Article

A Method for Segmenting Disease Lesions of Maize Leaves in Real Time Using Attention YOLACT++

Mingfeng Huang, Guoqin Xu, Junyu Li and Jianping Huang *

School of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China; 2019120551@nefu.edu.cn (M.H.); xuguoqin@nefu.edu.cn (G.X.); lijunyu6112@gmail.com (J.L.)

* Correspondence: jphuang@nefu.edu.cn

Abstract: Northern leaf blight (NLB) is a serious disease in maize which leads to significant yield losses. Automatic and accurate methods of quantifying disease are crucial for disease identification and quantitative assessment of severity. Leaf images collected with natural backgrounds pose a great challenge to the segmentation of disease lesions. To address these problems, we propose an image segmentation method based on YOLACT++ with an attention module for segmenting disease lesions of maize leaves in natural conditions in order to improve the accuracy and real-time ability of lesion segmentation. The attention module is equipped on the output of the ResNet-101 backbone and the output of the FPN. The experimental results demonstrate that the proposed method improves segmentation accuracy compared with the state-of-the-art disease lesion-segmentation methods. The proposed method achieved 98.71% maize leaf lesion segmentation precision, a comprehensive evaluation index of 98.36%, and a mean Intersection over Union of 84.91%; the average processing time of a single image was about 31.5 ms. The results show that the proposed method allows for the automatic and accurate quantitative assessment of crop disease severity in natural conditions.

Keywords: plant disease; northern maize leaf blight; lesions segmentation; attention mechanism; YOLACT++; instance segmentation; convolutional neural network

Citation: 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. <https://doi.org/10.3390/agriculture11121216>

Academic Editor: Maciej Zaborowicz

Received: 15 October 2021

Accepted: 29 November 2021

Published: 2 December 2021

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

Copyright: © 2021 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

Maize is an important economic crop, with the third-largest area sown and total production in the world, after rice and wheat. The northern leaf blight of maize, caused by the fungus *S. turcica*, is a major disease impacting maize in wet climates and typically shows symptoms of oblong, “cigar-shaped” tan or greyish lesions. Leaf lesions result in a reduction in the leaf area where photosynthesis takes place. The more lesions on maize leaves and the earlier in the season the lesions occur, the greater the loss of photosynthetic area and the reduction in maize yield [1]. The annual yield loss of maize grown in the United States and Canada due to northern leaf blight reached approximately 14 million tons between 2012 and 2015, accounting for a quarter of the total loss caused by the disease globally [2]. Therefore, a timely grasp of the severity of crop diseases is of great significance for effective disease prevention and the formulation of scientific prevention and control strategies.

The segmentation of disease lesions in maize leaf images directly affects the recognition effect of crop diseases and the accuracy of a quantitative assessment of the severity of crop diseases [3,4]. How to segment the diseased leaves of crops with high efficiency and high quality is a research hotspot. In the last two decades, traditional image processing techniques, such as edge detection, color space transformation, feature space transformation, etc., were used to achieve the extraction and recognition of lesions [5,6]. Using the gray-scale structure intensity histogram of channel H (from the HSV color space) and channel a (from the L*a*b* color space), one can find the pixel value that can best separate healthy and diseased tissues, and segment the lesions [7]. On the basis of image enhancement, a strong correlation-based approach was applied to segment the apple leaf lesions, and the extracted

features were optimized with a genetic algorithm. Furthermore, the segmentation results were improved by fusing expectation-maximizing segmentation [8]. A joint framework of feature fusion and selection technique frameworks was used to classify cucumber diseases. In this case, feature selection used the Manhattan Distance Control Entropy (MDcE) technique to select strong features that can be more accurate in classifying cucumber diseases [9]. In recent years, deep learning has been successfully applied in many fields [10], including computer vision [11], natural language processing [12], bioinformatics [13], and medical image analysis [14–16], and has also been employed in many applications of agriculture and forestry, such as plant species recognition, plant disease detection [17–20], and plant disease lesions segmentation [21,22]. Using the framework of deep learning, a platform for the semantic segmentation of small-area crops based on DeepLabv3+ was constructed, which was superior to other semantic segmentation models [23], along with a framework based on best feature selection for identifying multiple classes of leaf diseases in cucumbers, with 96.5% recognition accuracy [24]. Huang et al. used a neural network search algorithm to automatically and accurately identify 54,306 public data in the PlantVillage dataset. This method could automatically learn a suitable deep neural network structure according to a specific dataset [25].

