Method for obtaining phenotypic traits of weeds in complex field environments based on instance segmentation

Longzhe Quan*, Bing Wu[1], Shouren Mao[1], Huaiqu Feng[1], Chunjie Yang[1], Wei Jiang[1], Hengda Li[1]

1. College of Engineering, Northeast Agricultural University, Harbin, 150030, China

Abstract

Background:

Weeds are the biggest threat to crop growth, and leaf age and central area of weeds are important phenotypic traits of weeds. They have an important role in understanding the morphological structure of weeds, guiding precision for target weeding and reducing the use of herbicides. However, it is still a substantial challenge to obtain weed types, leaf age and central area in the complex field conditions of light changes, variation in appearance of plants, leaf occlusion. The latest developments in deep learning provide new tools for solving challenging computer vision tasks.

Results:

In this study, we present a weed phenotype segmentation method based on Mask R-CNN that obtains weed types, leaf age and central area in the complex field conditions. By shooting three different angles of the main weeds (Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus) in the field through mobile devices, two datasets (data enhancement and without data enhancement) were produced and used as input to the network, two backbone networks was tested, namely ResNet50 and ResNet101, and the detection results and instance segmentation results of the model were evaluated. The results showed that data enhancement can improve the performance of the model. In the case of data enhancement, the F_1 value

with ResNet101 as the backbone network was 0.9214, the mAP scores were 0.6932 and 0.5244 (for IOU thresholds of 0.5 and 0.7, respectively), the mIOU reached 0.585, and the best segmentation performance example was obtained. Furthermore, the weed image taken from the top view angle compared to the other two angles achieved the highest detection accuracy.

Conclusion:

Mask R-CNN can achieve accurate segmentation of weeds to obtain the types, leaf age and central area of weeds. Data enhancement and the weed image taken from the top view angle can help to improve the performance of the model. This dataset and research results may provide important resources for the development of precision agriculture in the future.

Keywords: Leaf age; Mask R-CNN; weed segmentation; Deep learning; Machine vision.

Nomenclature

Symbols

AP	Average precision
CNN	Convolutional neural network
DCNN	Deep convolutional neural network
F_1	Metric function of balance P and R
Faster R-CNN	Faster Regions with convolutional neural network features
FC	Fully convolutional
FCN	Fully convolutional network
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 thermometer

62	Tmin	Minimum thermometer
63	TN	True negative
64	TP	True positive

65

1. Background

Weeds are the largest threat to crop growth, hindering the growth of crops and promoting the use of herbicides [1], which can lead to a large reduction in crop yield [2], and because weeds compete with major cash crops for space, water, light, nutrients and other resources, the loss of plants is greater in the early stages of plant growth [3]. Therefore, removing weeds as early as possible is critical to ensure good crop yield. The use of herbicides has increased crop yields, and reduced the amount of labour [4]; however, more than 90% of the crops in the United States have had herbicides applied to them [5], and the global use of pesticide is estimated to be 3.5 billion kg/year [4]. Excessive use of herbicides has brought about a series of problems such as environmental pollution. Therefore, there is an urgent need to reduce the use of herbicides without affecting the yield of crops, which will also reduce the cost of weed management and is also a valuable goal 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 phenotype information of weeds is closely related to the dosage of herbicide. For example, the herbicide dosage required for weeds of different leaf ages is different [8] [9]. The upper limit of herbicide application can kill weeds of different ages, but it can also cause excessive use of herbicides. Leaf age refers to different stages in the development of a plant. Generally speaking, the number of complete leaves grown by the plant is the leaf age [10]. The leaves of the weed are connected to the stem through the petiole. When we look down, it will show the central area formed by the overlapping of the petiole and stem of the top leaves of the weed. This area is also the intersection centre area of weed top leaves [7]. In this study we simply referred to as the central area. Most of the central area is new tissue. Because the epicuticular waxes of plant leaves are

closely related to the absorption of herbicides, but the composition of epicuticular waxes in different parts of the same plant is different and will vary with the season, location and age of plants [11] [12], resulting in different organs of the same plant having different sensitivities to herbicides. The new tissue has a large number of stomata and a thin waxy layer, which is beneficial to the absorption of herbicides. Therefore, the central area is more sensitive to herbicides, which is beneficial to herbicide absorption, and the central area has better retention capacity, which is a particularly obvious characteristic especially for weeds over four in leaf age. Therefore, the central area is more sensitive to herbicides and is beneficial to herbicide absorption. Therefore, an effective method is to reduce the dosage of herbicide by guiding the use of herbicide according to the close relationship between the leaf age of the weeds and the central area and the absorption and conduction of the herbicide.

The premise of reducing herbicide use is to accurately identify 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, proving the feasibility of machine vision in the agricultural field. Researchers have developed a new algorithm that can segment plants according to the soil background under uncontrolled lighting conditions in the wild, separating weeds from crops [15]. However, this method is less effective when there is no change in the field image under lighting conditions. Wavelet transform was used to distinguish weeds from crops in the image [16], but when the number of weeds is large, this traditional visual method may not be useful. These methods only distinguish crops from weeds and do not obtain the type and phenotype information of weeds. Although Shirzadifar et al. [17] classified the weeds based on the canopy spectral information of the plants and

obtained the species of weeds, achieving good results, in the complex field environment, wind, soil background, and shadows will change the spectral characteristics of plants, affect the performance of the model [18] [19]. A substantial body of related research has also examined the phenotype information of plants. Minervini et al.[20] established finely grained datasets for image-based plant phenotyping, which provided a great contribution to the study of plant phenotypes, and leaf segmentation is also a very important challenge in the field of the plant phenotype analysis. Bell et al. [21] segmented leaves by edge classification and achieved good results for plant overlap. Dobrescu et al. [22] proposed a multi-task deep learning framework for plant phenotypes and achieved good results in leaf counting. But the plant images of these studies were collected under indoor conditions, and images collected indoors tend to have a pure background and light uniformity [23]. These studies are mainly to calculate the number of leaves for leaf segmentation, but an image may have multiple weeds, and our research needs to obtain the leaf age, weed type and the central area of each weed. It can be seen that segmentation of plant phenotypes in a complex farmland environment is still an area that has not been fully studied. Due to the complex environment of farmland, the differences between the plants, and the mutual occlusion of the leaves, it is very difficult to segment the leaf age and the central area, making this study a certain challenge. In this study, the weed species, growth stage and central area were segmented by machine vision in a complex field environment. The purpose is to prepare for guiding the use of herbicides.

Deep learning is an emerging field of machine learning that is employed to solve 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. An agricultural survey has shown that deep learning technology has higher accuracy than traditional image processing technology [24]. Because in the complex field environment, the illumination is uncertain, the plants overlap, the climate changes, and the soil background is complex and changeable [25] [26], which makes the DCNN model perform better in this respect. Although the DCNN model can overcome these difficulties, the farmland environment is complex, it needs a large enough dataset to train the deep learning model, which is helpful to overcome the complex field environment and improve the accuracy of the model [27]. One method to solve this problem is data enhancement. Data enhancement is also a common method in the field of image recognition. 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 six methods of data enhancement were used to improve the performance of the model, achieving 96.4% classification accuracy. Piedad et al. [29] used the Mask R-CNN model for non-invasive classification of clustered horticultural crops. Due to the limited dataset, the dataset was expanded to improve the accuracy of the model. It can be seen that data enhancement is an important method to enrich the training samples and improve the performance of the model, and can also make the dataset more suitable for the complex farmland environment.

When we collect the dataset, the shooting angle in the field [30], the growth stage of the weeds may affect the accuracy of the dataset. Quan and Feng realized the detection of seedlings in different growth cycles and different angles in the field. It was proposed that when the angle between the camera and the vertical direction is 0°,

the detection accuracy is 0.95% lower than other angles [27], It could be seen that data collection from different angles would affect the performance of the model. The position and shape of weeds in the field are complex and changeable, the shape of the same object is different under different shooting angles, which affects the accuracy of the dataset. Therefore, in this study, we collected data from three angles: front view, side view and top view to explore the impact of different angles on the model under study. In this study, an instance segmentation algorithm based on deep learning is proposed to obtain the weed phenotype in a complex field environment. From the previous research, we can know that the DCNN model performs better in dealing with complex environmental problems in the field. Instance segmentation based on deep learning is a new challenge of computer vision [31]. The model used in this study is a new instance segmentation model; its purpose is to detect each object in a weed image and classify each pixel of each instance. The output is the mask and bounding box of the target object [32], which is particularly suitable for solving the problem of leaf adhesion and occlusion [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 prove 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%. It is worth noting that this method identifies a stem-base-centred square region at the plant level, which corresponds to the protected area image of mechanical weeding, and this area is also the central area of the plant. It can be seen that the central area is also of great significance for plant research. The instance segmentation algorithm Mask R-CNN proposed by He et al. [35] can not only identify the bounding box but also mask the

target contour, performing better than 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%. Some scholars also used the Mask R-CNN model to complete the fruit detection of the strawberry picking robot in a non-structural environment, which overcame the conditions of overlapping and hidden fruits in a non-structural environment [38].

