

A novel deep learning-based method for detection of weeds in vegetables

Xiaojun Jin,^a Yanxia Sun,^b Jun Che,^a Muthukumar Bagavathiannan,^c Jialin Yu^{c,d} and Yong Chen^{a*}

Abstract

BACKGROUND: Precision weed control in vegetable fields can substantially reduce the required weed control inputs. Rapid and accurate weed detection in vegetable fields is a challenging task due to the presence of a wide variety of weed species at various growth stages and densities. This paper presents a novel deep-learning-based method for weed detection that recognizes vegetable crops and classifies all other green objects as weeds.

RESULTS: The optimal confidence threshold values for YOLO-v3, CenterNet, and Faster R-CNN were 0.4, 0.6, and 0.4/0.5, respectively. These deep-learning models had average precision (AP) above 97% in the testing dataset. YOLO-v3 was the most accurate model for detection of vegetables and yielded the highest F_1 score of 0.971, along with high precision and recall values of 0.971 and 0.970, respectively. The inference time of YOLO-v3 was similar to CenterNet, but significantly shorter than that of Faster R-CNN. Overall, YOLO-v3 showed the highest accuracy and computational efficiency among the deep-learning architectures evaluated in this study.

CONCLUSION: These results demonstrate that deep-learning-based methods can reliably detect weeds in vegetable crops. The proposed method avoids dealing with various weed species, and thus greatly reduces the overall complexity of weed detection in vegetable fields. Findings have implications for advancing site-specific robotic weed control in vegetable crops.

© 2022 Society of Chemical Industry.

Keywords: precision weed management; deep learning; YOLO-v3; CenterNet; faster R-CNN

1 INTRODUCTION

Vegetable production represents a large share of the agricultural production in the world, and vegetables greatly contribute to human nutrition. China is the largest consumer of vegetables in the world, and vegetables account for 35% of *per capita* food consumption in the country.¹ The total area under vegetable production in China is 19.64 million hectares, which accounts for approximately 12% of the total cropping area of the country.² Agricultural production needs to increase to meet the growing demand for food produce, fueled by rapid population growth. Weed control is a notable challenge in agricultural systems, especially in vegetable production. Weeds compete with vegetables for light, water, nutrients, and space and can significantly reduce vegetable yields if not controlled adequately.^{3–7} For example, a mixture of grass and broadleaf weeds at a density of 65 weeds m^{-2} reduced the yield of lettuce (*Lactuca sativa* L.) by over 90%, whereas purple nutsedge (*Cyperus rotundus* L.) at a density of 200 shoots m^{-2} reduced the yield of tomato (*Solanum lycopersicum* L.) by 44%.^{7,8} In China, weed control in vegetable fields still relies largely on hand-weeding. Labor shortages and rising wages have increased production costs, which in turn have led to increased market price of vegetables.^{9,10} In addition, there are a limited number of herbicides registered for weed control in vegetable crops.¹¹ The increasing occurrence of weed populations with

resistance to popular herbicides, along with the diminishing discovery and commercialization of new effective herbicides, creates demand for alternative weed control practices in vegetables.¹²

With the recent advances in mechanization and robotics, precision weed control with mechanical methods appears to be a promising option for weed management in row crops and vegetables.¹³ Robotic weeders can also improve the sustainability of weed management while reducing herbicide inputs and thwarting herbicide resistance evolution. Previous researchers have invested great effort to develop mechanical weeding actuators, but selective weed control without damaging the

* Correspondence to: Y Chen, College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China. E-mail: chenrongjsnj@163.com

a College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, China

b School of Rail Transportation, Nanjing Vocational Institute of Transport Technology, Nanjing, China

c Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, USA

d College of Forestry, Nanjing Forestry University, Nanjing, China

desired crops cannot be achieved without assistance from a target detection system.¹⁴ The lack of robust sensing technology to discriminate between crops and weeds is the primary limitation to the commercial development of intelligent robotic weeders.⁶

