

Weed Detection in The Field Environment Based on Faster R-CNN

Jianing Liu^{1†}, Zhiheng Qin^{1†}, Ruiyan Wang^{1,2*}, Yuhuan Li², Xiaoyan Xu³, Steve Finch³

¹College of Resources and Environment, Shandong Agricultural University, Tai'an 271018, Shandong, China

²National Engineering Laboratory for Efficient Utilization of Soil and Fertilizer Resources, Tai'an 271018, Shandong, China

³Royal Agricultural University, Cirencester GL7 6JS, Gloucestershire, England

[†]These authors contributed equally to the work.

*Author to whom correspondence should be addressed.

Copyright: © 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Accurate weed identification in farmland is crucial for enhancing intelligent weeding precision. This study focuses on weeds in maize seedling fields and builds an accurate identification model using the Faster-RCNN deep-learning algorithm. An image database is created, and the VGG-16 network extracts labeled datasets for weed feature extraction. By calculating random candidate region scores, a neural network training model is established to determine weed positions and types. This model achieves an average accuracy of 81.25% and an identification rate of 94.3% in weed identification. To test the model's performance in the field, it is evaluated under different conditions, such as lighting, field of view, and occlusion. Occlusion has the most significant impact on the identification rate. Without occlusion, the precision is 94.4%, dropping to 79.2% when the occlusion rate exceeds 50%. However, adjusting the shooting angle can increase the precision to 97.1%. In real-world conditions, considering all factors, the weed identification precision is 94.3%. The results show that this technology is highly adaptable in the field with fast image detection. With GPU acceleration, the average detection speed per image is 50 milliseconds, and the video stream can reach 20 frames per second. This technology can adapt to complex environments, detect accurately, and has a short calculation time. It provides key support for automated mechanical weeding and holds great promise for practical applications in agricultural production.

Keywords: Weed detection; Computer vision; Maize; Deep learning; Field environment

Online publication: 5 June, 2025

1. Introduction

Nowadays, large amounts of pesticides, especially herbicides, are used in China's agriculture. Herbicides, being environmental pollutants, leave 20%–80% residues^[1], harming the environment and human health.

Maize, with an annual sown area of about 42.129 million hectares (36% of the total) and an output of around 257 million tons (39% of total crops), is crucial for food supply in China^[1]. But weeds in maize fields compete

for resources, affecting growth. Chemical weeding, the main approach, often causes maize phytotoxicity due to improper use, reducing yields by 10%–30% [2]. So, cutting reliance on herbicides is urgent.

In the information age, intelligent weeding robots, integrating key techs, are vital for modern agriculture [3]. They can boost productivity, solve labor issues, and protect from pesticides. Weed identification is key for these robots. There are two main weed-identification methods. The spectral-difference-based one, using hyperspectral instruments, has speed and cost problems, remaining experimental. The other, using ordinary-camera photos, is cheap and has low-light needs. But current research mainly focuses on single-plant, plain-background photos, and real-field factors like light, terrain, and plant occlusion limit its use [4,5].

Deep-learning-based computer vision has advanced. The Faster R-CNN network, with RPN for speed, shows potential in farmland target detection [6–8].

This paper, using Faster R-CNN and ordinary-camera photos in maize fields, aims to develop a real-time, accurate weed-identification tech for complex fields, support automatic weeding machines, and analyze environmental impacts with solutions.

2. Materials and methods

2.1. Image acquisition

In this study, an experimental plot of ‘DengHai 605’ maize, one of the most prevalent maize varieties in China, was utilized in an experiment. The plot was located at the Huang-Huai-Hai Maize Experimental Station (36°9′48″N, 117°9′5″E, altitude 134 m). Maize and weed images were captured from a top-down view to mimic the perspective of an automatic weeding machine. An ordinary SLR camera was positioned approximately 100 cm above the ground surface, set to “AUTO” mode with a resolution of 3000×4000 to clearly show the morphological features of weeds in the field. RGB images were taken under diverse weather conditions, such as sunny and cloudy days, during the sowing season in June 2019. Images were acquired at different times of the day (500 each in the morning and afternoon) under varied lighting conditions and at different angles, as depicted in **Figure 1**.