Therefore, the above research provides a feasible basis and reference for the application of DCNN in plant segmentation. It also proves that DCNN can overcome the shortcomings of traditional image segmentation methods. The Mask R-CNN model shows good performance when dealing with complex field environments. The weeds we selected are as follows: Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus; these three weeds are commonly found in fields in Northeast China. It can be seen from the above research that we need a sufficient number of well-defined weed datasets to train DCNN models. Weed images should be obtained from real scenes in fields so that they contain the morphological characteristics of the weeds at different growth stages in the complex field environment and cover more variables in the input of the model. We collected three different angle images (front view, side view, top view) of three weeds. The classic DCNN network is modified to improve the accuracy of the model. Therefore, we created two datasets: one containing 4000 weed images without data enhancement, and the other one containing 6000 data-enhanced weed images.

Based on the above problems, this study proposes a weed phenotype segmentation method based on the improved Mask R-CNN to obtain the weed species, leaf age and central area of weeds. The main objectives of the present study are

(1) Evaluate the feasibility of Mask R-CNN to obtain weed species, leaf age and

central area in weed phenotype segmentation through 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 the complex field environment, and select the most suitable angle.

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

(4) To explore whether using Resnet101 combined with the FPN architecture for feature extraction can improve model performance.

2. Materials and methods

2.1. Overview

In this section, we select three typical weeds of Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus from Northeast China. The leaf age and central area of weeds were segmented by the Mask R-CNN model. The dataset needed to train the model was collected in the actual field environment. The datasets were collected at different angles (front view, side view, top view) and were enhanced. The datasets produced were annotated, and the generated file was used as the input of the network to train the network model. The backbone network of the Mask R-CNN initialization model is the residual network combined with the feature pyramid network (FPN). This study uses different backbone networks (ResNet50 and ResNet101) combined with the FPN architecture. The feature extraction performance based on the weed leaf age and central area was evaluated. Figure 1 shows the workflow of the experiment.

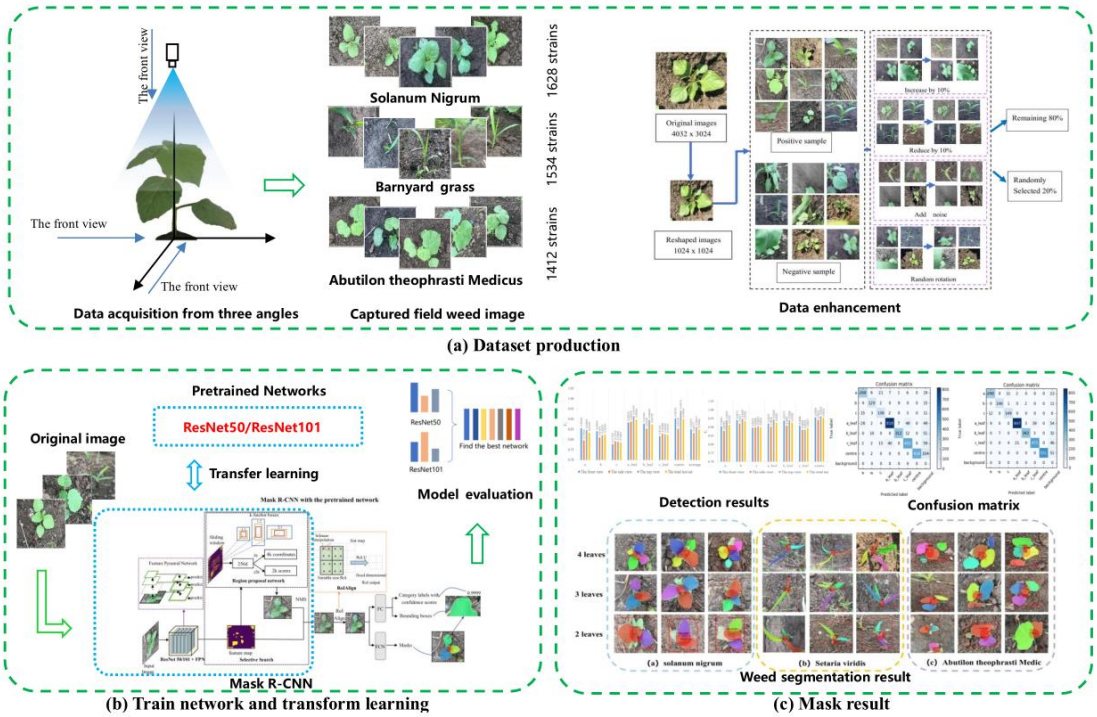


Fig. 1 Work process of using Mask R-CNN to segment leaf age and central area.

2.2. Image acquisition

The three kinds of weeds selected in this study are: Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus. Solanum nigrum is an annual dicotyledon, and Barnyard grass is an annual herb, an annual subshrub weed of Abutilon theophrasti Medicus. These three kinds of weeds are common in fields of Northeast China. At present, the data image source is the field weed image. Because the background of greenhouse weeds is single and the image of field weeds is more complex, the ability of the model to recognize weeds in the natural state can be verified. The field data images were collected in the Xiangfang District from May to June 2018. 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 above weed images were collected. The weeds in the field are mainly weeds of leaf ages of 2-5; thus, we collected only weed images before the 5 leaf stage.

Since the weed information obtained from a single shooting angle is not comprehensive, in order to better show that the weed information obtained from different angles is different, the field weed background is removed, as shown in Figure 2. In addition, the shooting weather [39] and acquisition angle [30] had a greater impact on the dataset, affecting the segmentation precision [40]. Therefore, from the two-leaf period after crop planting, an iPhone 6s Plus camera with a 4.2 mm focal length, a maximum aperture of f/2.2 and a maximum resolution of 4032 x 3024 pixels was used to capture images. From May 20, 2019, to June 29, 2019, weed images of different leaf ages were collected every 2 to 5 days under different weather conditions, different angles, and different growth stages, to obtain the data of weeds at each leaf age stage in the growth cycle, as shown in Table 1, and the weed images were stored in the JPEG file format. The purpose of generating the dataset is to study the recognition performance of the deep learning model for the leaf age and central area of individual weeds at different growth stages in the natural state.

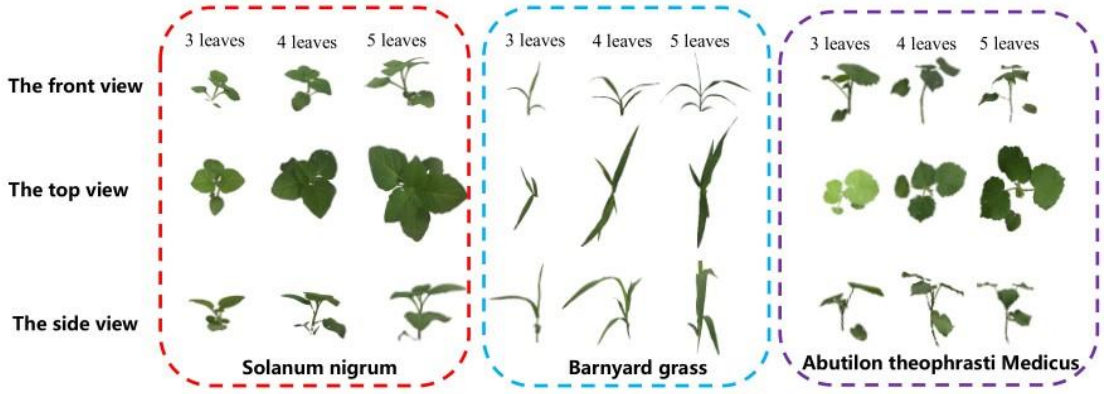


Fig. 2 Front view, top view, and side view of the three weeds.