A variety of visual characteristics have been used in weed detection,^{15,16} classification¹⁷ and weed mapping¹⁸ through image-processing techniques. Generally, these characteristics can be divided into three categories, namely color, morphological, and textural features.⁶ Color features are appropriate for vegetation segmentation, but they cannot reliably discriminate between crops and weeds. A large number of studies have investigated the fundamental feasibility of using biological morphology (shape and structure) to distinguish between plant species.¹⁹ However, this method is limited to only the situation where the plant leaves are well displayed without occlusion, and the shape of the entire leaf is undamaged.⁵ Texture is normally regarded as a similarity grouping in an image. By extracting and analyzing textural characteristics, weeds and crops can be discriminated.²⁰ However, one of the common challenges in weed detection and classification is that crops and weeds may exhibit similar morphological characteristics.¹¹ In recent years, significant advancements have been made with deep learning for weed detection.²¹ Kamilaris *et al.* reviewed deep-learning techniques for weed detection and concluded that these techniques have generally outperformed traditional image-processing methods.²²

The rise of deep learning has been fostered by improvements in graphics processing units (GPUs).^{23,24} GPUs allow deep-learning models to learn from a large amount of data, which is vital to achieve high accuracy levels.²⁵ Deep learning has proven to be a promising approach in computer vision,^{26–28} natural language processing,^{29,30} and speech recognition.^{25,31} One of the most significant benefits of deep learning is that it can automatically learn representations from raw data without introducing hand-coded rules or human domain knowledge.²⁵ Among the various deep-learning models, convolutional neural networks (CNNs) have a remarkable ability to extract desirable features from images.^{32–34} CNNs are currently being employed as a powerful tool for image classification and object detection.^{35–37}

In agriculture, a considerable amount of research has been reported on the application of various deep-learning techniques for yield prediction,^{38,39} disease detection,^{40,41} insect damage recognition,^{42,43} weed detection,^{44,45} and crop quality examination.^{46,47} Olsen *et al.* presented deep-learning models that accurately classified 16 different types of weeds with an overall accuracy of >95%.⁴⁸ dos Santos Ferreira *et al.* reported that a CNN model reliably identified various broadleaf and grassy weeds in soybean (*Glycine max* L.).⁴⁹ Yu *et al.* compared three image classification neural networks, including AlexNet, GoogleNet, and VGG16, as well as an object detection neural network, DetectNet, and found the highest overall accuracy for weed detection with VGG16.^{50,51} Recently, Arun *et al.* performed weed detection in an unmanned aerial vehicle imagery using Faster R-CNN and Single Shot Detection Detector (SSD).⁵² These studies utilized deep learning to detect weeds and distinguish crops and weeds directly. With this approach, deep-learning models need to be trained with various weed species.^{53–55} However, a wide variety of weed species may present at various growth stages and densities in vegetable fields. Setting up such huge training datasets is time-consuming and

labor-intensive. Moreover, it is difficult for the neural networks to maintain high accuracy of detection for every weed species.

In the present work, the vegetable crops were detected by CNNs, while nonvegetables (weeds) were extracted and discriminated by image processing utilizing color features and returning visual detection. The target vegetable species used in this research was bok choy (*Brassica rapa* ssp. *chinensis*), which is a fast-growing vegetable and typically matures within 25 days after planting. The objectives of this research were to (i) evaluate and compare the performance of different deep-learning models for vegetable detection and (ii) investigate the feasibility of using deep-learning-based methods for discriminating the vegetable crop and weeds for robotic mechanical weeding.

2 MATERIALS AND METHODS

2.1 Overview

The method proposed in this paper for the detection of weeds in vegetables can be divided into the following general steps:

- (1) A trained CNN was used for detecting the vegetable crop and drawing bounding boxes around it.
- (2) Plants growing outside the bounding boxes were marked as weeds and extracted via image processing utilizing color features.

Three state-of-the-art CNN-based architectures, You Only Look Once-v3 (YOLO-v3),⁵⁶ CenterNet,⁵⁷ and Faster R-CNN,⁵⁸ were evaluated for detection of bok choy, also known as Chinese white cabbage. YOLO-v3 is an inheritance and improvement over YOLO⁵⁹ and YOLO-v2.⁶⁰ In YOLO, object classification and localization are unified into a regression problem. A YOLO does not require a feature pyramid network (FPN), and it generates bounding box coordinates and probabilities of each class directly at the output layer. YOLO-v3 is a single-stage detection model and provides much faster detection. For feature extraction, YOLO-v3 utilizes Darknet-53 as the backbone architecture, and the softmax loss in the old version has been replaced by a logistic loss. CenterNet is an anchor-free object detector. It models objects as key points and uses heatmaps to predict the centers of objects. The heatmaps are generated through a Gaussian kernel and a fully convolutional network (FCN). Object properties, including size, dimension, orientation, and pose, can be regressed directly from the center localization without any prior anchor.⁶¹ Faster R-CNN is a two-stage object detection algorithm with fast detection speed. It consists of a region proposal network (RPN) and a Fast R-CNN detector. The RPN is an FCN, which extracts candidate-bound boxes by setting anchor boxes with different proportions.⁶² The RPN is trained end-to-end to generate high-quality region proposals, which are adopted by Fast R-CNN for detection. The Fast R-CNN detector shares a set of convolutional features with the RPN. In the region of interest (RoI) pooling layer, bounding box regions are resized into uniform-sized feature vectors using max pooling. Finally, a regressor is utilized to produce accurate coordinate values of the bounding boxes.

