

## Article

# Cucumber Leaf Segmentation Based on Bilayer Convolutional Network

Tingting Qian <sup>1,2</sup> , Yangxin Liu <sup>3</sup>, Shenglian Lu <sup>3</sup> , Linyi Li <sup>1,2</sup>, Xiuguo Zheng <sup>1,4</sup>, Qingqing Ju <sup>4</sup>, Yiyang Li <sup>3</sup> , Chun Xie <sup>2</sup> and Guo Li <sup>3,\*</sup>

- <sup>1</sup> Shanghai Academy of Agricultural Sciences, Institute of Agricultural Science and Technology Information, Shanghai 201403, China; qiantingting@saas.sh.cn (T.Q.); lly@saas.sh.cn (L.L.); zhengxiuguo@saas.sh.cn (X.Z.)
- <sup>2</sup> Key Laboratory of Smart Agricultural Technology (Yangtze River Delta), Ministry of Agriculture and Rural Affairs, Shanghai 201403, China; 20221513129@stu.sspu.edu.cn
- <sup>3</sup> Key Lab of Multi-Source Information Mining and Security, School of Computer Science & Engineering, Guangxi Normal University, Guilin 541004, China; ac237550857@163.com (Y.L.); lsl@gxnu.edu.cn (S.L.); leonlee@stu.gxnu.edu.cn (Y.L.)
- <sup>4</sup> Shanghai Engineering Research Center of Information Technology in Agriculture, Shanghai 201403, China; 1000525995@smail.shnu.edu.cn
- \* Correspondence: liguo@gxnu.edu.cn

**Abstract:** When monitoring crop growth using top-down images of the plant canopies, leaves in agricultural fields appear very dense and significantly overlap each other. Moreover, the image can be affected by external conditions such as background environment and light intensity, impacting the effectiveness of image segmentation. To address the challenge of segmenting dense and overlapping plant leaves under natural lighting conditions, this study employed a Bilayer Convolutional Network (BCNet) method for accurate leaf segmentation across various lighting environments. The major contributions of this study are as follows: (1) Utilized Fully Convolutional Object Detection (FCOS) for plant leaf detection, incorporating ResNet-50 with the Convolutional Block Attention Module (CBAM) and Feature Pyramid Network (FPN) to enhance Region of Interest (RoI) feature extraction from canopy top-view images. (2) Extracted the sub-region of the RoI based on the position of the detection box, using this region as input for the BCNet, ensuring precise segmentation. (3) Utilized instance segmentation of canopy top-view images using BCNet, improving segmentation accuracy. (4) Applied the Varifocal Loss Function to improve the classification loss function in FCOS, leading to better performance metrics. The experimental results on cucumber canopy top-view images captured in glass greenhouse and plastic greenhouse environments show that our method is highly effective. For cucumber leaves at different growth stages and under various lighting conditions, the Precision, Recall and Average Precision (AP) metrics for object recognition are 97%, 94% and 96.57%, respectively. For instance segmentation, the Precision, Recall and Average Precision (AP) metrics are 87%, 83% and 84.71%, respectively. Our algorithm outperforms commonly used deep learning algorithms such as Faster R-CNN, Mask R-CNN, YOLOv4 and PANet, showcasing its superior capability in complex agricultural settings. The results of this study demonstrate the potential of our method for accurate recognition and segmentation of highly overlapping leaves in diverse agricultural environments, significantly contributing to the application of deep learning algorithms in smart agriculture.

**Keywords:** dense target detection; attention mechanism; instance segmentation; canopy; greenhouse



**Citation:** Qian, T.; Liu, Y.; Lu, S.; Li, L.; Zheng, X.; Ju, Q.; Li, Y.; Xie, C.; Li, G. Cucumber Leaf Segmentation Based on Bilayer Convolutional Network. *Agronomy* **2024**, *14*, 2664. <https://doi.org/10.3390/agronomy14112664>

Academic Editor: Yuxing Han

Received: 10 October 2024

Revised: 4 November 2024

Accepted: 7 November 2024

Published: 12 November 2024



**Copyright:** © 2024 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

Plant phenomics has recently been a hot research topic between agriculture and information science. The rapid and accurate measurement of plant phenotypic traits can be realized by graphics and image processing, machine vision and other technologies. As one of the most important organs of plants, leaves have undoubtedly become the object of extensive research.

Image analysis in agricultural settings presents numerous challenges, particularly regarding the segmentation of objects under complex field conditions. Agricultural images are often affected by factors such as variable lighting, cluttered backgrounds, significant object overlap and morphological variability. Many scholars have conducted research on plant leaf image analysis for a wide range of applications, such as identification of leaf diseases, extraction of leaf phenotypic parameters, leaf segmentation and growth prediction [1–5]. The segmentation of plant images has been studied for decades. Traditional image segmentation algorithms include threshold segmentation, edge detection, regional segmentation, segmentation based on mathematical morphology, segmentation based on clustering, etc. [6–11]. For example, Omrani et al. [12] applied the K-means clustering method to recognize and segment leaves of various plants. They divided the leaves of the target region according to the characteristic distribution of the cluster center before mapping back to the original image to complete segmentation under the condition of field planting. Hu et al. [13] proposed a segmentation algorithm based on a two-dimensional histogram to segment weed images. This method enhances the information association between adjacent pixels in the segmentation algorithm, improves the integrity of the connected area of the blade and reduces the influence of light intensity on the image. This traditional image segmentation method is generally used in plant image processing with less background interference and fewer leaves. However, in complex plant leaf groups, such algorithms cannot achieve ideal segmentation effect and accuracy, because the edge features of leaf overlap are not obvious [14].

Recent advancements in Convolutional Neural Networks (CNNs) have significantly propelled research the field of plant phenomics, particularly in the segmentation of plant leaves [15]. Ferro et al. [16] assessed the performance of different methodologies, including Mask R-CNN, U-Net, OBIA and unsupervised methods, in identifying pure canopy pixels. Vayssade [17], based on recent Convolutional Neural Network mechanisms, proposes a pixelwise instance segmentation to detect leaves in dense foliage environment. It combines “Deep Contour Aware” (to separate the inner area of big leaves from their edges), “Leaf Segmentation Through Classification of Edges” (to separate instances with specific inner edges) and “Pyramid CNN for Dense Leaves” (to consider edges on different scales).

After Mask R-CNN emerged, Xu et al. [18] utilized the model to explore leaf instance segmentation and counting. Compared with traditional image segmentation methods, Mask R-CNN greatly improves the segmentation accuracy of plant leaves in complex backgrounds. The outstanding results for leaf instance segmentation and counting show that object detection and segmentation algorithms based on deep convolutional networks may be promising tools for plant phenotype studies. Kuznichov et al. [19] proposed a method for recording geometric structure information about specific plant leaves to make the target leaves resemble the leaves in the real scene as closely as possible. Subsequently, based on Mask R-CNN, they adopted the non-maximum suppression method to solve the blocking problem in training the full CNN model and improved the accuracy of plant leaf instance segmentation in dense scenes.