Table 1 List of images containing environmental information for the experiment

Date	Images	Tmax	Tmin	Weather	The front view	The top view	The 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	28	16	\	1411	1862	1301

2.3. Dataset construction and annotation

When training the network and conducting network testing, the input image size needs to match the input size of the network [41], thus, the images are adjusted to a pixel size of 1024 x 1024 to construct the image dataset of the DCNN. Due to the need to annotate the weeds in the picture, some images that are not suitable for annotation are discarded. Therefore, 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 such that they are more representative and reflect the real situation of field data more accurately [27], while additionally improving the training precision of the model [42], expanding the dataset and reducing overfitting [43] (Figure 3).

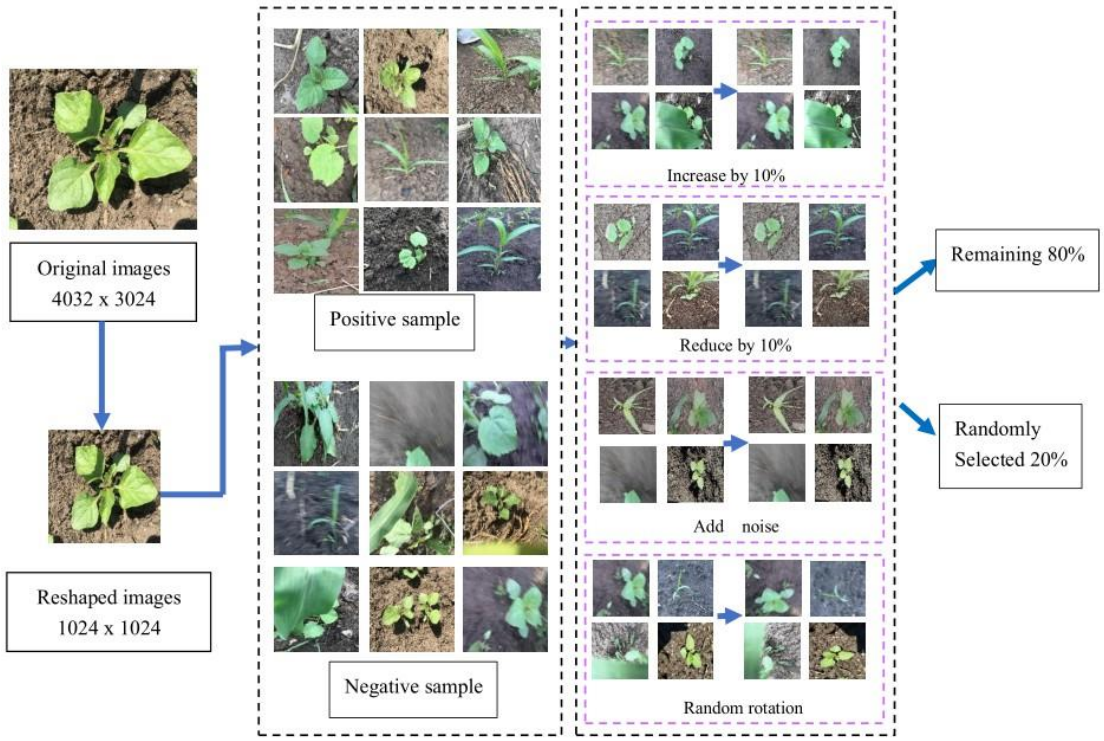


Fig. 3 Data enhancement.

The images were randomly rotated, and noise was added. Since illumination is an important reason for segmentation, to make the DCNN more robust to the illumination caused by environmental changes, the datasets were further enhanced by simulating the illumination change [44]. The brightness was adjusted to be 10% brighter and 10% darker. At the same time, some blurred, occluded and incomplete images were retained as the dataset of negative samples, and a total of 6000 data-enhanced pictures were obtained. The structure and proportion of the original dataset remain unchanged when data enhancement is carried out. Two datasets were made, one with data enhancement and the other without data enhancement. We reserve 600 of the 4000 images without data enhancement and use them to test the model for model evaluation. Both datasets were randomly divided into a training set and a verification set with a ratio of 8:2. After the training was completed, we used the 600 images previously reserved to test the model to evaluate the model. The VGG

Image Annotator labelling tool was used for annotation, as shown in Figure 4, and weed leaves were surrounded by irregular polygons, while the centre areas of the weeds were marked with a circle. Because the number of weeds in a picture is uncertain under actual working conditions and may contain multiple weeds, the number of masked leaves in the picture cannot be used to calculate the leaf age of a single weed, so 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 is the leaf age of the weed. The rectangular frame is not masked. The labels are divided into 7 categories (Figure 4).

Fig. 4 Image annotation example of segmentation.

Note: Leaf of Solanum nigrum (a_leaf), leaf of Barnyard grass (b_leaf), leaf of Abutilon theophrasti Medicus (c_leaf), central area (centre), Solanum nigrum (a), Barnyard grass (b), and Abutilon theophrasti Medicus (c).

2.4. Structure of Mask R-CNN mode

Mask R-CNN extends the target detection framework of Faster R-CNN [45] by adding a masking branch at the end of the model [38]. This process ensures that each output instance segments the proposal box using a fully connected (FC) layer to ensure that the segmentation is parallel to the target detection. To better detect small targets, ROI pooling is changed to ROIAlign. The Mask R-CNN model flow chart is

shown in Figure 5. The output consists of three branches: bounding boxes, target classifications and segmentation masks. The Mask R-CNN backbone networks we selected are ResNet50 and ResNet101 combined with FPNs. First, the backbone network extracts the feature map from the input image and then 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 the fully connected layer and segments the target region through the fully convolutional neural network (FCN). The instance segmentation task of the target is finally completed.

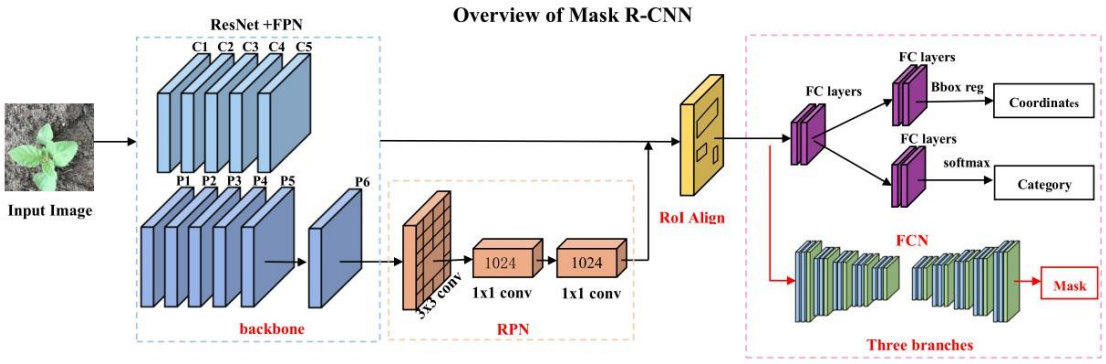


Fig. 5 The model structure of Mask R-CNN.

The FPN is an important module of Mask R-CNN [46] that detects objects of different scales in object detection. There are top-down and bottom-up paths as well as lateral connections. As shown in the FPN part of Figure 6, the thicker the contour is, the stronger the semantics. While the RPN is a sliding window in target detection, Figure 6 shows that the RPN method is used to generate a region proposal network. ROIAlign is a very important part of Mask R-CNN. It is an improvement over ROIPooling. As shown in the ROIAlign part of Figure 6, it first traverses each candidate area while keeping the floating point number unchanged and then divides

the candidate area into $k \times k$ units. Four fixed coordinate values are calculated in each unit, and the values of the four positions are calculated by a bilinear interpolation method. Then, a maximum pooling operation is performed.

