Instance segmentation based method to obtain the phenotypic information of weeds in complex field environments

Quan longzhe (quanlongzhe@163.com)
Northeast Agricultural University https://orcid.org/0000-0002-6391-0365
Bing Wu
Northeast Agricultural University
Shou ren Mao
Northeast Agricultural University
Huaiqu Feng
Northeast Agricultural University
Chunjie Yang
Northeast Agricultural University
Jiang Wei
Northeast Agricultural University
Hengda Li
Northeast Agricultural University

Research

Keywords: Leaf age, Mask R-CNN, Image segmentation, Deep learning, Computer vision.

Posted Date: November 3rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-51763/v2

Abstract

Background: Weeds pose a critical threat to crop growth. The leaf age and plant centre, which represent the key phenotypic information of weeds, can help understand the morphological structure of weeds, thereby facilitating precise targeted spraying and a reduction in the herbicide usage. However, determining the weed types, leaf age and plant centre under complex field conditions involving variations in the light and plant appearance along with leaf occlusion is challenging. With the advancement in the application of deep learning with computer vision, such approaches can likely overcome these challenges, as demonstrated in other complex agricultural applications.

Results: We developed a weed segmentation method based on BlendMask, which could obtain the weed types, leaf age and plant centre under complex field conditions. Mobile devices were used to capture digital images at different angles (front, side, and top views) of certain weeds (Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus) in the field. Subsequently, two datasets (with and without data enhancement) were produced and input to the network. Moreover, two backbone networks, ResNet50 and ResNet101, were compared, along with six instance segmentation algorithms, and the instance segmentation results of the model under different angles were evaluated. The results indicated that data enhancement could enhance the model performance. In the case with data enhancement, the F_1 value, AP50 and AP70 scores, and mIOU with ResNet101 as the backbone network were 0.9479, 0.720, 0.592, and 0.607, respectively, corresponding to the highest segmentation performance. Furthermore, the top view images of the weeds corresponded to the highest detection accuracy, compared to that for the other two angles.

Conclusion: BlendMask can realize accurate segmentation of weeds to obtain the types, leaf age and plant centre of weeds. Data enhancement and use of the weed image corresponding to the top view angle can help enhance the model performance. The dataset and research results can provide guidance to further develop precision agriculture practices.

1. Background

Weeds pose a critical threat to crop growth; specifically, weeds can hinder the growth of crops, thereby necessitating an increased use of herbicides [1], which can lead to a large reduction in the crop yield [2]. Moreover, because weeds compete with major cash crops for space, water, light, nutrients and other resources, the loss of plants is considerable in the early stages of the plant growth [3]. Therefore, removing weeds as early as possible is critical to ensure a satisfactory crop yield. In general, the use of herbicides can increase the crop yields and reduce the amount of labour [4]. Therefore, more than 90% of the crops in the United States are subjected to herbicide application [5], and the global use of pesticides is estimated to be 3.5 billion kg/y [4]. However, such excessive use of herbicides can induce a series of problems such as environmental pollution. Therefore, it is necessary to reduce the use of herbicides without affecting the crop yield, thereby reducing the cost of weed management and facilitating the realization of precision agriculture [6].

The size, shape and growth stage of field weeds are usually uncertain. According to the principle of herbicide and plant physiology [7], the phenotypic information of weeds is closely related to the herbicide dosage. For example, the herbicide dosage required for weeds with different leaf ages is different [8] [9]. Although the upper limit of herbicide application can help eliminate weeds of different ages, this scenario may involve excessive use of herbicides. In general, the leaf age refers to different stages in the plant development, and the number of complete leaves in a plant corresponds to the leaf age [10]. The leaves of a weed are connected to the stem through the petiole. When the weed canopy is viewed from the top, the plant centre formed by the overlapping of the petiole and stem of the top leaves of the weed can be observed. This centre represents the intersection centre of the weed top leaves [7]. In this study, this region is referred to as the plant centre. Most of the plant centre is composed of new tissue. Although the epicuticular waxes of plant leaves are closely related to the absorption of herbicides, the composition of the epicuticular waxes in different parts of the same plant is different and varies with the season, location and age of plants [11] [12]; therefore, different organs of the same plant may exhibit different sensitivities to herbicides. The new tissue involves many stomata and a thin waxy layer, which facilitates the absorption of the herbicides. Therefore, the plant centre is more sensitive to herbicides, with a higher capacity for herbicide absorption. In particular, the plant centre exhibits a higher retention capacity, especially for weeds with a leaf age of more than four. In this regard, the herbicide usage can be effectively reduced by implementing the herbicide use considering the close relationship between the leaf age of the weeds and the plant centre and absorption and conduction of the herbicide.

Nevertheless, the premise of reducing herbicide use is to accurately identify the weeds. In recent years, machine vision [13] has been widely used in the agricultural field. Brivot et al. [14] first used machine vision to segment weeds and identify crop rows, thereby demonstrating the feasibility of machine vision in the agricultural field. Moreover, researchers developed an algorithm to segment plants according to the soil background under uncontrolled lighting conditions in the wild, thereby separating the weeds from the crops [15]. However, this method was not entirely effective when the field image did not exhibit any changes under certain lighting conditions. The wavelet transform has been used to distinguish weeds from crops in images [16]; however, this traditional visual method is not effective when the number of weeds is large. Moreover, these methods only distinguish the crops from the weeds and do not specify the type and phenotypic information of the weeds. Shirzadifar et al. [17] classified the weeds based on the canopy spectral information of the plants and determined the weed species; however, in a complex field environment, the wind, soil background, and presence of shadows may change the spectral characteristics of the plants, thereby affecting the model performance [18] [19]. Furthermore, many researchers have examined the phenotypic information of plants. Minervini et al. [20] established fine grained datasets for image-based plant phenotyping, which was of notable significance for the study of plant phenotypes. In addition, the realization of leaf segmentation is a key challenge in the field of plant phenotype analyses. Bell et al. [21] segmented leaves through edge classification and achieved satisfactory results for the plant overlap. Dobrescu et al. [22] proposed a multi-task deep learning framework for plant phenotypes and achieved promising results for leaf counting. However, the plant images in these studies were collected under indoor conditions, and such images usually exhibit a pure

background and light uniformity [23]. Moreover, these studies were aimed at calculating the number of leaves for leaf segmentation; however, an image may have multiple weeds, and it is necessary to identify the leaf age, weed type and plant centre of each weed. Therefore, the segmentation of plant phenotypes in a complex farmland environment is a relatively unexplored research domain. Due to the complex environment conditions of farmlands, differences among plants, and the mutual occlusion of leaves, segmenting the leaf age and plant centre is challenging and often limits such analyses. In this study, the weed species, growth stage and plant centre were segmented through machine vision in a complex field environment to guide the use of herbicides.

Deep learning is an emerging field of machine learning, aimed at solving big data analysis problems. The DCNN is a deep learning method that is especially suitable for computer vision problems. In this study, an instance segmentation algorithm based on deep learning is proposed to obtain the weed phenotype in a complex field environment. According to an agricultural survey, deep learning technology is more accurate than the traditional image processing technology [24]. Moreover, in the complex field environment, the illumination, weather conditions, and soil background are complex and variable, and the plants may overlap [25] [26], and the DCNN model can address these aspects. Nevertheless, although the DCNN model can overcome these difficulties, the farmland environment is complex, and a sufficiently large dataset is required to train the deep learning model, to effectively manage the complex field environment aspects and increase the model accuracy [27]. To this end, data enhancement can be performed, which is a common method in the field of image recognition. In this approach, the image is expanded by randomly flipping the image, adding noise, and adjusting the brightness. Geetharamani et al. [28] used a nine-layer deep convolutional neural network to identify plant leaf diseases and employed six methods of data enhancement to enhance the model performance, which resulted in a classification accuracy of 96.4%. Piedad et al. [29] used the Mask R-CNN model to realize the non-invasive classification of clustered horticultural crops. Due to the limited dataset, the dataset was expanded to increase the model accuracy. In general, data enhancement is a key method to enrich the training samples and enhance the model performance; moreover, this approach can help enhance the suitability of a dataset for complex farmland environments.

When collecting the dataset, the shooting angle in the field [30] and growth stage of the weeds may affect the dataset accuracy. Quan et al. used deep learning methods to detect maize seedlings under different growth stages, angles and weather conditions in a complex field environment. It was proposed that when the angle between the camera and vertical direction is 0°, the detection accuracy is 0.95% lower than that for the other angles [27]; therefore, the model performance varied when the data were collected from different angles. However, Quan et al. primarily considered different oblique angles, and the information pertaining to different oblique angles is the fusion of that corresponding to different orthogonal angles [27]. Moreover, the position and shape of weeds in the field are complex and changeable, and the shape of the same object is different under different shooting angles, which affects the accuracy of the dataset. Therefore, we collected data from three angles, corresponding to the front, side and top views, which could clarify the comprehensive information of weeds and enable the model to cope with the requirements of operations from different angles. Moreover, an instance segmentation algorithm based

on deep learning was developed to obtain the weed phenotype in a complex field environment. According to the existing research, the DCNN exhibits a high performance in solving complex environmental problems in the field. Among the relevant approaches, instance segmentation based on deep learning is a new challenge for computer vision applications [31]. The model used in this study is a state-of-the-art method, aimed at detecting each object in a weed image and classifying each pixel of each instance. The output is the mask and bounding box of the target object [32], and the problem of leaf adhesion and occlusion can be solved in this manner [21].

Yu et al. [33] proposed an exemplar-based recursive instance segmentation framework to segment plant phenotypes and conducted experiments on a public benchmark to demonstrate the effectiveness of the method. Huang et al. [34] proposed a deep learning model for in-row crop detection in rice fields and constructed a field rice detection dataset with a detection accuracy of 93.22%. This method identified a stem-base-centred square region at the plant level, which corresponds to the protected area image of mechanical weeding and is the plant centre. The plant centre is of significance for plant research. Moreover, the instance segmentation algorithm based on Mask R-CNN, proposed by He et al. [35], could not only identify the bounding box but also mask the target contour, thereby outperforming the other models [36]. Jia et al. [37] used an improved Mask R-CNN model to segment overlapping apples, with an accuracy rate of 97.31%. In particular, the Mask R-CNN is representative of a two-stage segmentation network. Although many scholars have studied and applied the Mask R-CNN technique and achieved satisfactory results, the BlendMask [38] model proposed by Chen Hao et al. exhibited a higher segmentation performance on the COCO dataset [39] than the Mask R-CNN. BlendMask combines the ideas of top-down and bottom-up methods. Moreover, BlendMask employs the fully convolutional one-stage object detection (FCOS) [40], which eliminates the calculation of the position-sensitive feature map and mask feature. Thus, the inference time does not increase with the number of predictions, as in the traditional two-stage method.

The aforementioned studies provide a feasible basis and reference for the application of the DCNN in plant segmentation. Moreover, it is noted that the DCNN can overcome the shortcomings of traditional image segmentation methods. The excellent performance of the BlendMask model indicates that in can well address the complex environment in the field. In this study, the following weeds, which are commonly found in fields in Northeast China, were selected: Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus. According to the existing studies, a sufficient number of well-defined weed datasets are required to train DCNN models. Moreover, the weed images should be obtained from real scenes in fields to ensure that the images contain the morphological characteristics of the weeds at different growth stages in the complex field environment and cover more variables as the model input. Therefore, we collected the weed images at three different angles (front, side, and top views). In addition, the classic DCNN network can be modified to increase the model accuracy. Therefore, we created two datasets containing 4000 and 6000 weed images, without and with data enhancement, respectively.

In particular, considering the aforementioned problems, this paper proposes a weed phenotype segmentation method based on the BlendMask to obtain the weed species, leaf age and plant centre of

weeds. The main objectives were as follows:

(1) To evaluate the feasibility of BlendMask in obtaining the weed species, leaf age and plant centre through weed phenotypic segmentation, considering seven evaluation indicators.

(2) To explore the influence of data collection from different angles (front view, side view, top view) on the phenotypic segmentation of weeds in a complex field environment, and to identify the optimal angle.

(3) To explore whether data enhancement can enhance the model performance.

(4) To explore whether the combination of ResNet101 with the FPN architecture for feature extraction can enhance the model performance.