Figure 1. Corn and weed images under natural light exposure

2.2. Image datasets

The image dataset comprises 1000 images with a resolution of 3000×4000 pixels, each featuring 3–10 target plants. Manual ground-truth annotation was performed using LabelImg software. Considering practical scenarios, samples were classified into four categories: Maize seedlings, *Eleusine indica*, *Portulaca oleracea*, and *Amaranthus tricolor*. The annotated maize dataset was evenly split into a training set and a testing set. The training

set was further divided into a training subgroup and a validation subgroup. Training images were randomly selected through independent and uniform sampling of the entire dataset. The validation and test images are mutually exclusive, ensuring the reliability of subsequent evaluation criteria. During model training, horizontal inversion transformation was applied to the images to augment the training samples and mitigate over-fitting risks. Edge operations were conducted on the images to incorporate labeled edge areas into the training model.

2.3. Construction of weeding detection model

2.3.1. Construction of faster R-CNN

Faster R-CNN provides three feature-extraction networks. To boost detection accuracy, VGG16 with the deepest network convolution layers is chosen to build the model (**Figure 2**). In the model, the extracted feature map is used for RPN and the fully-connected layer. RPN generates tight candidate anchors in the original image, classifies them as positive or negative via Softmax, and refines anchors by bounding-box regression to get the ground-truth of target plants. The proposal feature map, from the feature map and region proposals, goes to the fully-connected layer to determine the plant category. The plant classifier at the end of the network uses this map to identify the category, and the final precise location of the target is obtained by bounding-box regression.

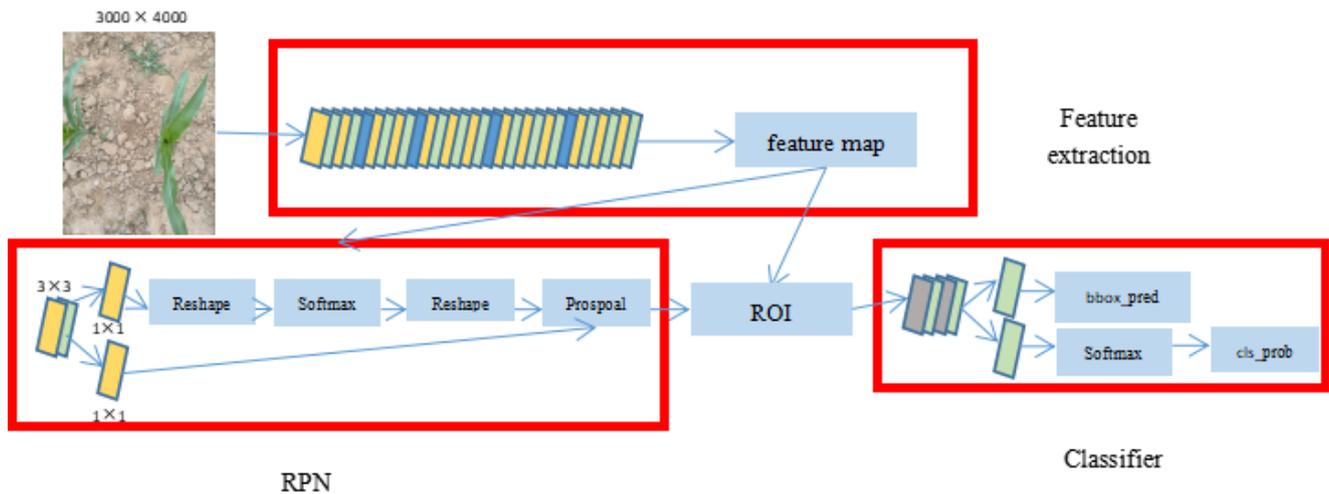


Figure 2. Faster R-CNN network

2.3.2. Region proposal network

Without a traditional sliding window and Selective Search, Faster R-CNN uses RPN to generate detection boxes. This is a major advantage, greatly speeding up box generation. **Figure 3** shows the RPN network structure.