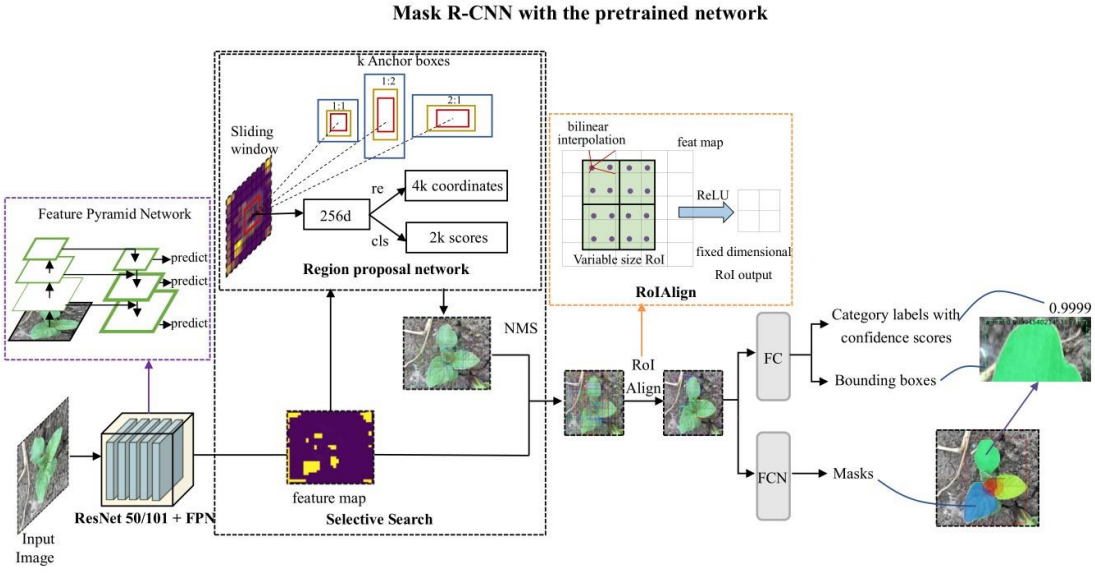


Fig. 6 Mask R-CNN with a pretrained network.

2.4.1 Image Feature Extraction (ResNet50 and ResNet101 + FPN)

Mask R-CNN can establish deep neural network models of different depths by designing different weight layers. Currently, the deep learning network models are AlexNet, ZF, GoogLeNet, VGG, and ResNet. Although a deeper number of network layers may lead to higher accuracy, the deeper network layers will result in lower model training and detection speeds. However, since the residual network does not increase the parameters of the model, it can effectively reduce the problem of training degradation and improve the model convergence [38]. Therefore, this paper uses ResNet50 and ResNet101 combined with the FPN as the backbone network to extract the features of weed images.

2.4.2 ROI Generation and ROIAlign (RPN+ROIAlign)

The RPN uses the convolutional feature map output from the backbone network

as its own input, traversing the feature map in a fixed-size sliding-window fashion and finding images containing weed blades and central areas. This area of the RPN scans is called an anchor [47]. The weed target in this study is small, especially in the central area. The size of the weeds changes throughout the growth cycle [38]. Therefore, according to the total number of pixels of the target weed in the image, this study designed 32×32, 64×64, 128×128, 256×256, and 512×512 anchors. The length-width ratios were designed to be 1:1, 1:2, and 2:1. A total of fifteen anchor points were selected, and the most likely target area was selected for segmentation and detection.

Faster R-CNN [45] is a recognition algorithm proposed for the regional convolutional network R-CNN to perform a large number of repetitive operations in each ROI and then transfer the features extracted by the RPN to the convolution of the last layer. The ROIPool layer is then added later. However, when performing the quantization operation in ROIPool, misalignment between the ROI and the extracted features can easily occur, which will have a great negative impact on the results, making this layer unsuitable for this study. Mask R-CNN [35] uses ROIAlign instead of the traditional pooling of interest (ROIPool), which is very suitable for addressing smaller targets and solves the problem of spatial position misalignment caused by ROIPool. ROIAlign uses bilinear interpolation to extract the corresponding features of each ROI on the feature map [38], calculates the exact value of each position, summarizes it by the maximum pooling method, adjusts the dimensions of each ROI to meet the FC requirement, and finally sends it to the FC layer and FCN for target classification, the bounding boxes and the mask.

2.5 Mask R-CNN training model

Before training Mask R-CNN, we introduced a pretraining model based on the

COCO dataset [48] using transfer learning. The COCO dataset has 328k images, including 91 categories. The pretraining model extracted the weights after training on COCO. On this basis, the datasets established by themselves will be trained again. By means of this transfer learning, the manpower and cost of training can be reduced, the training efficiency can be improved, and the parameters of the model can be better adjusted. The Mask R-CNN model of this experiment was carried out under the deep learning framework of TensorFlow-gpu 1.14.0 and Keras 2.1.5. This study is based on the Windows 10 64-bit (DirectX 12) operating system, 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), the NVIDIA GeForce RTX 2080 Ti graphics card. The parameters of the pretraining network are shown in Table 2.

Table 2 Characteristic parameters of the pretraining network.

Network	Depth	Size (MB)	Parameters (Millions)	Image Input Size	Feature Extraction Layer	ROI Pooling Output Size
ResNet50	50	96	25.6	224 x 224	block_13_expand_relu	[14 14]
ResNet101	101	167	44.6	224 x 224	mixed7	[17 17]

Insert ROIAlign after the feature extraction layer, extract the corresponding features of each ROI, calculate the exact value of each position, and finally transport it to the fully connected layer and FCN to predict the target class, bounding box and segmentation mask. We put the marked dataset into the model for training and obtain a training model based on Mask R-CNN for weed segmentation.

2.6 Training and evaluation

The weight decay coefficient of Mask R-CNN was set to 0.0005, the momentum was set to 0.9, the initial learning rate was set to 0.0001, and the training BatchSize was set to 1. After the parameters were set, training was conducted for 100 rounds, with training being conducted 1000 times per round. The basic framework of Mask

R-CNN involved either ResNet50 or ResNet101.

The purpose of the evaluation was to test the ability of the algorithm to identify the weed leaf age and the central area on the image. We used seven key indexes: precision rate (P) (Eq. (1)), recall rate (R) (Eq. (2)), F_1 (Eq. (3)), intersection over union (IOU) (Eq. (4)), average precision (AP), mean average precision (mAP) (Eq. (5)), and mean intersection over the union (mIOU), for evaluation. The Mask R-CNN model completes object detection and object segmentation. For the object detection part, we used the precision rate (P), recall rate (R), F_1, AP, and mean average precision (mAP) for evaluation. For the object segmentation part, we use the mean intersection over the union (mIOU) error for evaluation.

The precision rate and recall rate are defined by the following equation:

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

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

where "true positive (TP)" indicates the number of positive results detected as positive; "false positive (FP)" indicates the number of negative results detected as positive; and "false negative (FN)" indicates the number of positive results detected as negative. The metric function (F_1) [49] of the precision and recall rates is defined by the following equation:

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

To better meet the problem of multiclass imbalance in this paper, we averaged the seven classification indicators [50]. To better evaluate the model algorithm, the IOU is used to conduct measurements [51]. The IOU measures the overlap between two bounding boxes. As shown in Figure 7, the overlap degree between the weed prediction box and the real box on the ground was calculated. The value of the IOU

can be divided into three regions. When the IOU threshold is set to 0.7, anchors with IOU values less than or equal to 0.3 are considered to be negative anchors, while anchors with IOU values between 0.3 and 0.7 are neutral anchors, and we do not need to consider these cases. In addition, IOU values greater than or equal to 0.7 are positive anchors. The system will identify the positive anchors and bounding box and match them 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 to 0.5, the IOU values are greater than 0.5, and a positive ROI is observed; when the IOU values are less than 0.5, a negative ROI is observed. The positive ROI is allocated to the mask and ground truth by the system. The detection performance of the model is evaluated by the mean accuracy (mAP) [51]. The mAP has excellent evaluation performance when it is related to the target position information and the category information of the target in the image. The AP can be calculated for each category separately, and then each category can be averaged to calculate the mAP. The larger the mAP value is, the better it is. It can be calculated by the AP. The thresholds were 0.5 and 0.7 in this study. The IOU and mAP are defined by the following equation:

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

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

Note: N represents the number of images.