2. Materials And Methods

2.1. Overview

Three typical weeds, Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus, which are commonly found in Northeast China were selected. The leaf age and plant centre of the weeds were segmented using the BlendMask model. First, datasets from different angles (front, side, top views) were collected in the actual field environment. Second, two datasets (with and without data enhancement) were created. Third, the produced datasets were annotated, and the generated file was input to the network to train the network model. The backbone network of the BlendMask initialization model is the residual network combined with the feature pyramid network (FPN). This study employed different backbone networks (ResNet50 and ResNet101) in combination with the FPN architecture. The feature extraction performance based on the weed leaf age and plant centre was evaluated. Figure 1 shows the workflow of the experiment.

2.2. Image acquisition

The following three kinds of weeds were selected in this study: Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus. Solanum nigrum is an annual dicotyledon, Barnyard grass is an annual herb, and Abutilon theophrasti Medicus is an annual subshrub weed. The three kinds of weeds are commonly found in the fields of Northeast China, as shown in Figure 2. The data image source was the field weed image. Because the greenhouse weeds exhibit a single background, whereas the images of field weeds are more complex, the ability of the model to recognize weeds in the natural state could be verified. The field data images were acquired in the Xiangfang District from May to June 2019. The Xiangfang District is located in the northeast plain and is the main planting area for maize, soybean and rice. The main weeds in the cornfields of the Xiangfang District are Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus. The images of these weeds were collected. Since most the weeds in the field had a leaf age of two–five, the images of only the weeds with a leaf age of less than five were acquired.

Because the weed information obtained from a single shooting angle is not comprehensive, to more clearly illustrate the difference in the weed information obtained from different angles, a Python code was implemented to remove the background of the field weed images, as shown in Figure 3. In contrast, the images used for training the model corresponded to the complex environment of the field, and the background was not removed. Moreover, in general, the shooting weather [41] and acquisition angle [30] considerably influence the dataset and affect the segmentation precision [42]. Data collection was initiated from the two-leaf period after crop planting, from May 20, 2019, to June 29, 2019, and images of weeds with different leaf ages were collected every 2 to 5 d under different weather conditions, angles, and growth stages, to obtain the data of the weeds at each leaf age stage in the growth cycle, as shown in Table 1. The camera of the iPhone 6s Plus device, with a focal length, maximum aperture, and maximum resolution of 4.2 mm, f/2.2 and 4032 × 3024 pixels, respectively, was used to capture images, and the weed images were stored in the JPEG file format. When collecting data, we marked the sample variety, leaf age, collection time, collection angle, collection weather, and temperature into the sample data. All the datasets were collected randomly in the farmland, and the images corresponded to a relatively clean soil background and complex field background covered by straw and leaves; such disturbances were treated as part of the background.

Table 1 List of images containing the environmental information for the experiment

Date	Images	Tmax	Tmin	Weather	Front view	Top view	Side view
25/05/2019	441	30	17	Cloud	201	140	100
30/05/2019	449	22	11	Cloud	112	226	111
04/06/2019	481	24	16	Cloud	132	200	149
08/06/2019	538	24	18	Rain	213	221	104
10/06/2019	422	25	16	Cloud	136	161	125
14/06/2019	426	24	14	Rain	119	203	104
17/06/2019	458	25	13	Cloud	102	254	102
21/06/2019	420	26	14	Cloud	125	107	188
26/06/2019	412	26	17	Cloud	119	136	157
29/06/2019	527	23	15	Cloud	152	214	161
Total	4574	\	\	\	1411	1862	1301

2.3. Dataset construction and annotation

When training the network and conducting network testing, the input image size must match the input size of the network [43]; thus, the images were adjusted to a pixel size of 1024 × 1024 to construct the image dataset of the DCNN. The size of the field shot image was 4032 × 3024; without disturbing the morphology of the plants in the image, the image was cropped to a size of 3024 × 3024, and these

images were resized to 1024 × 1024. As the weeds in the images were required to be annotated, certain images not suitable for annotation were discarded. Finally, 4000 images were selected from 4574 images. Due to the limited number of datasets, a data enhancement scheme was adopted to further enrich the images to ensure that the images were highly representative and could reflect the real situation of the field data more accurately [27]; moreover, the training precision of the model could be increased [44], the dataset could be expanded and overfitting could be reduced [45] (Figure 4).

The images were randomly rotated, and noise was added. Since the illumination is a critical aspect in the segmentation process, to enhance the robustness of the DCNN against the illumination variations owing to environmental changes, the datasets were further enhanced by simulating illumination changes [46]. The brightness was increased and decreased by 10%. Moreover, certain blurred, occluded and incomplete images were included in the dataset as negative samples, and 6000 data-enhanced pictures were obtained. The structure and proportion of the original dataset remain unchanged when data enhancement was implemented. Two datasets were prepared, with and without data enhancement. Both the datasets were randomly divided into training and verification sets, with a ratio of 8:2. The test set is selected from the images without data enhancement. The VGG Image Annotator labelling tool [47] was used for the annotation, as shown in Figure 5; the leaves and centre of weeds were surrounded by irregular polygons. Because the number of weeds in an image is uncertain under the actual working conditions, and the image may contain multiple weeds, the number of masked leaves in the picture could not be used to calculate the leaf age of a single weed. In this study, we used a rectangular frame to mark the outline of the outermost layer of a single weed and calculated the number of leaf masks in the rectangular frame, which corresponded to the leaf age of the weed. The rectangular frame was not masked. The labels were divided into seven categories (Figure 5).

2.4. Weed instance segmentation model

2.4.1 Two-stage instance segmentation model

Instance segmentation is one of the most challenging tasks in computer vision, because it involves not only the classification at the pixel level of semantic segmentation, but also certain characteristics of target detection. The two-stage Mask R-CNN is a representative algorithm. The Mask R-CNN extends the target detection framework of the faster R-CNN [48] by adding a masking branch at the end of the model [49]. This process ensures that each output instance segments the proposal box through a fully connected layer, to ensure that the segmentation is parallel to the target detection. To better detect small targets, the ROI pooling is changed to ROIAlign. The process flow of the Mask R-CNN model is shown in Figure 6. The output consists of three branches: bounding boxes, target classifications and segmentation masks. The selected backbone networks for the Mask R-CNN were ResNet50 and ResNet101 combined with FPNs. In this configuration, first, the backbone network extracts the feature map from the input image and outputs the features from the backbone network. The map is sent to the region proposal network (RPN) and ROIAlign to generate the region of interest (ROI). Finally, the ROI predicts the target category and bounding box through the convolutional layer and fully connected layer and segments the

target region through the fully convolutional neural network (FCN). The instance segmentation task of the target is thus completed.

2.4.2 One-stage instance segmentation model

BlendMask is a one-stage dense instance segmentation algorithm that combines the instance-level information with lower-level fine-granularity semantic information. BlendMask is composed of a one-stage target detection network FCOS [40] and a mask branch. Figure 7 shows the model structure of BlendMask. The mask branch has three parts: The bottom module is used to process the bottom features to generate the score maps, the top layer is attached to the box head of the detector to generate the top level attention corresponding to the base, and the blender module is used to fuse the base and attention.

BlendMask adds the bottom module to extract the low-level detailed features, based on the anchor-free detection model FCOS, and predicts the attention on the instance-level. BlendMask draws on the fusion method of the fully convolutional instance-aware semantic segmentation (FCIS) [50] and YOLACT [51] and incorporates the blender module to better integrate these features. Moreover, BlendMask combines the concepts of the top-down and bottom-up methodologies, thereby combining the rich instance-level information with accurate dense pixel features [38].

The structure of BlendMask is similar to that of the Mask R-CNN; however, in contrast to the RPN used by the Mask R-CNN, BlendMask chooses the one-stage detector FCOS, which eliminates the calculation of the position-sensitive feature map and mask feature. BlendMask uses the attention guided blender module to calculate the global map representation. Compared with the more complex hard alignment used in FCN and FCIS, the amount of calculation is considerably reduced under the same resolution. BlendMask is a method of dense pixel prediction, and the output resolution is not limited by the top level sampling. In the Mask R-CNN framework, to achieve more accurate mask features, the resolution of RoIPooler must be increased, thereby increasing the calculation time and network depth of the head.

BlendMask can establish deep neural network models of different depths by implementing different weight layers. The deep learning network models applied at present include AlexNet, ZF, GoogLeNet, VGG, and ResNet. Although a larger number of network layers may lead to a higher accuracy, the deeper network layers may result in degraded model training and detection speeds. Nevertheless, since the residual network does not increase the number of model parameters, the problem of training degradation can be alleviated, and the model convergence can be accelerated [49]. Therefore, in this study, ResNet50 and ResNet101 combined with the FPN were used as the backbone networks to extract the features of the weed images.

2.5 BlendMask training model

Before training the BlendMask, we introduced a pretraining model based on the COCO dataset [39] through transfer learning. The COCO dataset has 328,000 images, including 91 categories. The pretraining model extracted the weights after training on the COCO dataset, based on which, the

established datasets were retrained. Through the transfer learning, the labour and cost of training could be reduced, the training efficiency could be enhanced, and the model parameters could be better adjusted. The BlendMask model was implemented using the AdelaiDet open source toolbox based on detectron2. The experiment was performed on the Ubuntu 18.04 operating system, over a six-core Intel Core i7-8700K @ 3.70 GHz processor, 32 GB of memory, and a GPU built by NVIDIA GeForce (Santa Clara, CA, USA), along with the NVIDIA GeForce RTX 1080 Ti graphics card. The pretraining network parameters are listed in Table 2.

Table 2 Characteristic parameters of the pretraining network.

Network	Depth	Size (MB)	Parameters (millions)	Image Input Size	Feature Extraction Layer
ResNet50	50	96	25.6	224 × 224	block_13_expand_relu
ResNet101	101	167	44.6	224 × 224	mixed7

2.6 Training and evaluation

The momentum and initial learning rate for BlendMask was set as 0.9 and 0.01, respectively, and the training BatchSize was set as 4. After the parameters were set, training was conducted for 12 rounds, with 10000 iterations being implemented in each round. The basic framework of the BlendMask involved either ResNet50 or ResNet101.

The evaluation was aimed at examining the ability of the algorithm to identify the weed leaf age and plant centre in the image. For the evaluation, we used seven key indexes: precision rate (P) (Eq. (1)), recall rate (R) (Eq. (2)), F_1 (Eq. (3)), intersection over the union (IOU) (Eq. (4)), average precision (AP), mean average precision (mAP) (Eq. (5)), and mean intersection over the union (mIOU).

The precision and recall rates can be defined as follows:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

where "true positive (TP)" and “false positive (FP)” indicate the number of positive and negative results detected as positive, respectively, and "false negative (FN)" indicates the number of positive results detected as negative. The metric function (F_1) [52] of the precision and recall rates can be defined as follows:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

To address the problem of multiclass imbalance, we averaged the seven classification indicators [53]. To more extensively evaluate the model algorithm, the IOU was considered to examine the measurements [54]. In general, the IOU measures the overlap between two bounding boxes. Figure 8 illustrates the calculation of the overlap degree between the weed prediction box and real box on the ground. The value of the IOU can be divided into three regions. When the IOU threshold is set as 0.7, anchors with IOU values less than or equal to 0.3 are considered to be negative anchors, and anchors with IOU values between 0.3 and 0.7 are neutral anchors; these cases are not considered. The anchors with IOU values greater than or equal to 0.7 are positive anchors. The system identifies the positive anchors and bounding box and matches these boxes to the ground-truth boxes to optimize the RPN output of the model. The maximum value of the anchor overlap with the ground-truth boxes is retained by the system. When the IOU threshold is set as 0.5, and the IOU values are greater or smaller than 0.5, a positive or negative ROI is observed, respectively. The positive ROI is allocated to the mask and ground-truth by the system. The detection performance of the model is evaluated considering the mean accuracy (mAP) [54]. The mAP can clearly reflect the performance when it is related to the target position information and category information of the target in the image. The AP can be calculated for each category separately, and the value for each category can be averaged to calculate the mAP. A larger mAP is desirable. This value can be calculated considering the AP. The thresholds were set as 0.5 and 0.7 in this study. When the IOU threshold was equal to or greater than 0.5 and 0.7, the mAP were defined as AP50 and AP70, respectively. The IOU and mAP were defined as follows:

$$IOU = \frac{Area\ Overlap}{Area\ Union} \quad (4)$$

$$mAP = \frac{1}{N} \cdot \sum_{i=1}^{N} AP_i \quad (5)$$

Note: N represents the number of images.