Beyond instance segmentation, other approaches utilizing deep learning have been explored. Praveen et al. [20] proposed a plant leaf segmentation algorithm based on the Deep Convolution Neural Network (DCNN) to extract leaf information in the target region and perform rosette leaf segmentation using orthogonal transformation. Jin et al. [21] proposed a CNN based on plant voxel clustering for the classification and segmentation of corn stems and leaves. Lai et al. [22] proposed a 3D point cloud segmentation method for plant leaves based on deep learning. Lu et al. [23] proposed a fast plant leaf segmentation method based on CenterNet. By optimizing the network structure and detection loss function in CenterNet, the network can accurately extract edge information about plant leaves in dense scenes. However, the model lacks an instance segmentation network, which increases the workload of subsequent plant phenotypic parameter extraction. Lou et al. [24] proposed a plant stem and leaf instance segmentation method combining multi-view time-series images and depth images using Mask R-CNN to solve the difficulty of extracting

phenotypic parameters caused by the dynamic changes of plants at different time points in the growth process.

Another innovative approach is the Bilayer Convolutional Network (BCNet), which aims to tackle the segmentation of dense and overlapping leaves under various lighting conditions. BCNet serves as a central component in the image processing workflow, leveraging the capabilities of Fully Convolutional Object Detection (FCOS) for plant leaf detection and incorporating advanced techniques such as the Convolutional Block Attention Module (CBAM) and Feature Pyramid Network (FPN) [25]. Ke et al. [26] validate the efficacy of bi-layer decoupling on both one-stage and two-stage object detectors with different backbones and network layer choices. Despite its simplicity, extensive experiments on COCO and KINS show that occlusion-aware BCNet achieves significant and consistent performance gains, especially in heavy occlusion cases.

The segmentation of plant leaves in images holds substantial importance in agricultural research. Although image segmentation technology has a long history, the complexity of agricultural scenes, characterized by dense distributions and significant overlaps of plant organs, makes it challenging to establish a universal method for plant images. Consequently, the choice of algorithm must be tailored to specific application scenarios.

This study addresses the significant challenge posed by the incomplete shapes of organs due to mutual occlusion in complex agricultural production environments, which greatly hampers segmentation algorithms. To enhance the segmentation accuracy of densely packed and heavily occluded cucumber leaves in real-world production settings, we propose the following methods:

- (1) **Multi-Scale Strategy and Dilated Convolutions:** Considering the significant morphological changes in plant leaves during growth, we employ a multi-scale strategy based on deep learning to divide the original images into blocks. Subsequently, dilated convolutions are used for fusion, allowing better extraction of features from leaves of varying shapes.
- (2) **Attention Mechanism for Edge Features:** The model's ability to represent the edges of overlapping plant leaves is often insufficient. By incorporating an attention mechanism after the feature maps, the model can more accurately capture critical edge features of each leaf, facilitating improved leaf segmentation.
- (3) **Varifocal Loss for Dense Regions:** In images with densely distributed plant leaves, CNNs often overlook many target features. We replace the Focal Loss in BCNet with the Varifocal Loss from VarifocalNet, which is more effective for detecting dense targets.

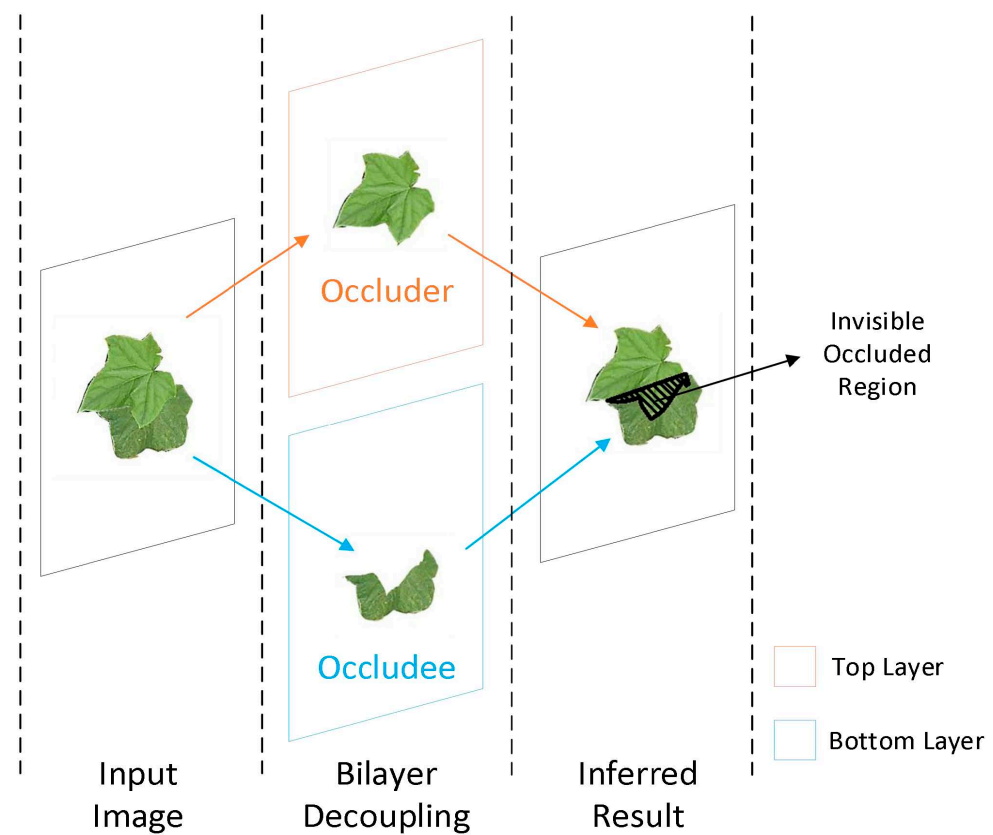
We anticipate that these techniques will enable precise identification and instance segmentation of cucumber plant leaves in greenhouse conditions. This approach is expected to lay a solid foundation for subsequent tasks, such as leaf phenotypic parameter extraction and further advancements in agricultural research.

## 2. Materials and Methods

### 2.1. BCNet

BCNet is a two-layer instance segmentation network based on image occlusion perception. It replaces the traditional single mask prediction branch network with a two-graph hierarchical associative graph neural network [26]. The main idea is to model the RoI in the image as two graph convolution network (GCN) layers. The schematic diagram of two-layer decomposition is shown in Figure 1. The upper GCN is responsible for detecting occluder objects. The underlying GCN is responsible for estimating the occludee part of overlapping targets. Most instance segmentation errors arise from the presence of two overlapping objects within the same region of interest (RoI), which can obscure each other's true contours. This issue is particularly pronounced when the overlapping objects belong to the same class and have similar texture features, leading to increased segmentation errors. BCNet improves the performance of the instance segmentation model in the face

of complex overlapping objects in the image by modeling the interfaces of overlapping regions and improving the effect of mask branch prediction.