Convolutional Neural Networks (CNNs) are some of the most significant networks in deep learning, which are currently state-of-the-art in many computer vision tasks, including object detection and segmentation [26]. Using images taken by unmanned aerial vehicles to train CNNs, the output results were fed into a conditional random field (CRF). Ultimately, the maize leaf images were segmented into lesion and non-lesion areas, and the CNN could detect millimeter-level plant diseases through deep learning and crowdsourced data, which is the most accurate aerial plant disease detection scale achieved so far [27]. CNNs are increasingly being used in agriculture. Part of this success is due to the ability of CNNs to automatically extract features instead of using manual feature extraction. High-resolution images of northern maize leaf blight were obtained using drones and trained a CNN model on low-resolution sub-images to classify disease lesions, obtaining an accuracy of 95.1% [28]. After integrating multiple CNN structures, the united model could extract the complementary lesions features, which enhanced the model's ability to recognize lesions [29]. Combining drone technology with deep-learning-based methods, for instance, segmentation, the Mask R-CNN model could accurately detect and segment a single lesion in the test set for NLB lesions in the UAV images [30]. To overcome the problem of slow neural network recognition, Fuentes et al. proposed a real-time tomato pests and diseases identification model, which can identify nine kinds of diseases [31]. To discard redundant and irrelevant features more accurately, Khan et al. proposed a classification method based on deep convolutional neural networks to classify the leaf diseases of five different fruits in the Plant Village Dataset with an accuracy of 97.8% [32]. To raise the detection effect of small targets, they added a retinex model with low-pass output to preprocess the dataset. Furthermore, the multi-scale feature fusion and the fine-tuning network of an anchor box was used in the detection network. It provided a reference for the accurate detection of maize leaf blight [33]. There were also many studies that use CNNs to identify and segment plant diseases [34–38].

According to the research of Liu et al., identifying plant disease symptoms in images may belong to one of three types of tasks: classification, detection or segmentation. Object detection methods determine the location and range of symptoms in an image at a rough spatial level, and the most common one is to describe them with bounding boxes. Semantic segmentation methods describe the boundaries of features and assign each pixel of the image to a given region [39]. The instance segmentation methods have more tasks than semantic segmentation. While segmenting the lesions, they also depict the boundary of the feature. Instance segmentation can be defined as a technology that simultaneously solves the problem of object detection and semantic segmentation [40]. In this study, we are responsible for the final task of identifying and delineating each diseased area in the image.

In recent years, the attention model has been widely used in various types of deep learning tasks—for example, natural language processing, image recognition, and speech recognition. The association of attention mechanisms and deep learning makes the task of plant disease lesion recognition and segmentation more interesting and in-depth. The main common attention modules are SE-Net (Squeeze-and-Excitation Networks) [41], CBAM (Convolutional Block Attention Module) [42], and VSG-Net (Visual-Spatial-Graph Network) [43], etc. Zhong et al. proposed a grouped attention module based on a grouped activation strategy, which used high-order features to guide the enhancement of low-order features. Meanwhile, the enhancement coefficients within groups were calculated by grouping to reduce the suppression between different groups and enhance the ability of feature expression. The pixel accuracy of segmentation was 93.9%, and the mIoU was 78.6% [44].

Images of crop leaves collected under complex backgrounds are affected by various factors, such as weeds, soil, light intensity, etc. More importantly, each disease spot has its own shape, color, and texture. The presence of these factors presents the researcher with a great challenge. Based on the above problems, in this study, a novel model named Attention YOLACT++ for maize NLB lesion recognition and segmentation is proposed. It can better detect and segment the edges of diseased spots. It also provides a technical tool for subsequent accurate identification and quantitative assessment of the severity of diseased maize leaves. The main contributions of this study are as follows:

- We proposed a new instance segmentation architecture. We adopted the YOLACT++ model for the segmentation task and applied the convolutional block attention module to the ResNet-101 module and the FPN module to improve segmentation performance and model robustness.
- Our model achieved more accurate segmentation speed and accuracy on a maize northern leaf blight dataset collected in a complex background, wherein its performance was superior to that of current instance segmentation models.

2. Materials and Methods

2.1. Dataset Description

The maize images used for evaluating the performance of the proposed method corresponded to aerial images of northern leaf blight, which can be downloaded from the website <https://osf.io/vfawp/> (accessed on 10 October 2021). This study took the northern leaf blight of maize as the research object. All trials were captured at Cornell University's Musgrave Research Farm in Aurora, New York, in the summer of 2017. The dataset was captured by mounting the camera on a UAV flying at an altitude of 6 m [45]. Examples of the maize images collected by the drone are shown in Figure 1.