3. Results

The results of weed segmentation are shown in Figure 9. The leaf age was determined based on the number of complete leaves in the weed, and the accuracy of the leaf age recognition was evaluated by comparing this value with the corresponding label at the time of collection. The plant centre was determined based on the intersection area of the top leaves of the weed [7]. Six instance segmentation algorithms were compared, the algorithm with the highest performance was selected, and adjust different

hyperparameters for this algorithm. Two datasets were trained using two different backbones networks (ResNet50 and ResNet101), and the network that could realize the optimal balance between mIOU and mAP was selected. and the segmentation performance of the best-performing algorithm under different shooting angles and different leaf ages was evaluated. The results indicate that the data enhancement can enhance the model performance. The AP50 and mIOU of the BlendMask model using ResNet101 as the backbone combined with FPN are 0.720 and 0.607, respectively; these values are better than those for the ResNet50 framework, and thus, the former model can be used for weed segmentation. Moreover, the weed image captured at the top view angle exhibited a higher detection accuracy than that of the other two angles.

3.1 Comparison of instance segmentation models

To verify the effectiveness of the proposed method for weed segmentation, six instance segmentation algorithms, including Mask R-CNN, SOLO [55], PolarMask [56], CentreMask [57], YOLACT [51] and BlendMask were compared. The six algorithms were tested on two data sets (with and without data enhancement). To examine the recognition effect of the model in a complex field environment, the images in the test were those without any data enhancement. The test results are shown in Figure 10.

Figure (a) shows the F_1 values of six instance segmentation networks. Two data sets (enhanced and unenhanced) were applied to train the instance segmentation model, and subsequently, the verification set was used to evaluate the performance of the algorithm to analyse the weed segmentation performance. According to figure (a), the performance of the 6 instance segmentation models on the dataset with data enhancement was generally higher than that on the dataset without data enhancement; thus, the implementation of the data enhancement can effectively improve the model accuracy. The F_1 values of the Mask R-CNN, SOLO, PolarMask, CentreMask, YOLACT, and BlendMask after training on the data enhancement dataset were 0.9214, 0.9432, 0.8873, 0.9297, 0.9023, 0.9479, respectively. It can be noted that the F_1 value of the BlendMask model was higher than that of the other five instance segmentation models in the weed segmentation task. The F_1 value of the SOLO model was smaller than that of BlendMask by 0.0047. PolarMask exhibited the lowest weed segmentation performance among the six models.

Figures (b) and (c) show the AP50 and AP70 values of the six instance segmentation networks, respectively. The AP50 values of the R-CNN, SOLO, PolarMask, CentreMask, YOLACT and BlendMask after training on the dataset with enhanced data were 0.6932, 0.7131, 0.6542, 0.7085, 0.6721 and 0.7200, and the corresponding AP70 values were 0.5244, 0.5796, 0.4982, 0.5737, 0.5032, 0.5921, respectively. According to Figure (b), data enhancement can lead to higher average precision values than those corresponding to the dataset without data enhancement. In the case of data enhancement, the AP50 value of the BlendMask model is 0.7200, which is the highest among the six instance segmentation models. The AP50 value of the SOLO model is 0.7131, and the corresponding instance segmentation performance is similar to that of the BlendMask model. Moreover, the AP50 value of the PolarMask model is 0.6542, corresponding to the lowest segmentation performance. In the case of data

enhancement, the AP50 of the five models ranges from 62% to 72%, and these models can thus satisfy the requirements for weed instance segmentation. According to Figure (c), in the case of data enhancement, the AP70 value of the BlendMask model is 0.5921, corresponding to the highest value among the 6 instance segmentation models. The AP70 value of the SOLO model is 0.5796, and the instance segmentation performance is similar to that of the BlendMask model.

Overall, among the 6 instance segmentation algorithms, the F_1, AP50 and AP70 values of the BlendMask model were the highest in the task of weed segmentation in a complex field environment, and the AP50 value was 0.0268, 0.0069, 0.0685, 0.0115 and 0.0479 higher than that of the Mask R-CNN, SOLO, PolarMask, CentreMask and YOLACT, respectively. The AP50 value of the SOLO model was 0.0199, 0.0589, 0.0046, 0.041 higher than that of the Mask R-CNN, PolarMask, CentreMask, and YOLACT models, respectively. The model results show that the BlendMask and SOLO models can better exploit the image features than the other four models, thereby exhibiting a higher segmentation performance.

Since BlendMask and SOLO outperformed the other four models in terms of the segmentation performance, the BlendMask and SOLO models were replaced with two underlying networks (ResNet50 and ResNet101), and the prediction time under the different underlying networks was compared. The corresponding results are presented in Table 3

Table 3 Prediction time of different models under different backbone networks

Network	Times (ms)	
	ResNet50	ResNet101
SOLO	102.7	128.5
BlendMask	89.3	114.6

Table 3 lists the prediction durations of the model for a single picture, and it can be noted that when ResNet50 is used as the backbone network, BlendMask has the smallest prediction duration, which is 13.4 ms lower than that of SOLO. Figure 10 indicated that the segmentation performance of BlendMask was comparable to that of SOLO. However, according to the comparison in Table 3, under both the backbone networks, the prediction time of BlendMask for a single image is lower than that for SOLO. Therefore, considering both the segmentation performance and prediction time, it can be considered that BlendMask exhibits a satisfactory segmentation performance; the model is feasible and can realize prompt and accurate weed segmentation.

3.2 Comparison of different hyperparameters of BlendMask

We have changed the following four hyper-parameters of BlendMask:

- *R*, the resolution of bottom-level RoI,
- *M*, the resolution of top-level prediction,
- *K*, the number of bases,
- The bottom module is composed of the feature of the backbone network or FPN

The Table 4 is a comparison of different resolutions when the *K*=4 and the bottom module is C3 and C5. We set the resolution *R* of the bottom-level RoI to 28 and 56, with *R/M* ratio from 14 to 4.

Table 4 Comparison of different resolutions

R	*M*	AP50	AP70	Time(ms)
28	2	0.652	0.505	85.7
	4	0.664	0.517	86.1
	7	0.677	0.521	88.3
56	4	0.685	0.532	86.3
	7	0.693	0.535	88.2
	14	0.698	0.538	91.1

Note: Set the bases (K) of the model to 4, and use C3 and C5 for the bottom module from the backbone network ResNet101. By changing the resolution of the bottom-level RoI and the top-level prediction to compare the performance of the model.

It can be seen from Table 4 that increasing the resolution *R* of bottom-level RoI will lead to longer operation time of the model. Compared to *R* is 28, the performance of the model is generally higher when the *R* is 56. When *R* is set to 56, the *M* is set to 14, the AP50 value is 0.005 higher than when it is set to 7, and AP70 is 0.003 higher. But the prediction time is 2.9ms longer. Comprehensive consideration of prediction speed and accuracy, we set *R* to 56 and *M* to 7 in the next ablation experiment.

Table 5 Comparison of different bases

K	1	2	4	8
AP50	0.645	0.672	0.693	0.663
AP70	0.497	0.504	0.535	0.524

Note: We set R to 56 and M to 7.

Table 5 lists the comparison of different bases when *R* is set to 56 and *M* is set to 7. We set the number of bases from one to eight, looking for the best performance of the model. From Table 5, we can see that four bases achieve the best performance. In the next ablation experiments, we set the number of bases to 4.

Table 6 Comparison of the bottom feature locations from backbone or FPN

	Feature	M	Time(ms)	AP50	AP70
Backbone	C3,C5	7	88.3	0.653	0.507
		14	91.1	0.657	0.519
FPN	P3,P5	7	84.9	0.664	0.523
		14	89.5	0.667	0.523

Note: the resolution of top-level prediction is set to , the resolution of bottom-level RoI is set to 7 and the number of bases is set to 4.

Table 6 lists the feature extraction performance comparison of different bottom modules when R is set to 56, M is set to seven and K is set to four. From Table 6, we can see that using FPN features as the input of the bottom module is effective for the performance of the model. In the next experiment, we use backbone combined with FPN to extract the features of weeds.

3.3 Segmentation results of weeds with different shooting angles and leaf ages

We compared the segmentation results of BlendMask with those of two different backbone networks (ResNet50 and ResNet101) combined with FPN under different leaf ages and shooting angles. The test set, which included 600 images, was used to verify the generalization ability of the model; therefore, 600 images without data enhancement were selected for testing. The total test set included 200 images each for the front view, side view, and top view. As shown in Figure 11, the labels a, b, c, a_leaf, b_leaf, c_leaf, and centre were not recognized, and we considered that these labels were identified as the background.

Figure 11 shows the confusion matrices of the detection results of the model in the case of data enhancement. This matrix calculates the statistics pertaining to the number of classified images by comparing the actual labels in the validation set data with the predicted types and indicates whether the model can differentiate among different classes. As shown in Figure 11, both ResNet50 and ResNet101 exhibit intuitive common features: The prediction for the leaves of Solanum nigrum is highly accurate; however, the prediction for Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus is not satisfactory. The centre prediction for the total test set for ResNet101 is second only to that for the leaves of Solanum nigrum; however, the prediction accuracy of Barnyard grass is the lowest. We can adjust the number of Barnyard grass data images in the model appropriately. Moreover, a part of the leaves of Solanum nigrum is predicted as that of the leaves of Abutilon theophrasti Medicus, likely because of the similar shape of these two leaves. ResNet101 exhibited a slightly higher predictive accuracy than that of ResNet50 in all the test sets. Thus, the performance of ResNet101 is higher than that of ResNet50.

Figure 12 shows the detection results of the BlendMask model under two backbone networks, three angles, and seven types of labels in the case of data enhancement. The precision rate, recall, and F_1 value of the ResNet101 network for the total test set are 0.9710, 0.9271, and 0.9479, respectively.

According to Figure 12, under the condition of data enhancement, the precision, recall, and F_1 values of ResNet50 in the front, side, and top views and all the test sets were greater than or equal to 0.7619, 0.7463, and 0.7634, respectively. In comparison, the precision, recall, and F_1 values of ResNet101 were greater than or equal to 0.8983, 0.8267, and 0.8983, respectively, in the front, side, and top views and all the test sets. In other words, the precision, recall and F_1 values of ResNet101 in the front, side, and top views and all the test sets were considerably higher than those of ResNet50, and ResNet101 exhibited a consistently higher performance than ResNet50. The F_1 values of ResNet101 in the front, side, top, and total test sets were 0.9445, 0.9371, 0.9643 and 0.9479, respectively. When ResNet101 was used as the backbone network, the recall values of the top, front, and side view test sets were greater than or equal to 0.9149, 0.9051, and 0.8983, respectively. The top view test set exhibited the highest performance in all the classifications. For the total test set, the F_1 values were 0.9661 and 0.9725, and the recall rates were 0.9679 and 0.9648 when Solanum nigrum and the leaves of Solanum nigrum were detected, respectively. Moreover, on the front, side, and top test sets, the classification performance of Solanum nigrum was higher than that of the other two kinds of weeds. For the plant centre, the precision values of ResNet101 in the front, side, and top views and total test sets were 1.0000. Since Figure 11 and Figure 12 only indicate the classification performance of the model, the recognition accuracy of the model cannot be determined, and the actual environment in the field is complex, which is expected to influence the weed identification. Therefore, the model recognition accuracy is critical to evaluate the model performance. Table 7 presents the detection results for the weeds under different networks and angles with data enhancement.

Table 7 Detection results for the weeds under different networks and angles with data enhancement.