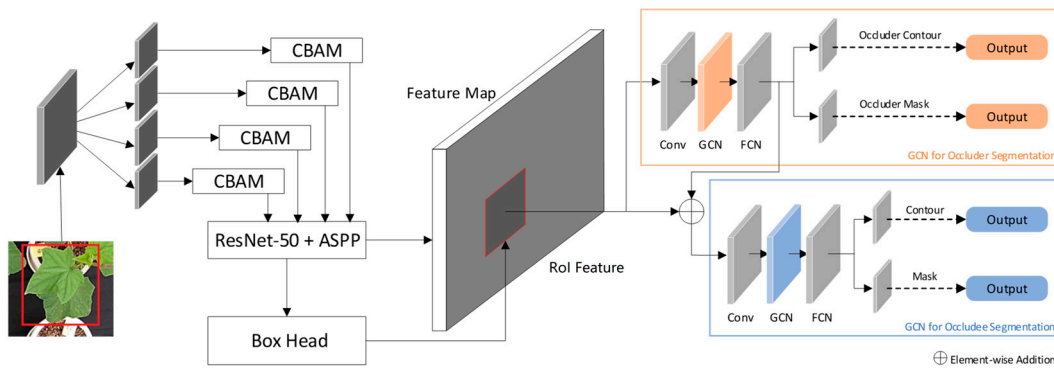


**Figure 1.** Bilayer decomposition diagram.

## 2.2. Algorithm Pipeline

The algorithm structure framework used in this experiment is shown in Figure 2. The algorithm flow is divided into the following steps:

- (1) Image feature extraction and ROI selection: The object detection algorithm based on FOCS [27] is used to predict the ROI in the image. ResNet-50 [28] is used as the backbone, and FPN [29] is used to extract the features of the whole image.
- (2) According to the position of the target object detection frame, the ROI Align [30] algorithm is used to obtain the ROI sub-region in the feature map in the image, and this region is used as the input of BCNet.
- (3) Instance segmentation through BCNet: Firstly, the target characteristics of ROI are inputted through the graph convolution network layer at the top, the appearance of the upper target is modeled, and the mask and edge of the upper target in the box of interest are outputted. Secondly, combined with the upper target data extracted from the convolution network of the top graph and the features of the ROI sub-region obtained using the ROI Align algorithm, the new features of the occluded target are obtained by adding and inputting the convolution network layer into the bottom graph. Finally, according to the above characteristics of the occluder object and the occludee object in the region, the instance segmentation task of the target is completed.



**Figure 2.** Flow chart of image segmentation based on improved BCNet.

### 2.3. Improved Scheme

Due to problems related to dense targets and serious mutual occlusion in the leaf images of plant groups in the real planting environment, this experiment has made two improvements to the algorithm:

- (1) In order to accurately extract edge information on the overlapped parts of plant leaves, this study introduces an attention mechanism module after the last convolution layer of the ResNet-50 backbone network to allow the model to pay more attention to key information concerning the edge features of each leaf, to realize the accurate segmentation of plant leaves.
- (2) Due to the presence of numerous densely packed areas of plant leaves in the images, CNNs often overlook a lot of target feature information. To address this issue, we use the Varifocal Loss function of VarifocalNet to replace the classification loss function, Focal Loss, in FCOS, which is conducive to detecting dense targets.

#### 2.3.1. CBAM

In the design of a deep neural network, the function of the attention mechanism is to reallocate the object’s weight. Cucumber leaves in real planting environments have a large number of overlaps, and the leaf features in different growth periods are obviously different. The introduction of an attention mechanism can effectively suppress the non-target features in the image, enhance the appearance and shape features of the target leaves, and improve the accuracy of model detection and segmentation. The attention mechanism we use is CBAM, a lightweight module that combines channel and spatial attention mechanisms [31].

The channel attention module compresses the characteristic graph in the spatial dimension and obtains two one-dimensional vectors after being processed by the average value pooling and maximum value pooling functions. The realization of this module is to obtain the channel attention characteristic graph by adding a multilayer neural network and activating the sigmoid function. The specific formula is as follows:

$$\begin{aligned}
 M_c(F) &= \sigma(MLP(AvgPool(F)) \\
 &\quad + MLP(MaxPool(F))) \\
 &= \sigma(W_1(W_0(F_{avg}^c) + W_1(W_0(F_{max}^c))))
 \end{aligned}
 \tag{1}$$

where  $\sigma$  represents sigmoid activation function.  $F$  represents the input characteristic diagram.  $F_{avg}^c$  and  $F_{max}^c$  are the characteristic graphs obtained after global average pooling and maximum pooling, respectively.  $W_0$  and  $W_1$  respectively represent two parameters under the multi-layer sensing mechanism.

Due to the different levels of importance of spatial position information about the target area in the image, we use the spatial attention module to focus on the key information when we segment the image. The spatial attention module compresses the channel and pools the average and maximum values to obtain the spatial attention matrix. It then determines the target location in the matrix through the semantic information on the

feature graph. This module divides the feature map of a single dimension channel into several sub-features. The specific calculation formula is as follows:

$$\begin{aligned} M_s(F) &= \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\ &= \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \end{aligned} \quad (2)$$

where the convolution layer indicated by  $f$  in this section uses a size of  $7 \times 7$  Convolution kernel. We added the CBAM module behind the last convolution layer of ResNet-50. We then weighted the results to obtain the improved feature map.

### 2.3.2. Loss Function

In the training process of the model, the target detection and instance segmentation framework are monitored by the multi process loss function, which is defined as follows:

$$\begin{aligned} L &= aL_{Detect} + L_{Occluder} + L_{Occludee} \\ L_{Occluder} &= a_1L'_{Occ-B} + a_2L'_{Occ-S} \\ L_{Occludee} &= a_3L'_{Occ-B} + a_4L'_{Occ-S} \end{aligned} \quad (3)$$

$L_{Occluder}$  and  $L_{Occludee}$  respectively represent the loss function for each instance segmentation subtask of the occluder object and the occludee object.  $L'_{Occ-B}$  and  $L'_{Occ-S}$  are nonlinear transformation functions for modeling occluded objects. They represent boundary detection and mask segmentation loss of the underlying GCN, respectively. The parameters  $a$ ,  $a_1$ ,  $a_2$ ,  $a_3$  and  $a_4$  are used to balance the weight of the loss functions. The target detection loss function  $L_{Detect}$  in FCOS includes the classification loss function branch  $L_{Class}$ , center point offset value branch loss function  $L_{Centerness}$  and regression branch loss function  $L_{Regression}$ .  $L_{Detect}$  is defined as follows:

$$L_{Detect} = L_{Class} + L_{Centerness} + L_{Regression} \quad (4)$$

The classification loss function  $L_{Class}$  employs Focal Loss to address the issue of class imbalance in target prediction indices through a balanced cross-entropy approach. During the down-sampling process at certain magnifications, dense target pixels can be lost, diminishing the effectiveness of Focal Loss in extracting features from these targets. Intersection over Union (IoU) serves as a standard metric for measuring the accuracy of object detection within a specific data set. To enhance this, Varifocal Loss predicts IoU-aware classification scores by training dense target detectors on the foundation of Focal Loss, subsequently sorting the detection results [32]. This approach aims to rectify the imbalance between foreground and background in dense target detector training, thereby improving feature extraction for dense targets. In cucumber leaf images captured under real planting conditions, targets are densely distributed, and many small targets are present. By replacing the original category branch loss function  $L_{Class}$  with Varifocal Loss, the training process can converge more easily, target information loss during down-sampling can be mitigated and pixel outliers can be more effectively regressed during the target detection process. Varifocal Loss is defined as follows:

$$VFL(p, q) = \begin{cases} -q(q \log(p) + (1 - q) \log(1 - p)), & q > 0 \\ -bp^\gamma \log(1 - p), & q = 0 \end{cases} \quad (5)$$

where  $p$  is the predicted IoU-aware classification score,  $p^\gamma$  is the weight of different samples,  $q$  is the target IoU score and  $b$  is the parameter used to balance the sample weight.

### 2.4. Data Acquisition and Model Training

In order to verify the effectiveness of the improved algorithm described in this paper for plant leaf segmentation in real planting environments, we observed several different growth stages of planting observed in glass and plastic greenhouses, and two different light-intensity cucumber plant leaf images taken on sunny and cloudy days.