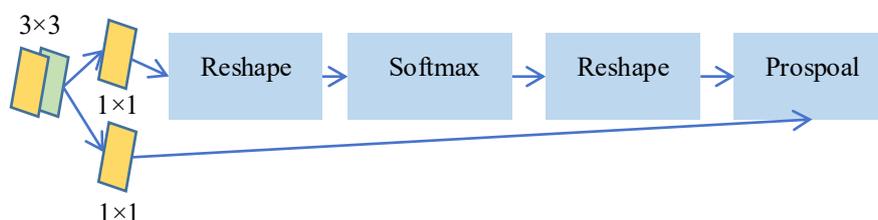


Figure 3. Region proposal network

First, the RPN network uses Softmax to determine anchor attributes and calculates the bounding-box regression offset of anchors to get proposal regions. Then, it removes those too small or out of bounds. The result of anchor points on the original image is shown in **Figure 4**, indicating they're spread across the image. Finally, a series of proposal regions, like the red-framed areas in **Figure 5**, are obtained.



Figure 4. The anchor figure

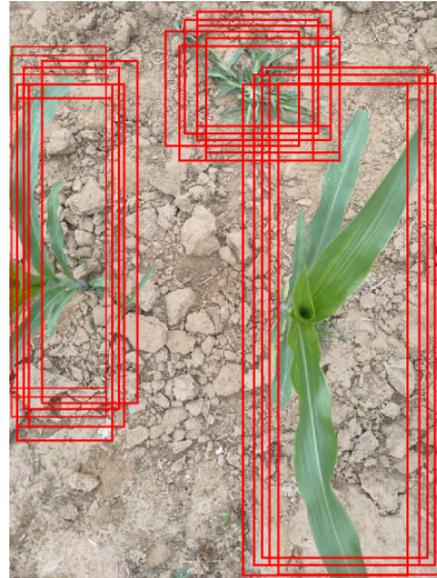


Figure 5. Proposal feature map

2.3.3. Classifier

As shown in **Figure 6**, the classifier uses the fully-connected layer and Softmax to calculate weed categories in each proposal area, outputting a probability vector (cls_prob). By getting the position offset of each region proposal (bbox_pred) with cls_prob, a more accurate target bounding box is obtained. The blue-bordered area in **Figure 6**, the highest-scoring in the proposal area, is the target identification area.