Network		ResNet50	ResNet101
Front view	AP50	0.573	0.732
	AP70	0.485	0.602
	mIOU	0.482	0.597
Side view	AP50	0.564	0.645
	AP70	0.472	0.540
	mIOU	0.472	0.583
Top view	AP50	0.637	0.784
	AP70	0.521	0.633
	mIOU	0.553	0.642
Total test set	AP50	0.591	0.720
	AP70	0.493	0.592
	mIOU	0.502	0.607

The mAP is a commonly used index in target detection. Table 7 indicates that the mAP of ResNet101 is higher than that of ResNet50, indicating that the ResNet101 exhibits a high target detection performance. For the total test set, when ResNet101 is used as the backbone network, the AP50 and AP70 values are 0.720 and 0.592, respectively. Thus, when the threshold is equal to or greater than 0.5, ResNet101 exhibits

a high detection performance. When ResNet101 is used as the backbone network, the AP50 values for the top, front, and side views and total test set are 0.784, 0.732, 0.645, and 0.720, respectively. The top view corresponds to a high detection performance.

The mIOU a valuable index to evaluate the segmentation results [31] and is commonly used to evaluate the segmentation performance of the BlendMask model. As indicated in Table 7, according to the test images of 600 weeds, for the total test set, the mIOU of ResNet50 and ResNet101 is 0.502 and 0.607, respectively. Thus, the ResNet101 model exhibits a higher network performance and can be applied to the segmentation of small target objects. In other words, this model can satisfy the needs of weed instance segmentation. When ResNet101 is used as the backbone network, the mIOU of the top view is 0.642, which is higher than those of the other datasets. Therefore, this configuration can achieve satisfactory segmentation results. Therefore, we choose ResNet101 combined with FPN to extract the features of weeds, as shown in Figure 13 is the recognition accuracy of different leaf ages in the case of data enhancement and ResNet101 combined with FPN.

In order to obtain the recognition accuracy of different leaves, we used three test sets, namely Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus datasets. Each dataset has 300 weed images, a total of 900 images, all images without data enhancement. In the Solanum nigrum data set, there are 100 weeds with 2 leaves, 3 leaves and 4 leaves respectively, and the remaining two data sets were also set in the same way. The accuracy of leaves identification is determined by comparing the leaves value calculated by the computer with the leaves value on the label when the data was collected. From Figure 13, we can see that the accuracy of leaves identification of the three weeds was higher than 80%. The recognition accuracy of 2 leaves and 3 leaves of Solanum nigrum is generally higher than that of the other two weeds. The recognition accuracy of 3 leaves was 0.957, which was the highest among all leaves. The recognition accuracy of Barnyard grass at 4 leaves was 0.017 and 0.022 lower than that at 2 leaves and 3 leaves, respectively. The recognition accuracy of 2 leaves of Barnyard grass was 0.005 lower than that of 3 leaves, which is the closest. The recognition accuracy of 2 leaves of Abutilon theophrasti Medicus is the lowest among all categories, with a value of 0.887.

4. Discussion

BlendMask combines top-level and bottom-level information. The top-level corresponds to a broader receptive field, such as the posture of the whole weed plant. Because the top-level is a rough prediction, the resolution of the top-level ROI is relatively small, and the general maximum is 14. Compared to Mask RCNN fixed output of , BlendMask output resolution can be higher, because its backbone is not limited by FPN. The bottom-level corresponds to more detail information, such as the position and centre of the weed, which can retain better position information. In our study, we need to accurately segment leaves and plant centre, which belong to some detail information, so we set the resolution of bottom-level to , and increasing the resolution did not cause the prediction speed to be too slow. There is a good balance between prediction speed and precision.

In the case of data enhancement, in terms of the F_1 value, the recognition accuracy for the Solanum nigrum leaf was the highest. According to the confusion matrix, most of the Solanum nigrum leaves are classified as Solanum nigrum leaves, and only a small part are considered as Abutilon theophrasti Medicus leaves, Barnyard grass leaves and background, and thus, the corresponding recall rate is high. Furthermore, only a small part of Abutilon theophrasti Medicus is considered as Solanum nigrum, leading to a high precision rate. In summary, the F_1 value is the highest. Solanum nigrum is an annual herb with oval leaves. Moreover, this weed has a large number of leaves, which allows the model to learn more features. Because the model can extract sufficient features from the Solanum nigrum leaves, Solanum nigrum exhibits a high recognition accuracy. In ResNet50, the recognition accuracy of Barnyard grass leaves is higher than that of Abutilon theophrasti Medicus; however, in ResNet101, the corresponding accuracies are comparable. We propose the following hypothesis: Because ResNet50 has fewer convolutional layers than ResNet101, it may not be able to extract sufficient features. ResNet101 increases the depth of the network; therefore, the corresponding F_1 values for the Barnyard grass and Abutilon theophrasti Medicus leaves are high, and the accuracy is comparable. The less satisfactory detection result for Abutilon theophrasti Medicus occurs because the leaves of Abutilon theophrasti Medicus are elliptical, similar to those of Solanum nigrum. Moreover, a blurred image may result in insufficient image information extraction. Hence, misjudgement may occur, which is undesirable because this phenomenon may introduce errors in leaf age identification. ResNet101 exhibits a higher F_1 score in terms of the plant centre, because the characteristics of the plant centre are more notable than those in other categories when the model is classified. In the total test set, when the IOU threshold of ResNet101 is 0.5 or 0.7, the mAP value is greater than that of ResNet50. However, the mIOU of ResNet101 is 0.720 in the total test set, indicating that this model can satisfy the needs of instance segmentation. Thus, the BlendMask model using ResNet101 as the backbone can reliably segment weeds.

The front, side, and top view images were obtained to comprehensively clarify the information of the weeds in the data set to obtain the leaf age and plant centre of the weeds. The method of acquiring three views is a common practice in the study of plant phenotype. The field plant perspective in the images contains front, top, and side views, and the information of the oblique angle is fused by the information of the orthogonal angles. Among the three orthogonal angles, more comprehensive weed phenotype information can be obtained from the perspective of the top view; specifically, the plant centre of the weeds can be identified more clarify. In contrast, the side and front views cannot clearly observe the plant centre. Consequently, the detection accuracy for the top view angle is higher than that for the other angles. Nevertheless, when intelligent agricultural equipment is employed in the field, the camera is usually fixed at an angle, although the position and shape of the weeds in the field are complex and changeable. When the machine is moving, the imaging angle of the weeds changes, and the imaging angle differs owing to the different positions of the weeds. The information of the side and front views is exposed at certain angles; therefore, obtaining the images of the side and front views can help the model in accurately segmenting the weeds. Constructing datasets from different perspectives can enable the model to adapt to the job requirements of different scenarios. The presented findings can provide reference for future vision acquisition systems of intelligent agricultural robots.

In this study, we identified the individual plant images of three weeds in the field; however, weeds are visual objects with complex structures and rich texture features, and even the same species may be considerably different in terms of the morphology and colour. Solanum nigrum has a higher recognition accuracy of 3 leaves. Because the recognition accuracy of Solanum nigrum leaves is higher, the calculated leaves value is also higher. And the Solanum nigrum leaves mostly grow from the same centre, and there are fewer leaves shaded below. The recognition accuracy of 4 leaves of Barnyard grass is lower than that of 2 leaves and 3 leaves. Because when Barnyard Grass was 4 leaves, the bottom of the main stem would mostly have some small leaves, which were small and easily concealed by the upper leaves, bringing great difficulties to the counting of leaves. The leaves of Abutilon theophrasti Medicus are elliptical, which are similar to the leaves of Solanum nigrum, causing some leaves to be misidentified. The recognition accuracy of Abutilon theophrasti Medicus is higher than the other two weeds at 4 leaves, because there are more leaves at 4 leaves, and the petiole of Abutilon theophrasti Medicus is longer and the lower leaves are not easily blocked.

At present, the treatment of weeds in the field mostly involves weed classification and detection. However, weed classification can determine only the species of weeds, and the specific position coordinates of the weeds cannot be obtained; thus, the exact target cannot be sprayed. Weed detection can facilitate the drawing of the bounding box of the weeds. However, weeds exhibit an irregular shape and size, which may cause the machine to be inaccurate with respect to the target, resulting in certain herbicide falling to the ground and not being absorbed by the weeds; this aspect may lead to environmental pollution and wastage of the herbicide. As a kind of deep learning model, instance segmentation can detect the target pixel by pixel, thereby solving the problems of blade adhesion and occlusion. Moreover, the leaf age of weeds and the position of the pant centre can be obtained accurately. In the Northeast Plain of China, the main economic crops are maize, soybeans, and wheat, which are susceptible to annual and perennial weeds. Controlling the annual and perennial weeds can increase the crop yields and reduce the likelihood of damage caused by the weeds in the second year [58]. Moreover, studying the interaction between the plant phenotype and vision through effective phenotypic analysis can help provide information regarding the plant growth and morphological changes.

The limitation of this study is that the recognition speed of the proposed method is low, but tensorRT can be considered for acceleration. The efficiency needs to be further improved before the approach can be applied in engineering practice. The employed DCNN model was used to segment only three kinds of weeds. If we consider more kinds of weeds, collect and segment the images of field crops, and increase the number of datasets, the model can achieve a higher segmentation accuracy. Moreover, the obtained leaf age of the economic crops can provide a basis for crop fertilization. For certain plants, the plant centre is the pollination area of flowers, and segmentation of this part can provide valuable guidance for subsequent studies. Future research will focus on evaluating the image datasets covering a wider range of weeds and crop varieties. In addition, most of the images used for model testing contained only single-plant weeds, and only a few images contained multiple weeds. The BlendMask framework failed to segment weeds near the edges in a few test images containing multiple weeds. In this case, continuous video input can help eliminate the edge effects when applied in the field.

The results show that the combination of weed phenotypic information and computer vision can effectively address the complex field conditions such as light changes, leaf occlusion, and mixed leaf age. The proposed models and methods can be applied to the study of different types of plants. The data for this study were obtained from a complex field environment, whereas in the previous studies on plant phenotypes, the data were obtained in an indoor environment, in which the image background is often pure and the illumination is uniform. Studying the field environment can help make the model more suitable for practical applications, and the shooting angle of the dataset determines the amount of target image information obtained; therefore, it is meaningful to study the segmentation results under different shooting angles. Only a few of the existing studies on plant phenotypes are specific to weed phenotypes. However, the weeds of different leaf ages require different doses of herbicides; therefore, it is of significance to obtain the information of weed leaf ages to reduce the amount of herbicides. To facilitate practical application, in future work, we can deploy the trained model on the mobile platform of the spray system used for weeding, which can promote the development of precision agriculture and intelligent agriculture.

5. Conclusions

This paper proposes a weed phenotype segmentation method based on the BlendMask to determine the weed species, leaf age and plant centre of weeds. In the field of plant phenotype research, the determination of weed phenotypes under complex field environments is a substantial challenge. According to the research status at home and abroad, the leaf age and plant centre are key phenotypic information of weeds. In this study, we identified the leaf age and plant centre, which are of significance to realize targeted weeding. The model performance could be enhanced through data enhancement. In addition, the weed image obtained from the top view angle corresponded to an enhanced model performance. Weed datasets were constructed through data collection from three angles and data enhancement, and the datasets contained the weed information, corresponding to the different growth stages, angles, and types. The dataset and research results can function as valuable resources for future plant phenotype research.

Because the DCNN can extract features from complex environments, it can effectively address the problems of complex images in the field. The experimental results show that despite the interference of the straw and crop leaves in the background of the weeds in the field, the BlendMask model using ResNet101 as the backbone network can realize accurate segmentation of the weeds with a satisfactory segmentation performance. Future research will be focused on evaluating image datasets that cover a wider range of weeds and crop varieties. Moreover, the identification efficiency of the proposed approach is low; thus, the model efficiency needs to be enhanced, and the trained model must be applied to the mobile platform of the spray system used for weeding. The proposed study combines artificial intelligence technology with agronomic research concepts, and the findings can facilitate the development of intelligent agriculture.

Abbreviations

AP	Average precision
AP50	mAP with the IOU threshold of 0.5
AP70	mAP with the IOU threshold of 0.7
CNN	Convolutional neural network
DCNN	Deep convolutional neural network
F_1	Metric function to balance P and R
Faster R-CNN	Faster regions with convolutional neural network features
FC	Fully convolutional
FCIS	Fully convolutional instance-aware semantic segmentation
FCN	Fully convolutional network
FCOS	Fully convolutional one-stage object detection
FN	False negative
FP	False positive
FPN	Feature pyramid network
GPU	Graphic processing unit
IOU	Intersection over the union
mAP	Mean average precision
Mask R-CNN	Mask regions with convolutional neural network features
mIOU	Mean intersection over the union
P	Precision rate
R	Recall rate
ROI	Region of interest
RPN	Region proposal network

Tmax	Maximum temperature
Tmin	Minimum temperature
TN	True negative
TP	True positive

Declarations

Availability of data and materials

Given that the data used in this study were self-collected, the dataset is being further improved. Thus, the dataset is unavailable at present.

