Prunus* replant disease in the areas where fumigants are most needed. Gnyp et al. [9] tested the tractor and UAV sensors to determine the nitrogen status of crops through the results of spectral reflectance measurement.

Many studies conducted disease segmentation from the aspects of using vegetation indices (VIs) drawn from multispectral and hyperspectral images [10–12]. Su et al. [13] performed a temporal–spatial monitoring for stripe rust of winter wheat at the field scale and presented the results of the sensitive degree of different VIs in the different growth periods of crops regarding the disease monitoring. Since there are a large number of bands in hyperspectral images, more VIs can be extracted as features for disease segmentation. For example, Bohnenkamp et al. [14] selected sensitive feature bands from hyperspectral UAV images to identify winter wheat yellow rust using three classification algorithms. In addition, the information transfer probability between the ground level and the UAV level was discussed. The acquisition of hyperspectral images usually involves significant efforts and a complicated process; thus, it is difficult to widely use hyperspectral images at present.

With the increased interest in deep learning, more and more studies have investigated its potential for disease segmentation due to the advantages of end-to-end learning, strong feature-extracting capabilities, and the lack of a need to manually design feature templates, such as using the backpropagation (BP) method to classify wheat *Fusarium* head blight [15] and applying an artificial neural network to evaluate remote sensing technologies for segmenting *Ganoderma*-infected oil palms [16]. A deep convolutional neural network (CNN) was used to detect radish wilt disease from images captured by a low-altitude UAV [17]. With the development of semantic segmentation, many scholars have applied various semantic segmentation networks (e.g., FCN, UNet, and SegNet) to the study of disease segmentation. For example, Kerkech et al. [18] proposed a UAV image registration method to register visible and infrared UAV images, and used SegNet to segment the registered images into four types: shadow, ground, healthy and diseased zones. In another example, FCN was applied to segment six types of corn leaf disease, each type of which contained 125 pictures, and the final mean accuracy was 96% [19]. PSPNet was used to realize the segmentation of wheat yellow rust, bare land, and healthy wheat, and the recognition accuracy reached 98% [20]. ResNet-34 was used as the classification network to determine the occurrence of maize NLB, and the sliding window method was used to determine the pixel-wise segmentation [21]. To improve the performance of deep learning, scholars have studied the methods of weight initialization for deep learning models [22]. For example, Wadii et al. [23] studied four weight initialization methods for deep learning models in remote sensing. Liu et al. proposed an improved weight initialization method termed GLIT for asymmetric activation function neural networks. Formal methods could provide strict assurance of correctness for hardware and software systems [24]. As machine learning and deep learning become more popular, the application of formal methods to validate systems that incorporate machine learning has recently been considered [25]. Many studies have considered verifiable AI from the perspective of formal methods [26]. For example, Katz et al. [27] proposed a Reluplex method extending the processing of non-convex ReLU activation function to verify the characteristics of deep neural networks.

According to the scene of agricultural disease segmentation, the typical semantic networks have been improved from the aspects of architecture design and attention mecha-

nism fusion. For instance, Zhang et al. [28] proposed Ir-Unet based on the improvements of UNet to monitor winter wheat. Additionally, a spatial content attention network was proposed, which was composed of spatial-information-retained modules preserving more spatial information and content information modules expanding the receptive field to segment pine wood nematode disease, and the F1 score of the experiment achieved 0.88 [29]. However, the design of the disease segmentation network equipped with the attention mechanism relies on the analyses of the scene with a certain amount of expert knowledge. From the perspective of sample augmentations, Hu et al. [30] used the conditional generative adversarial network to generate more samples whose leaves were segmented by SVM method, and finally used the VGG-16 network to classify three types of diseases according to the generated samples. As for the implementation of lightweight models, Huang et al. [31] proposed a network based on YOLACT++ integrating attention modules to segment corn NLB, and the accuracy of the segmentation reached 84.91%.

Object detection technique has also been used in disease detection, such as using YoLoV3 and Faster R-CNN with different backbone networks to detect the occurrence of pine nematode disease [32] as well as the Mask RCNN model with the ResNet101 to segment the UAV images of corn NLB [33]. Although accurate pixel-level information could not be provided by YoLoV3 and the other object detection model, re-segmenting according to the detected zone were in demand for pixel-level segmentation results. Zhou et al. [34] proposed an adaptive active location method for *Camellia oleifera* fruit based on the YOLOv7 model. Tang et al. [35] developed the YOLO-Oleifera model based on the YOLOv4-tiny model to find camellia fruits in the complex orchard environment, and the boundary boxes generated by the YOLO-Oleifera model were used to extract the regions-of-interest-containing fruits. In summary, Table 1 lists recent studies on plant disease segmentation using UAV images.

Table 1. Summary of plant disease segmentation with UAV imagery in the last few years.

Ref.	Crop	Disease	Data Type	Method	Accuracy
[3]	Soybean	Foliar diseases	RGB UAV images	SLIC	98.3% for height between 1 and 2 m
[5]	Tea	Diseased yellow leaves	Infrared thermal canopy UAV Images	HSV transform and thresholding	R2 of 0.97
[10]	Potato	Vascular wilt	Multispectral UAV images	Generalized linear model and supervised random forest classification	Classification accuracy of 73.5–82.5% in Plot 1
[11]	Banana	Fusarium wilt	Multispectral images	Binary logistic regression	More than 80%
[12]	Pear	Fire Blight	Aerial multispectral imagery of trees crown	Support vector machine	Classification accuracy of 95%
[15]	Winter wheat	Fusarium head blight	Hyperspectral UAV image	BP neural network	Accuracy of 98%
[16]	Oil palms	Basal stem rot	Infrared UAV images	Artificial neural network	Classification accuracy of 97.5%
[17]	Radish	Fusarium	RGB UAV images	CNN	Accuracy of 93.3%
[20]	Winter wheat	Yellow rust disease	RGB UAV images	PSPNet	Accuracy of 98%
[21]	Maize	NLB	RGB UAV images	CNN	Accuracy of 95.1%
[28]	Winter wheat	Yellow rust disease	Red-Edge multispectral UAV images	Ir-Net	Accuracy of 97.1%
[30]	Tea	Tea red scab, red leaf spot, and leaf blight	Hand-held digital camera images and UAV images	Generative adversarial networks	Accuracy of 90%
[33]	Maize	NLB	RGB UAV images	Mask R-CNN model	IoU of 0.73

1.1.2. Weakly Supervised Segmentation

Even though satisfactory accuracies of disease segmentation have been achieved at present, there are still many difficulties in the acquisition of training samples and pixel-wise labeling. Therefore, achieving reliable segmentation accuracy with only a small number of samples has recently received extensive attention and become a hot topic for research. Nowadays, weakly supervised semantic segmentation has made great progress,

from the CAM method [36] to the recently salient regions erase [37,38], receptive field expansion [39,40], and prototype estimation [41], etc.

There are limited studies on weakly supervised segmentation of agricultural disease to date. Yi et al. [42] applied the weakly supervised training strategy to the encoder–decoder semantic segmentation network to obtain the seed region, and used the seed region to train the segmentation network. Although the PlantVillage dataset was used in their study, those images were captured in a laboratory, and the background was relatively single. Kim et al. [43] used weakly supervised segmentation to obtain a planting area and applied it to the automatic driving of a combine harvester, whose data were collected by a camera installed inside the cabin of the combine harvester. These data were the results of near-ground measurements, and are different from the UAV data used in this study.

2. Materials and Methods

2.1. Materials

The drone dataset used in the study [44] was selected and the data captured from an experimental cornfield located at Cornell University’s Musgrave Research Farm in Aurora, NY, in the summer of 2017. As described in [44], the trials consisted of maize hybrids from Te Genomes to Fields Initiative, arranged in two-row plots with a length of 5.64 m and inter-row spacing of 0.76 m, and the trials were rainfed and managed with conventional maize cultivation practices. The maize was inoculated at the V5–V6 stage with both a liquid suspension of *Setosphaeria turcica* (isolate NY001) spores and sorghum grains colonized by the fungus. The UAV images in the dataset were collected by a DJI Matrice 600 UAV with an altitude of 6 m and a velocity of 1 m/s, and they were captured with a nadir view every 2 s. The dataset is the largest corn NLB drone image dataset at present.