2.2. Dataset Annotation

In this study, images of maize leaf taken by the drone were cropped using Photoshop. To maintain the aspect ratio of the lesions, the pixel size of the maize leaf images at the time of cropping was 550×550 . Some example images are shown in Figure 2a. We used 1200 images of maize northern blight leaf, randomly divided in a 2:1:1 ratio into a training set (600), a validation set (300), and a test set (300), manually annotated as the reference (ground truth) of the diseased areas (Figure 2b). To increase the diversity of the dataset and to avoid overfitting, in this study we used photometric distortion, random contrast, random cropping, flipping, and random rotation operations for data enhancement in both the training and validation sets. The above enhancement operations expanded the training set to 2000, and the validation set to 700. For the test set, no augmentation was performed and it was used directly for model evaluation to ensure the authenticity of the test set.

Figure 1. Examples of maize images collected by the UAV.

Figure 2. Reference of maize images and their corresponding labels. (a) cropped images, (b) labeled images.

2.3. Model Architecture

To improve the accuracy and time of the lesion segmentation of maize leaf under natural conditions, we propose an image segmentation method based on YOLACT++ [46] with an attention module. First, we introduce a Convolutional Block attention module (CBAM) [42] between the multi-scale output of ResNet-101 [47] and the input of the Feature Pyramid Networks (FPN) [48]. Additionally, we add it to the output of the FPN.

The attention module can obtain the importance of each feature channel through automatic learning and assign different weights to different feature channels so that the network can focus on the most relevant features and improve the segmentation performance of the network. Focusing on the diseased areas of the maize leaves during feature extraction improves the accuracy of network recognition and detection.

The architecture of the proposed method is illustrated in Figure 3. The detailed structure consists of five parts: feature extraction, attention module, FPN architecture,

segmentation network, and image post-processing. Attention module 1 and attention module 2 are both CBAM.

Figure 3. The proposed attention-based YOLACT++ architecture.

(1) Feature extraction.

In this study, we use a residual network with 101 layers (ResNet-101). ResNet-101, which is less computationally intensive and preforms better, is applied as the feature extraction network. Furthermore, Deformable convolutional neural networks (DCNs) [49,50] are deployed on the last three (C_3 to C_5) ResNet-101 stages with an interval of three (i.e., this network replaces the 3×3 convolutional layers in the ResNet module with a 3×3 deformable convolutional layer at intervals of 3 convolutional layers.). The reason for adding the DCNs structure to this network is that YOLACT [51] is a one-shot sampling method without a resampling process and that DCNs can enhance the network's ability to handle different scales, rotations, and aspect ratios. The sampling method of the ResNet network was changed by using free-form sampling rather than the rigid grid sampling found in traditional CNNs.

(2) Attention module.

The CBAM [42] used in this study, is shown in Figure 4. CBAM consists of two separate sub-modules: the channel attention module (CAM) and the spatial attention module (SAM).

Figure 4. Convolutional block attention module.

The CAM structure is displayed in Figure 5. Its structure takes the input feature map C_i or P_j ($C_i \in \mathbb{R}^{H \times W \times C}$, $i = 3, 4, 5$, $P_j \in \mathbb{R}^{H \times W \times C}$, $j = 3, \dots, 7$) through global maximum pooling and global average pooling, respectively, to obtain two feature maps, where C_i is the input of attention module 1 and P_j is the input of attention module 2 (Figure 3). These are then sent to a two-layer neural network (MLP); the activation function is Relu. The

sigmoid activation function is used to generate the final channel attention feature M_{cc} or M_{cp} ($M_{cc}, M_{cp} \in \mathbb{R}^{1 \times 1 \times C}$).

Figure 5. Channel attention module.

M_{cc} or M_{cp} is multiplied with the input feature map C_i or P_j to generate the feature maps F_c or F_p . Of these, M_{cc} and F_c are the results of attention module 1; M_{cp} and F_p are the results of Attention module 2, respectively.

The structure of the SAM is shown in Figure 6, taking F_c or F_p as the input feature map of this module. Firstly, two $H \times W \times 1$ feature maps are obtained through global maximum pooling and global average pooling based on the channel. Additionally, the two feature maps are spliced based on the channel. Through a 7×7 convolutional layer, the dimension is then reduced to 1 channel, i.e., $H \times W \times 1$. Next, the spatial attention feature M_{sc} or M_{sp} ($M_{sc}, M_{sp} \in \mathbb{R}^{H \times W \times 1}$) is generated through the sigmoid activation function, in which M_{sc} and M_{sp} are the outputs of the SAM after attention module 1 and attention module 2, respectively.

Figure 6. Spatial attention module.

Finally, M_{sc} or M_{sp} and the input feature maps F_c or F_p of the module are multiplied to obtain the final feature F'_c or F'_p , i.e., C'_i or P'_j .

(3) FPN architecture.