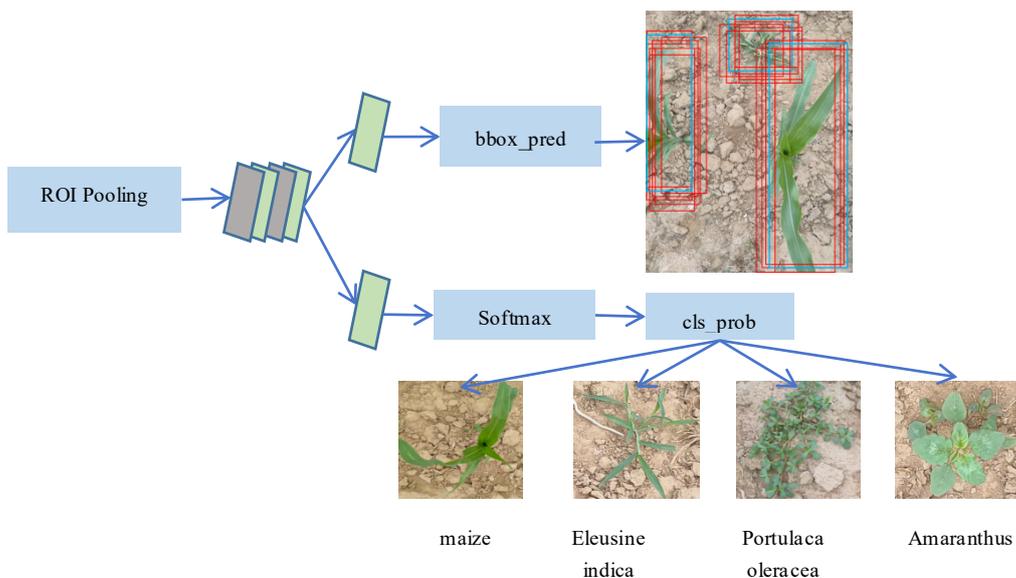


Figure 6. Classification network

2.3.4. Network training

Weed detection is achieved by segmenting and locating plants' rectangular boundaries (candidate areas) in an image, then extracting features and classifying them via a recognition network. The network is end-to-end trained and iteratively optimized by back-propagation and random gradient descent. Random weight initialization may slow model convergence or cause a local minimum. Transfer learning adapts to new tasks quickly with insufficient data. Using a large dataset pre-trained model, share underlying weights and fine-tune the model to handle dataset differences. The overall weed identification framework is in **Figure 7**.

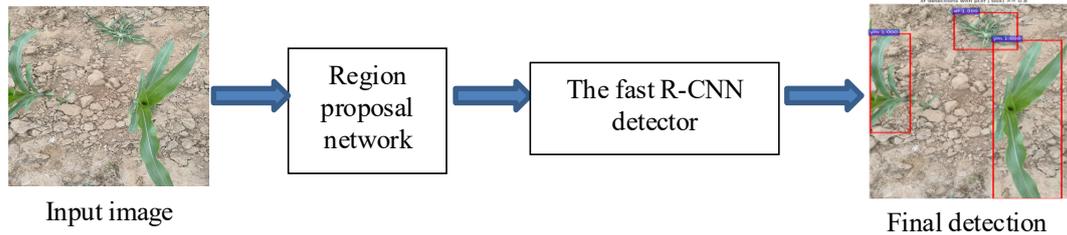


Figure 7. Frame diagram of weed identification based on faster R-CNN

Weed detection segments and locates plants' rectangular boundaries (candidate areas) in an image, then uses a recognition network for feature extraction and classification. The network is end-to-end trained and optimized by back-propagation and random gradient descent. Random weight initialization may slow model convergence or cause a local minimum. Transfer learning adapts quickly to new tasks with insufficient data^[9]. Using a pre-trained model, share underlying weights and fine-tune it to handle dataset differences. The overall framework is shown in **Figure 7**.

This paper randomly samples images to train Faster R-CNN. Network momentum is set to 0.9, weight decay to 0.0005, and learning rate for each layer to 0.001. The training step size is 30000, and the learning rate becomes 10% of the current value every 30000 iterations. 70,000 iterations on the training set take about 24 hours. Loss variation is shown in **Figure 8**.

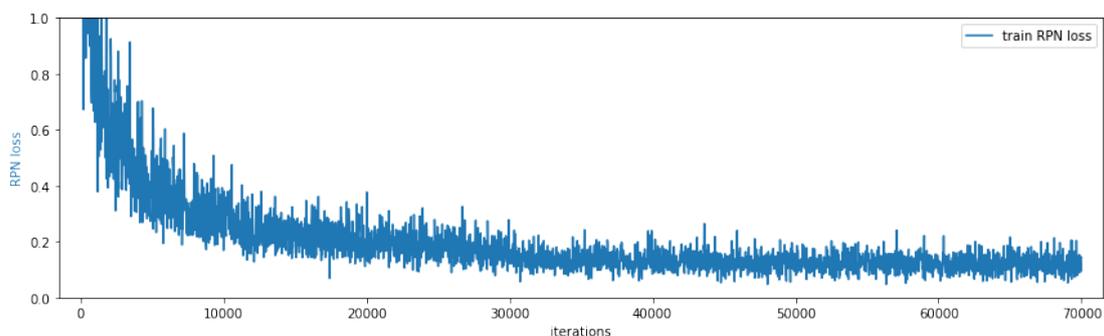


Figure 8. Curve of loss

From **Figure 8**, the loss curve shows that the model fits rapidly at the stage of 8,000 iterations, and tends to be stable after reaching 30,000 iterations, with a small amplitude of oscillation.