The diseased zones and healthy zones in the leaves from the UAV image were cut out in a range from 448×448 to 1568×1568 , and all the samples were uniformly resized to the size of 448×448 pixels. The numbers of samples in the training dataset, validation dataset, and test dataset were 1700, 500, and 300, respectively. The examples of the samples in the experiment are given in Figure 2. The training dataset contained only 850 healthy and 850 diseased samples, and the training set was used to train the network for pseudo-labels generation and a disease segmentation network using diseased samples and their corresponding pseudo-labels. The validation dataset included 250 healthy samples, 250 diseased samples, and ground truths corresponding to diseased samples. The ground truths of diseased samples were used to measure the accuracy of pseudo-labels. The test dataset contained 300 diseased samples and ground truths of diseased samples, which were used to measure the accuracy of the disease segmentation model. As shown in Figure 1, the diseased areas were red, and they were labeled by the open-source LabelMe software [45].

2.2. Methods

The whole process was divided into two parts: pseudo-label extraction and segmentation model training. The overall architecture is given in Figure 3. The capability of feature learning in the proposed method was enhanced by adding auxiliary branches and feature reusing to improve the accuracy of weakly supervised segmentation. More details of the proposed method will be detailed below.

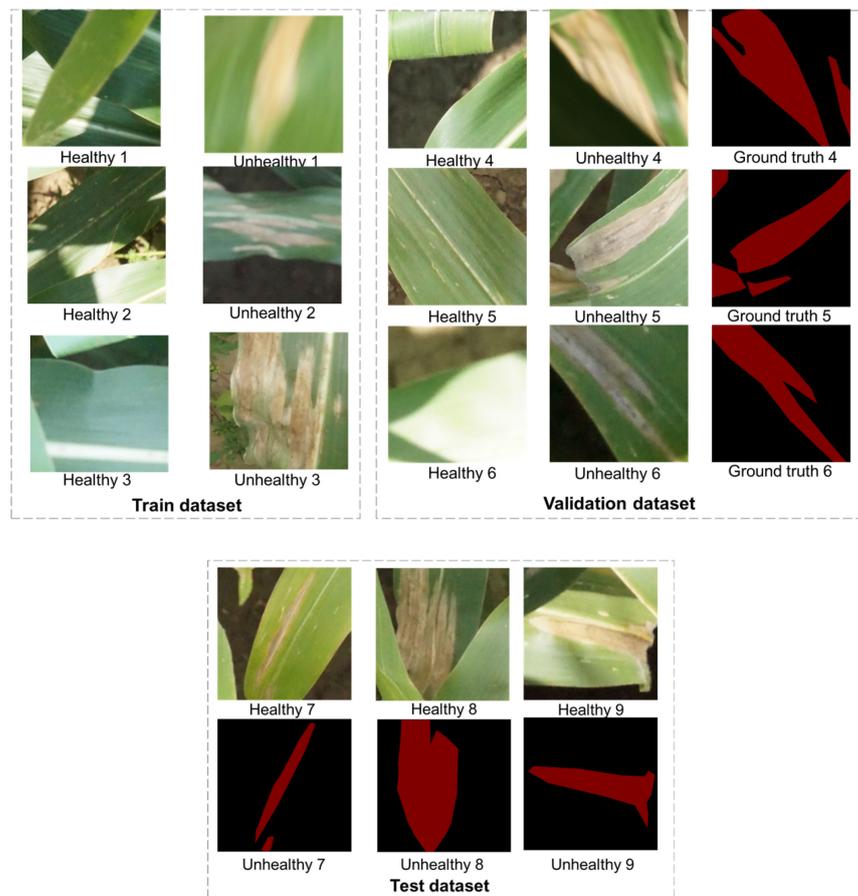


Figure 2. Examples of samples in the train, validation, and test dataset. The red area represents the corn blight area shown in the middle row. The black area refers to the background in the images, containing leaves and soil, etc.

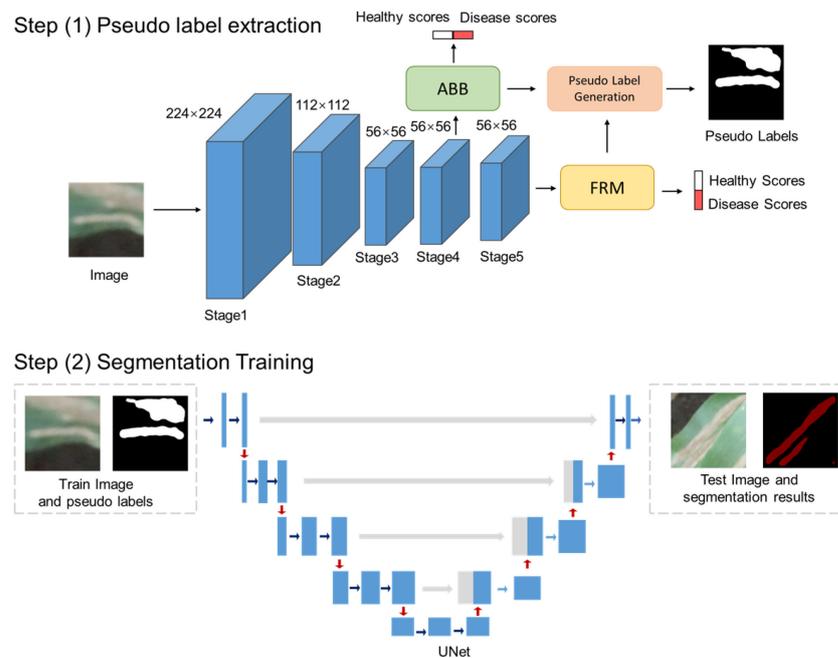


Figure 3. Architecture of the proposed method.

2.2.1. Pseudo-Label Extraction

The VGG-16 classification network [46] was utilized in this extraction process. The network was divided into five stages, and the size of the feature maps became 1/2 of its previous stage after the completion of each stage. To expand the scope of the discriminate areas, this study removed the maximum pooling operation at the last position of the 4th and 5th stages to increase the size of the feature maps. As a result, the size of the final feature maps was 56×56 . The feature dimensions of the 1st stage feature, 2nd stage feature, 3rd stage feature F_{s3} , the 4th stage feature F_{s4} , and the 5th stage feature F_{s5} were 64, 128, 256, 512, and 512, respectively. In order to improve the ability of discriminating the healthy and unhealthy leaves, feature reuse module and auxiliary branch module were proposed in the proposed method. When the network was trained, the pseudo-labels were extracted from the CAMs.

- Auxiliary Branch Block (ABB)

The scale of disease areas in UAV images varies greatly, and it is not enough to depend on high-level features to generate pseudo-labels. Features of the early feature extraction stage are abundant in spatial information, which has a positive effect on the extraction of the small disease areas. Therefore, an auxiliary branch block was constructed for F_{s4} to obtain auxiliary classification scores S_{s4} , and its computation process was given in Equation (1). The processes were to use a grouped 3×3 convolution φ_1 and an ordinary 1×1 convolution φ_2 , which served as the classifier here. The output channels and group number of φ_1 were set to 512 and 512, respectively, due to the efficient properties of group convolution. The output channel number of φ_2 was the same as the number of categories, i.e., 2. The feature maps, processed by φ_1 and φ_2 , were averagely pooled to obtain the S_{s4} . The specific structure of the auxiliary branch block is given in Figure 4.

$$S_{s4} = AvgPool(\varphi_2(\varphi_1(F_{s4}))) \tag{1}$$

φ_1 is a 3×3 group convolution, and φ_2 is a 1×1 convolution, which served as the classifier here.