### 2.4.1. Image Acquisition

The experimental material, cucumber canopy, was grown in both glass and plastic greenhouses at the experimental base at the Shanghai Academy of Agricultural Sciences, China from 21 May 2019–4 July 2019. In the glass greenhouse, we cultivated 16 potted cucumber plants arranged in a  $4 \times 4$  grid. In the plastic greenhouse, we selected 10 continuously arranged cucumber plants planted in soil, arranged in a  $2 \times 5$  grid. The potted plants in the glass greenhouse were placed on a black background cloth, while the soil surface in the plastic greenhouse was covered with black plastic film, leaving the pathways as exposed soil. Image monitoring began at the 5-leaf stage and continued through three growth stages: the 5-leaf to 8-leaf vining stage (Early Growth Stage), the 8-leaf to 12-leaf flowering stage (Metaphase Growth Stage) and the 12-leaf to 20-leaf fruiting stage (Terminal Growth Stage).

To capture the images, a network camera (DS-2CD40C5F-AP, Hikvision Co., Ltd., Hangzhou, China) was positioned 5 meters above the ground directly over the cucumber canopy, ensuring a full view of all the cucumber leaves. Photographs were taken every two hours from 7:50 a.m. to 3:50 p.m. daily. A total of 290 images were collected in the glass greenhouse and 110 images in the plastic greenhouse, covering the early, middle and late stages of cucumber plant development. These images included various lighting conditions such as sunny and cloudy days, as well as strong light at noon and weak light in the morning. According to the classification of cucumber growth stages and the light intensity during shooting, the classification of picture data under the two planting environments is shown in Table 1. We divided the images of cucumber canopy taken in the glass greenhouse into six datasets, selecting 230, 38 and 22 images as the training set, validation set and test set, respectively. Similarly, the images of cucumber canopy taken in the plastic greenhouse were divided into four datasets, with 78, 20 and 12 images selected as the training set, validation set and test set, respectively.

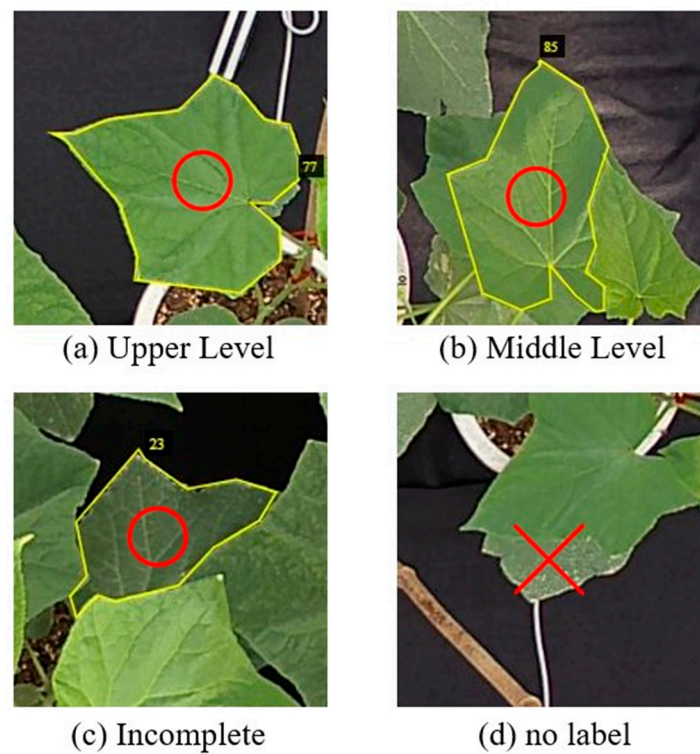
**Table 1.** Classification scheme of image data set.

Picture Category	Image in Glass Greenhouse	Image in Plastic Greenhouse
Early Growth Stage, Sunny	38	30
Early Growth Stage, Cloudy	34	36
Metaphase Growth Stage, Sunny	50	-
Metaphase Growth Stage, Cloudy	61	-
Terminal Growth Stage, Sunny	41	23
Terminal Growth Stage, Cloudy	66	21
Total	290	110

### 2.4.2. Image Annotation

VGG Image Annotator (VIA) was used to annotate images. As shown in Figure 3, according to the approximate estimation of the sheltered area, cucumber leaves are divided into four categories for marking:

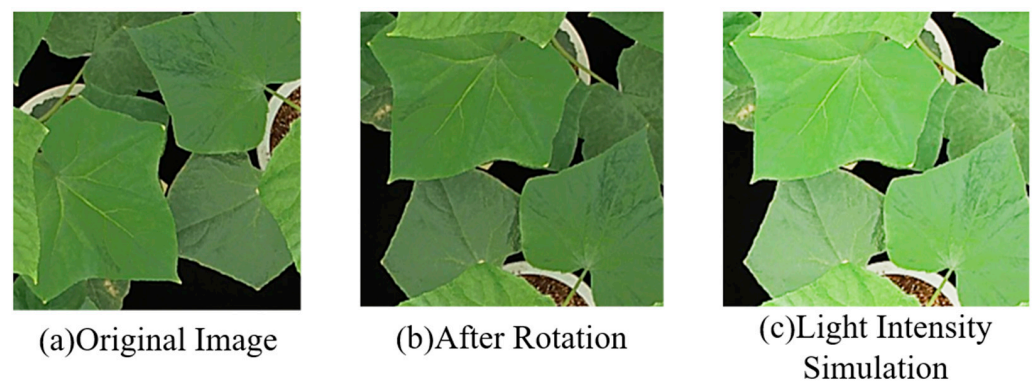
- (1) If the covered area of the blade is less than 20%, they are marked as Upper Level.
- (2) If the covered area of the blade accounts for about 20–60% of the whole blade area, they are marked as Middle Level.
- (3) If the covered area of leaves exceeds 60%, but the shape characteristics of cucumber leaves are retained, they are marked as Incomplete.
- (4) Targets with an occluded area of more than 60% and no cucumber leaf shape features will not be marked.



**Figure 3.** Schematic diagram of image annotation method. ○ means labeled, × means unlabeled.

#### 2.4.3. Image Augmentation

After completing the data annotation, in order to improve the richness of the experimental data set, a total of 308 images in the training set were expanded. As shown in Figure 4, the specific operations include flipping the image and adjusting the contrast. By this amplification method, 1150 and 390 images of cucumber plants planted in the glass greenhouse and the plastic greenhouse were obtained, respectively, which were used as the data set for model training for subsequent experiments.



**Figure 4.** Schematic diagram of image expansion scheme.

#### 2.4.4. Model Training

The experimental environment for cucumber population leaf segmentation in this experiment is shown in Table 2. The operating system is Ubuntu 18.04. The CPU is Intel (R) Xeon(R) gold 6230; the GPU is Tesla V100s-PCIE-32GB, and the basic software environment is Python 3.7, Pytorch 1.4.0, OpenCV 4.4.0, CUDA 10.1 and CUDNN 7.6.5.



**Table 2.** Experimental environment configuration.

Project	Configuration
Operating System	Ubuntu 18.04
CPU	Intel(R) Xeon(R) Gold 6230
GPU	Tesla V100S-PCIE-32GB
Video Memory	32 G
Memory	72 G
Programing Language	Python 3.7

The data set format used in this experiment is the COCO data set format converted after VIA annotation. When training the model, it is necessary to generate a double-layer mask annotation file from the annotation file of the training set through the layer division dataset conversion script under the network. The annotation file of the verification set remains unchanged. ResNet-50 is used as the backbone network. Model training and evaluation using a single GPU. The initial learning rate of the model is set to 0.0005. With the passage of training time, the learning rate will become half the current value every 20 epochs. In order to compare with Faster R-CNN, Mask R-CNN, YOLOv4, PANet and other deep learning models [30,33–36], it is necessary to make data sets that meet the requirements of these algorithms. The structure of data sets and directories is similar to that of COCO data sets. The detection framework parses COCO annotations into corresponding formats, avoiding the process of changing scripts.