Acknowledgements

The authors appreciate the financial support provided by the Overseas Study and Return Foundation of Heilongjiang LC2018019.

Funding

The authors appreciate the financial support provided by the Overseas Study and Return Foundation of Heilongjiang LC2018019.

Contributions

QLZ and WB prepared the manuscript, WB FHQ and MSR collected the data, MSR developed the codes, and YCJ, LHD and JW supervised the project. All the authors read and approved the manuscript.

Conflict of interest

The authors declare no conflict of interest.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Hashim S, Jan A, Fahad S, Ali HH, Mushtaq MN, Laghari KB et al. WEED MANAGEMENT AND HERBICIDE RESISTANT WEEDS: A CASE STUDY FROM WHEAT GROWING AREAS OF PAKISTAN.

Pakistan Journal of Botany. 2019;51(5):1761-7. doi:10.30848/pjb2019-5(28).

2. Slaughter DC, Giles DK, Downey D. Autonomous robotic weed control systems: A review. Comput Electron Agric. 2008;61(1):63-78. doi:10.1016/j.compag.2007.05.008.
3. Bakhshipour A, Jafari A, Nassiri SM, Zare D. Weed segmentation using texture features extracted from wavelet sub-images. Biosystems Engineering. 2017;157:1-12. doi:10.1016/j.biosystemseng.2017.02.002.
4. Partel V, Kakarla C, Ampatzidis Y. Development and evaluation of a low-cost and smart technology for precision weed management utilizing artificial intelligence. Computers and Electronics in Agriculture. 2019;157:339-50. doi:10.1016/j.compag.2018.12.048.
5. Gianessi LP, Reigner NP. The value of herbicides in US crop production. Weed Technology. 2007;21(2):559-66. doi:10.1614/wt-06-130.1.
6. Pretty J, Bharucha ZP. Integrated Pest Management for Sustainable Intensification of Agriculture in Asia and Africa. Insects. 2015;6(1):152-82. doi:10.3390/insects6010152.
7. Jeranyama P, Ndlovu F, Morgan J. Plant Physiology Research. 2011.
8. Qinghu L, Guoqi C, Yuhua Z, Zhonghua S, Liyao D. Sensitivities of Leptochloa chinensis,Echinochloa crusgalli and Digitaria sanguinalis at different leaf stages to cyhalofop-butyl and penoxsulam. Journal of Nanjing Agricultural University. 2016;39(05):771-6.
9. Xiu L, Yu-quan D, Jing-bo L, Yun-yun Z, Chen-zhong J. Sensitivity of Barnyard Grass at Different Leaf Stage to Bispyribac-Sodium and Cyhalofop-Butyl. Journal of Weeds. 2017;35(03):22-6.
10. YiMing P. RESEARCH ON ALGORITHM OF RICE DISEASES IDENTIFICATION AND LEAFAGE DETECTION BASED ON MACHINE LEARNING [Master of Engineering]: Harbin Institute of Technology; 2019.
11. Jetter R, Schaffer S. Chemical composition of the Prunus laurocerasus leaf surface. Dynamic changes of the epicuticular wax film during leaf development. Plant physiology. 2001;126(4):1725-37. doi:10.1104/pp.126.4.1725.
12. Zeisler-Diehl V, Muller Y, Schreiber L. Epicuticular wax on leaf cuticles does not establish the transpiration barrier, which is essentially formed by intracuticular wax. Journal of Plant Physiology. 2018;227:66-74. doi:10.1016/j.jplph.2018.03.018.
13. Garcia-Santillan ID, Pajares G. On-line crop/weed discrimination through the Mahalanobis distance from images in maize fields. Biosyst Eng. 2018;166:28-43. doi:10.1016/j.biosystemseng.2017.11.003.
14. Brivot R, Marchant JA. Segmentation of plants and weeds for a precision crop protection robot using infrared images. IEE Proceedings Part K. 1996;143(2).
15. Jeon HY, Tian LF, Zhu HP. Robust Crop and Weed Segmentation under Uncontrolled Outdoor Illumination. Sensors. 2011;11(6):6270-83. doi:10.3390/s110606270.
16. Bossu J, Gee C, Jones G, Truchetet F. Wavelet transform to discriminate between crop and weed in perspective agronomic images. Comput Electron Agric. 2009;65(1):133-43.

doi:10.1016/j.compag.2008.08.004.

17. Shirzadifar A, Bajwa S, Mireei SA, Howatt K, Nowatzki J. Weed species discrimination based on SIMCA analysis of plant canopy spectral data. Biosyst Eng. 2018;171:143-54. doi:10.1016/j.biosystemseng.2018.04.019.
18. Ozdogan M, Yang Y, Allez G, Cervantes C. Remote Sensing of Irrigated Agriculture: Opportunities and Challenges. Remote Sensing. 2010;2(9):2274-304. doi:10.3390/rs2092274.
19. Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment. 2002;83(1-2):195-213.
20. Minervini M, Fischbach A, Scharr H, Tsaftaris SA. Finely-grained annotated datasets for image-based plant phenotyping. Pattern Recognit Lett. 2016;81:80-9. doi:10.1016/j.patrec.2015.10.013.
21. Bell J, Dee HM. Leaf segmentation through the classification of edges. 2019.
22. Dobrescu A, Giuffrida MV, Tsaftaris SA. Doing More With Less: A Multitask Deep Learning Approach in Plant Phenotyping. Frontiers in Plant Science. 2020;11. doi:10.3389/fpls.2020.00141.
23. Weng Y, Zeng R, Wu C, Wang M, Wang X, Liu Y. A survey on deep-learning-based plant phenotype research in agriculture. Scientia Sinica Vitae. 2019;49(6):698-716.
24. Kamilaris A, Prenafeta-Boldu FX. Deep learning in agriculture: A survey. Comput Electron Agric. 2018;147:70-90. doi:10.1016/j.compag.2018.02.016.
25. Le VNT, Ahderom S, Apopei B, Alameh K. A novel method for detecting morphologically similar crops and weeds based on the combination of contour masks and filtered Local Binary Pattern operators. Gigascience. 2020;9(3). doi:10.1093/gigascience/giaa017.
26. Gao JF, French AP, Pound MP, He Y, Pridmore TP, Pieters JG. Deep convolutional neural networks for image-based Convolvulus sepium detection in sugar beet fields. Plant Methods. 2020;16(1). doi:10.1186/s13007-020-00570-z.
27. Quan LZ, Feng HQ, Li YJ, Wang Q, Zhang CB, Liu JG et al. Maize seedling detection under different growth stages and complex field environments based on an improved Faster R-CNN. Biosystems Engineering. 2019;184:1-23. doi:10.1016/j.biosystemseng.2019.05.002.
28. Geetharamani G, Pandian JA. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. Computers & Electrical Engineering. 2019;76:323-38. doi:10.1016/j.compeleceng.2019.04.011.
29. Le TT, Lin CY, Piedad E. Deep learning for noninvasive classification of clustered horticultural crops - A case for banana fruit tiers. Postharvest Biology and Technology. 2019;156. doi:10.1016/j.postharvbio.2019.05.023.
30. Mccool CS, Perez T, Upcroft B. Mixtures of Lightweight Deep Convolutional Neural Networks: applied to agricultural robotics. IEEE Robotics & Automation Letters. 2017;2(3):1344-51.
31. Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Garcia-Rodriguez J. A Review on Deep Learning Techniques Applied to Semantic Segmentation. 2017.

32. Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. Ieee Transactions on Pattern Analysis and Machine Intelligence. 2017;39(4):640-51. doi:10.1109/tpami.2016.2572683.
33. Yu JG, Li YS, Gao CX, Gao HX, Xia GS, Yu ZL et al. Exemplar-Based Recursive Instance Segmentation With Application to Plant Image Analysis. Ieee Transactions on Image Processing. 2020;29:389-404. doi:10.1109/tip.2019.2923571.
34. Huang SP, Wu SH, Sun C, Ma X, Jiang Y, Qi L. Deep localization model for intra-row crop detection in paddy field. Comput Electron Agric. 2020;169. doi:10.1016/j.compag.2019.105203.
35. He K, Gkioxari G, Dollar P, Girshick R, editors. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV); 2017.
36. He KM, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. IEEE Trans Pattern Anal Mach Intell. 2020;42(2):386-97. doi:10.1109/tpami.2018.2844175.
37. Jia WK, Tian YY, Luo R, Zhang ZH, Lian J, Zheng YJ. Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot. Comput Electron Agric. 2020;172. doi:10.1016/j.compag.2020.105380.
38. Chen H, Sun K, Tian Z, Shen C, Huang Y, Yan Y, editors. BlendMask: Top-Down Meets Bottom-Up for Instance Segmentation. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020.
39. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D et al. Microsoft COCO: Common Objects in Context. 2014.
40. Tian Z, Shen C, Chen H, He T, editors. FCOS: Fully Convolutional One-Stage Object Detection. 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 2020.
41. Garcia-Santillan ID, Montalvo M, Guerrero JM, Pajares G. Automatic detection of curved and straight crop rows from images in maize fields. Biosystems Engineering. 2017;156:61-79. doi:10.1016/j.biosystemseng.2017.01.013.
42. Ho D, Tong M, Ienco D, Gaetano R, Maurel P. Deep Recurrent Neural Networks for mapping winter vegetation quality coverage via multi-temporal SAR Sentinel-1. IEEE Geoscience & Remote Sensing Letters. 2017;PP(99):1-5.
43. Ienco D, Gaetano R, Dupaquier C, Maurel P. Land Cover Classification via Multitemporal Spatial Data by Deep Recurrent Neural Networks. Ieee Geoscience and Remote Sensing Letters. 2017;14(10):1685-9. doi:10.1109/lgrs.2017.2728698.
44. Lee SH, Chan CS, Wilkin P, Remagnino P. Deep-plant: Plant identification with convolutional neural networks. 2015.
45. Krizhevsky A, Sutskever I, Hinton GE, editors. ImageNet Classification with Deep Convolutional Neural Networks. International Conference on Neural Information Processing Systems; 2012.
46. Ma J, Li Y, Chen Y, Du K, Zheng F, Zhang L et al. Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network. European Journal of Agronomy. 2019;103:117-29. doi:10.1016/j.eja.2018.12.004.

47. Dutta A, Zisserman A. The VGG Image Annotator (VIA). 2019.
48. Ren SQ, He KM, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Ieee Transactions on Pattern Analysis and Machine Intelligence. 2017;39(6):1137-49. doi:10.1109/tpami.2016.2577031.
49. Yu Y, Zhang KL, Yang L, Zhang DX. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. Computers and Electronics in Agriculture. 2019;163:9. doi:10.1016/j.compag.2019.06.001.
50. Li Y, Qi H, Dai J, Ji X, Wei Y, editors. Fully Convolutional Instance-Aware Semantic Segmentation. Computer Vision & Pattern Recognition; 2017.
51. Bolya D, Zhou C, Xiao F, Lee YJ, editors. YOLACT: Real-Time Instance Segmentation. International Conference on Computer Vision.
52. Hripcsak G, Rothschild AS. Agreement, the F-measure, and reliability in information retrieval. Journal of the American Medical Informatics Association. 2005;12(3):296-8. doi:10.1197/jamia.M1733.
53. Al-Najjar HAH, Kalantar B, Pradhan B, Saeidi V, Halin AA, Ueda N et al. Land Cover Classification from fused DSM and UAV Images Using Convolutional Neural Networks. Remote Sensing. 2019;11(12). doi:10.3390/rs11121461.
54. Zhang E, Yi Z. Average Precision. 2016.
55. Wang X, Kong T, Shen C, Jiang Y, Li L. SOLO: Segmenting Objects by Locations. 2019.
56. Xie E, Sun P, Song X, Wang W, Luo P, editors. PolarMask: Single Shot Instance Segmentation With Polar Representation. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020.
57. Lee Y, Park J, editors. CenterMask: Real-Time Anchor-Free Instance Segmentation. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020.
58. Lehoczky E, Nagy P, Lencse T, Toth V, Kismanyoky A. Investigation of the Damage Caused by Weeds Competing with Maize for Nutrients. Communications in Soil Science and Plant Analysis. 2009;40(1-6):879-88. doi:10.1080/00103620802694944.