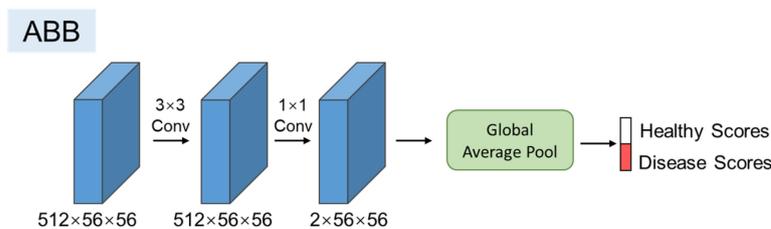


Figure 4. Architecture of the proposed ABB.

- Feature Reuse Module (FRM)

Convolution operations have the characteristics of local perception and weight sharing, and the size of the convolution kernel is much smaller than the size of the image. Therefore, when a small number of convolutions are carried out, the local spatial information from the image is still preserved, and the semantic features of the whole image are relatively few. In other words, low-level features have rich spatial information, while their semantic information is insufficient. High-level features have sufficient semantic information but lack spatial information related to boundary description [47]. Both types of features are complementary. To improve the capability of feature learning of structural and semantic information from UAV images, FRM was proposed. The abovementioned F_{s3} , F_{s4} , and F_{s5} were concatenated along the channel dimensions, and the number of the channels became 1280 for the fused features F_{fusion} , as given in Equation (2). F_{fusion} was enhanced by applying two 3×3 ordinary convolution operations φ_1 and φ_2 , and the out channel number of φ_1 and φ_2 were 512 and 256, respectively. The enhanced features of F_{fusion} were named F'_{fusion} , which can be computed in Equation (3). Another 1×1 convolutional

operation, φ_3 , was used as a classifier to transform the channel number of F'_{fusion} into 2, which is equal to the category number in the study, and the feature maps of the main branch, termed as F_{main} , were obtained. Equation (4) represents the process of calculating the F_{main} . The main classification score S_{main} was calculated through a global average pooling operation on F_{main} . Equation (5) describes the generation of S_{main} , and the architecture diagram of feature reuse module is illustrated in Figure 5.

$$F_{fusion} = F_{s3} \oplus F_{s4} \oplus F_{s5} \tag{2}$$

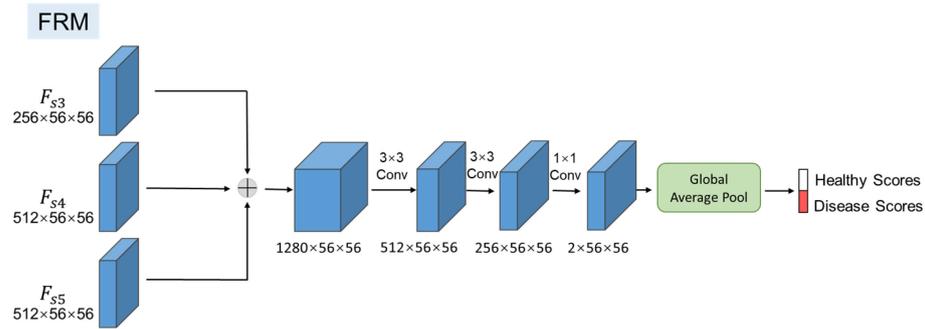


Figure 5. Structure of the FRM.

\oplus refers to the concatenate operation along with channels of features.

$$F'_{fusion} = (\varphi_2(\varphi_1(F_{fusion}))) \tag{3}$$

$$F_{main} = \varphi_3(F'_{fusion}) \tag{4}$$

$$S_{main} = AvgPool(F_{main}) \tag{5}$$

- Loss Function for Pseudo-Label Extraction

The cross-entropy loss function was selected as the specific loss function for the classification in the proposed method, and it is given in Equation (6). The term y is 1 if the classification result of an image that belongs to the true category n ; otherwise, it was 0. K is the total number of categories, which were two in this study. The \hat{y} are the predicted scores in each branch, i.e., S_{main} and S_{s4} .

$$Loss = -\frac{1}{K} \sum_{n=1}^K [y_n \log \hat{y}] \tag{6}$$

A multi-task learning approach was used to construct the total loss function, which consists of two parts (as shown in Equation (7) below). In Equation (7), L_{main} and L_{s4} were the loss values for the main classification branch, and for the ABB, respectively. They were computed using Equation (6). The weight of λ was empirically set to 0.5.

$$L_{total} = L_{main} + \lambda L_{s4} \tag{7}$$

- Pseudo-Label Generation

Referred to [36], the weight values of classifiers in each proposed module, i.e., φ_3 in FRM and φ_2 in ABB, were used to generate CAMs, respectively. The computation equation of CAM was given in Equation (8).

$$CAM = \sum_k w_k^c f_k(x, y) \tag{8}$$

The term w_k^c was the weight corresponding to class c for channel k in the classifier; f_k was the feature map for channel k ; x and y were the row and column position in f_k .

For each classification branch, Equation (9) was used to generate the normalized CAMs.

$$CAM_{norm}^{c,i} = \frac{CAM_{value}^{c,i} - CAM_{min}^c}{CAM_{max}^c - CAM_{min}^c} \quad (9)$$

where $CAM_{norm}^{c,i}$ is the normalized CAM for class c in i position; $CAM_{value}^{c,i}$ is the original CAM value in the i th pixel of the CAM for class c ; CAM_{min}^c and CAM_{max}^c are the minimum and maximum values in CAM for class c , respectively, in each channel of CAMs, where the total number of channels is two for background and diseased zones, respectively.

Since two CAMs were generated in this study, the final CAMs were the normalized summary of the two CAMs, and the formula is the following:

$$CAM_{final} = CAM_{norm}(CAM_{main} + CAM_{stage4}) \quad (10)$$

where CAM_{stage4} and CAM_{main} are the normalized CAMs from the main classification branch and ABB branch.

The pseudo-labels were generated under the condition that the threshold value was set to CAM_{final} . Specifically, if the pixel value of CAM_{norm} in the disease channel is larger than the threshold and the pixel in the background channel is not larger than the threshold, the pixel will be assigned to the disease label; otherwise, the pixel will be regarded as the background label, as shown in Equation (11).

$$\text{If } CAM_{final}^{1,i} > T \text{ and } CAM_{final}^{0,i} \leq T, \text{ Label}_p^i = 1, \text{ otherwise } \text{Label}_p^i = 0 \quad (11)$$

where $CAM_{final}^{c,i}$ is the final CAM for class c at position i , specifically; $CAM_{final}^{1,i}$ is the disease CAM value at position i ; Label_p^i is the pseudo-label for at position i ; and T is the threshold value to confirm pseudo-labels, where it is set to 0.5.

2.2.2. Segmentation Model

UNet [48] was selected as the segmentation network. The reason for utilizing UNet is based on two factors: being is the disease zones are dispersed and their size varies greatly, including both small and large disease spots; the other being the effective structure of UNet. With their encoder and decoder structure and skip connection, features at all levels could be fused by the UNet; thus, good segmentation performance and generalization performance can be achieved by the UNet, which has been widely applied in various types of semantic segmentation tasks such as automatic driving, land cover classification, and disease segmentation. To compare the segmentation accuracies of using pseudo-labels generated by different benchmark networks, generated pseudo-labels served as the true label along with their corresponding original images from the test dataset used to train the segmentation network. Afterwards, the segmented diseased images from the UNet could be acquired. The cross-entropy loss function was used for the loss function of the segmentation model, as shown in Equation (6). It should be note that all pixels were involved in the segmentation loss function. The loss values were obtained from the predicted pixel-wise results corresponding to the manual labeled zones.

2.2.3. Evaluation Indicator

Precision, accuracy, F1 score, and the intersection over union (IoU) value were used to reflect the segmentation quality and measure the segmentation results, whose equations were given in Equations (12)–(15), respectively.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

$$F1 = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (14)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (15)$$

True positives (TPs), false positives (FPs), and false negatives (FN) were used to express the precision, accuracy, and F1-score, respectively. TP represents the number of positive samples for which the prediction results are positive; FP is the number of negative samples wrongly classified as positive; TN is the number of negative samples accurately classified; and FN denotes the number of positive samples erroneously classified as negative.