2.3.5. Evaluation standard

Excessive model iterations may cause over-fitting, so evaluating the trained model is essential. Average Precision (AP) measures network sensitivity to targets and reflects overall network performance. Intersection over Union (IoU) is a common criterion for object-detection accuracy. Thus, AP and IoU evaluated the corn-field weed-detection

method. AP depends on the recall rate and precision, and mAP is the mean of AP. The formulas are as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{IoU} = \frac{DR \cap GT}{DR \cup GT} \quad (3)$$

$$\text{mAP} = \frac{1}{C} \sum_{k=i}^N P(k) \Delta R(k) \quad (4)$$

Where TP is true positive rate, TN is true negative rate, FP is false positive rate, FN is false negative rate, DR is the detection result, and GT is the ground truth. In the mAP formula, C is the number of categories, N is the number of reference thresholds, k is the threshold, P(k) is the precision, and R(k) is the recall.

First, select the best recognition model by finding the model file with the highest mAP from all saved models. Then, adjust the Non-Maximum Suppression threshold in the selected model. Next, measure precision, recall, and IoU to fit target detection and framing. As the focus is on accurate weed identification, precision is prioritized over recall. Crops have regular shapes, and since only their center points need to be determined in this paper, the requirement for IoU is not high.

3. Results and discussion

3.1. Average accuracy of the network

In the experiment, 70,000 iterations were run, with a model generated every 2,000 iterations, yielding 35 models in total. Their weed-detection performance was evaluated. The model with the highest mAP was found, with a mean accuracy of 81.25%. The P-R curve is in **Figure 9**. The data shows the model can accurately distinguish maize from weeds. For weed identification, it works better for *Eleusine indica* and *Portulaca oleracea*, but worse for *Amaranthus tricolor*.

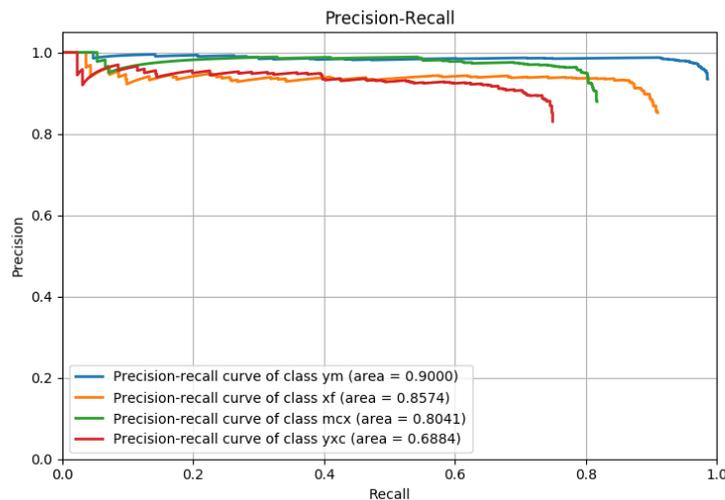


Figure 9. P-R curve of the model detection. ym (maize), xf (*Eleusine indica*), mcx (*Portulaca oleracea*), yxc (*Amaranthus tricolor*).

3.2. Validation of the model on the original images

To further verify the model’s effectiveness and test the algorithm under various real-world conditions, this experiment uses illumination, field size, occlusion, and continuous-image overlap rate as condition variables to test the model.