Figures

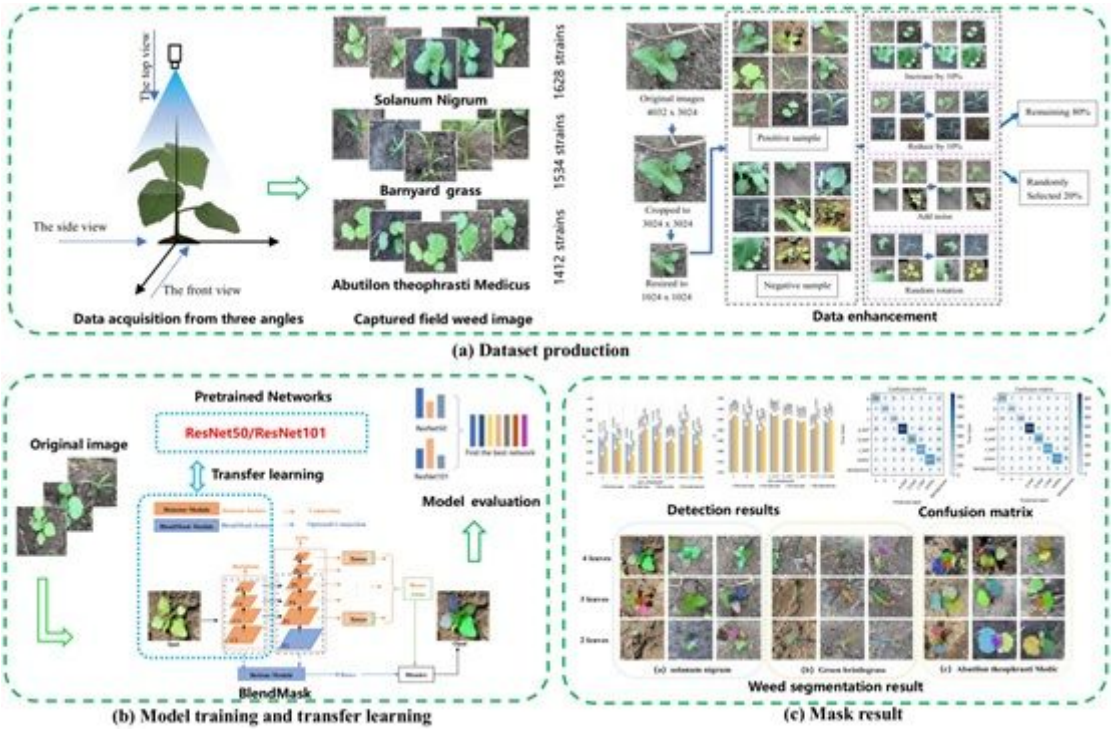


Figure 1

Process flow of using BlendMask to segment the leaf age and plant centre.

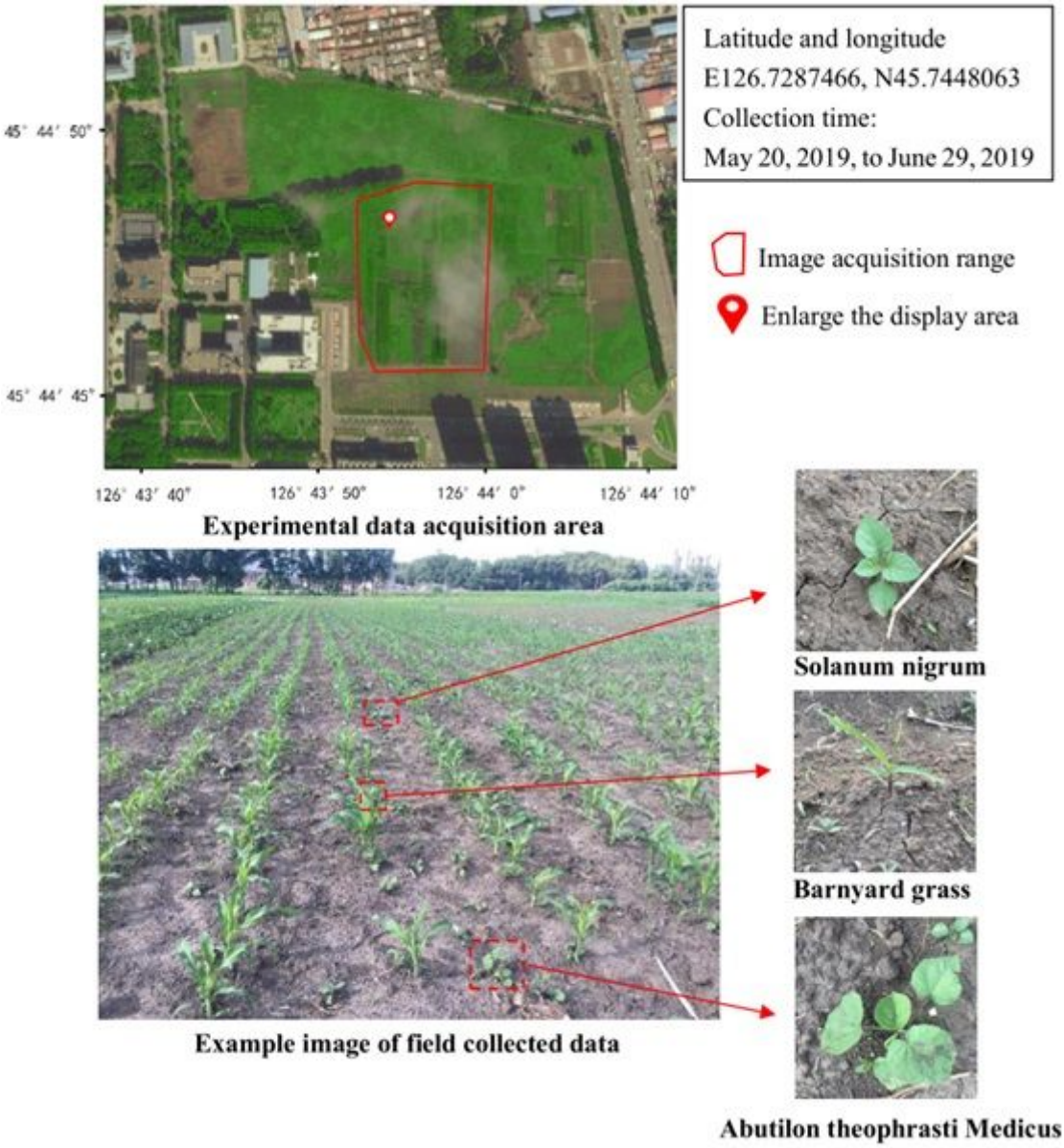


Figure 2

Image collection area and collection varieties

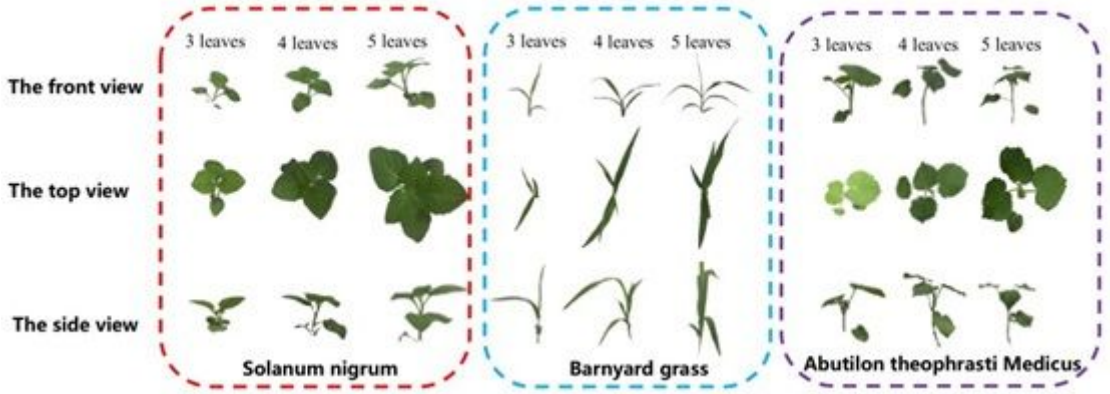


Figure 3

Front, top, and side views of the three weeds.

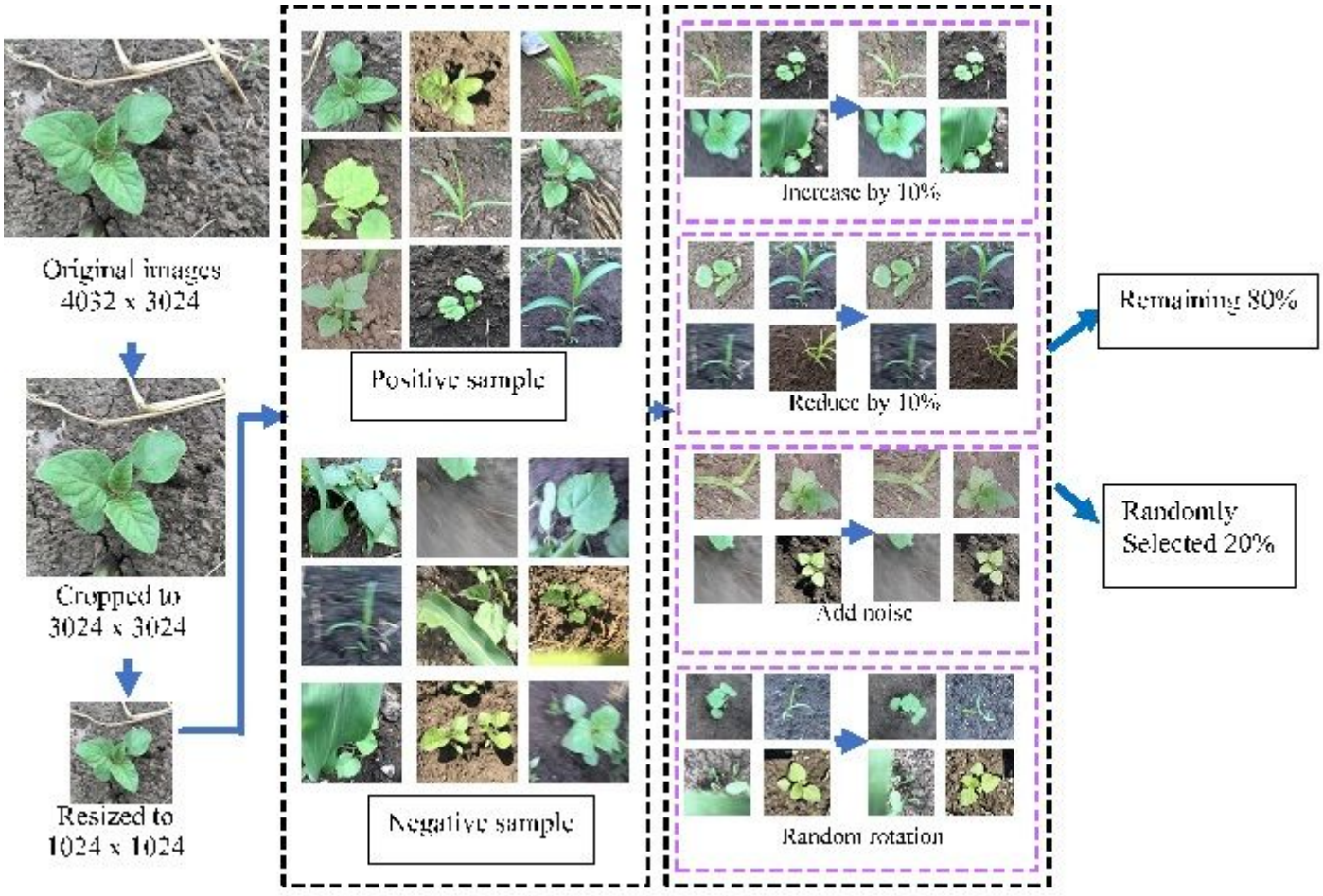


Figure 4

Data enhancement. Note: The size of the original images is 4032×3024. After cropping the images to a size of 3024×3024, the images are resized to 1024×1024. The collected images include positive and negative samples. The acquired images are brightened and darkened by 10% and subjected to increased noise and random rotation. Finally, the dataset is randomly divided into training and verification sets with a ratio of 8:2.