2.2 Image acquisition

Images of bok choy growing under field conditions were acquired using a digital camera (HV1300FC, DaHeng Image, Inc., Beijing, China) at a ratio of 4:3, with an original dimension of

2048 × 1536 pixels. During image acquisition, the camera was approximately 0.6 m above the ground. The training and testing images were acquired at multiple times between September 2020 and January 2021. The images of bok choy were collected from multiple vegetable fields around Nanjing, Jiangsu, China (32°12' N, 118°48' E). Bok choy did not emerge uniformly in the fields. Thus, the training images were taken evenly under various growth stages. Furthermore, the images were taken from 10:00 a.m. to 5:00 p.m. under various illumination conditions, including cloudy, partly cloudy, and sunny days.

2.3 Training and testing

For vegetable detection training, all images were resized to 1400 × 1050 pixels using ImageJ (version 2.1.0, an open-source software available at <https://github.com/imagej/imagej>). The training dataset contained a total of 920 images, and these images were expanded to 9200 images using data augmentation methods to enrich the experimental dataset and enhance the robustness of the detectors in the training step. The training images were pre-processed in terms of brightness, rotation, flips, color variations, and image definition. The validation or testing dataset contained a total of 115 images. For detection of the vegetable crop, the training dataset, validation dataset, and testing dataset contained a total of 11 339, 1552, and 1580 individual bok choy plants, respectively. Manual annotation was performed by drawing bounding boxes onto the bok choy in the training images using the custom software Labellmg (version: 1.8.1, an open source software available at <https://github.com/tzutalin/labellmg>). The program generated Extensible Markup Language label files used for model training.

The training and testing processes were performed in the PyTorch (Facebook, Inc., CA, USA) open-source deep-learning environment using a graphic processing unit (NVIDIA GeForce RTX 2080 SUPER; NVIDIA, Santa Clara, CA, USA). Transfer learning aims to utilize previously acquired knowledge accumulated from data in auxiliary domains to solve new but similar problems in the current domain.⁶³ ImageNet is a large dataset with more than 14 million labeled images.⁶⁴ In this work, three types of CNNs were pre-trained using ImageNet to initialize the weights and bias through the transfer learning approach. To ensure a fair comparison between the results of all deep-learning models, all three CNNs were converted to the PyTorch version by modifying their weights and models to the corresponding PyTorch compositions. Furthermore, the parameters were fine-tuned to improve the performance of each CNN model before training. The hyperparameters across experimental configurations are presented in Table 1.

The results of validation and testing for all CNN models can be divided into four types: true positive (tp), true negative (tn), false positive (fp), and false negative (fn). In this context, tp represents the total number of vegetable instances successfully detected by the model, tn indicates the total number of nonvegetable instances successfully excluded by the model, fp indicates the total number of predictions

that are incorrectly identified as the vegetable crop, and fn represents the total number of vegetable instances that are incorrectly not identified as the vegetable crop. For each model, the precision, recall, and F_1 (defined in Eqn 6) score were computed over the whole period of training at the end of 120 epochs.

Precision measures the ability of the model to accurately detect the target, which was defined using the following equation:

$$\text{precision} = \frac{tp}{tp + fp} \quad (1)$$

Recall measures the effectiveness of the model to detect the target and was defined by the following equation:

$$\text{recall} = \frac{tp}{tp + fn} \quad (2)$$

During the training process, an intersection-over-union (IoU) was used to determine if the object detected was a true positive, with a threshold of 0.5.⁶⁵ IoU is the overlap rate between the ground truth box and the predicted box, which is used to denote the coincidence between the reference and the predicted bounding box. IoU was defined as follows:

$$\text{IoU} = \frac{\text{area}(\text{Box}_T \cap \text{Box}_P)}{\text{area}(\text{Box}_T \cup \text{Box}_P)} \quad (3)$$

where Box_T is the ground truth box based on the training label, Box_P is the predicted bounding box, and function *area* indicates the area of intersection region.