### 2.5. Evaluating Indicator

The *Precision* and *Recall* evaluation indexes for cucumber leaf recognition and segmentation results are used, while, according to the size of the target in the image, the Average Precision (AP), the average precision of IoU = 0.50 ( $AP_{50}$ ) and the average precision of IoU = 0.75 ( $AP_{75}$ ) evaluation metric from the COCO dataset are used to evaluate segmentation precision.

## 3. Results

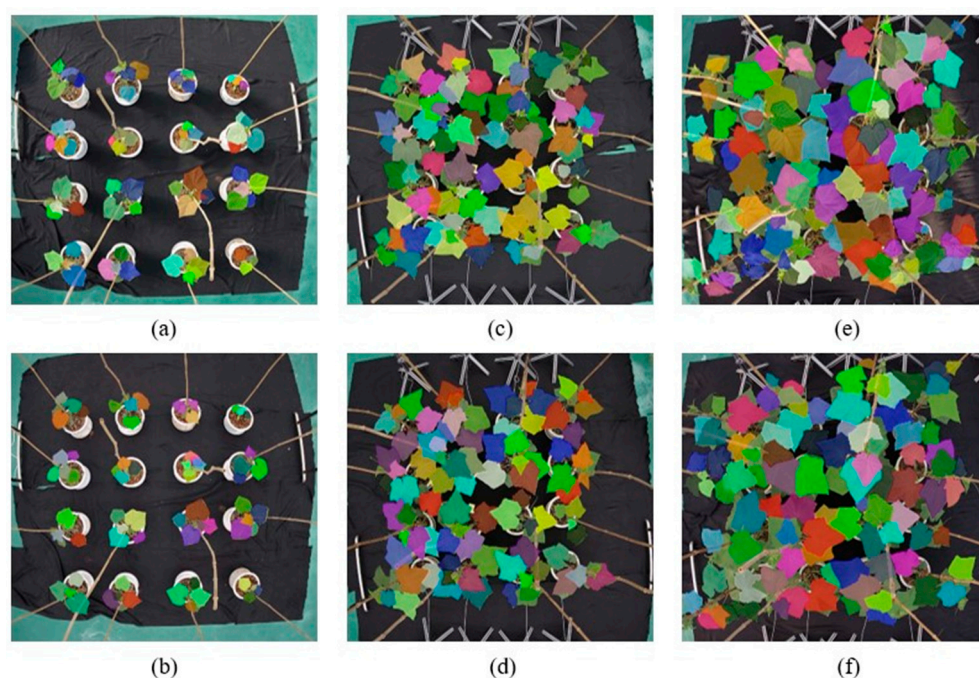
### 3.1. Results and Analysis of Glass Greenhouse Image Segmentation

Segmentation experiments were conducted on cucumber plant population leaves obtained from glass greenhouses under different growth states and different light intensities.

$AP^b$  and  $AP^s$  represent the average accuracy of detection and the average accuracy of instance segmentation, respectively. The experimental results are shown in Table 3. The test results show the average recognition accuracy value of cucumber leaves in glass greenhouse that  $AP^b$  is 96.57%,  $AP_{50}^b$  is 95.05% and  $AP_{75}^b$  is 93.96%. The accuracy of leaf recognition can reach 99% for cucumber plants in the early growth stage under sunny conditions. The leaf recognition accuracy of cucumber plants at the end of cloudy days can also reach 92.66%. Regarding the average instance segmentation accuracy,  $AP^s$  is 84.71%,  $AP_{50}^s$  is 95.05% and  $AP_{75}^s$  is 93.96%. The model has the highest segmentation accuracy of 87.54% for the leaves of cucumber plants in the early growth stage on sunny days and 81.29% for the leaves of cucumber plants in the late growth stage on cloudy days. The segmentation result is shown in Figure 5.

**Table 3.** Segmentation accuracy of cucumber plant image in glass greenhouse.

Picture Category	$AP^b$	$AP_{50}^b$	$AP_{75}^b$	$AP^s$	$AP_{50}^s$	$AP_{75}^s$
Early Growth Stage, Sunny	0.9900	0.9900	0.9630	0.8754	0.9304	0.9304
Early Growth Stage, Cloudy	0.9874	0.9871	0.9627	0.8730	0.9307	0.9304
Metaphase Growth Stage, Sunny	0.9857	0.9894	0.9795	0.8653	0.9295	0.9287
Metaphase Growth Stage, Cloudy	0.9831	0.9872	0.9795	0.8579	0.9202	0.9198
Terminal Growth Stage, Sunny	0.9370	0.9208	0.9107	0.8235	0.8908	0.8857
Terminal Growth Stage, Cloudy	0.9266	0.9117	0.9028	0.8129	0.8907	0.8759
Average	0.9657	0.9504	0.9396	0.8431	0.9723	0.9078

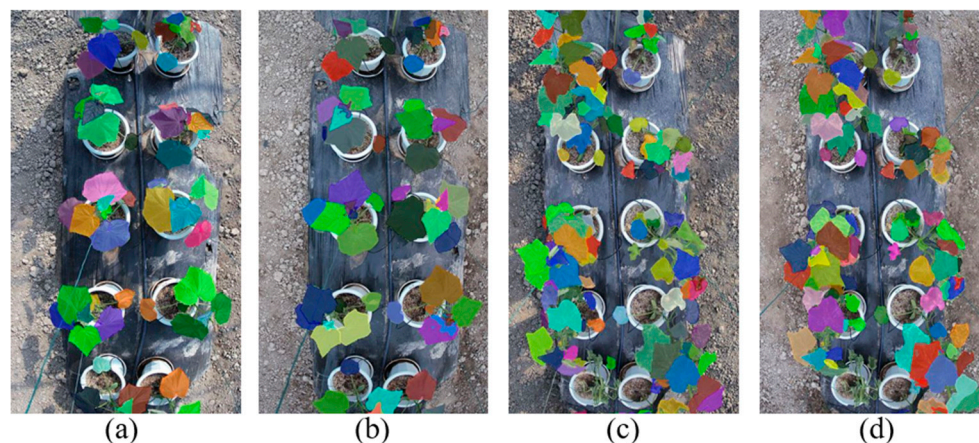
**Figure 5.** Segmentation effect of cucumber plant images in glass greenhouses. (a) Early Growth Stage, Sunny. (b) Early Growth Stage, Cloudy. (c) Metaphase Growth Stage, Sunny. (d) Metaphase Growth Stage, Cloudy. (e) Terminal Growth Stage, Sunny. (f) Terminal Growth Stage, Cloudy.

### 3.2. Results and Analysis of Plastic Greenhouse Image Segmentation

The parameter statistics of the experimental results are shown in Table 4. It can be seen that the improved algorithm in this study has the same excellent effect on the leaf segmentation of cucumber plants in different growth environments, in which the best value of the target recognition accuracy is 98.33% and the average value is 96.86%. The most accurate instance segmentation is 86.75% and the average is 83.27%. Figure 6 shows the experimental results of cucumber plant population leaves in plastic greenhouses, including the images of early growth stage and late growth stage on sunny and cloudy days.