Figure 5

Sample of image annotation in the segmentation process. Note: Leaf of Solanum nigrum (a_leaf), Barnyard grass (b_leaf), and Abutilon theophrasti Medicus (c_leaf). Plant centre (centre). Solanum nigrum (a), Barnyard grass (b), and Abutilon theophrasti Medicus (c).

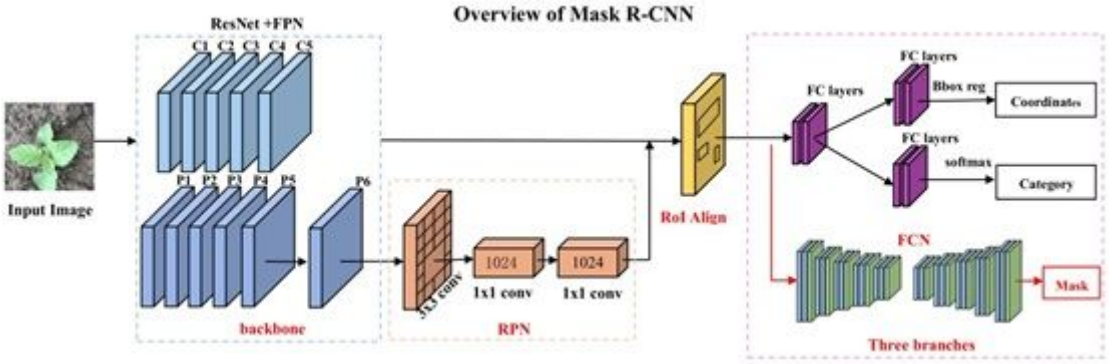


Figure 6

Model structure of Mask R-CNN.

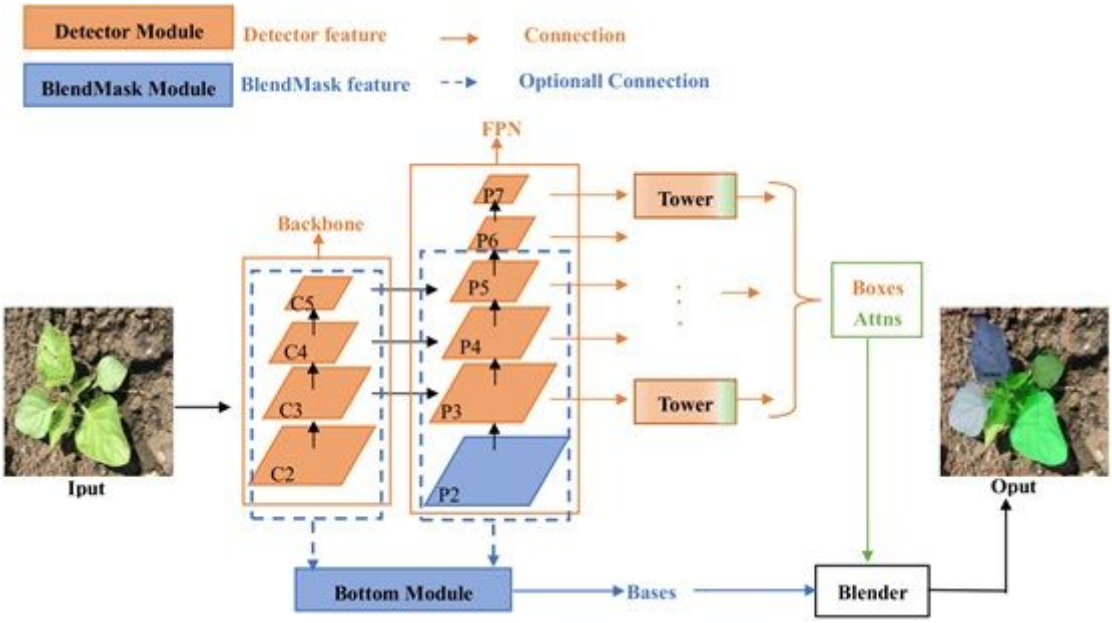


Figure 7

Model structure of BlendMask

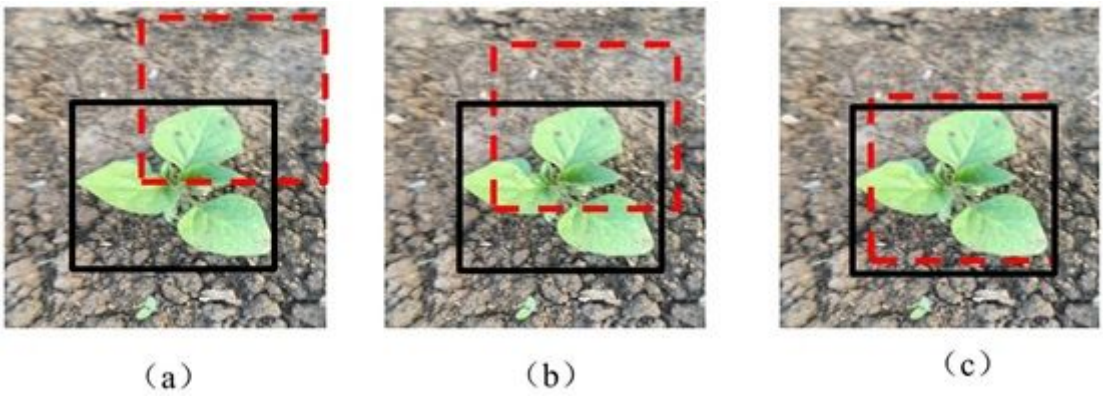


Figure 8

Visual example of intersection over the union. Note: The red dashed box and black solid-line box represent the prediction of the detection result and ground-truth, respectively. The overlap between the two boxes is visible, and the different IOU values can be determined. (c) indicates the optimal case.

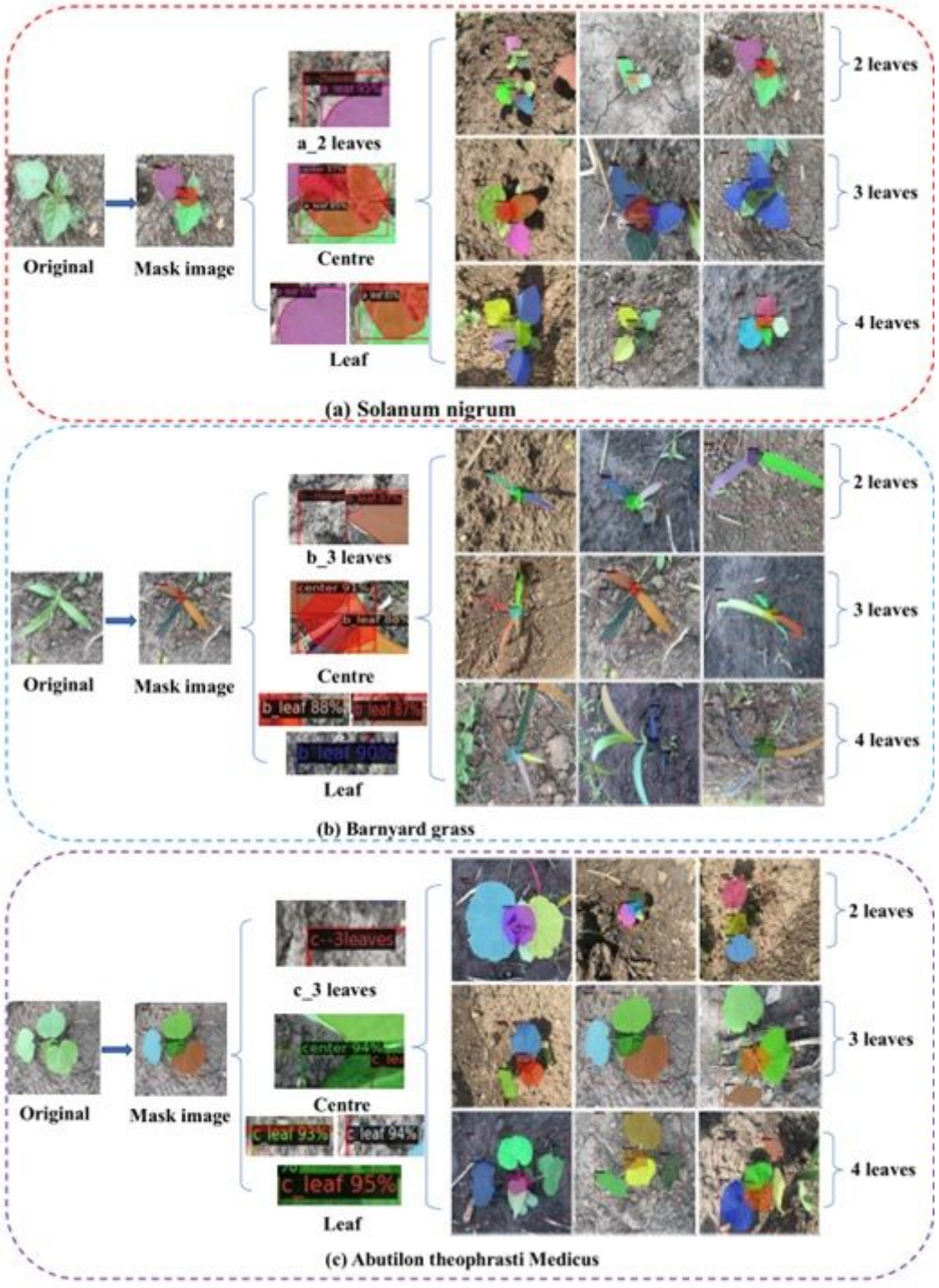


Figure 9

Segmentation results for the three weeds with different leaf ages Note: The left area of the image shows the leaves, plant centre and leaf age for each type of weed mask. (a), (b) and (c) show the recognition

results for the different types of weeds with different leaf ages.