2.2.4. Three Comparison Methods

The three benchmark models, i.e., ACoL [38], MDC [39], and RCA [41], were selected as weakly supervised segmentation networks for comparisons with the model proposed in this study. Each of the three weakly supervised segmentation methods have shown excellent performance in the publicly available benchmark nature image datasets, i.e., PASCAL VOC dataset and CoCo2014 dataset.

ACoL, as a representative in weakly supervised segmentation methods, uses the anti-erasure method. It has two classification branches to process the features extracted from the backbone networks for generating CAMs. As shown in Figure 6, features extracted by the VGG-16 network were put into two classification branches, respectively. For the branch where classifier A was located, the regions with high activation value of category were filtered. The features of another branch were erased using the abovementioned regions, and the erased features were classified by classifier B to obtain the classification score. The two branches optimized the weights of the network together. The final CAMs were the combination of the CAMs from A and B, respectively.

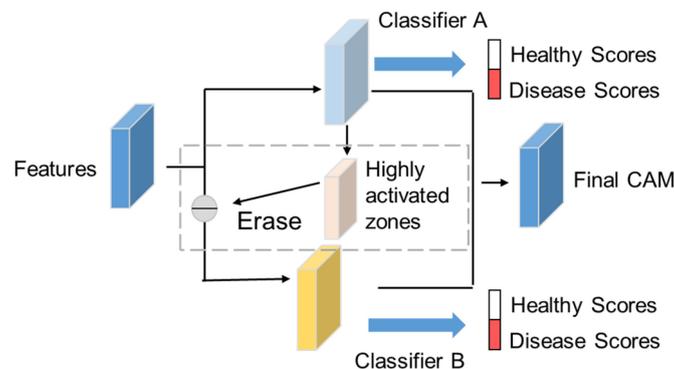


Figure 6. Main diagram of ACoL network.

MDC employed atrous convolutions to expand the receptive field of the classification branch network. The features extracted from the VGG-16 backbone were sent into four 3×3 convolutions with different sizes of dilation rate, i.e., 1, 3, 6, and 9, respectively. After the classification scores of the four branches were computed, the parameters of the MDC model were optimized. Four CAMs are obtained from each branch, and the calculation formula of the final CAM is shown in Figure 7, which is the sum of CAM from convolution with dilation rate 1 and the mean value of other CAMs from other branches.

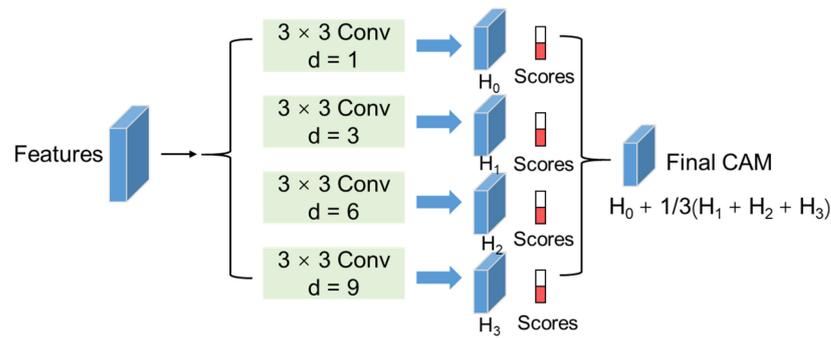


Figure 7. Structure of MDC network.

RCA uses the prototype estimation method to generate the prototypes of each category. It stores the estimated prototypes of each picture in each class queue in the memory bank. The whole process is as follows: with the features extracted by the VGG-16 network, the rough CAMs were obtained through convolutions. The areas of high value in the corresponding label from rough CAMs were regarded as rough pseudo-labels, and the rough pseudo-labels were used to averagely pool the features to obtain the feature vectors representing disease and health. To improve the distinguishing ability, the feature vectors of each category were enhanced after the operators were mixed up, and then they were stored in the memory bank. The compressed features, standing for the class prototypes, were obtained through K-means clustering of the feature vectors. In the final step, the refined CAMs and the classification score were computed by enhancing the semantics of the class prototypes and the input features of the network. Figure 8 gives the simplified architecture of the RCA model.

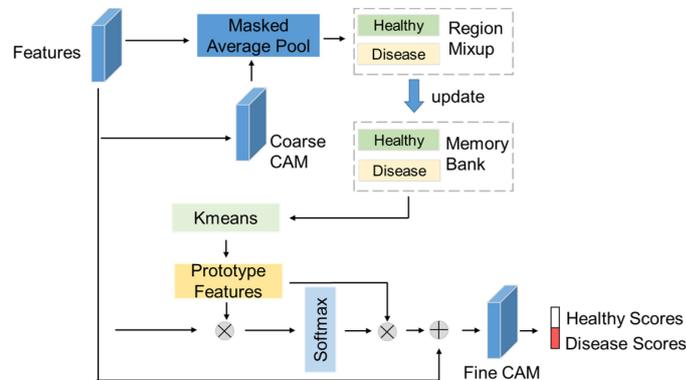


Figure 8. Main architecture of RCA network.

2.2.5. Implementation Details

All models were implemented using the open-source deep learning framework PyTorch. The VGG-16 network was used as the backbone network in the label extraction process of the proposed and comparison networks, and the pre-trained weight from ImageNet was used. The SGD and Adam optimizer were selected as the optimizer for the classification and segmentation models, respectively. The poly learning rate was used as the learning rate adjustment method, as shown in the equation below.

$$learning\ rate = base_{lr} \cdot 1 - \left(\frac{iteration}{num_iterations} \right)^{0.9} \quad (16)$$

where $base_{lr}$ is the initial learning rate of the classification model, with a value set to 0.0001; $iteration$ is the current number of iterations during the training process; and $num_iterations$ is the total number of iterations in the training process. In addition, a RTX 2080 graphic processing unit (GPU) was used in the study.

For the process of pseudo-label extraction, the number of training epochs was 20, and the batch size was set to 8, while the batch size of 4 and the training epoch of 30 were assigned to the segmentation model due to the limited capacity of the GPU. Moreover, random horizon flip, random vertical flip, and random rotation 90 degrees were used as the sample augmentation method in the study.

3. Results

3.1. Pseudo-Labels Quality Comparison

Good segmentation results depend on pseudo-labels that are of a good quality. To measure the quality of pseudo-labels generated by different methods, the degrees of fitness between pseudo-labels obtained by different methods and ground-truth labels in the validation dataset are illustrated in Table 2. It can be seen in Table 2 that the IoU of disease from the proposed model is 1.5%, 10.1%, and 14.3% higher than those extracted by ACoL, RCA, and MDC, respectively. The precision and F1 of the proposed model were the highest among the four models. The precision of the proposed model was 19.3% higher than that of the recall. The precision and recall of MDC were close; however, their overall values were low. Although the recall of RCA, which reached 60%, was the highest among the four models, the precision was the lowest among the four models, indicating that the predicted disease areas were too large, i.e., large amounts of false alarms occurred. The F1-score and IoU of pseudo-labels from the ACoL model reached 58.7% and 41.6%, respectively, and both of them ranked second out of the four models. Therefore, it can be said that in the comparison of the quality of pseudo-labels in the validation dataset, the proposed model had the best effect, and it was robust even in the data not involved in training.

Table 2. Accuracies of pseudo-labels extracted from different models.