**Table 4.** Segmentation accuracy of cucumber plant images in plastic greenhouses.

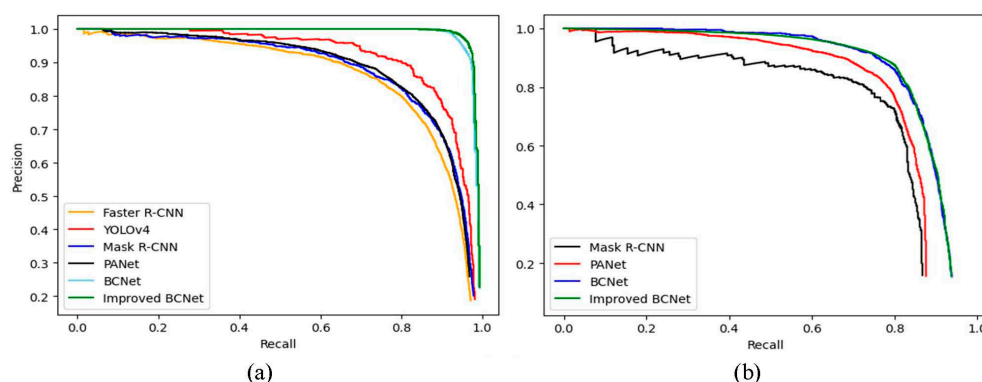
Picture Category	$AP^b$	$AP_{50}^b$	$AP_{75}^b$	$AP^s$	$AP_{50}^s$	$AP_{75}^s$
Early Growth Stage, Sunny	0.9828	0.9810	0.9550	0.8675	0.9274	0.9170
Early Growth Stage, Cloudy	0.9833	0.9804	0.9527	0.8628	0.9279	0.9185
Terminal Growth Stage, Sunny	0.9616	0.9499	0.9398	0.8123	0.8793	0.8728
Terminal Growth Stage, Cloudy	0.9608	0.9423	0.9418	0.8101	0.8760	0.8721
Average	0.9686	0.9520	0.9423	0.8327	0.8710	0.8708



**Figure 6.** Effect of cucumber plant image segmentation in plastic greenhouses. (a) Early Growth Stage, Sunny. (b) Early Growth Stage, Cloudy. (c) Terminal Growth Stage, Sunny. (d) Terminal Growth Stage, Cloudy.

### 3.3. Comparison and Analysis of Different Models

In order to verify the effect of the improved algorithm, we compared it with other deep learning algorithms. Cucumbers grown in a glass greenhouse were selected as the experimental subjects. Comparative experiments were performed using detection algorithms, such as Faster R-CNN and YOLOv4, as well as instance segmentation algorithms, including Mask R-CNN, PANet and the original BCNet. The test effect of each algorithm is shown in Figure 7, where Figure 7a is the P-R curve of target detection effect and Figure 7b is the P-R curve of instance segmentation effect.

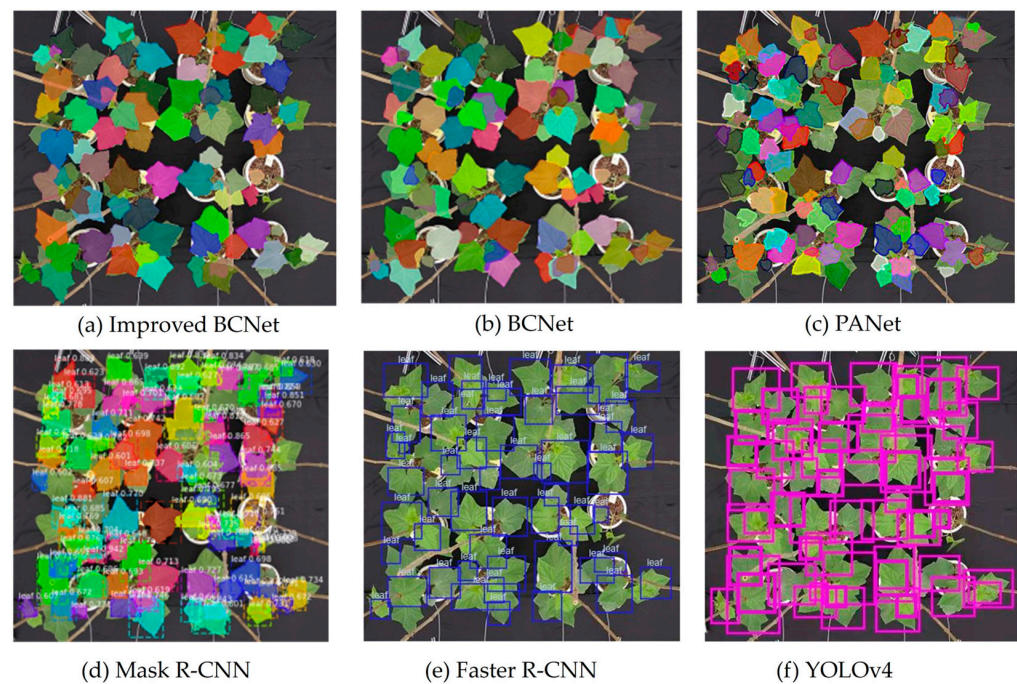


**Figure 7.** Target recognition and instance segmentation P-R curve of six models: (a) Object detection P-R curve; (b) Example split P-R curve.

The experimental results of each method are shown in Table 5. From the data in the table, it can be seen that the average detection accuracy of the improved BCNet for cucumber plant population leaves is improved by nearly 1.5% compared with the original BCNet [37]. It is 5–15% higher than YOLOv4, PANet, Mask R-CNN and Faster R-CNN. The average instance segmentation accuracy of our improved algorithm for cucumber leaves is slightly improved compared with the original BCNet and is 10% and 22% higher than that of PANet and Mask R-CNN, respectively. This shows that our improved BCNet is better than other deep learning algorithms in the segmentation of cucumber plant population leaves. The effect of each method on the identification and example segmentation of cucumber plant population leaves is shown in Figure 8.

**Table 5.** Comparison of parameters of different models.

Model	Object Detection			Instance Segmentation		
	Precision	Recall	AP	Precision	Recall	AP
Faster R-CNN	0.84	0.81	0.8200	-	-	-
Mask R-CNN	0.87	0.82	0.8467	0.68	0.64	0.6614
YOLOv4	0.93	0.88	0.9100	-	-	-
PANet	0.88	0.86	0.8770	0.76	0.73	0.7467
BCNet	0.96	0.92	0.9513	<b>0.87</b>	0.82	0.8389
Improved BCNet	<b>0.97</b>	<b>0.94</b>	<b>0.9657</b>	<b>0.87</b>	<b>0.83</b>	<b>0.8471</b>

**Figure 8.** Effect of cucumber plant image segmentation in glass greenhouse: (a) improved BCNet; (b) BCNet; (c) PANet; (d) Mask R-CNN; (e) YOLOv4; (f) Faster R-CNN.

The FLOPs (floating point operations) and detection time for one image are also two important indicators for evaluating the model. As Table 6 shows, our improved model has relatively small FLOPs and fast inference speed.