The feature maps $P_3 - P_7$ in the FPN structure were obtained from the convolutional layer C'_5 through one convolutional layer to obtain feature map P_5 . The bilinear interpolation method was then used to double the size of feature map P_5 , the convolution of C'_4 was added to obtain feature map P_4 , and the same method was used to obtain feature map P_3 . Finally, feature maps P_5 and P_6 were convolved and down-sampled to obtain the feature maps P_6 and P_7 . These feature maps obtained from FPN network were fed into CBAM. P'_j ($j = 3, \dots, 7$), which was obtained as the input of segmentation network.

(4) Segmentation network.

The Protonet (Figure 3) was used to generate k -many prototype masks of the same size as the original image by means of a fully convolutional neural network (FCN) [52]. It takes the feature map P'_3 as input, and the dimensions of the output mask are $138 \times 138 \times k$; that is, k -many prototype masks are obtained, and the size of each mask is 138×138 .

The Prediction Head structure (Figure 3) uses a shared convolutional network to improve the segmentation speed. It takes the five feature maps $P'_3 - P'_7$ from the feature extraction network as input, and completes the three tasks of target classification prediction, bounding box prediction and mask coefficient prediction. The fast non-maximum suppression (NMS) algorithm then obtains the mask coefficients with the highest confidence.

The outputs of the Protonet branch and the Prediction Head branch are derived from the mask by basic matrix multiplication and sigmoid function, as shown in Equation (1).

$$M = \text{sigmoid}(PC^T) \quad (1)$$

where P is the prototype mask, and C is a $n \times k$ matrix of mask coefficients.

(5) Image post-processing.

Image post-processing mainly includes crop, fast mask re-scoring and threshold. First, the terminal masks are cropped with the predicted bounding box, i.e., the pixels outside the box region are zeroed out.

Second, the branch of fast mask re-scoring network is composed of six convolutional layers and one global average pooling layer, as shown in Figure 7. The function is to re-score the mask based on the Intersection over Union (IoU) between the predicted mask and the original leaf mask. The specific steps are as follows: (I) the mask of leaf disease image with size $138 \times 138 \times 1$ —obtained after cropping, as input and output, the IoU of the original leaf mask of the corresponding category is taken; (II) the IoU of the mask corresponding to the category predicted by the classification branch is multiplied with the corresponding category confidence as the final score of the mask.

Figure 7. Fast mask re-scoring network.

Finally, thresholding performs image-thresholding processing on the re-scored mask to obtain the final segmented image.

2.4. Loss Function

The loss function used in this study includes three parts, i.e., training classification loss function L_{cls} , prediction frame loss function L_{box} , and mask-generation loss function L_{mask} . The total loss function L of the network is shown in Equation (2):

$$L = \alpha L_{cls} + \beta L_{box} + \gamma L_{mask} \quad (2)$$

where the mask generation loss function L_{mask} is defined as:

$$L_{mask} = \text{sigmoid}(-p_i^* \log[p_i^* p_i + (1 - p_i^*)(1 - p_i)]) \quad (3)$$

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

The classification loss function L_{cls} and the prediction frame loss function L_{box} are expressed, respectively, as:

$$L_{cls} = -\sum p_i^* \log(p_i) \quad (5)$$

$$L_{box} = \text{smooth}_{L1} \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{other} \end{cases} \quad (6)$$

where i is the index number of the anchor, p_i is the predicted probability of the target, and p_i^* is the original leaf probability; α , β and γ are the weights of each loss.

2.5. Experimental Setup

The network training and testing hardware environment was Intel(R) Core i7-9700k 3.60 GHz processor, 16 GB RAM. NVIDIA GeForce RTX 2080Ti GPU with 11 GB graphics memory. CUDA Toolkit 10.0 and CUDNN V7.6.5 were used as the network model training acceleration toolkit. The software environment was Ubuntu 18.04 LTS 64-bit operating

system, Python (3.7). PyTorch (1.4.0) deep learning opensource framework was used to build the network model.

In this study, the feature extraction network adopted the mode of transfer learning to fine-tune the pre-trained parameters of the ImageNet classification model [53]. The training parameters in this study were set as follows: eight images per batch were trained using the stochastic gradient descent method with a momentum factor of 0.9 [54–56]. The initial learning rate was set as 10^{-3} , and the maximum number of iterations was 400,000. The learning rate decreased by one-tenth at 180,000, 220,000 and 350,000 times, respectively. The weight decay parameter was 5×10^{-4} .

The weights α , β and γ of the loss function Equation (2) were set to 1, 1.5, and 6.125, respectively. The classification loss function L_{cls} , the prediction frame loss function L_{box} , the mask loss function L_{mask} , and the overall loss value L of the network for different iterations are shown in Figure 8. It can be seen from the figure that after 400,000 iterations, the losses of the network started to converge and gradually stabilize.