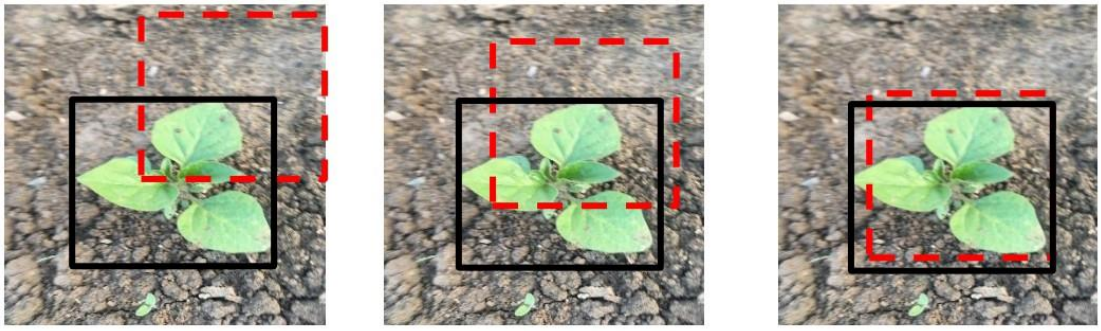


Fig.7 Visual example of intersection over union. The red dashed box is the prediction of the detection result. The black solid-line box is the ground truth. The overlap between the two is visible, and different IOU values are depicted from left to right. (c) is the best.

3 Results

In the following three stages, the development of a segmentation model based on Mask R-CNN for weed leaf age and central area was completed. First, the image acquisition work was completed in three angles of front view, left view and top view and under different weather conditions, and 4574 weed images were collected. Second, the data pre-processing and image annotation of the acquired image was completed. Two datasets were created: one dataset contains 4000 weed images without data enhancement, and the other dataset contains 6000 data-enhanced weed images. Third, two selected pretrained networks were used to replace the backbone network of the classic Mask R-CNN, network construction and parameter optimization, and an improved Mask R-CNN model processed by the pretraining network was proposed. The training and evaluation of the model were then completed. Data enhancement can improve the performance of the model. The mAP of the Mask R-CNN model using ResNet101 as the backbone network is 0.693, and the mIOU is 0.585, which is better than ResNet50 and can be used for weed segmentation. Moreover, the weed image taken from the top view angle compared to the other two angles has the highest detection accuracy.

The result of weed segmentation is shown in Figure 8. A method based on Mask R-CNN for segmentation of weed leaf age and central area is proposed. Two datasets are trained in two different backbones networks (ResNet50 and ResNet101), and the network is chosen that achieves the best balance between mIOU and mAP. The leaf age is determined based on the number of complete leaves of the weed, and the centre area is determined based on the intersection area of the top leaves of the weed [7]. The evaluation of the Mask R-CNN model is carried out from two aspects: first, the

detection results of the model are evaluated, and then the segmentation performance of the model is evaluated.

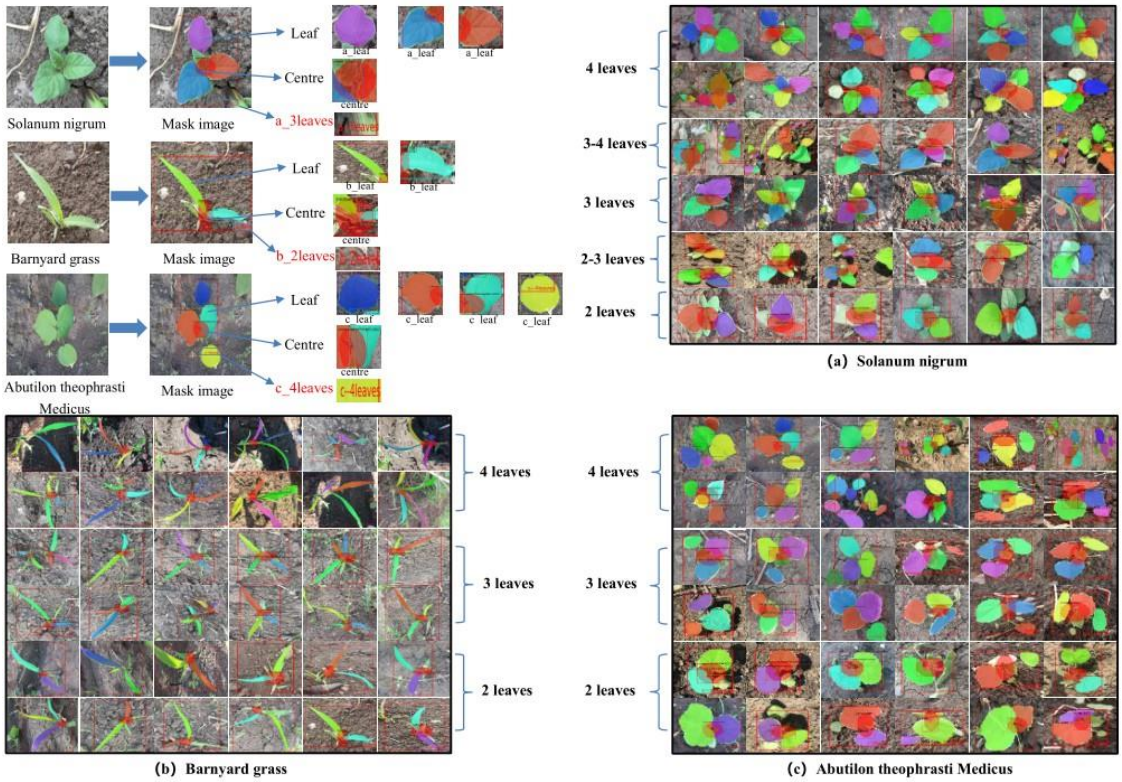


Fig. 8 Weed segmentation results.

3.1 Weed test results and evaluation

The precision rate (P), recall rate (R), F_1, IOU, AP, and mAP are important indicators used to evaluate the performance of Mask R-CNN detection. Table 3 lists the F_1 and mAP values of different datasets (without data enhancement and data enhancement) on the total test set. The number of the total test set is 600. The total test set includes 200 front view images, 200 side view images, and 200 top view images. It can be seen from Table 3 that when data enhancement is used, the F_1 value of the ResNet101 network in the total test set is 0.9214. Without data enhancement, the F_1 value of the ResNet101 network on the total test set was 0.8965. The F_1 value of ResNet50 in the case of data enhancement is 0.0413 higher than the F_1 value without data enhancement. It can be seen that for the dataset with data enhancement,

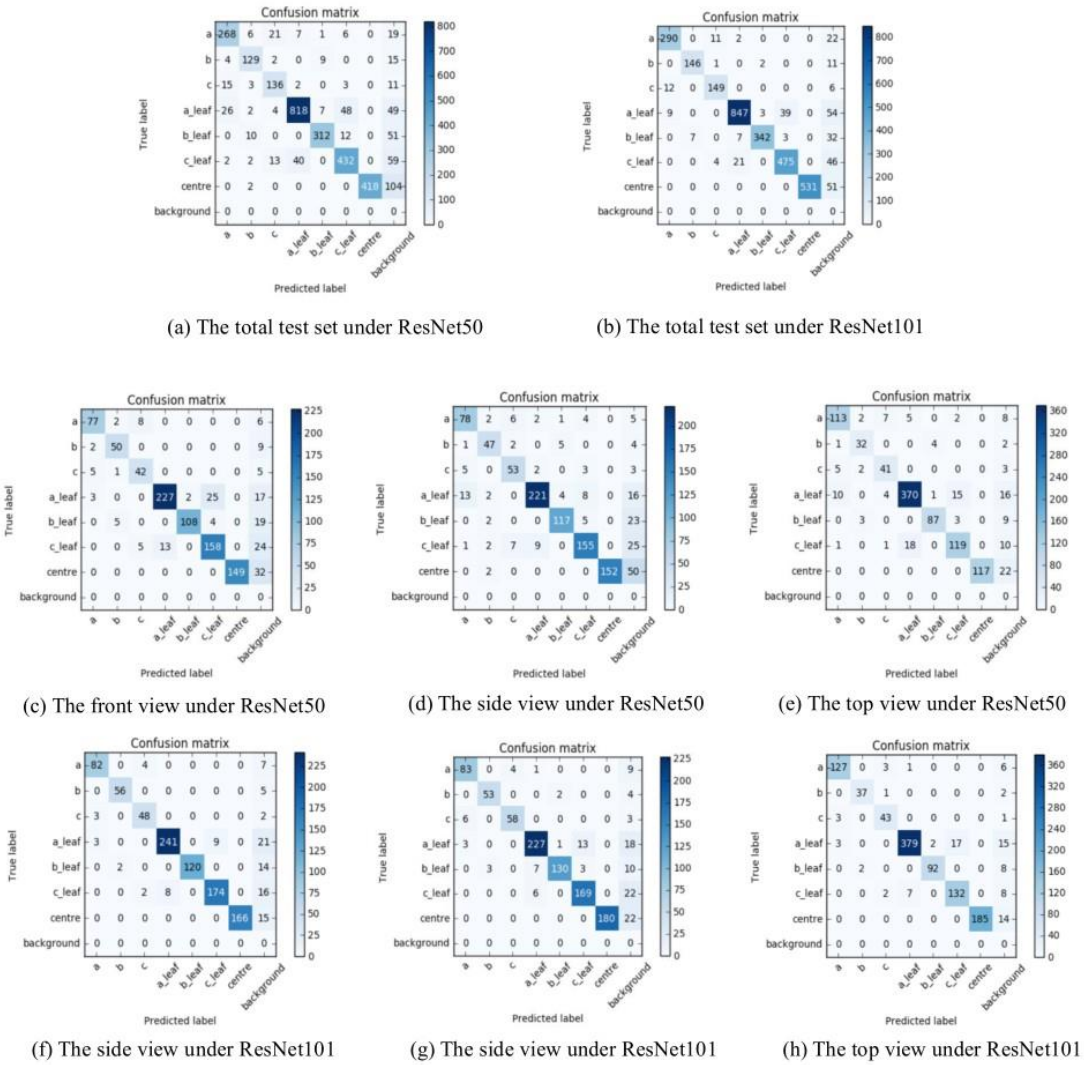
the detection results of the two backbone networks ResNet50 and ResNet101 on the total test set are higher than those without data enhancement. It can be seen that the data enhancement effect is better than that without data enhancement. Table 3 also lists the values of mAP under the two thresholds of data enhancement and without data enhancement in the total test set. According to Table 3, for the total test set, when the IOU threshold of ResNet101 is greater than or equal to 0.5, the mAP value is 0.6512 without data enhancement, while the mAP value is 0.6932 with data enhancement. It can be seen that no matter which backbone network is used, the map value of the model is greater than that without data enhancement. Although it is only increased by 0.042, it also improves the performance of the network and can reduce the impact of overfitting.