Mean average precision (mAP) is the mean of the AP over all classes at all recall values at different IoU thresholds from 0.5 to 0.95 for Common Objects in Context (COCO)-based evaluation. In Pascal Visual Object Classes (Pascal VOC)-based evaluation, mAP is calculated for an IoU threshold of 0.5.⁶⁶ The mAP is a commonly used metric in object detection domain and was defined using the following equation:

$$\text{mAP} = \frac{\sum_{i=1}^N \text{AP}_i}{N} \quad (4)$$

where N is the number of object classes. AP_i is the average precision for each class, which is computed as the area under precision-recall curve (PRC):

$$\text{AP} = \int_0^1 P(R) dR \quad (5)$$

In this study, we used mAP_{50} , which represents the average precision calculated at an IoU threshold of 0.5. Since we only have one

Table 1. Values of the hyperparameters for the three different convolutional neural network models evaluated in the study

Model	Batch	Momentum	Initial learning rate	Optimizer	Decay	Training epochs
YOLO-v3	4	0.937	0.01	SGD	0.0005	120
CenterNet	4	0.900	1.25e−4	Adam	0.0001	120
Faster R-CNN	4	0.900	0.02	SGD	0.0001	120

SGD, stochastic gradient descent.

class (vegetable), the mAP of the model is therefore equal to the AP of the vegetable detection.

The F_1 score is also one of the essential measures to evaluate the model. It is defined as the harmonic means of the precision and recall, and was calculated using the following equation:

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

Frames per second (FPS) measures the number of images that are processed on the model per second. The higher the FPS value, the faster the image processing is. The FPS was used as a quantitative metric to evaluate the speed of different object detection models.

2.4 Weed segmentation

To extract weeds (pixels outside the bounding boxes) from the background (i.e., soil, straws, and residues), the widely used excess green (ExG) index introduced by Woebbecke *et al.*⁶⁷ was adopted with minor modifications:

$$\text{ExG} = \begin{cases} 0, & \text{if } (g < r \parallel g < b) \\ 2g - r - b, & \text{otherwise} \end{cases} \quad (7)$$

where r , g , and b are the normalized RGB coordinates ranging from 0 to 1. Normalized RGB coordinates were defined as follows:

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B} \quad (8)$$

Since the modified ExG index is based on normalized RGB coordinates, it is insensitive to the intensity of the light source as well as illumination angle.⁶⁸ The resulting grey image was then transformed into a binary image using a method previously reported by Otsu.⁶⁹ Finally, an area filter was applied to eliminate random noises, returning visual detection regarding the presence of weeds in the image.

3 RESULTS AND DISCUSSION

3.1 Model performance with vegetable detection

The performance of the trained CNN models was tuned by specifying a threshold for the output confidence scores. To determine the threshold, 1552 ground truths from the validation dataset were used to evaluate the CNN models. Precision, recall, and F_1 score of the models at nine threshold values (0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, and 0.1) were gathered (Table 2).

Among all prediction boxes, the one with a confidence score greater than the threshold was considered as a *tp*. The best precision value for YOLO-v3 was 1.0 in the validation dataset when the confidence threshold value was 0.9; however, in this case, a low recall value of 0.440 was obtained. In contrast, when the confidence threshold value was 0.1, the model achieved high recall and precision values of 1.0 and 0.953, respectively. The optimal confidence threshold for YOLO-v3 was 0.4 as it achieved a balanced performance with precision and recall. Using this confidence threshold, we obtained precision, recall, and F_1 score of 0.983, 0.980, and 0.981, respectively. A recall value of 0.980 indicated that 98.0% of the vegetable instances were detected. Likewise, CenterNet achieved the highest F_1 score when the threshold value was 0.6. The precision value with this confidence

threshold was 0.978 in the validation dataset, indicating that the model achieved 97.8% accuracy in successfully predicting the correct vegetables. When the confidence threshold was 0.4/0.5, Faster R-CNN achieved the highest F_1 score of 0.972 along with high precision and recall values of 0.974 and 0.970, respectively.