Model	Pseudo-Label IoU(%)	Precision	Recall	F1-Score
MDC	28.8	48.5	41.6	44.8
ACoL	41.6	64.7	53.8	58.7
RCA	33.0	42.3	60.0	49.6
Proposed Model	43.1	71.4	52.1	60.2

Figure 9 shows examples of pseudo-labels generated by different models. It can be observed in Figure 9 that multiple types of disease occurrence were included. The presence of a single lesion ((a) and (b)), multiple disease areas (c), as well as obvious illumination interference (d) increases the extraction difficulty of the disease. However, the results of Figure 9 shows that the pseudo-labels extracted by the proposed model were close to their corresponding ground-truth labels, even the only image label given. Moreover, the boundaries of the disease from the proposed method were more clear and more accurate compared with other methods. As for the performance of RCA, although a high accuracy in natural image processing was achieved, it has a complex network structure and may not suit the disease segmentation scenario in UAV images well. The accuracy of the pseudo-label of ACoL is the second in all the models, indicating that the salient erasure and random hide played roles in the performance improvements in the methods. According to the defined threshold, most salient parts of the feature maps were masked by the salient erasure, and then the erased feature maps were put into another classifier in sequence. However, there might be some mistakes in the generation of the pseudo-labels due to the same threshold being used for all data, which might be unreasonable. The reason for the poorest performance of MDC is likely that the classification results of the dilated convolution were added together, which acts as an inhibitor. For the characteristic of the relatively complex background of UAV remote sensing images, simply expanding the receptive field may have limited effects. It will inevitably introduce noise such as soil background and leaves with similar colors to disease, which hinder the quality of pseudo-labels and further segmentation results. The reason for the passable accuracy is likely to be related to the fact that the features of agricultural remote sensing images used

to generate the prototype might be confusing, namely there being some noise, e.g., a few yellow leaves and bare soil potentially being involved. Therefore, the quality of the entry queues for the generation of the prototype needs to be inspected. In general, the proposed model had a simple structure and was easy to implement, and the comprehensive use of different levels of features improves the accuracy of the pseudo-label.

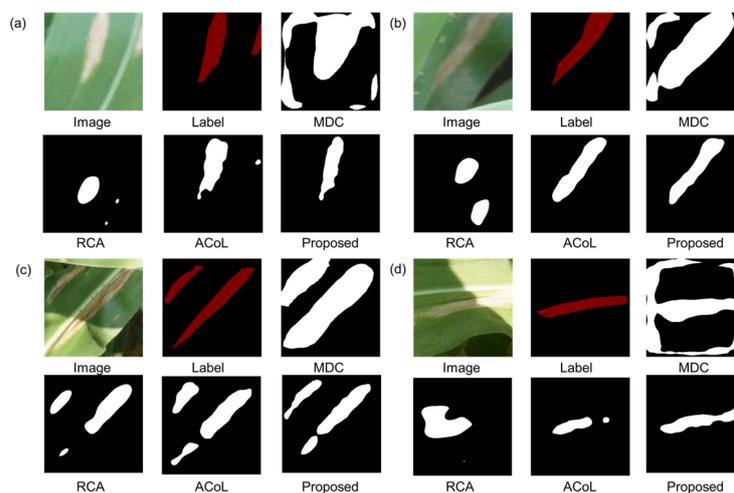


Figure 9. Examples of generated pseudo-labels from four methods. (a–d) are four group results for images, labels and generated pseudo labels. Black and white represent background and diseased zones, respectively. The red zones in the labels refer to manually labeled diseased zones. (a).

3.2. Pseudo-Labels' Generation Time

The time on the generation of pseudo-labels was another indicator used to measure the performance of the aforementioned five models. The trained classification model was loaded (one image was loaded at a time) to obtain the total time. The time on generating a pseudo-label was the average of all the durations of the pseudo-labels from the 250 samples contained in the validation dataset. Table 3 lists the results of the four models. The generation time from the proposed model was only 0.06 s longer than that of the shortest time-consuming model ACoL, but the accuracy of pseudo-labels was nearly 1.5% higher than that of ACoL.

Table 3. Average time on generating a pseudo-label from each of four models.

Model Name	Running Time/s
ACoL	0.02
MDC	1.06
RCA	0.36
Proposed Model	0.08

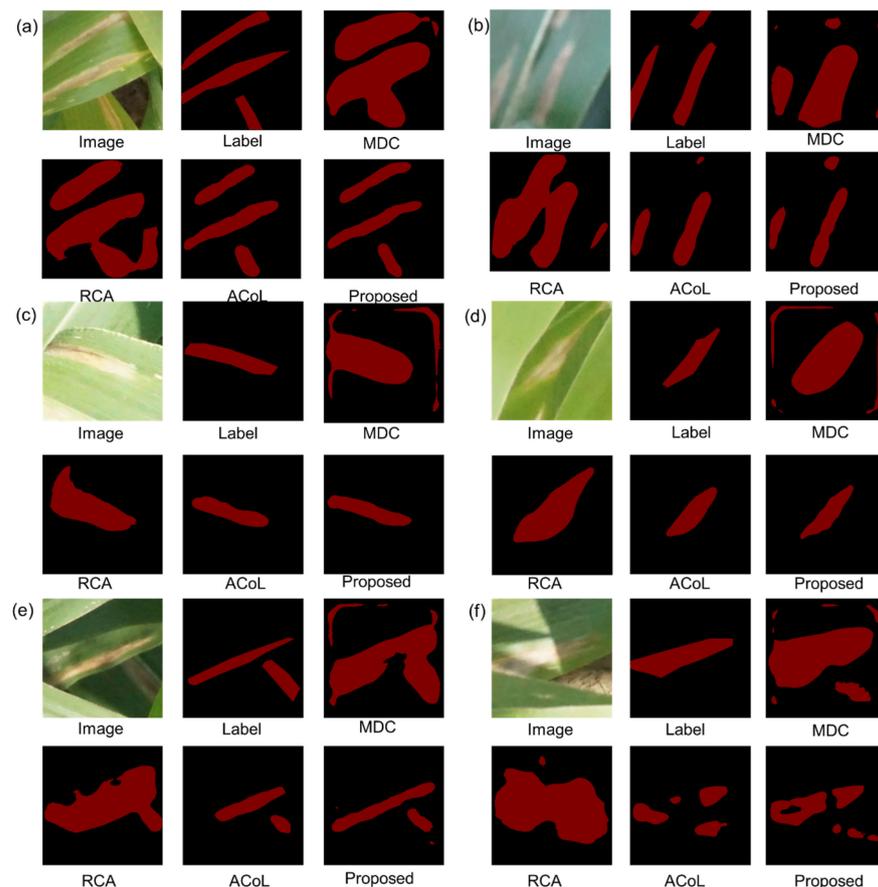
3.3. Segmentation Results

The accuracies of segmentation model, trained by pseudo-labels extracted from the training dataset, were tested by using the test dataset, and the segmentation results are given in Table 4; the proposed method achieved the optimal segmentation accuracy of disease. It can be observed that the proposed model achieved the best segmentation IoU of 50.8%, while the segmentation accuracies from the other three models ranged from 34% to 45.5%, exhibiting the superior of the proposed method. In addition, the proposed model had the highest F1 score of 67.4%, while other models had low values ranging from 50.8% to 62.5%. Therefore, it can be concluded that the segmentation results from the proposed method significantly outperformed other methods.

Table 4. Accuracies of segmentation using pseudo-labels from different models.

Model	Disease IoU(%)	Precision	Recall	F1-Score
MDC	34.0	43.8	60.4	50.8
ACoL	45.5	81.4	50.7	62.5
RCA	36.5	40.6	78.2	53.5
Proposed Model	50.8	84.5	56.1	67.4

Figure 10 is for a qualitative comparison of the segmentation results. Figure 10a,b had more than one diseased zone. Figure 10c,d shows the situation where the diseased zones were easily confused with the leaves. The leaves in Figure 10e,f were influenced by the shadow and the light, respectively. The segmented disease zones from the proposed model were more accurate than those from other methods, and the edges of diseased zones from the proposed model were sharper. These results imply that the segmentation results of the proposed model can resist noise and background interference, no matter the shape, number, and light condition of the diseased zones in an image. Good pseudo-labels result in good segmentation results. For other models, it could be seen from Figure 10 that the segmentation results of MDC and RCA models were too large because of the noise of pseudo-labels introduced in the model training. The results of ACoL were close to ground truths; however, dispersion states existed in the segmentation results (Figure 10f).

**Figure 10.** Examples of segmentation results from four methods. (a–f) are six example groups for segmentation results. Black and red represent background and diseased zones, respectively.