Table 3 List of the detection results of different datasets on the total test set

dataset		Without data augmentation	Data augmentation
ResNet50	F_1	0.8073	0.8486
	mAP (IOU≥0.5)	0.5421	0.5702
	mAP (IOU≥0.7)	0.4534	0.4791
ResNet101	F_1	0.8965	0.9214
	mAP (IOU≥0.5)	0.6512	0.6932
	mAP (IOU≥0.7)	0.5113	0.5244

Note: The total test set consisted of 600 images, including 200 front view images, 200 side view images, and 200 top view images.

Figure 9 shows the confusion matrices of the detection results of the model in the case of data enhancement. Figure 10 lists the detection results of the Mask R-CNN model under two backbone networks, three angles, and seven types of labels in the case of data enhancement. The precision rate of the ResNet101 network in the total test set is 0.9523, the recall is 0.8929, and the F_1 value is 0.9214.



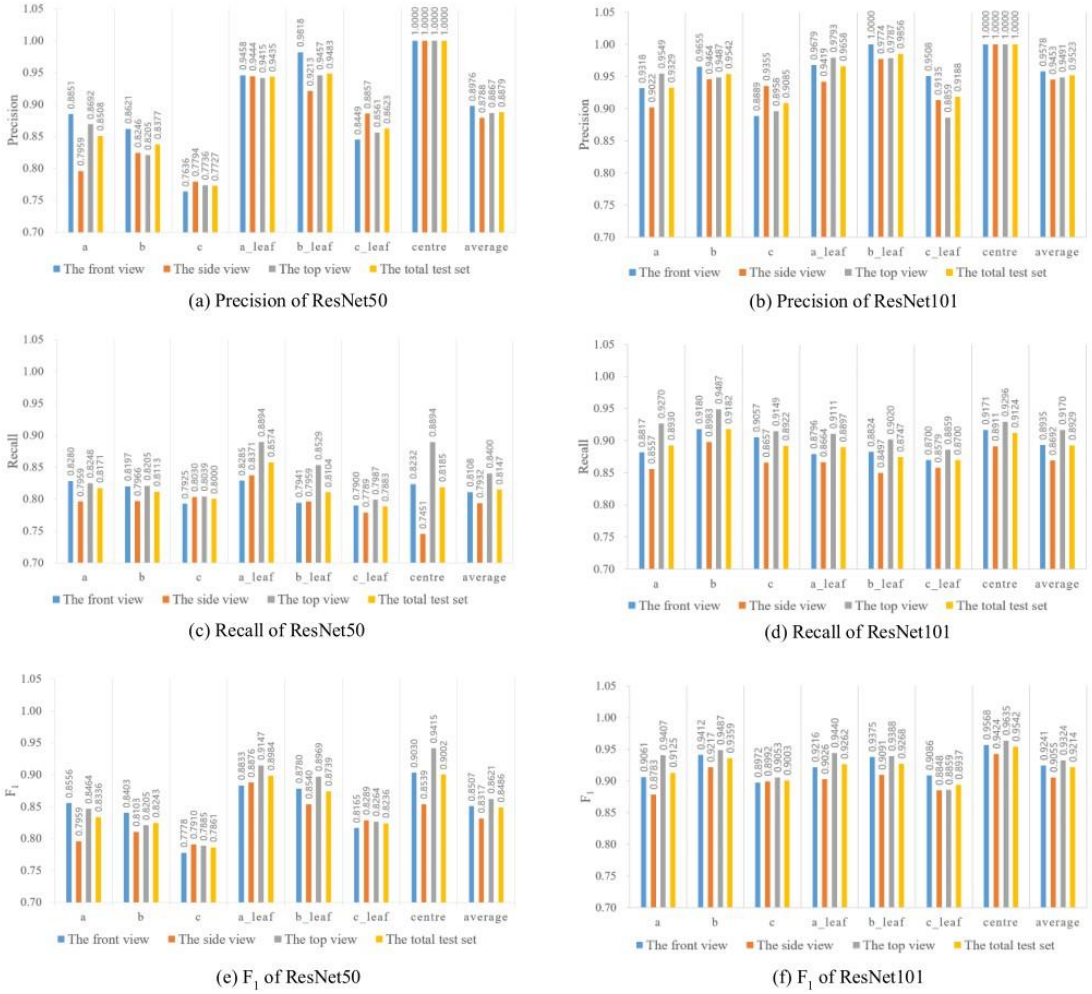
(a) The total test set under ResNet50 (b) The total test set under ResNet101

(c) The front view under ResNet50 (d) The side view under ResNet50 (e) The top view under ResNet50

(f) The side view under ResNet101 (g) The side view under ResNet101 (h) The top view under ResNet101

Fig. 9 Confusion matrix of the detection results of ResNet50 and ResNet101 in the case of data enhancement

Note: The Mask R-CNN with ResNet50 is abbreviated as ResNet50, and Mask R-CNN with ResNet50 is abbreviated as ResNet101. a_leaf represents the leaves of Solanum nigrum, b_leaf represents the leaves of Barnyard grass, c_leaf represents the leaves of Abutilon theophrasti Medicus, a represents Solanum nigrum, b represents Barnyard grass, c represents Abutilon theophrasti Medicus, and centre represents the central area of each weed. The total test set consisted of 600 images, including 200 front view images, 200 side view images, and 200 top view images.



(a) Precision of ResNet50 (b) Precision of ResNet101

(c) Recall of ResNet50 (d) Recall of ResNet101

(e) F_1 of ResNet50 (f) F_1 of ResNet101

Fig. 10 Detection results of the Mask R-CNN with pretrained networks in the case of data enhancement.

Note: The Mask R-CNN with ResNet50 is abbreviated as ResNet50, and Mask R-CNN with ResNet50 is abbreviated as ResNet101. a_leaf represents the leaves of Solanum nigrum, b_leaf represents the leaves of Barnyard grass, c_leaf represents the leaves of Abutilon theophrasti Medicus, a represents Solanum nigrum, b represents Barnyard grass, c represents Abutilon theophrasti Medicus, and centre represents the central area of each weed. The total test set consisted of 600 images, including 200 front view images, 200 side view images, and 200 top view images.