The final performances of YOLO-v3, CenterNet, and Faster R-CNN (with threshold values of 0.4, 0.6, and 0.4/0.5 for the confidence score, respectively) were evaluated using 115 images from the testing dataset. Table 3 shows the metrics comparison results in terms of precision, recall, F_1 score, AP, and FPS. All three models had the AP above 97% in the testing dataset, illustrating that these models were able to obtain a favorable performance. YOLO-v3 performed better than CenterNet and demonstrated higher precision, recall, AP, and F_1 score values. Faster R-CNN achieved a higher recall value with a lower precision value compared to YOLO-v3. YOLO-v3 and CenterNet had similar AP values. YOLO-v3 demonstrated the highest F_1 score (0.971) among the three models, along with high precision and recall values of 0.971 and 0.970, respectively. Overall, these results indicate that YOLO-v3 with the confidence score of 0.4 was the best model in detecting the vegetable crop.

The speed of weed detection, in terms of FPS, is shown in Table 3. YOLO-v3, with 55.56 images detected per second, was 6.94 slower than CenterNet, but noticeably faster than Faster R-CNN. YOLO-v3 and CenterNet demonstrated a significant speed advantage of the one-stage model over the two-stage model. CenterNet exhibited a fast inference rate because it is an anchor-free detection model and does not require nonmaximum suppression and eliminates the need for designing anchor boxes.⁵⁷ In this study, CenterNet outperformed the other detectors on detection efficiency, which demonstrated the speed advantage of the anchor-free model. Typically, two-stage models tend to achieve higher accuracy, but with a higher computational cost than one-stage and/or anchor-free models.⁷⁰ Consequently, the low detection speed of the two-stage detector Faster R-CNN model may limit its potential applications.⁷¹

By jointly analyzing the F_1 score, AP, and FPS, YOLO-v3 from the one-stage family demonstrated superiorities in both accuracy and computational efficiency compared to CenterNet and Faster R-CNN. This competitive result mainly comes from its inherent power in the deeper feature extraction network and the one-stage architecture.⁷² Overall, these results demonstrated that YOLO-v3 was the most efficient and accurate model for detecting bok choy among the three models.

YOLO-v3 achieved excellent performances in detecting the vegetable crop under various conditions, including the crop growing with grasses (Fig. 1(a)), the crop with broadleaf weeds (Fig. 1(b)), the crop with grasses and desiccated plant residues (Fig. 1(c)), and the crop planted at high densities (Fig. 1(d)). The bounding boxes generated by YOLO-v3 failed to cover every single instance of the vegetable due to occlusion, which reduced the recall values. As shown in Fig. 2(a), some vegetables were planted too close and were totally occluded. Although the occurrence of such cases resulted in missed detection, this is unlikely to be an issue in field applications because all vegetable areas have been detected. As shown in Fig. 2(b), the similarity in plant morphology resulted in erroneous detection and a loss of precision. The occurrence of this type of erroneous detection can be minimized by increasing the number of training images containing such weed species.

Table 2. Model performance metrics at different confidence scores in the validation dataset for the three convolutional neural networks evaluated in the study

Model	Confidence score	True positive	False positive	Precision	Recall	F_1 score
YOLO-v3	0.9	683	0	1.000	0.440	0.611
	0.8	1195	5	0.996	0.770	0.868
	0.7	1335	12	0.991	0.860	0.921
	0.6	1428	14	0.990	0.920	0.954
	0.5	1490	23	0.985	0.960	0.972
	0.4	1521	26	0.983	0.980	0.981
	0.3	1536	48	0.970	0.990	0.980
	0.2	1536	48	0.970	0.990	0.980
	0.1	1552	77	0.953	1.000	0.976
	CenterNet	0.9	1288	5	0.996	0.830
0.8		1412	11	0.992	0.910	0.949
0.7		1474	19	0.987	0.950	0.968
0.6		1505	34	0.978	0.970	0.974
0.5		1521	57	0.964	0.980	0.972
0.4		1536	76	0.953	0.990	0.971
0.3		1536	76	0.953	0.990	0.971
0.2		1552	208	0.882	1.000	0.937
0.1		1552	208	0.882	1.000	0.937
Faster R-CNN		0.9	1428	17	0.988	0.920
	0.8	1474	24	0.984	0.950	0.967
	0.7	1474	24	0.984	0.950	0.967
	0.6	1490	37	0.976	0.960	0.968
	0.5	1505	40	0.974	0.970	0.972
	0.4	1505	40	0.974	0.970	0.972
	0.3	1521	63	0.960	0.980	0.970
	0.2	1521	63	0.960	0.980	0.970
	0.1	1536	150	0.911	0.990	0.949

3.2 Performance of weed segmentation

While the vegetable crop instances were detected, the plants (image pixels) outside the bounding boxes of the detected vegetable instances were marked as weeds. A color index-based segmentation utilizing the ExG + Otsu method was employed to identify weeds in an RGB color space and to extract weeds from the background.