3.2.1. Under the different conditions of illumination

Taking illumination during photography as the control variable, 50 images each under sunny and cloudy conditions were selected from the image database for testing. **Figure 10** shows two photos taken in sunny and cloudy conditions. The identification accuracies of maize and weeds are shown in **Table 1**.



Figure 10. Weeds under the different conditions of illumination: (a) Sunday; (b) Cloudy

As seen in **Figure 10**, compared to cloudy days, sunny-day images are brighter, with plant shadows on the ground covering nearby weeds. But **Table 1** shows little difference in experimental results for the two illumination conditions. This indicates that shadows and illumination have no impact on weed detection. The reason is the significant color differences among plants, shadows, and soil. Even with shadow coverage, maize and weed colors and boundaries are visible, and the recognition effect is not greatly affected.

Table 1. Accuracy under different lighting conditions

Average precision	Maize	<i>Eleusine indica</i>	<i>Portulaca oleracea</i>	<i>Amaranthus</i>	Weeds (total)
Sunny	91.2%	84.1%	80.4%	69.1%	89.7%
Cloudy	90.5%	86.6%	81.3%	67.8%	90.6%

Note: Weeds (total) are referred to as weeds, including *Eleusine indica*, *Portulaca oleracea*, and *Amaranthus tricolor*.

3.2.2. Under the different sizes of field view

In the detection model, the camera’s field of view was set as the control variable. Given maize’s row-growth characteristic, images were classified into two field of view types: single-row crop views (**Figure 11a**) and double-row crop views (**Figure 11b**). 50 images of each type were selected from the database for testing. The identification accuracy results are in **Table 2**.



Figure 11. Weeds under the different sizes of field view: (a) Single row vision; (b) Double row view

As **Table 2** shows, the model's detection effect weakens as the number of objects in the image increases. The decline in the identification of *Amaranthus tricolor* is the most notable. Its leaves are scattered and are similar in size and shape to soil clods. Young *Amaranthus tricolor*, being small, is hard to detect in a wide field of view (red box in **Figure 12b**). *Eleusine indica* and *Portulaca oleracea* grow in clusters and are distinct from the soil (yellow and blue boxes in **Figure 12b**), so they are less affected. Overall, there is no difference in the recognition accuracy of weeds. Thus, for the sake of efficiency, wide-field weed detection can be applied in practice.

Table 2. Accuracy under different field sizes

Average precision	Maize	<i>Eleusine indica</i>	<i>Portulaca oleracea</i>	<i>Amaranthus</i>	Weeds (total)
Single	91.3%	86.7%	80.3%	73.8%	91.6%
Double	89.2%	84.7%	80.7%	67.8%	89.3%

3.2.3. Under the different conditions of occlusion

Most maize grows in rows, while weeds are randomly distributed in the field, either solitary or in groups. As seen in **Figure 13**, maize and weeds vary in distance. When far from plants, weeds are unobscured, with clear morphological features contrasting with the background (**Figure 12a**). When near other plants (**Figure 12b**), they are partially shaded, with incomplete features. To study occlusion's impact on weed identification, plant occlusion in images was set as a control variable for the detection model, with occlusion rate calculated as occluded area divided by total area. 50 images, each with no occlusion, less than 50% occlusion rate, and more than 50% occlusion rate, were selected from the image database for testing. The identification accuracy results are in **Table 3**.



Figure 12. Weeds under the different conditions of occlusion: (a) Uncovered grass; (b) Covered grass

The table shows that object occlusion significantly impacts target detection. Compared to no occlusion, recognition precision drops by 6.76% with an occlusion rate < 50% and 16.7% with an occlusion rate > 50%. *Amaranthus tricolor* is most affected by occlusion, while *Eleusine indica* and *Portulaca oleracea* are less so.