It can be seen from Figure 10 that the labels a, b, c, a_leaf, b_leaf, c_leaf, and centre were not recognized, we surmise that they were recognized as the background. It can be seen from Figure 10 that under the condition of data enhancement, the precision values of ResNet50 in the front, side, top, and total test sets were greater than or equal to 0.7636, the recall values were greater than or equal to 0.7451, and the

F_1 values were greater than or equal to 0.7778. In comparison, the precision values of ResNet101 were greater than or equal to 0.8859, the recall values were greater than or equal to 0.8497, and the F_1 values were greater than or equal to 0.8783 in the front, side, top, and total test sets. As a result, the precision, recall and F_1 values of ResNet101 in the front, side, top, and total test sets were considerably higher than those of ResNet50, and ResNet101 consistently performed better than ResNet50. The F_1 values of ResNet101 in the front, side, top, and total test sets were 0.9241, 0.9055, 0.9324 and 0.9214, respectively. When ResNet101 was used as the backbone network, the recall values of the top view test set were greater than or equal to 0.8859, the recall values of the front view and the total test sets were greater than or equal to 0.8700, and the recall values of the side view test set were greater than or equal to 0.8497. The top view test set shows the best performance in all classifications. For the total test set, the F_1 values were 0.9359 and 0.9268, and the recall rates were 0.9182 and 0.8747 when Barnyard grass and the leaves of Barnyard grass were detected, and on the front, side, and top test sets, the classification performance of Barnyard grass was better than the other two kinds of weeds. For the central area, the precision values of ResNet101 in the front, side, top, and total test sets were 1.0000. Since Figure 9 and Figure 10 can show only 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 has a certain impact on the identification of weeds. Therefore, the model recognition accuracy is very important for evaluating the model performance. Table 4 lists the mAP values for ResNet50 and ResNet101 for different test sets when IOU values are greater than or equal to 0.5 and greater than or equal to 0.7 under data enhancement.

Table 4 List of detection results of weeds under different networks and different

592 angles with data augmentation.

Networks		ResNet50	ResNet101
The front view	mAP (IOU≥0.5)	0.552	0.701
	mAP (IOU≥0.7)	0.456	0.522
The side view	mAP (IOU≥0.5)	0.533	0.625
	mAP (IOU≥0.7)	0.468	0.458
The top view	mAP (IOU≥0.5)	0.625	0.753
	mAP (IOU≥0.7)	0.513	0.592
The total test set	mAP (IOU≥0.5)	0.570	0.693
	mAP (IOU≥0.7)	0.479	0.524

593 The mAP is a commonly used index in target detection. Table 4 shows that the
594 mAP of ResNet101 is higher than that of ResNet50, indicating that ResNet101 has
595 good target detection performance. For the total test set, when the IOU threshold is
596 equal to or greater than 0.5, the mAP value is 0.693, and when the IOU threshold is
597 equal to or greater than 0.7, the mAP value is 0.524. Thus, when the threshold is equal
598 to or greater than 0.5, ResNet101 has good detection performance. When ResNet101
599 is used as the backbone network and the IOU threshold is greater than or equal to 0.5,
600 the mAP value in the top view test set is 0.753, the mAP value in the front view test
601 set is 0.701, the mAP value in the side view test set is 0.625, and the mAP value in the
602 total test set is 0.693. The top view achieved good detection performance.

603 ## 3.2 Instance segmentation results and evaluation

604 For the instance segmentation phase, two backbone networks ResNet50 and
605 ResNet101 were executed. From the above comparison, we can observe that the effect
606 of data enhancement is better than that without data enhancement. Therefore, we
607 tested the data-enhanced dataset, and the results of weed segmentation are shown in
608 Figure 8. The mIOU is an important index for evaluating the segmentation results [31]
609 and is commonly used to evaluate the segmentation performance of the Mask R-CNN
610 model. Table 5 shows the mIOU of Mask R-CNN under different backbone networks
611 and different test sets. The test image of 600 weeds shows that for the total test set, the

mIOU of ResNet50 is 0.491 and that the mIOU of ResNet101 is 0.585. The value for ResNet101 is significantly higher than that of ResNet50. Thus, the ResNet101 model will result in better network performance and can be applied to the segmentation of small target objects. Hence, it can meet the needs of weed instance segmentation. When ResNet101 was used as the backbone network, the mIOU of the top view was 0.603, which is higher than those of other datasets. Therefore, good segmentation results were achieved.

Table 5 Segmentation results of weeds under different networks and different angles in the case of data enhancement.

Angles	mIOU	
	ResNet50	ResNet101
The front view	0.476	0.585
The side view	0.463	0.566
The top view	0.535	0.603
The total view	0.491	0.585

4 Discussion

In this study, we proposed a weed phenotype segmentation method based on Mask R-CNN to obtain the growth stage and central area of weeds in a complex environment. In this study, images of Solanum nigrum, Barnyard grass, and Abutilon theophrasti Medicus were collected from three angles of front, side, and top view, and two datasets (data enhancement and without data enhancement) were produced. Two backbone networks were used, namely, ResNet50 and ResNet101. The results of detection and instance segmentation are evaluated. Evaluation of this method can detect weeds at the object level and segment weeds at the pixel level. It is concluded that data enhancement can improve the performance of the model. The Mask R-CNN model using ResNet101 as the backbone network has the highest mIOU value, which is better than ResNet50 and can be used for weed segmentation. Furthermore, the

weed image taken from the top view angle compared to the other two angles has the highest detection accuracy.

We identified individual plant images of three weeds in the field, but weeds are visual objects with complex structures and rich texture features, and even the same species can have large differences in morphology and colour. However, the deep learning model can automatically learn and extract the features of complex objects, and instance segmentation is a kind of deep learning model that can detect the target pixel by pixel, which solves the problems of blade adhesion and occlusion [52]. Good results have been obtained in the study of plant phenotypes [53]. It can be seen that the instance segmentation model is very suitable for dealing with complex environmental problems in the field. However, the deep learning model relies on a large number of datasets to train the network. Due to the limited number of datasets, we use data enhancement methods to expand the dataset from 4000 to 6000. The experimental results (Table 3) show that the use of data enhancement methods can help improve the performance of the model and help reduce the impact of overfitting. In the case of data enhancement, the leaves of Barnyard grass compared with most broadleaf weeds showed the best performance in all classifications; therefore, because Barnyard grass is an annual herb and its leaves are narrow and long, it is distinct in terms of characteristics from the other two weeds. The less appealing 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 of the results can easily occur, which is undesirable because it can lead to errors in leaf age identification. ResNet101 showed a higher F_1 score in the central area, because the characteristics of the centre area are more obvious than 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 was 0.585 in the total test set, indicating that it can meet the needs of instance segmentation. Thus, the Mask R-CNN model using ResNet101 as the backbone can reliably segment weeds.

In the research status at home and abroad, we found that in the research of Quan and Feng, when the angle between the camera and the vertical direction is 0°, the detection accuracy is 0.0095 lower than other angles [27]. But our research found that when ResNet101 is used as the backbone network, the average F_1 value of the top view is 0.0269 and is 0.0083 higher than that of the side view and the front view. I think the reason should be that his research is to obtain maize seedlings and weeds in the image. At the shooting Angle of 0°, the maize in the image contains only one plant, and only contains the top view, but 30° and 75° will produce maize seedlings and weeds from different angles, and also contains the front view and top view. For the detection of maize seedlings and weeds, different angles contain different information, and more comprehensive information is conducive to improving the performance of the model. But in my research, the purpose is to obtain the leaf age and central area of the weed, as shown in Figure 2, we can obtain more comprehensive weed phenotype information from the perspective of the top view, especially given that the central area of weeds will be obtained more, while the side view and the front view will see less from the central area. This may be the cause of the different results.

At present, the treatment of weeds in the field mostly involves weed classification and detection. Weed classification can determine only the species of weeds, while the specific position coordinates of weeds cannot be obtained; thus, it is impossible to spray the exact target. Weed detection can facilitate drawing the

bounding box of weeds, but weeds are irregular in shape and size, causing the machine to be inaccurate relative to the target, which will result in some herbicide falling to the ground and not being absorbed by the weeds, leading to environmental pollution and wasted herbicide. As a kind of deep learning model, instance segmentation can detect the target pixel by pixel, which solves the problems of blade adhesion and occlusion. The leaf age of weeds and the position of the centre area can be obtained more accurately. In the Northeast Plain of China, the main economic crops are maize, soybeans, and wheat, which are susceptible to annual and perennial weeds. On the one hand, controlling annual and perennial weeds can increase crop yields and reduce the likelihood of damage caused by weeds in the second year [54]. On the other hand, studying the interaction between plant phenotype and vision through effective phenotypic analysis can obtain information on plant growth and morphological changes.