Figure 8. Loss value corresponding to different iterations.

In the experiment, in addition to qualitative assessment consisting of visually comparing segmentation results, several indices were calculated to quantitatively evaluate the performance for disease lesion segmentation with different methods. They are the comprehensive evaluation indexes F_1 score and the mean Intersection over Union (mIoU).

The F_1 score reflects the overall segmentation accuracy of the lesions. The larger the fraction is, the more stable the model is. The F_1 score is given by:

$$F_1 = 2 \times \frac{P \times R}{P + R} \quad (7)$$

where, P and R represent precision and recall, respectively.

The mIoU is the ratio of intersection and union of two sets of ground truth and prediction of leaf lesion area. The larger the mIoU value is, the better the segmentation effect will be. The mIoU is defined as:

$$mIoU = \frac{TP}{FP + FN + TP} \quad (8)$$

In addition, the segmentation time of each leaf image in the test set is calculated, and the average time is used as the performance index to evaluate the real-time performance of the model.

3. Results and Discussion

3.1. Results

The IoU threshold represents the degree of overlap between the true and predicted values. In the experiment, only $\text{IoU} > 0.5$ can be considered as a correctly predicted value. The segmentation precision at different IoU thresholds is shown in Figure 9. When the IoU threshold was 0.5, the segmentation precision was 99.06%, indicating that the model performs well in the loose IoU threshold range. The mean precision (mP) of the maize leaf lesion images in the test set was 86.2% when the IoU threshold range was 0.5 to 0.95.

Figure 9. Segmentation precision at IoU thresholds of 0.50–0.95.

In this paper, 0.7 was chosen as the IoU threshold for maize leaf segmentation. When the number of iterations was 400,000, the training time of the proposed method was approximately four to five days. To improve the robustness of the model, 10-fold cross-validation was used in this study. After five cross-validation experiments, the average was taken. The Attention YOLACT++ network correctly segmented 296 out of 300 maize NLB images with precision of 98.71%, recall of 98.02%, and mIoU of 84.91%. Table 1 shows the precision, recall, F_1 scores, mIoU, and total segmentation time of the network for maize leaf segmentation.

Table 1. Evaluation parameters after 400,000 iterations.

Model	Precision (%)	Recall (%)	F1 (%)	mIoU (%)	Total Time (s)
Attention YOLACT++	98.71	98.02	98.36	84.91	9.44

Figure 10 shows the segmentation results of maize leaf lesions under different influencing factors, such as weeds, light intensity, soil, and mutual covering of leaves, etc. It was discovered that the proposed method provided satisfying segmentation results with clear edges. However, due to the influence of soil and other factors, the problem of owe-segmentation occurred in the segmentation of maize leaf lesions, as shown in Figure 10d.

Figure 10. Example diagrams of segmentation results of maize leaf lesions. (a) weeds; (b) mutual covering of leaves; (c) light intensity; (d) soil and other factors.

3.2. Prediction Results Comparison

To further validate the performance of the proposed method in this paper, we compared it with the state-of-the-art instance segmentation model Mask R-CNN [57] and YOLACT++ [46]. As observed from Table 2, the proposed method achieved a better segmentation performance than the Mask R-CNN and YOLACT++ methods for the quantitative indices. The segmentation precision of the proposed method was about 15.14% and 1.27% higher than that of Mask R-CNN and YOLACT++ models. The mIoU of the proposed method reached 84.91%, which was 11.91% and 6.26% higher than Mask R-CNN and YOLACT++ models. The main reason for this was that the proposed method added an attention module, which could accurately extract the features of a lesion. The segmentation time in the test set was obtained by averaging the prediction time of all images. The prediction time for our model was slightly longer than YOLACT++, and shorter than Mask R-CNN, but the segmentation mIoU achieved by our model was the highest. This suggested that our model was well-suited for real-time NLB lesion segmentation.

Table 2. Comparison of segmentation performance of different models.

Model	Precision (%)	mIoU (%)	Single Image Segmentation Time (s)
Mask R-CNN	83.57	73.00 [30]	3
YOLACT++	97.44	78.65	0.0242
Attention YOLACT++	98.71	84.91	0.0315

Figure 11 shows a comparison of the mask quality of Mask R-CNN, YOLACT++ and the proposed method in this study for the segmentation of maize leaf disease. It could be observed that the method proposed in this paper provided better segmentation results when visually compared to other methods. The Mask R-CNN segmentation model was affected by uneven illumination and a complex background, and could segment the approximate area of the lesions, but there was still inaccurate segmentation. In particular, the reason why Mask R-CNN segmentation was inaccurate is that the feature extraction of maize lesions was not accurate enough. It led to poor detection and segmentation of maize lesion edge regions, as seen in Figure 11a.