3.4. Ablation Studies

In this section, ablation experimental results for the improvements made by each of the two extra parts—ABB and FRM—used in the proposed model were analyzed. The accuracies of the segmentation results were used as the performance indicator. The VGG-16

network without two added blocks was utilized as the baseline to evaluate the component effectiveness of each proposed block in the model. The detailed results are shown in Table 5. From Table 5, it is indicated that the overall performance of our network has been greatly improved. The ABB increased by 11.5% in IoU, and it supported the idea of using another classification branch to achieve better weakly supervised segmentation. With FRM, the IoU increased by 30.7% compared with that of baseline, and it showed the powerful ability of multi-level feature fusion. When the two proposed modules were added, features of different levels are comprehensively used to make the information complementary, and the structural texture and semantic information of the disease were enhanced. Therefore, the proposed method achieved the optimal IoU performance of 50.9% on the test dataset.

Table 5. Accuracies of disease segmentation in ablation studies.

Baseline	ABB	FRM	IoU	Precision	Recall	F1
✓			14.2	94.8	14.3	24.9
✓	✓		25.7	89.4	26.5	40.9
✓		✓	44.9	85.3	48.7	62.0
✓	✓	✓	50.9	84.6	56.1	67.4

ABB: auxiliary branch block. FRM: feature reuse module.

4. Discussion

4.1. Effect of the Proposed Modules for Label Generation

The proposed model achieved the best results in the generation of pseudo-labels and the segmentation results of the disease, which showed the effectiveness of the proposed model. A high-performance segmentation model is the result of good initial labels. To explain why the proposed model can generate high-quality pseudo-labels, the ablation experiment models were used to display the features before the main classification in the form of heat maps. Since the channel number of visualized features is 256, the summation along the channels was carried out and the summarized feature maps were normalized for visualization. The visualization results are shown in Figure 11. The red part indicates the larger value of the normalization feature maps, which represents the region to which the feature maps paid more attention. Compared with the labels and the original maps, the proposed model has larger values in the diseased areas, and the range is more closed to their labels compared with other methods, especially (a) and (c). These results demonstrate the effectiveness of the proposed method for generating high-quality labels. The heat maps generated by the original VGG-16 were relatively scattered, and the highlighted areas occupied less diseased areas compared to those of other models. The highlighted areas of the heat map were expanded with the addition of ABB module. Furthermore, the ranges of heat maps with FRM module were more complete than those with only ABB added, indicating the importance of FRM for the generation of pseudo-labels, and they also reflected the significance of the comprehensive use of features at all levels in the segmentation of NLB zones.

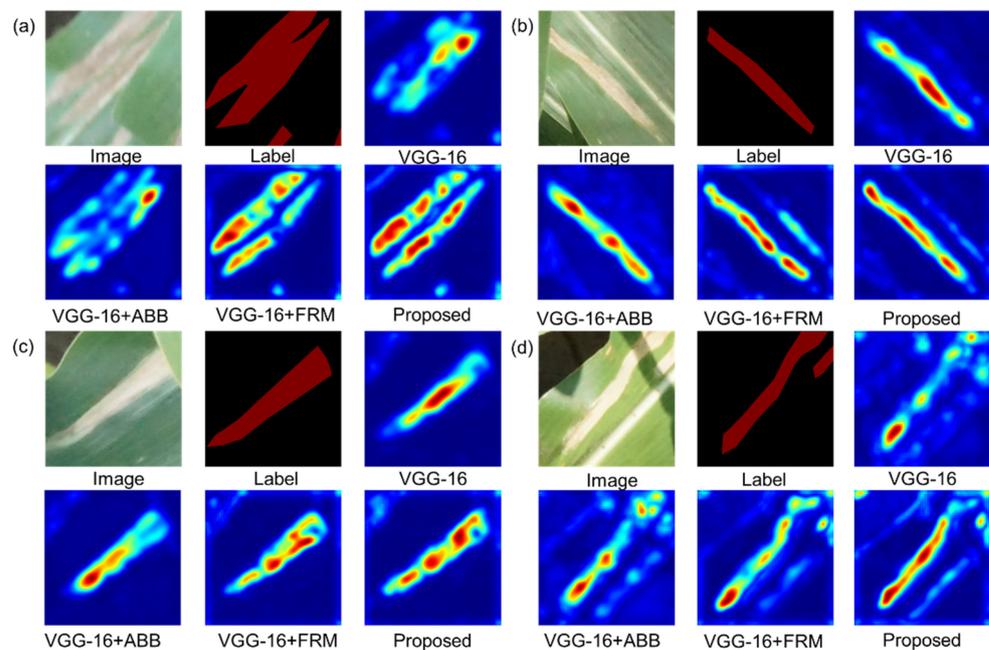


Figure 11. Heat maps from models in ablation studies. (a–d) are four example groups for heat maps generated by different combinations of proposed modules.

4.2. Analysis of Coefficient in Loss Function

The adjustment coefficient λ was added to the total loss function in the proposed method. To further analyze the function of λ in the loss function, the proposed method was trained by different λ , and the segmentation results of the disease are given in Table 6. It can be observed in Table 6 that when the λ was 0.3 or 0.7, recall in its result was high, indicating that there were a lot of false alarms in the segmentation results of the disease. The possible reason is that when the λ was 0.7, the ABB branch played a large role. Since the features extracted by the ABB branch contained more spatial information, there would be a certainty of false alarms. As for the λ at 0.3, the effect from the ABB branch and the main classification branch could not be fused well, so the high recall values occurred. Therefore, λ was set to 0.5 in the proposed method, and the two branches can be balanced well to achieve good segmentation effect.

Table 6. Segmentation results from different coefficient in loss function.

Coefficient	IoU	Precision	Recall	F1
0.3	36.6	45.2	65.9	53.6
0.5	43.1	71.4	52.1	60.2
0.7	34.3	38.1	77.4	51.1

4.3. Segmentation Results with Part of Full Labels

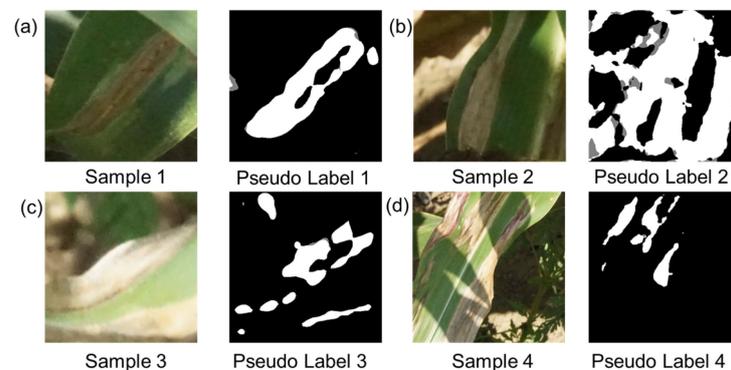
Totals of 100, 150, 200, and 400 pseudo-labels were randomly replaced with their corresponding ground truths, respectively, and the accuracies of the disease segmentation are shown in Table 7. As can be seen from Table 7, the IoU of disease spot segmentation has been improved with ground truths, especially when nearly half of the pseudo-labels were replaced with the ground truths, indicating that the quality of pseudo-labels needs to be further improved to achieve the same effect from the ground truths as supervision information. Nevertheless, the pseudo-label saves a lot of annotation costs, and the ground truth needs pixel-level manual delineation. Therefore, how to further make the pseudo-label close to the ground truth is the next consideration.

Table 7. Segmentation results with some ground truths as supervision information.