The limitation of this study is that the identification efficiency of this study is low, and further improvement of efficiency is needed before it can be used in engineering practice. The DCNN model of this study was used to segment only three kinds of weeds. If we expand the 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. According to the obtained leaf age of economic crops, it can also provide an important basis for crop fertilization. For some plants, the central area is the pollination area of flowers, and segmentation of this part will provide an important basis for subsequent studies. Future research will focus on evaluating image datasets covering a wider range of weeds and crop varieties. Most of the images used for model testing contained only single-plant weeds, and only a few contained multiple weeds. Mask R-CNN failed to segment weeds near the edges in a few test

images containing multiple weeds, but continuous video input will eliminate edge effects when applied in the field.

The results of this study show that the combination of weed phenotype and computer vision is very suitable for dealing with complex field conditions such as light changes, leaf occlusion, and mixed leaf age. The models and methods proposed in this study can be applied to the study of many different types of plants. The data of this study were taken from a complex field environment, and previous studies on plant phenotypes were mostly taken in an indoor environment, in which the image background is often very pure and the illumination is very uniform. Studying the field environment can make the model more suitable for practical applications, and the shooting angle of the dataset determines how much we obtain from the target image information, so it is meaningful to study the segmentation results at different shooting angles. At present, there are few studies on plant phenotypes that are specific to weed phenotypes, but weeds of different leaf ages require different doses of herbicides, so it is of great significance to obtain the information of weed leaf ages to reduce the amount of herbicides. To make this research practical, we can deploy the trained model on the mobile platform of the spray system used for weeding in the future, which can promote the development of precision agriculture and intelligent agriculture.

5 Conclusions

In this paper, we proposed a weed phenotype segmentation method based on the improved Mask R-CNN to obtain the weed species, leaf age and central area of weeds. In the field of plant phenotype research, in the context of complex field environments, obtaining weed phenotypes is still a substantial challenge. According to the research status at home and abroad, we can know that leaf age and central area are important

phenotypic traits of weeds. We obtained leaf age and central area, which are of great significance for targeted weeding. Improve the performance of the model through data enhancement. In addition, we found that the weed image taken from the top view angle can help to improve the performance of the model. Weed datasets were constructed by data collection from three angles and data enhancement, the dataset can contain weed information of different growth stages, different angles, and different types. The dataset and research results may provide important resources for future plant phenotype research.

Because the DCNN has the ability to extract features from complex environments, it is very suitable for addressing complex image problems in the field. The experimental results show that despite the interference of straw and crop leaves in the background of weeds in the field, Mask R-CNN model using Resnet101 as the backbone network still achieves accurate segmentation of weeds. Good segmentation performance has been achieved. We hypothesize that the Mask R-CNN example segmentation model used here is suitable for all weeds and cash crops in the field, such as Setaria viridis, Cirsium setosum, maize, and wheat, and that the segmentation of different parts of plants will provide more help in the study of plant phenotypes. Future research will focus on evaluating image datasets that cover a wider range of weeds and crop varieties. The identification efficiency of this study is low, so we need improve model efficiency, and deploy the trained model on the mobile platform of the spray system used for weeding in the future. The model combines artificial intelligence technology with agronomic research and applies it to the development of intelligent agriculture.

6 Declarrations

Availability of data and materials

Given that the data used in this study were acquired by self-collection, the dataset is being further improved. Thus, the dataset is unavailable for the time being.

Acknowledgments

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

Funding

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

Contributions

QLZ and WB wrote this paper, WB FHQ and MSR collected data, MSR developed the codes, YCJ, LHD and JW supervised the project. All 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.

7 REFERENCES

1. Hashim S, Jan A, Fahad S, Ali HH, Mushtaq MN, Laghari KB, Jabran K, Chauhan BS: **WEED MANAGEMENT AND HERBICIDE RESISTANT WEEDS: A CASE STUDY FROM WHEAT GROWING AREAS OF PAKISTAN**. *Pakistan Journal of Botany* 2019, **51**(5):1761-1767.
2. Slaughter DC, Giles DK, Downey D: **Autonomous robotic weed control systems: A review**. *Comput Electron Agric* 2008, **61**(1):63-78.
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.
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**:339-350.
5. Gianessi LP, Reigner NP: **The value of herbicides in US crop production**. *Weed Technology* 2007, **21**(2):559-566.
6. Pretty J, Bharucha ZP: **Integrated Pest Management for Sustainable Intensification of Agriculture in Asia and Africa**. *Insects* 2015, **6**(1):152-182.
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):771-776.
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):22-26.
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):1725-1737.
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.
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.
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):6270-6283.
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):133-143.
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**:143-154.

18. Ozdogan M, Yang Y, Allez G, Cervantes C: **Remote Sensing of Irrigated Agriculture: Opportunities and Challenges**. *Remote Sensing* 2010, **2**(9):2274-2304.
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**:80-89.
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**.
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.
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).
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).
27. Quan LZ, Feng HQ, Li YJ, Wang Q, Zhang CB, Liu JG, Yuan ZY: **Maize seedling detection under different growth stages and complex field environments based on an improved Faster R-CNN**. *Biosystems Engineering* 2019, **184**:1-23.
28. Geetharamani G, Pandian JA: **Identification of plant leaf diseases using a nine-layer deep convolutional neural network**. *Computers & Electrical Engineering* 2019, **76**:323-338.
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**.
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):1344-1351.
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):640-651.
33. Yu JG, Li YS, Gao CX, Gao HX, Xia GS, Yu ZL, Li YQ: **Exemplar-Based Recursive Instance Segmentation With Application to Plant Image Analysis**. *Ieee Transactions on Image Processing* 2020, **29**:389-404.
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**.
35. He K, Gkioxari G, Dollar P, Girshick R: **Mask R-CNN**. In: *2017 IEEE International Conference on Computer Vision (ICCV): 2017*; 2017.
36. He KM, Gkioxari G, Dollar P, Girshick R: **Mask R-CNN**. *IEEE Trans Pattern Anal Mach Intell* 2020, **42**(2):386-397.

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**.
38. 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.
39. 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.
40. 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.
41. 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):1685-1689.
42. Lee SH, Chan CS, Wilkin P, Remagnino P: **Deep-plant: Plant identification with convolutional neural networks**. 2015.
43. Krizhevsky A, Sutskever I, Hinton GE: **ImageNet Classification with Deep Convolutional Neural Networks**. In: *International Conference on Neural Information Processing Systems: 2012*; 2012.
44. Ma J, Li Y, Chen Y, Du K, Zheng F, Zhang L, Sun Z: **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**:117-129.
45. 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):1137-1149.
46. Wu YT, Tang SM, Zhang SW, Ogai H: **An Enhanced Feature Pyramid Object Detection Network for Autonomous Driving**. *Applied Sciences-Basel* 2019, **9**(20).
47. Gonzalez S, Arellano C, Tapia JE: **Deepblueberry: Quantification of Blueberries in the Wild Using Instance Segmentation**. *Ieee Access* 2019, **7**:105776-105788.
48. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL: **Microsoft COCO: Common Objects in Context**. 2014.
49. Hripcsak G, Rothschild AS: **Agreement, the F-measure, and reliability in information retrieval**. *Journal of the American Medical Informatics Association* 2005, **12**(3):296-298.
50. Al-Najjar HAH, Kalantar B, Pradhan B, Saeidi V, Halin AA, Ueda N, Mansor S: **Land Cover Classification from fused DSM and UAV Images Using Convolutional Neural Networks**. *Remote Sensing* 2019, **11**(12).
51. Zhang E, Yi Z: **Average Precision**; 2016.
52. Yu J-G, Li Y, Gao C, Gaoa H, Xia G-S, Yub ZL, Lic Y: **Exemplar-Based Recursive Instance Segmentation With Application to Plant Image Analysis**. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society* 2019.
53. Ubbens JR, Stavness I: **Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks**. *Frontiers in Plant Science* 2017, **8**.
54. 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):879-888.