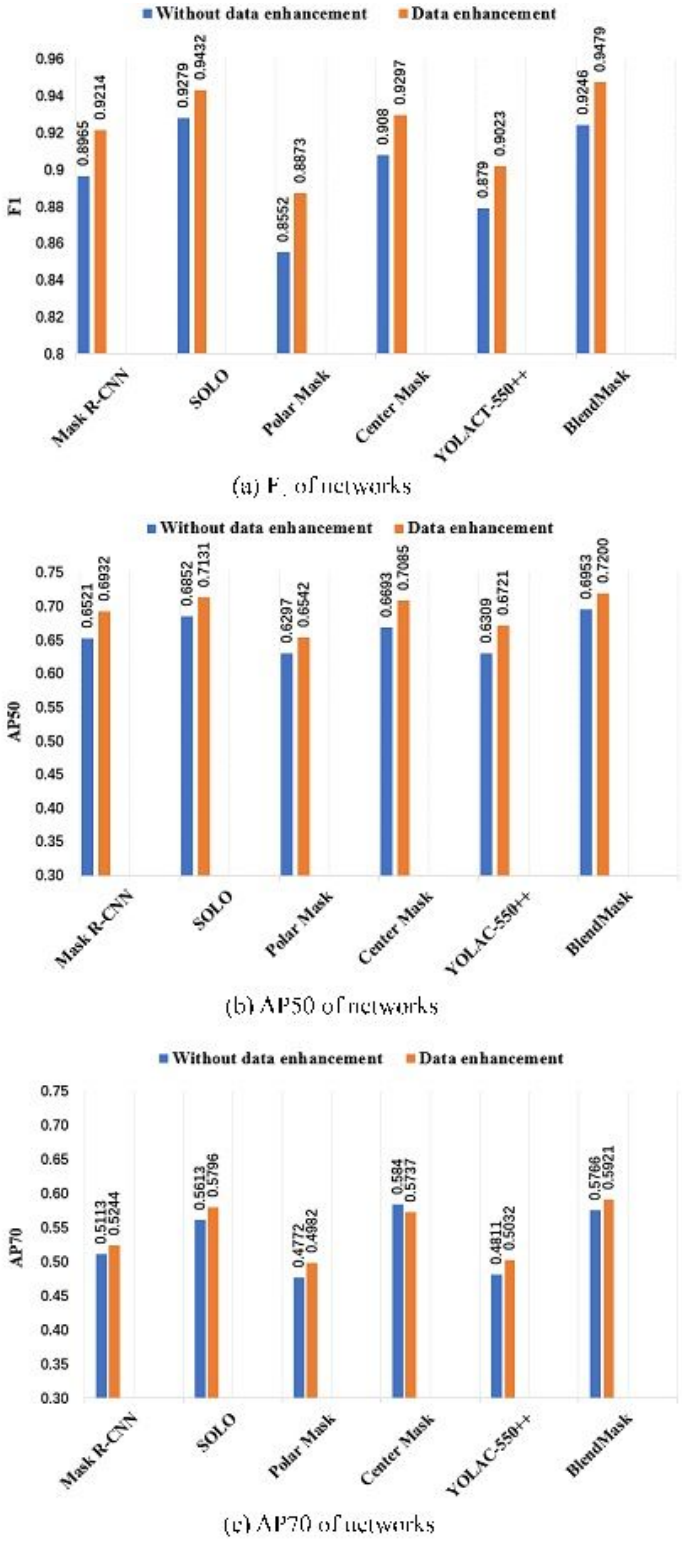
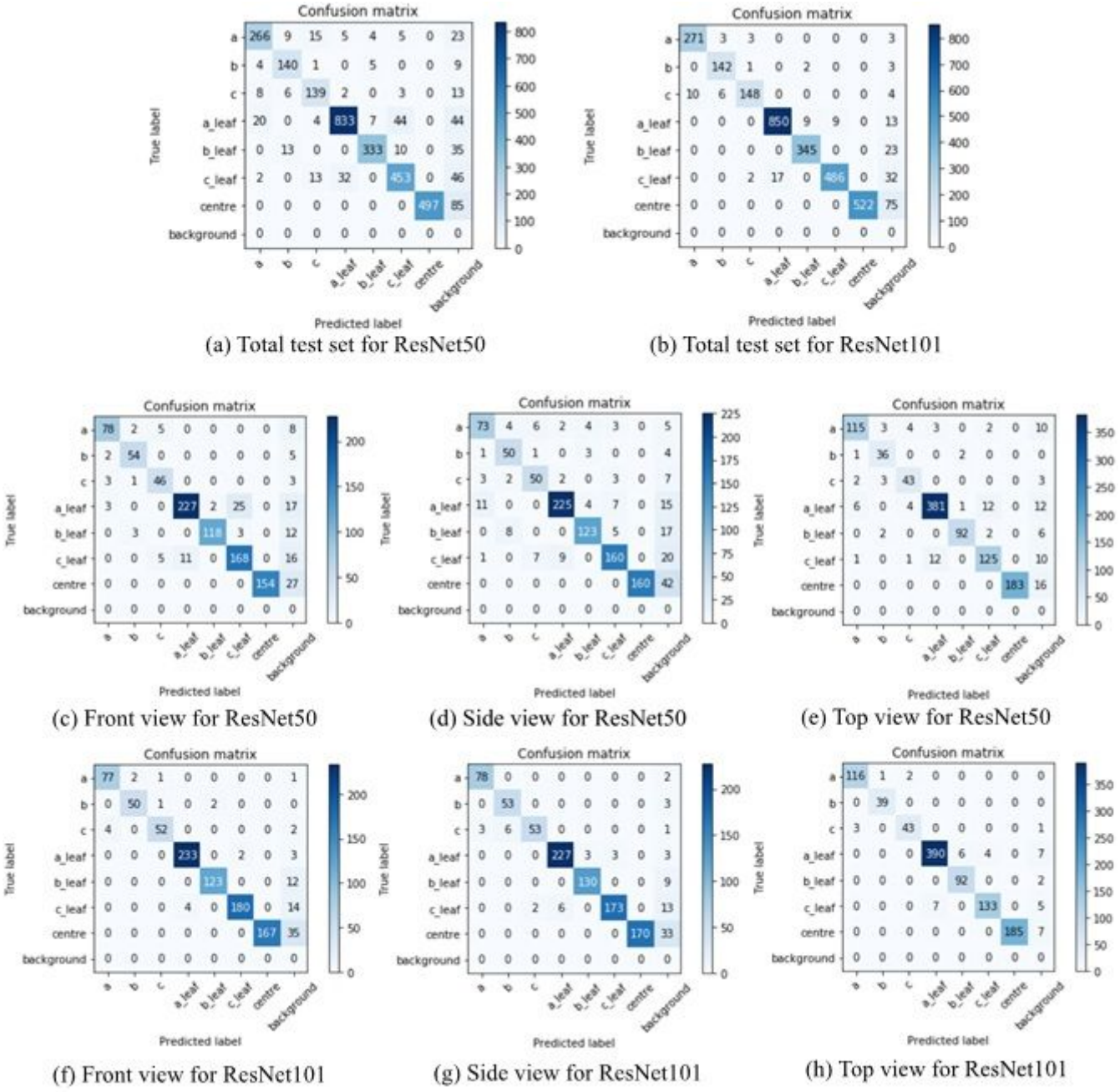


Figure 10

Detection results for different instance segmentation models Note: The backbone network of the six models(YOLACT, PolarMask, BlendMask, CentreMask, SOLO and Mask R-CNN) is ResNet101. “Without data enhancement” and “Data enhancement” refer to the models trained using 4000 unenhanced

datasets and 6000 enhanced datasets, respectively. When the IOU threshold is greater than or equal to 0.5 and 0.7, the mAP is defined as AP50 and AP70, respectively.



(a) Total test set for ResNet50 (b) Total test set for ResNet101

(c) Front view for ResNet50 (d) Side view for ResNet50 (e) Top view for ResNet50

(f) Front view for ResNet101 (g) Side view for ResNet101 (h) Top view for ResNet101

Figure 11

Confusion matrix of the detection results of ResNet50 and ResNet101 in the case of data enhancement Note: The BlendMask with ResNet50 and ResNet101 frameworks are expressed as ResNet50 and ResNet101, respectively. a_leaf, b_leaf, and c_leaf represent the leaves of Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus, respectively. Moreover, a, b, and c represent Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus, respectively, and centre represents the plant centre of each weed.

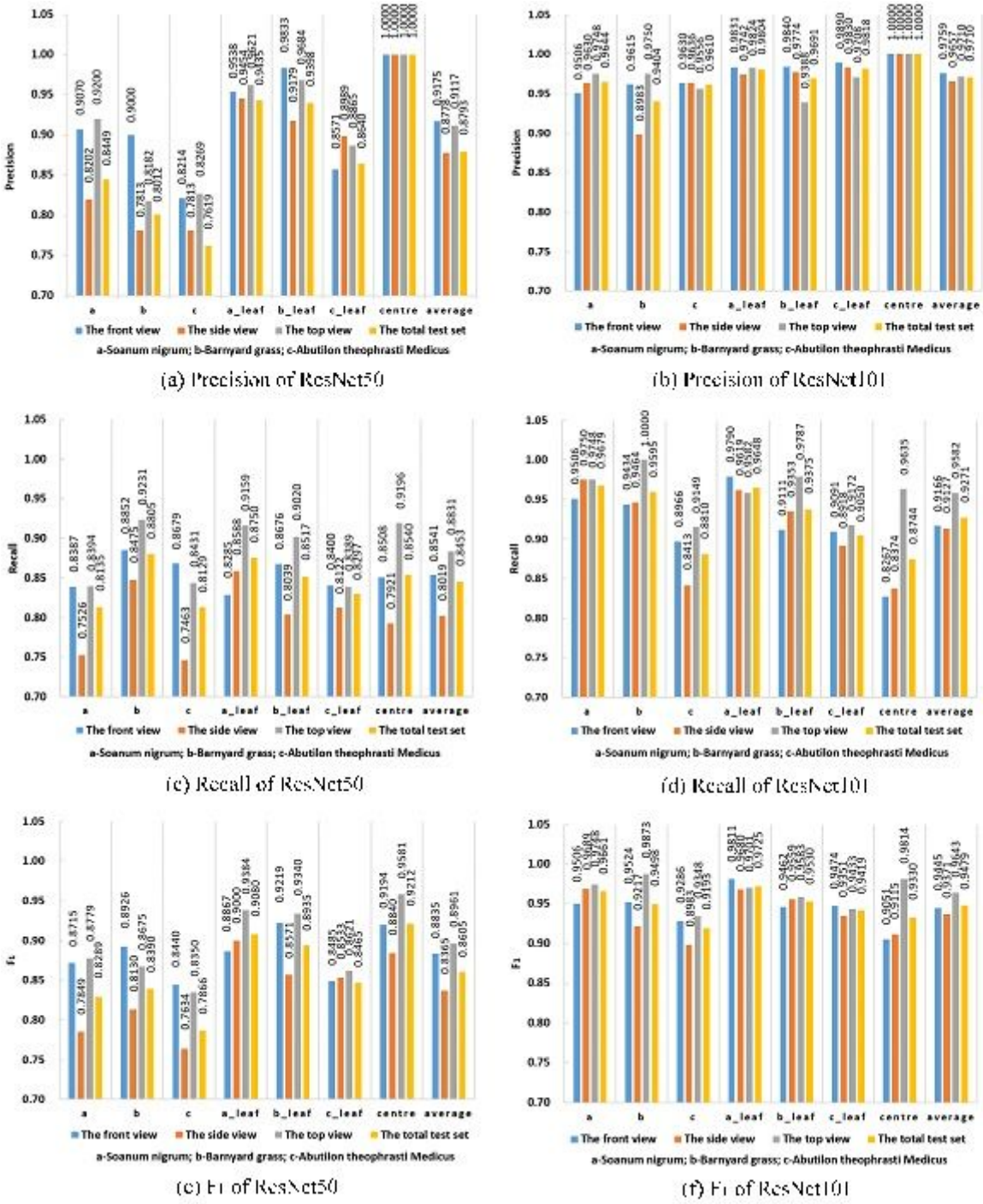


Figure 12

Detection results of the BlendMask with pretrained networks in the case of data enhancement. Note: The BlendMask with ResNet50 and ResNet101 frameworks are expressed as ResNet50 and ResNet101, respectively. Centre represents the plant centre of each weed.

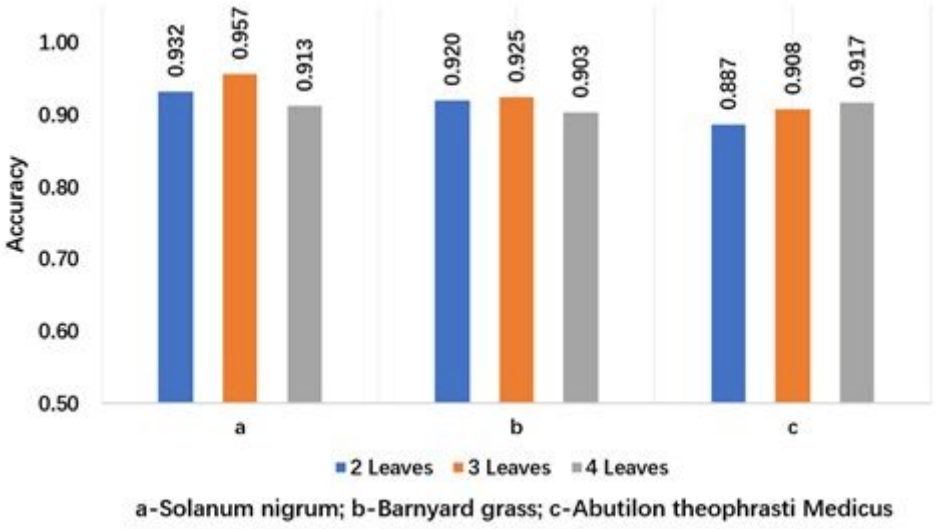


Figure 13

Accuracy for different leaf ages of the BlendMask in the case of data enhancement.