Pseudo-Label No.	Ground Truth No.	IoU	Precision	Recall	F1
750	100	52.8	86.2	57.7	69.1
700	150	55.2	86.9	60.2	71.1
650	200	56.4	87.4	61.4	72.1
450	400	64.9	87.9	71.3	78.7

4.4. Comparison with Related Studies and Room for Improvements

In related studies on the segmentation of maize NLB, various methods have been used, e.g., the data augmentation method and the crowded labeling used in [2], as well as the lightweight segmentation network designed in [31]. All of the mentioned studies require fine pixel-level labels to optimize the segmentation model. However, this study is the first attempt to investigate the potential of using weakly supervised segmentation for maize NLB from UAV images. Our study only requires image-level labels (one label for one image representing whether NLB occurs), which significantly reduces the cost of labeling. Although good results were obtained in the proposed method, the accuracy of the pseudo-label can be further improved. As shown in Figure 12, when the background of leaf was complex (Subfigure (c) and (d)), the pseudo-labels were not satisfied; therefore, generating more accurate pseudo-labels of images in a complex background is the next step. There exist some holes in pseudo-labels (Subfigure (a) and (b)), and a filling method to generate more complete pseudo-labels could be considered. Although the segmented edges from the proposed method are relative fine, the boundaries of the disease zones should be better depicted. In future work, more effort should be made to close the performance gaps between weakly supervised segmentation and fully supervised segmentation.

**Figure 12.** Examples of unsatisfied pseudo-labels. (a–d) are four groups of unsatisfied pseudo-labels.

Since the available data belonged to the single type of maize disease, the proposed method may not be suited for the segmentation of other types of diseases. Furthermore, although the proposed method could segment maize NLB, the severity degree of the disease could not be detected. Therefore, the criterion similar to the evaluation index used in [49] is worth paying attention to.

5. Conclusions

A weakly supervised segmentation method for the segmentation of maize NLB from UAV images was proposed in this study. This method can effectively overcome the problems that a large amount of manually labelled data are required in fully supervised methods. ABB and FRM were developed in the proposed method to improve the quality of pseudo-labels. The accuracy of the pseudo-labels in the validation dataset achieved 43.1%. Moreover, the time for generating a pseudo-label using the proposed method was only 0.08 s.

When the UNet was selected as the segmentation model trained by pseudo-labels, the segmentation accuracy of the NLB in the test dataset reached 50.8%. Compared with three

typically used weakly supervised segmentation methods in terms of both qualitative effects and quantitative results of disease segmentation, the proposed method was significantly better. The shapes and boundaries of the segmented disease were intact, and the IoU of the disease was nearly 5% higher than the second highest network. In addition, the effectiveness of ABB and FRM were explored in the study, and the results showed that the features of the proposed model gradually concentrate on the diseased zones with the ABB and FRM added. Therefore, more accurate pseudo-labels could be generated based on precise CAMs.

There is still room for further improvement, e.g., using other types of UAV image data (multispectral, hyperspectral, etc.), corn NLB images from different growth stages, integrating multi-source label information, and involving other weakly supervised information, e.g., bounding boxes and point annotations. These approaches will be considered in future work.

Author Contributions: Conceptualization, S.C. and K.Z.; methodology, S.C.; software, S.C.; validation, S.C., Y.S. and S.W.; formal analysis, Z.T., Y.Z. and Z.S.; investigation, Y.Z.; resources, K.Z.; data curation, S.C.; writing—original draft preparation, S.C.; writing—review and editing, K.Z. and S.W.; visualization, Z.T.; supervision, Y.S.; project administration, K.Z.; funding acquisition, K.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China (Grant No. 42274021), 2022 Jiangsu Provincial Science and Technology Initiative—Special Fund for International Science and Technology Cooperation (Grant No. BZ2022018), the Independent Innovation Project of “Double-First Class” Construction (Grant No. 2022ZZCX06). This work was supported by the Construction Program of Space-Air-Ground Well Cooperative Awareness Spatial Information Project (Grant No. B20046).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data link for data presented in this study are available on request from the corresponding author.

Acknowledgments: We would also thank Wiesner-Hanks et al. for providing the source data used in our study.

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

Abbreviations

ABB	auxiliary branch block
BP	backpropagation
CAM	classification activation map
CNN	convolutional neural network
FRM	feature reuse module
GPS	global positioning system
HSV	Hue, saturation and value
UAV	unmanned aerial vehicles
NLB	Northern Leaf Blight
RGB	red green and blue
SLIC	simple linear iterative cluster
SIFT	scale-invariant feature transform
VI	vegetation index

References

1. Bayraktar, E.; Basarkan, M.E.; Celebi, N. A low-cost UAV framework towards ornamental plant detection and counting in the wild. *ISPRS-J. Photogramm. Remote Sens.* **2020**, *167*, 1–11. [[CrossRef](#)]
2. Wiesner-Hanks, T.; Wu, H.; Stewart, E.; DeChant, C.; Kaczmar, N.; Lipson, H.; Gore, M.A.; Nelson, R.J. Millimeter-level plant disease detection from aerial photographs via deep learning and crowdsourced data. *Front. Plant Sci.* **2019**, *10*, 1550. [[CrossRef](#)] [[PubMed](#)]
3. Tetila, E.C.; Machado, B.B.; de Souza Belete, N.A.; Guimaraes, D.A.; Pistori, H. Identification of Soybean Foliar Diseases Using Unmanned Aerial Vehicle Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2190–2194. [[CrossRef](#)]
4. Yue, J.; Lei, T.; Li, C.; Zhu, J. The application of unmanned aerial vehicle remote sensing in quickly monitoring crop pests. *Intell. Autom. Soft Comput.* **2012**, *18*, 1043–1052. [[CrossRef](#)]
5. Yang, N.; Yuan, M.; Wang, P.; Zhang, R.; Sun, J.; Mao, H. Tea diseases detection based on fast infrared thermal image processing technology. *J. Sci. Food Agric.* **2019**, *99*, 3459–3466. [[CrossRef](#)]
6. Available online: www.deere.com/en/sprayers/see-spray-ultimate/ (accessed on 15 February 2023).
7. Dammer, K.H.; Garz, A.; Hobart, M.; Schirrmann, M. Combined UAV- and tractor-based stripe rust monitoring in winter wheat under field conditions. *Agron. J.* **2021**, *114*, 651–661. [[CrossRef](#)]
8. Gillis, M.; Shafii, M.; Browne, G.T.; Upadhyaya, S.K.; Coates, R.W.; Udometaikul, V. Tractor-mounted, GPS-based spot fumigation system manages Prunus replant disease. *Calif. Agric.* **2013**, *67*, 222–227. [[CrossRef](#)]
9. Gnyp, M.; Panitzki, M.; Reusch, S.; Jasper, J.; Bolten, A.; Bareth, G. Comparison between tractor-based and UAV-based spectrometer measurements in winter wheat. In Proceedings of the 13th International Conference on Precision Agriculture, St. Louis, MO, USA, 31 July–3 August 2016.
10. León-Rueda, W.A.; León, C.; Caro, S.G.; Ramírez-Gil, J.G. Identification of diseases and physiological disorders in potato via multispectral drone imagery using machine learning tools. *Trop. Plant Pathol.* **2021**, *47*, 152–167. [[CrossRef](#)]
11. Ye, H.; Huang, W.; Huang, S.; Cui, B.; Dong, Y.; Guo, A.; Ren, Y.; Jin, Y. Recognition of Banana Fusarium Wilt Based on UAV Remote Sensing. *Remote Sens.* **2020**, *12*, 938. [[CrossRef](#)]
12. Bagheri, N. Application of aerial remote sensing technology for detection of fire blight infected pear trees. *Comput. Electron. Agric.* **2020**, *168*, 105147. [[CrossRef](#)]
13. Su, J.; Liu, C.; Hu, X.; Xu, X.; Guo, L.; Chen, W.-H. Spatio-temporal monitoring of wheat yellow rust using UAV multispectral imagery. *Comput. Electron. Agric.* **2019**, *167*, 105035. [[CrossRef](#)]
14. Bohnenkamp, D.; Behmann, J.; Mahlein, A.-K. In-Field Detection of Yellow Rust in Wheat on the Ground Canopy and UAV Scale. *Remote Sens.* **2019**, *11*, 2495. [[CrossRef](#)]
15. Liu, L.; Dong, Y.; Huang, W.; Du, X.; Ma, H. Monitoring Wheat Fusarium Head Blight Using Unmanned Aerial Vehicle Hyperspectral Imagery. *Remote Sens.* **2020**, *12*, 3811. [[CrossRef](#)]
16. Ahmadi, P.; Mansor, S.; Farjad, B.; Ghaderpour, E. Unmanned Aerial Vehicle (UAV)-Based Remote Sensing for Early-Stage Detection of Ganoderma. *Remote Sens.* **2022**, *14*, 1239. [[CrossRef](#)]
17. Ha, J.G.; Moon, H.; Kwak, J.T.; Hassan, S.I.; Dang, M.; Lee, O.N.; Park, H.Y. Deep convolutional neural network for classifying Fusarium wilt of radish from unmanned aerial vehicles. *J. Appl. Remote Sens.* **2017**, *11*, 042621. [[CrossRef](#)]
18. Kerkech, M.; Hafiane, A.; Canals, R. Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach. *Comput. Electron. Agric.* **2020**, *174*, 105446. [[CrossRef](#)]
19. Wang, Z.; Zhang, S. Segmentation of corn leaf disease based on fully convolution neural network. *Acad. J. Comput. Inf. Sci.* **2018**, *1*, 9–18.
20. Pan, Q.; Gao, M.; Wu, P.; Yan, J.; Li, S. A Deep-Learning-Based Approach for Wheat Yellow Rust Disease Recognition from Unmanned Aerial Vehicle Images. *Sensors* **2021**, *21*, 6540. [[CrossRef](#)]
21. Wu, H.; Wiesner-Hanks, T.; Stewart, E.L.; DeChant, C.; Kaczmar, N.; Gore, M.A.; Nelson, R.J.; Lipson, H. Autonomous Detection of Plant Disease Symptoms Directly from Aerial Imagery. *Plant Phenome J.* **2019**, *2*, 190006. [[CrossRef](#)]
22. Narkhede, M.V.; Bartakke, P.P.; Sutaone, M.S. A review on weight initialization strategies for neural networks. *Artif. Intell. Rev.* **2021**, *55*, 291–322. [[CrossRef](#)]
23. Boulila, W.; Driss, M.; Alshantqiti, E.; Al-Sarem, M.; Saeed, F.; Krichen, M. Weight Initialization Techniques for Deep Learning Algorithms in Remote Sensing: Recent Trends and Future Perspectives. In *Advances on Smart and Soft Computing; Advances in Intelligent Systems and Computing*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 477–484.
24. Krichen, M.; Mihoub, A.; Alzahrani, M.Y.; Adoni, W.Y.H.; Nahhal, T. Are Formal Methods Applicable To Machine Learning And Artificial Intelligence? In Proceedings of the 2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH), Riyadh, Saudi Arabia, 9–11 May 2022; pp. 48–53.
25. Urban, C.; Miné, A. A review of formal methods applied to machine learning. *arXiv* **2021**, arXiv:02466.
26. Seshia, S.A.; Sadigh, D.; Sastry, S.S. Toward verified artificial intelligence. *Commun. ACM* **2022**, *65*, 46–55. [[CrossRef](#)]
27. Katz, G.; Barrett, C.; Dill, D.L.; Julian, K.; Kochenderfer, M.J. Reluplex: An efficient SMT solver for verifying deep neural networks. In Proceedings of the Computer Aided Verification: 29th International Conference, CAV 2017, Heidelberg, Germany, 24–28 July 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 97–117.
28. Zhang, T.; Xu, Z.; Su, J.; Yang, Z.; Liu, C.; Chen, W.-H.; Li, J. Ir-UNet: Irregular Segmentation U-Shape Network for Wheat Yellow Rust Detection by UAV Multispectral Imagery. *Remote Sens.* **2021**, *13*, 3892. [[CrossRef](#)]