Figure 11. Examples of lesion segmentation results for different methods. (a) Mask R-CNN, (b) YOLACT++, and (c) Attention YOLACT++.

The YOLACT++ segmentation model featured improved segmentation accuracy and segmentation speed. In cases where there were many lesion targets on maize leaves, blurred boundaries, and close proximity, results, such as missed detections and misjudgments, can easily occur. This model could not detect all targets in maize leaves very well, as shown in Figure 11b.

The segmentation method proposed in this study introduced the convolutional attention module, which can accurately extract maize lesion features, especially the edge features of lesions. It could quickly and accurately detect and segment the area where the lesions were located, with a fast segmentation speed and high segmentation accuracy. Compared with YOLACT++, although the segmentation time per image was reduced, the segmentation precision of maize lesions was improved to a certain extent, as shown in illustration Figure 11c.

4. Conclusions

This work has proposed an image segmentation method based on the YOLACT++ with an attention module for segmenting disease lesions of maize leaf. Since feature extraction is blind and uncertain, we introduce CBAM based on the YOLACT++ network to improve the segmentation performance of the network. While improving the segmentation accuracy, the attention-module-based feature-network gives more attention to the diseased part of maize leaves, making the detection and identification of the lesion edges more accurate.

To deal with the problem wherein current maize leaf blight detection and segmentation models are susceptible to interference from shadows, occlusions, and light intensity, in our work, we applied a model for maize NLB segmentation. The results of comparative experiments demonstrate that introducing an attention mechanism allows for better detection and segmentation of disease edges, thus improving the accuracy of disease segmentation, which outperforms current example segmentation models, e.g., Mask R-CNN, YOLACT++, etc. The proposed method can be adapted to complex natural environments and lays the foundation for subsequent quantitative assessment of disease severity.

However, northern leaf blight is only one of the most important fungal diseases of maize. It would be interesting to apply the proposed method for other diseases of maize and more types of plants and diseases. Furthermore, according to the disease type, it would be helpful to introduce more highly accurate and lightweight modules into the proposed approach to further improve the segmentation efficiency of agricultural mobile equipment used in fields.

Author Contributions: Conceptualization, M.H. and J.H.; Data curation, G.X. and M.H.; Formal analysis, G.X., M.H., J.L. and J.H.; Methodology, M.H. and J.H.; Software, M.H. and G.X.; Supervision, J.H.; Validation, M.H. and J.L.; Visualization, M.H.; Writing—original draft, M.H.; Writing—review and editing, M.H. and J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Heilongjiang Province of China (Grant No. TD2020C001), the Fundamental Research Funds for the Central Universities (Grant No. 2572019CP19), and the project funded by China Postdoctoral Science Foundation (Grant No. 2017M610199).

Data Availability Statement: The data presented in this study are available within the article.

Acknowledgments: We are thankful to Jiuqing Liu and Shanchun Yan for providing funding support and supervision in this study. We are also grateful for anonymous reviewers' hard work and review.

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

References

1. Razzaq, T.; Khan, F.; Awan, S. Study of Northern Corn Leaf Blight (NCLB) on Maize (*Zea mays* L.) Genotypes and its Effect on Yield. *Sarhad J. Agric.* **2019**, *35*, 1166–1174. [\[CrossRef\]](#)
2. Mueller, D.S.; Wise, K.A.; Sisson, A.J.; Allen, T.W.; Warner, F. Corn Yield Loss Estimates Due to Diseases in the United States and Ontario, Canada from 2012 to 2015. *Plant Health Prog.* **2016**, *17*, 211–222. [\[CrossRef\]](#)
3. Esgario, J.G.M.; Krohling, R.A.; Ventura, J.A. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Comput. Electron. Agric.* **2020**, *169*, 105162. [\[CrossRef\]](#)
4. Xiong, Y.; Liang, L.; Wang, L.; She, J.; Wu, M. Identification of cash crop diseases using automatic image segmentation algorithm and deep learning with expanded dataset. *Comput. Electron. Agric.* **2020**, *177*, 105712. [\[CrossRef\]](#)
5. Revathi, P.; Hemalatha, M. Homogenous Segmentation based Edge Detection Techniques for Proficient Identification of the Cotton Leaf Spot Diseases. *Int. J. Comput. Appl.* **2013**, *47*, 18–21. [\[CrossRef\]](#)
6. Wang, L.; Tao, Y.; Tian, Y. *Crop Disease Leaf Image Segmentation Method Based on Color Features*; Springer: Boston, MA, USA, 2007.
7. Barbedo, J.G.A. A novel algorithm for semi-automatic segmentation of plant leaf disease symptoms using digital image processing. *Trop. Plant Pathol.* **2016**, *41*, 210–224. [\[CrossRef\]](#)
8. Khan, M.A.; Lali, M.I.; Sharif, M.; Javed, K.; Aurangzeb, K.; Haider, S.I.; Altamrah, A.S.; Akram, T. An Optimized Method for Segmentation and Classification of Apple Diseases Based on Strong Correlation and Genetic Algorithm Based Feature Selection. *IEEE Access* **2019**, *7*, 46261–46277. [\[CrossRef\]](#)
9. Kianat, J.; Khan, M.A.; Sharif, M.; Akram, T.; Rehman, A.; Saba, T. A joint framework of feature reduction and robust feature selection for cucumber leaf diseases recognition. *Optik* **2021**, *240*, 166566. [\[CrossRef\]](#)