The excess green (ExG) index has performed well in separating the plants from the background. ExG converted a color image into a greyscale image, which was easy to transform into a binary image using the method of Otsu. Figure 3 represents the results when the color index was applied to Fig. 1(a)–(d), which shows that the ExG + Otsu method yielded high segmentation quality and weeds were effectively extracted from the background.

Due to color similarities between weeds and the background, some background pixels were segmented as weeds incorrectly (noises). These noises disjointedly occur throughout the images. The relatively small noise regions were eliminated by an area filter using a thresholding technique. The area of each connected region was calculated. Afterwards, areas smaller than a preset threshold (depending on the physical pixel size) were marked as noise and filtered. Final segmentation results with vegetable regions were marked with red boxes (Fig. 3).

Weed detection methods traditionally focused on identifying weeds directly,^{21,22} but various weed species with distinct plant morphological features and growth densities may occur in a field. Establishing such a dataset containing various weed species is time-consuming and labor-intensive, and may not achieve an effective performance of weed

Table 3. Performance metrics comparison of the three convolutional neural networks in the testing dataset

Model	Confidence score	Precision	Recall	F_1 score	AP	FPS
YOLO-v3	0.4	0.971	0.970	0.971	0.984	55.56
CenterNet	0.6	0.956	0.950	0.953	0.983	62.50
Faster R-CNN	0.4/0.5	0.955	0.980	0.967	0.975	14.24

AP, average precision; FPS, frames per second.