Table 3. Accuracy under different visual occlusion conditions

Average precision	<i>Eleusine indica</i>	<i>Portulaca oleracea</i>	<i>Amaranthus</i>	Weeds (total)
Uncovered	90.9%	85.3%	80.9%	94.4%
Less than 50%	86.4%	80.5%	66.8%	87.2%
More than 50%	80.6%	76.3%	57.9%	79.2%

When detecting the model's actual effect, occlusion is found to most severely affect its recognition accuracy. For a single image, this issue remains unsolved. However, considering the weeding robot's field-operation characteristics, a solution is feasible. As the robot moves during weeding, the target's observation angle changes with position, altering the occlusion rate. Usually, images can be captured when the occlusion rate drops, increasing the chance of recognizing occluded targets. To ensure a high occlusion rate, some weeds near corn were deliberately chosen as identification targets. The position where they were fully covered was set as the starting shooting point. Continuous vertical shooting was conducted while slowly moving between corn rows until the target was unoccluded and identifiable. A total of 250 target-identifying images from different positions were obtained. **Figure 13** shows two *Amaranthus* images with different occlusion rates. Occluded images were selected, and the model identified weeds in them. The identification accuracy is shown in **Table 4**.



Figure 13. Weeds at different angles: (a) The target is completely obscured; (b) The target is partially obscured. The red box is the target object. The white arrow is the view direction.

The table also shows that compared to the previous section’s detection results, the accuracy for each occlusion scenario has improved. Thus, continuous multi-position detection effectively addresses the impact of occlusion rate on recognition precision.

Table 4. Accuracy under continuous detection

Average precision	<i>Eleusine indica</i>	<i>Portulaca oleracea</i>	<i>Amaranthus</i>	Weeds (total)
	91.4%	87.3%	74.3%	97.1%

3.2.4. Recognition speed

After analyzing environmental conditions on detection accuracy, it was essential to test target detection time. As the Faster R-CNN algorithm adjusts image size before convolution, detection speed for images of different resolutions is linked to this adjustment. The tested image resolution was 3000×4000. With GPU acceleration, each image’s recognition took 50ms. Video detection reached 20 frames per second, achieving real-time results.

3.2.5. Comprehensive validation

A complex area in the cornfield, covering all plants, was chosen as the experimental zone to test the model in a complex environment. The shooting took place in the afternoon on a cloudy day, with a double-row camera view. The camera moved linearly and at a uniform speed above the area for continuous shooting, capturing multiple photos at different positions. Maize and weeds had varying degrees of concealment. Image data was input into the model for detection, and the identification results are shown in **Figure 14**.

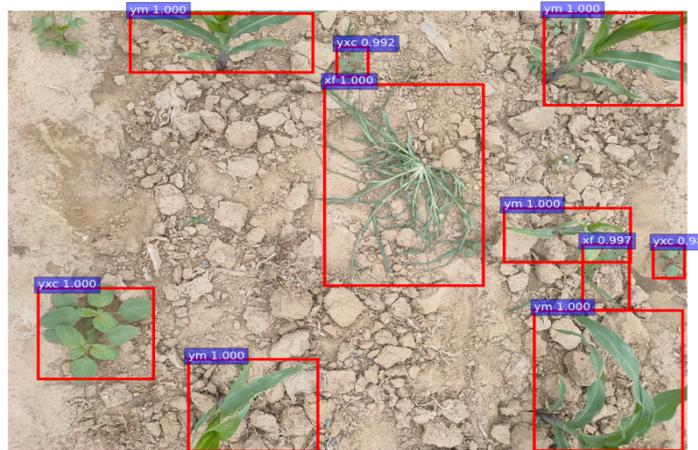


Figure 14. Result image of weed recognition

As depicted in **Figure 14**, the detection classifier exhibits excellent performance in field-scene imagery. Within the complex environment characterized by dense target distribution and cross-occlusion, the detection model demonstrates robustness. The plant-detection region (red box in **Figure 15**) accurately encompasses the actual location of weeds in the image. Ultimately, the weed recognition accuracy attains 94.3%. Specifically, the average precision for *Eleusine indica* is 91.4%, for *Portulaca oleracea* is 87.3%, and for *Amaranthus* is 73.3%. In summary, the detector can effectively discriminate and detect diverse weed species in the field.