29. Qin, J.; Wang, B.; Wu, Y.; Lu, Q.; Zhu, H. Identifying Pine Wood Nematode Disease Using UAV Images and Deep Learning Algorithms. *Remote Sens.* **2021**, *13*, 162. [[CrossRef](#)]
30. Hu, G.; Wu, H.; Zhang, Y.; Wan, M. A low shot learning method for tea leaf's disease identification. *Comput. Electron. Agric.* **2019**, *163*, 104852. [[CrossRef](#)]
31. Huang, M.; Xu, G.; Li, J.; Huang, J. A Method for Segmenting Disease Lesions of Maize Leaves in Real Time Using Attention YOLACT++. *Agriculture* **2021**, *11*, 1216. [[CrossRef](#)]
32. Wu, B.; Liang, A.; Zhang, H.; Zhu, T.; Zou, Z.; Yang, D.; Tang, W.; Li, J.; Su, J. Application of conventional UAV-based high-throughput object detection to the early diagnosis of pine wilt disease by deep learning. *For. Ecol. Manag.* **2021**, *486*, 118986. [[CrossRef](#)]
33. Stewart, E.L.; Wiesner-Hanks, T.; Kaczmar, N.; DeChant, C.; Wu, H.; Lipson, H.; Nelson, R.J.; Gore, M.A. Quantitative Phenotyping of Northern Leaf Blight in UAV Images Using Deep Learning. *Remote Sens.* **2019**, *11*, 2209. [[CrossRef](#)]
34. Zhou, Y.; Tang, Y.; Zou, X.; Wu, M.; Tang, W.; Meng, F.; Zhang, Y.; Kang, H. Adaptive Active Positioning of Camellia oleifera Fruit Picking Points: Classical Image Processing and YOLOv7 Fusion Algorithm. *Appl. Sci.* **2022**, *12*, 12959. [[CrossRef](#)]
35. Tang, Y.; Zhou, H.; Wang, H.; Zhang, Y. Fruit detection and positioning technology for a Camellia oleifera C. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision. *Expert Syst. Appl.* **2023**, *211*, 118573. [[CrossRef](#)]
36. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
37. Kumar Singh, K.; Jae Lee, Y. Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3524–3533.
38. Zhang, X.; Wei, Y.; Feng, J.; Yang, Y.; Huang, T.S. Adversarial complementary learning for weakly supervised object localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Lake City, UT, USA, 18–22 June 2018; pp. 1325–1334.
39. Wei, Y.; Xiao, H.; Shi, H.; Jie, Z.; Feng, J.; Huang, T.S. Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7268–7277.
40. Lee, J.; Kim, E.; Lee, S.; Lee, J.; Yoon, S. Ficklenet: Weakly and semi-supervised semantic image segmentation using stochastic inference. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 5267–5276.
41. Zhou, T.; Zhang, M.; Zhao, F.; Li, J. Regional semantic contrast and aggregation for weakly supervised semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, Louisiana, USA, 19–24 June 2022; pp. 4299–4309.
42. Yi, R.; Weng, Y.; Yu, M.; Lai, Y.-K.; Liu, Y.-J. Lesion region segmentation via weakly supervised learning. *Quant. Biol.* **2021**, *10*, 239–252. [[CrossRef](#)]
43. Kim, W.-S.; Lee, D.-H.; Kim, T.; Kim, H.; Sim, T.; Kim, Y.-J. Weakly Supervised Crop Area Segmentation for an Autonomous Combine Harvester. *Sensors* **2021**, *21*, 4801. [[CrossRef](#)] [[PubMed](#)]
44. Wiesner-Hanks, T.; Stewart, E.L.; Kaczmar, N.; DeChant, C.; Wu, H.; Nelson, R.J.; Lipson, H.; Gore, M.A. Image set for deep learning: Field images of maize annotated with disease symptoms. *BMC Res. Notes* **2018**, *11*, 440. [[CrossRef](#)] [[PubMed](#)]
45. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [[CrossRef](#)]
46. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
47. Zhang, Z.; Zhang, X.; Peng, C.; Xue, X.; Sun, J. Exfuse: Enhancing feature fusion for semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 269–284.
48. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
49. Wang, C.; Du, P.; Wu, H.; Li, J.; Zhao, C.; Zhu, H. A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net. *Comput. Electron. Agric.* **2021**, *189*, 106373. [[CrossRef](#)]

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