**Table 6.** Comparison of FLOPs and detection time for one image.

Model	FLOPs	Detection Time (s)
Faster R-CNN	181 M	3.52
Mask R-CNN	286 M	3.13
YOLOv4	90 M	0.46
BCNet	207 M	3.20
Improved BCNet	155 M	1.72

## 4. Discussion

### 4.1. Model Performance Evaluation

Segmenting highly-overlapping objects is challenging, because typically no distinction is made between real object contours and occlusion boundaries [27]. This difficulty is particularly pronounced in agricultural monitoring, where significant overlap and occlusion between similarly colored leaves, along with complex backgrounds and boundary

blurriness caused by leaf movement, present substantial challenges for independent segmentation. BCNet employs a dual-stream architecture that simultaneously learns to identify object boundaries and context features, enabling it to differentiate between overlapping leaves more effectively. The data in Table 5 indicate that while detection algorithms such as YOLOv4, as well as instance segmentation algorithms of the original BCNet, show improvements in object detection and instance segmentation, they may not match the overall robustness of our improved BCNet. This is achieved by incorporating ResNet-50 with the CBAM and FPN for RoI feature extraction and applying the Varifocal Loss Function to improve the classification loss function in FCOS. Recent studies have demonstrated that CBAM enhance the ability of BCNet to dynamically adjust focus to the most informative spatial regions, leading to more precise delineation of leaf boundaries [38], while the Varifocal Loss Function has robust performance in dense object detection [33,39].

The high average recognition accuracy of 96.57% underscores the model's capability to accurately identify cucumber leaves, achieving peak performance of 99% during early growth stages under sunny conditions, while still maintaining a commendable accuracy of 92.66% at the end of cloudy days. The average instance segmentation accuracy shows a maximum accuracy of 87.54% observed in early growth stages on sunny days, while accuracy slightly decreased to 81.29% in the later growth stages under cloudy conditions. In the study by Ferro et al. [16], a higher accuracy during the early stages of crop growth can also be observed compared to measurements taken in the later stages. Cucumber plants often have leaves that overlap and occlude one another. While the improved BCNet is adept at detecting individual leaves, accurately segmenting these overlapping regions remains challenging. Accurate segmentation relies heavily on precise boundary detection. Ngugi et al. [40] proposed a modified U-Net model '*KijaniNet*' for automatic background removal from tomato leaf images under complex field conditions, which demonstrates superior performance in leaf segmentation, achieving over 0.96 mIoU and 0.91 mBFScore on the test set. Talasila et al. [41] designed a customized deep learning network '*PLRSNet (Plant Leaf Region Segmentation Net)*' to accurately segment leaf regions from complex field images, addressing the challenge of varied morphology and natural artifacts. Despite the promising results achieved in leaf segmentation by these studies, they, like most segmentation tasks, primarily focus on images containing a single leaf. Segmenting multiple highly overlapping leaves at the canopy scale remains a significant challenge in segmentation tasks.

#### 4.2. Further Application of Leaf Segmentation in Phenotypic Data Collection

Leaf segmentation plays a critical role in advancing phenotypic data collection, which is essential for various agricultural and biological research applications. By accurately segmenting individual leaves from complex images, researchers can obtain precise measurements of leaf morphology, size and health status [42]. In leaf segmentation, it is easy to be affected by the surrounding environment, particularly lighting conditions. Shadows cast on the leaves by lighting can often be mistakenly detected as separate leaves [16]. The study results by Amean et al. [43] indicate that their segmentation model achieves an accuracy of 69% under sunny lighting conditions and 71% under shady and cloudy conditions. However, our results demonstrate no significant difference in accuracy between sunny and cloudy lighting conditions (Tables 3 and 4). The reason lies in the lighting conditions of glass and plastic greenhouses, which predominantly feature diffused light, resulting in more uniform illumination and smaller shadow areas. Additionally, the attention mechanism introduced in our method enhances boundary perception, making the shadowed parts of the boundaries more conducive to accurate segmentation on sunny days.

For corner-type plant structures like tomatoes and cucumbers, there is a noticeable height difference between the leaves. Segmentation methods that integrate spatial point cloud information will be more advantageous for separating plant organs. In the future, to achieve higher segmentation accuracy, efforts could be made to combine depth images with RGB images.

## 5. Conclusions

This study employs a segmentation algorithm based on a bilayer convolutional network to segment the leaves of cucumber plant populations. The experimental results demonstrate an average detection accuracy of 96.57% for cucumber leaves in a glass greenhouse and 96.86% in a plastic greenhouse, with corresponding instance segmentation accuracies of 84.71% and 83.27%, respectively. Additionally, our improved algorithm outperforms commonly used deep learning algorithms such as Faster R-CNN, Mask R-CNN, YOLOv4 and PANet, indicating its capability to accurately segment cucumber leaves under real planting conditions.

Despite these promising results, certain limitations must be acknowledged. The bilayer convolutional network may struggle with variability in leaf shapes and sizes, particularly in diverse crop environments. Moreover, the algorithm's performance may decline when applied to other plant species, such as tomatoes or eggplants, which possess different leaf morphologies. Future work will involve conducting segmentation experiments on plants with varying leaf shapes and exploring new object detection and instance segmentation algorithms. By applying these methods in more diverse and realistic planting environments, we aim to enhance the robustness and applicability of our approach in precision agriculture.

**Author Contributions:** Conceptualization, T.Q. and S.L.; methodology, Y.L. (Yangxin Liu); validation, Y.L. (Yangxin Liu), X.Z. and Q.J.; formal analysis, Y.L. (Yangxin Liu) and Y.L. (Yiyang Li); investigation, T.Q. and C.X.; resources, T.Q.; writing—original draft preparation, T.Q., Y.L. (Yangxin Liu), S.L. and L.L.; writing—review and editing, T.Q. and G.L.; supervision, G.L.; project administration, L.L.; funding acquisition, T.Q. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the Shanghai Agriculture Applied Technology Development Program, China (Grant No. 2023-02-08-00-12-F04621) and the Shanghai Science and Technology Committee Program (Grant No. 21N21900700).

**Data Availability Statement:** The data used to support the findings of this study are available from the corresponding author upon request.

**Acknowledgments:** The authors would like to thank the Phenotyping Innovation Team of the Shanghai Academy of Agricultural Sciences for their support regarding the experimental conditions.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Sudhesh, K.M.; Sowmya, V.; Kurian, S.; Sikha, O.K. AI based rice leaf disease identification enhanced by Dynamic Mode Decomposition. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105836.
2. Zhang, H.; Wang, L.; Jin, X.; Bian, L.; Ge, Y. High-throughput phenotyping of plant leaf morphological, physiological, and biochemical traits on multiple scales using optical sensing. *Crop J.* **2023**, *11*, 1303–1318. [[CrossRef](#)]
3. Pieruschka, R.; Schurr, U. Plant phenotyping: Past, present, and future. *Plant Phenomics* **2019**, *2019*, 7507131. [[CrossRef](#)] [[PubMed](#)]
4. Liu, X.; Hu, C.; Li, P. Automatic segmentation of overlapped poplar seedling leaves combining Mask R-CNN and DBSCAN. *Comput. Electron. Agric.* **2020**, *178*, 105753. [[CrossRef](#)]
5. Agustini, E.P.; Gernowo, R.; Wibowo, A.; Warsito, B. Bibliometric Analysis: Research Trends in Leaf Image Segmentation and Classification. In Proceedings of the 2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS), Virtual, 22–23 February 2024; pp. 1–5.
6. Guerra Ibarra, J.P.; Cuevas de la Rosa, F.J.; Arellano Arzola, O. Segmentation of Leaves and Fruits of Tomato Plants by Color Dominance. *AgriEngineering* **2023**, *5*, 1846–1864. [[CrossRef](#)]
7. Yang, R.; Wu, Z.; Fang, W.; Zhang, H.; Wang, W.; Fu, L.; Majeed, Y.; Li, R.; Cui, Y. Detection of abnormal hydroponic lettuce leaves based on image processing and machine learning. *Inf. Process. Agric.* **2023**, *10*, 1–10. [[CrossRef](#)]
8. Abdul-Nasir, A.S.; Mashor, M.Y.; Mohamed, Z. Colour image segmentation approach for detection of malaria parasites using various colour models and k-means clustering. *WSEAS Trans. Biol. Biomed.* **2013**, *10*, 41–55.
9. Zhang, X.; Li, M.; Liu, H. Overlap functions-based fuzzy mathematical morphological operators and their applications in image edge extraction. *Fractal Fract.* **2023**, *7*, 465. [[CrossRef](#)]
10. Nikbakhsh, N.; Baleghi, Y.; Agahi, H. A novel approach for unsupervised image segmentation fusion of plant leaves based on G-mutual information. *Mach. Vis. Appl.* **2021**, *32*, 5. [[CrossRef](#)]
11. Rauf, H.T.; Saleem, B.A.; Lali, M.I.; Khan, M.A.; Sharif, M.; Bukhari, S.A. A citrus fruits and leaves dataset for detection and classification of citrus diseases through machine learning. *Data Brief* **2019**, *26*, 104340. [[CrossRef](#)]