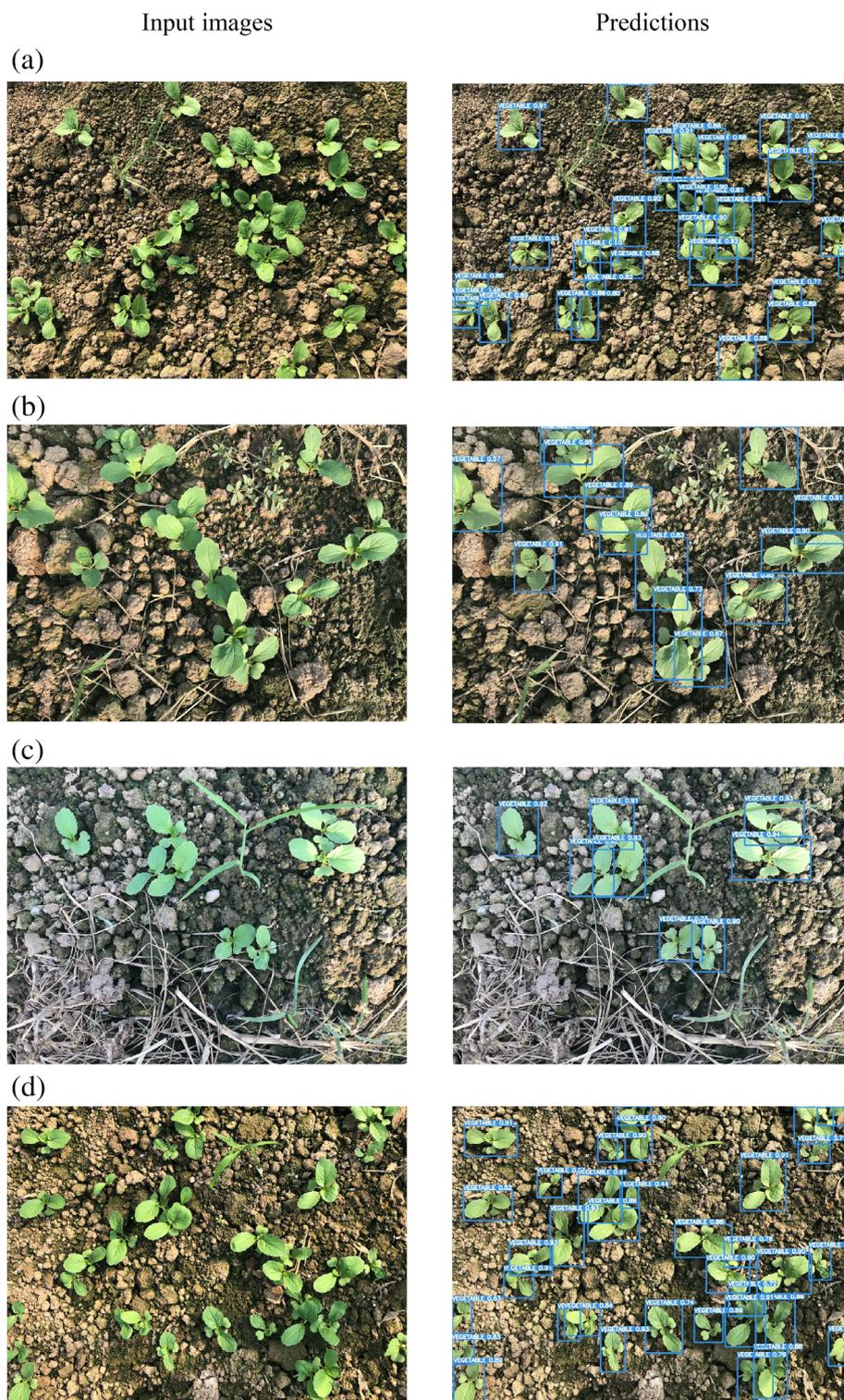


Figure 1. YOLO-v3 generated bounding box predictions for the vegetable crop under various conditions: (a) the vegetable crop growing with grasses, (b) the vegetable crop growing with broadleaf weeds, (c) the vegetable crop growing with grasses and desiccated plant residues, and (d) the vegetable crop planted at high densities.

detection. In contrast, our method trained the model to detect vegetables only and all vegetation growing outside the detected vegetable regions was marked as weeds. This strategy avoids dealing with various weed species and thus significantly reduces the complexity of weed detection.

Moreover, the proposed method can accurately detect weeds even for new and unseen weed species from a region or a country that are not included in the training dataset.

Based on the high-level performance, the proposed method is highly suitable for ground-based weed and vegetable

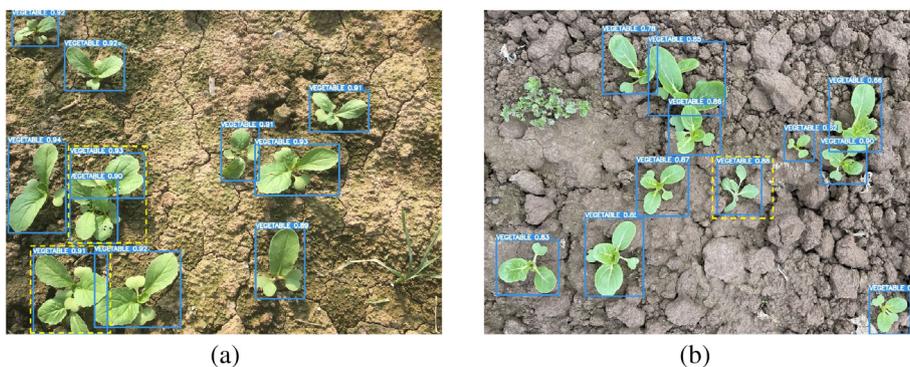


Figure 2. (a) YOLO-v3 failed to cover every single vegetable due to occlusion (the yellow dashed boxes represent missed detection). (b) The yellow dashed box represents erroneous detection.