10. Dargan, S.; Kumar, M.; Ayyagari, M.R.; Kumar, G. A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. *Arch. Comput. Methods Eng.* **2020**, *27*, 1071–1092. [\[CrossRef\]](#)
11. Li, X.; Grandvalet, Y.; Davoine, F.; Cheng, J.; Cui, Y.; Zhang, H.; Belongie, S.; Tsai, Y.-H.; Yang, M.-H. Transfer learning in computer vision tasks: Remember where you come from. *Image Vis. Comput.* **2020**, *93*, 103853. [\[CrossRef\]](#)
12. Otter, D.W.; Medina, J.R.; Kalita, J.K. A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 604–624. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Min, S.; Lee, B.; Yoon, S. Deep learning in bioinformatics. *Brief. Bioinform.* **2017**, *18*, 851–869. [\[CrossRef\]](#)
14. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [\[CrossRef\]](#)
15. Khan, M.A.; Muhammad, K.; Sharif, M.; Akram, T.; Albuquerque, V. Multi-Class Skin Lesion Detection and Classification via Tele dermatology. *IEEE J. Biomed. Health Inform.* **2021**, *1*. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Khan, M.A.; Kadry, S.; Zhang, Y.D.; Akram, T.; Sharif, M.; Rehman, A.; Saba, T. Prediction of COVID-19-Pneumonia based on Selected Deep Features and One Class Kernel Extreme Learning Machine. *Comput. Electr. Eng.* **2021**, *90*, 106960. [\[CrossRef\]](#)
17. Bansal, P.; Kumar, R.; Kumar, S. Disease Detection in Apple Leaves Using Deep Convolutional Neural Network. *Agriculture* **2021**, *11*, 617. [\[CrossRef\]](#)
18. Liu, J.; Wang, X. Plant diseases and pests detection based on deep learning: A review. *Plant Methods* **2021**, *17*, 22. [\[CrossRef\]](#)
19. Lu, J.; Tan, L.; Jiang, H. Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification. *Agriculture* **2021**, *11*, 707. [\[CrossRef\]](#)
20. Rehman, M.Z.U.; Ahmed, F.; Khan, M.A.; Tariq, U.; Jamal, S.S.; Ahmad, J.; Hussain, I. Classification of Citrus Plant Diseases Using Deep Transfer Learning. *CMC-Comput. Mater. Contin.* **2022**, *70*, 1401–1417. [\[CrossRef\]](#)
21. Chen, S.; Zhang, K.; Zhao, Y.; Sun, Y.; Ban, W.; Chen, Y.; Zhuang, H.; Zhang, X.; Liu, J.; Yang, T. An Approach for Rice Bacterial Leaf Streak Disease Segmentation and Disease Severity Estimation. *Agriculture* **2021**, *11*, 420. [\[CrossRef\]](#)
22. Khan, M.A.; Akram, T.; Sharif, M.; Alhaisoni, M.; Saba, T.; Nawaz, N. A probabilistic segmentation and entropy-rank correlation-based feature selection approach for the recognition of fruit diseases. *Eurasip J. Image Video Process.* **2021**, *2021*, 14. [\[CrossRef\]](#)
23. Du, Z.; Yang, J.; Ou, C.; Zhang, T. Smallholder Crop Area Mapped with a Semantic Segmentation Deep Learning Method. *Remote Sens.* **2019**, *11*, 888. [\[CrossRef\]](#)
24. Hussain, N.; Khan, M.A.; Tariq, U.; Kadry, S.; Yar, M.A.E.; Mostafa, A.M.; Alnuaim, A.A.; Ahmad, S. Multiclass Cucumber Leaf Diseases Recognition Using Best Feature Selection. *CMC-Comput. Mater. Contin.* **2022**, *70*, 3281–3294. [\[CrossRef\]](#)
25. Huang, J.; Chen, J.; Li, K.; Li, J.; Liu, H. Identification of multiple plant leaf diseases using neural architecture search. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 166–173.
26. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–21. [\[CrossRef\]](#)
27. 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. [\[CrossRef\]](#)
28. 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*, 1–9. [\[CrossRef\]](#)
29. Mj, A.; Lei, Z.B.; Qw, C. Automatic grape leaf diseases identification via UnitedModel based on multiple convolutional neural networks. *Inf. Process. Agric.* **2020**, *7*, 418–426.
30. 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. [\[CrossRef\]](#)
31. Fuentes, A.; Yoon, S.; Kim, S.C.; Park, D.S. A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition. *Sensors* **2017**, *17*, 2022. [\[CrossRef\]](#)
32. Khan, M.A.; Akram, T.; Sharif, M.; Saba, T. Fruits diseases classification: Exploiting a hierarchical framework for deep features fusion and selection. *Multimed. Tools Appl.* **2020**, *79*, 25763–25783. [\[CrossRef\]](#)
33. Sun, J.; Yang, Y.; He, X.; Wu, X. Northern Maize Leaf Blight Detection Under Complex Field Environment Based on Deep Learning. *IEEE Access* **2020**, *8*, 33679–33688. [\[CrossRef\]](#)
34. Lin, K.; Gong, L.; Huang, Y.; Liu, C.; Pan, J. Deep Learning-Based Segmentation and Quantification of Cucumber Powdery Mildew Using Convolutional Neural Network. *Front. Plant Sci.* **2019**, *10*, 155. [\[CrossRef\]](#)
35. Khan, M.A.; Akram, T.; Sharif, M.; Awais, M.; Javed, K.; Ali, H.; Saba, T. CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features. *Comput. Electron. Agric.* **2018**, *155*, 220–236. [\[CrossRef\]](#)
36. Shang, R.; Zhang, J.; Jiao, L.; Li, Y.; Stolkin, R. Multi-scale Adaptive Feature Fusion Network for Semantic Segmentation in Remote Sensing Images. *Remote Sens.* **2020**, *12*, 872. [\[CrossRef\]](#)
37. Xiong, X.; Duan, L.; Liu, L.; Tu, H.; Yang, P.; Wu, D.; Chen, G.; Xiong, L.; Yang, W.; Liu, Q. Panicle-SEG: A robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods* **2017**, *13*, 104. [\[CrossRef\]](#)
38. Ren, S.; Jia, F.; Gu, X.; Yuan, P.; Xue, W.; Xu, H. Recognition and segmentation model of tomato leaf diseases based on deconvolution-guiding. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 186–195.

39. Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.; Pietikainen, M. Deep Learning for Generic Object Detection: A Survey. *Int. J. Comput. Vis.* **2020**, *128*, 261–318. [[CrossRef](#)]
40. Hafiz, A.M.; Bhat, G.M. A survey on instance segmentation: State of the art. *Int. J. Multimed. Inf. Retr.* **2020**, *9*, 171–189. [[CrossRef](#)]
41. Jie, H.; Li, S.; Gang, S.; Albanie, S. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *42*, 7132–7141.
42. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
43. Ulutan, O.; Iftekhar, A.; Manjunath, B.S. Vsgnet: Spatial attention network for detecting human object interactions using graph convolutions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 13617–13626.
44. Zhong, C.; Hu, Z.; Li, M.; Li, H.; Yang, X.; Liu, F. Real-time semantic segmentation model for crop disease leaves using group attention module. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 208–215.
45. 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](#)]
46. Bolya, D.; Zhou, C.; Xiao, F.; Lee, Y.J. YOLACT++: Better Real-time Instance Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *1912*, 06218. [[CrossRef](#)]
47. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
48. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
49. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 764–773.
50. Zhu, X.; Hu, H.; Lin, S.; Dai, J. Deformable convnets v2: More deformable, better results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 15–20 June 2019; pp. 9308–9316.
51. Bolya, D.; Zhou, C.; Xiao, F.; Lee, Y.J. YOLACT: Real-Time Instance Segmentation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 9156–9165.
52. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 640–651.
53. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
54. Rakhlin, A.; Shamir, O.; Sridharan, K. Making gradient descent optimal for strongly convex stochastic optimization. In Proceedings of the 29th International Conference on Machine Learning, Edinburgh, UK, 26 June–1 July 2012; pp. 1571–1578.
55. Bottou, L. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*; Springer Nature: Cham, Switzerland, 2010; pp. 177–186.
56. Bottou, L. Stochastic gradient learning in neural networks. *Proc. Neuro-Nimes* **1991**, *91*, 687–706.
57. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. [[CrossRef](#)]