4. Conclusions

This paper gives an analysis of weed detection, recognition, and classification technology. This technology is Faster R-CNN-based and it can be applied in corn fields.

- (1) Research shows that deep learning is a detection algorithm for complex environments. With an average recognition speed of 20 frames per second and an image detection time of about 20ms, the model has a mAP of 81.25% and a weed recognition precision of 94.3%. This weeds identification, detection, and classification method balances timeliness and accuracy, and can locate weeds. It offers algorithmic support for intelligent weeding manipulator research, can classify weeds effectively, and provides technical support for precise herbicide application.
- (2) Real-world tests show the Faster R-CNN weed detection method can detect weeds in daytime images. It's robust under various conditions like light, weather, and leaf occlusion. Compared with methods such as support vector machine, principal component analysis, and Markov, it has higher robustness. It suits complex weed distributions and clusters, breaking the limits of traditional image-processing algorithms. Zheng *et al.* used a support vector machine for weed classification and detection^[10]. Its drawback is that recognition drops significantly when target colors are similar. Lottes *et al.* used a Markov random field with random forests for plant classification, achieving an 85% true-positive rate and 79% true-negative rate^[11], with accuracy affected by scene complexity. In contrast, deep learning is robust to complex scenarios during training.

- (3) Plant occlusion severely impacts identification accuracy. In video streams, image continuity reveals different weed angles, enabling target identification. This dynamic approach quickly identifies weeds and effectively boosts detection accuracy through continuous detection.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] The National Bureau of Statistics of China, 2018, China Statistical Yearbook 2018, China Statistical Publishing House, Beijing.
- [2] Wang C, Li Z, 2016, Weed Recognition Using SVM Model with Fusion Height and Monocular Image Features. *Transactions of the Chinese Society of Agricultural Engineering*, 32(15): 165–174.
- [3] Partel V, Charan Kakarla S, Ampatzidis Y, 2019, Development and Evaluation of a Low-cost and Smart Technology for Precision Weed Management Utilizing Artificial Intelligence. *Computers and Electronics in Agriculture*, 157: 339–350.
- [4] Barrero O, Perdomo SA, 2018, RGB and Multispectral UAV Image Fusion for Gramineae Weed Detection in Rice Fields. *Precision Agriculture*, 19(5): 809–822.
- [5] Louargant M, Villette S, Jones G, et al., 2017, Weed Detection by UAV: Simulation of the Impact of Spectral Mixing in Multispectral Images. *Precision Agriculture*, 18(6): 1–20 + 932–951.
- [6] Ren S, Girshick R, Girshick R, et al., 2017, Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 39(6): 1137–1149.
- [7] Barth R, Ijsselmuiden J, Hemming J, et al., 2019, Synthetic Bootstrapping of Convolutional Neural Networks for Semantic Plant Part Segmentation. *Computers and Electronics in Agriculture*, 161: 291–304.
- [8] Fu L, Feng Y, Majeed Y, et al., 2018, Kiwifruit Detection in Field Images Using Faster R-CNN with ZFNet. *IFAC-Papers Online*, 51(17): 45–50.
- [9] Zhen X, Chen J, Zhong Z, et al., 2017, Deep Convolutional Neural Network with Transfer Learning for Rectum Toxicity Prediction in Cervical Cancer Radiotherapy: A Feasibility Study. *Physics in Medicine and Biology*, 62(21): 8246–8263.
- [10] Zheng Y, Zhu Q, Huang M, et al., 2017, Maize and Weed Classification Using Color Indices with Support Vector Data Description in Outdoor Fields. *Computers and Electronics in Agriculture*, 141: 215–222.
- [11] Lottes P, Hoferlin M, Sander S, et al., 2016, Effective Vision-based Classification for Separating Sugar Beets and Weeds for Precision Farming. *Journal of Field Robotics*, 34(6): 1160–1178.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.