12. Omrani, E.; Khoshnevisan, B.; Shamshirband, S.; Saboohi, H.; Anuar, N.B.; Nasir, M.H.N.M. Potential of radial basis function-based support vector regression for apple disease detection. *Measurement* **2014**, *55*, 512–519. [[CrossRef](#)]
13. Hu, B.; Mao, H.; Zhang, Y. Weed image segmentation algorithm based on two-dimensional histogram. *Trans. Chin. Soc. Agric. Mach.* **2007**, *38*, 199–202.
14. Williams, D.; Macfarlane, F.; Britten, A. Leaf only SAM: A segment anything pipeline for zero-shot automated leaf segmentation. *Smart Agric. Technol.* **2024**, *8*, 100515. [[CrossRef](#)]
15. Fang, J.; Jiang, H.; Zhang, S.; Sun, L.; Hu, X.; Liu, J.; Gong, M.; Liu, H.; Fu, Y. BAF-Net: Bidirectional attention fusion network via CNN and transformers for the pepper leaf segmentation. *Front. Plant Sci.* **2023**, *14*, 1123410. [[CrossRef](#)]
16. Ferro, M.V.; Sørensen, C.G.; Catania, P. Comparison of different computer vision methods for vineyard canopy detection using UAV multispectral images. *Comput. Electron. Agric.* **2024**, *225*, 109277. [[CrossRef](#)]
17. Vayssade, J.A.; Jones, G.; Gée, C.; Paoli, J.N. Pixelwise instance segmentation of leaves in dense foliage. *Comput. Electron. Agric.* **2022**, *195*, 106797. [[CrossRef](#)]
18. Xu, L.; Li, Y.; Sun, Y.; Song, L.; Jin, S. Leaf instance segmentation and counting based on deep object detection and segmentation networks. In Proceedings of the Joint 10th International Conference on Soft Computing and Intelligent Systems(SCIS) and 19th International Symposium Symposium on Advanced Intelligent Systems(ISIS), Toyama, Japan, 5–8 December 2018; IEEE: Piscataway, NJ, USA, 2018.
19. Kuznichov, D.; Zvirin, A.; Honen, Y.; Kimmel, R. Data augmentation for leaf segmentation and counting tasks in rosette plants. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
20. Kumar, J.; Domic, S. Rosette plant segmentation with leaf count using orthogonal transform and deep convolutional neural network. *Mach. Vis. Appl.* **2020**, *31*, 6.
21. Jin, S.; Su, Y.; Gao, S.; Wu, F.; Ma, Q.; Xu, K.; Hu, T.; Liu, J.; Pang, S.; Guan, H.; et al. Separating the structural components of maize for field phenotyping using terrestrial LiDAR data and deep convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2644–2658. [[CrossRef](#)]
22. Lai, Y.; Lu, S.; Qian, T.; Chen, M.; Zhen, S.; Guo, L. Segmentation of Plant Point Cloud based on Deep Learning Method. *Comput.-Aided Des. Appl.* **2022**, *42*, 161–168. [[CrossRef](#)]
23. Lu, S.L.; Song, Z.; Chen, W.K.; Qian, T.T.; Zhang, Y.Y.; Chen, M.; Li, G. Counting Dense Leaves under Natural Environments via an Improved Deep-Learning-Based Object Detection Algorithm. *Agriculture* **2021**, *11*, 1003. [[CrossRef](#)]
24. Lou, L. Cost-Effective Accurate 3-D Reconstruction Based on Multi-View Images for Plant Phenotyping. Ph.D. Thesis, Aberystwyth University, Aberystwyth, UK, 2016.
25. Huang, W.; Gong, H.; Zhang, H.; Wang, Y.; Wan, X.; Li, G.; Li HShen, H. BCNet: Bronchus Classification via Structure Guided Representation Learning. *IEEE Trans. Med. Imaging* **2024**, *1*. [[CrossRef](#)] [[PubMed](#)]
26. Ke, L.; Tai, Y.W.; Tang, C.K. Deep occlusion-aware instance segmentation with overlapping bilayers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 19–25 June 2021; pp. 4019–4028.
27. Tian, Z.; Shen, C.; Chen, H.; He, T. Fcos: Fully convolutional one-stage object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019.
28. 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.
29. 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.
30. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
31. Woo, S.; Park, J.; Lee, J.; Kweon, I. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
32. Zhang, H.; Wang, Y.; Dayoub, F.; Sünderhauf, N. Varifocalnet: An iou-aware dense object detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021.
33. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
34. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
35. Mehta, S.S.; Ton, C.; Asundi, S.; Burks, T.F. Multiple camera fruit localization using a particle filter. *Comput. Electron. Agric.* **2017**, *142*, 139–154. [[CrossRef](#)]
36. Wang, K.; Liew, J.H.; Zou, Y.; Zhou, D.; Feng, J. Panet: Few-shot image semantic segmentation with prototype alignment. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019.
37. Jiang, B.; Zhang, J.; Hong, Y.; Luo, J.; Liu, L.; Bao, H. Bcnet: Learning body and cloth shape from a single image. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020.
38. Huang, X.Y.; He, R.J.; Dai, Y.C.; He, M.Y. Semantic Segmentation of Remote Sensing Images with Multi-scale Features and Attention Mechanism. In Proceedings of the 2023 IEEE 18th Conference on Industrial Electronics and Applications (ICIEA), Ningbo, China, 18–22 August 2023.

39. Li, Z.; Lin, Y.; Fang, Z.; Li, S.; Li, X. AV-GAN: Attention-Based Varifocal Generative Adversarial Network for Uneven Medical Image Translation. *arXiv* **2024**, arXiv:2404.10714.
40. Ngugi, L.C.; Abdelwahab, M.; Abo-Zahhad, M. Tomato leaf segmentation algorithms for mobile phone applications using deep learning. *Comput. Electron. Agric.* **2020**, *178*, 105788. [[CrossRef](#)]
41. Talasila, S.; Rawal, K.; Sethi, G. PLRSNet: A semantic segmentation network for segmenting plant leaf region under complex background. *Int. J. Intell. Unmanned Syst.* **2023**, *11*, 132–150. [[CrossRef](#)]
42. Weyler, J.; Magistri, F.; Seitz, P.; Behley, J.; Stachniss, C. In-field phenotyping based on crop leaf and plant instance segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022.
43. Amean, Z.M.; Low, T.; Hancock, N. Automatic leaf segmentation and overlapping leaf separation using stereo vision. *Array* **2021**, *12*, 100099. [[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.