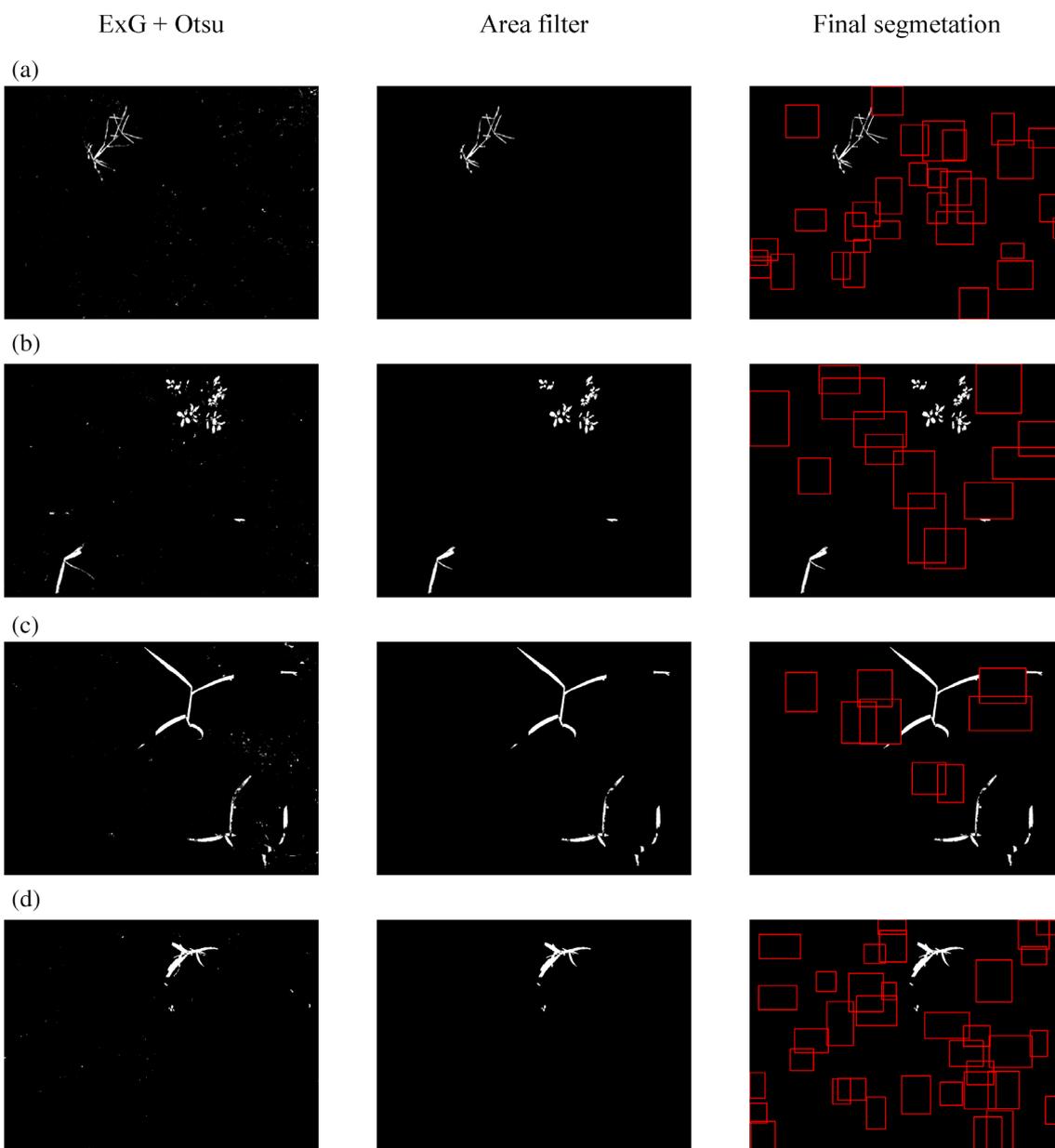


Figure 3. Results when the ExG + Otsu method and area filter were applied to the images in Fig. 1. Figure 3(a)–(d) represents the results when the proposed method was applied to Figs 1(a)–(d), respectively.

discrimination in bok choy fields. We anticipate that the same approach can be used to detect weeds in other vegetables as well as row crops.

4 CONCLUSIONS

This work demonstrated the feasibility of using deep learning for detection of vegetables, thereby detecting weeds indirectly. The proposed approach is composed of two stages. A deep-learning architecture based on CNN was used to detect vegetable instances and draw bounding boxes around them. Thereafter, the plants occurring outside of the bounding boxes were considered as weeds.

The performance of three state-of-the-art deep-learning models, CenterNet, YOLO-v3, and Faster R-CNN, were evaluated and compared in terms of AP, F_1 score, and inference speed. All of the three models had the AP above 97% in the testing dataset. YOLO-v3 outperformed CenterNet and Faster R-CNN with the highest F_1 score. Despite the higher accuracy of the two-stage detector Faster R-CNN, it is not effective at detecting vegetables primarily due to the unacceptable detection speed. Overall, the results of the present study demonstrated that the one-stage architecture YOLO-v3 was the best detector among the evaluated neural networks.

The proposed method combined deep learning and traditional image-processing technology to identify the vegetable crop and then discriminate between the crop and weeds. The neural network developed in this work can be used in the machine vision subsystem of robotic weeders for site-specific weed control. Additional studies will be conducted to detect weeds using the models obtained in this work for *in situ* video input under field conditions. To further improve the accuracy of weed detection, the deep-learning models obtained in the present study may need to be further optimized in the future.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 32072498) and the Key Research and Development Program of Jiangsu Province (Grant No. BE2021016).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- Ryder E, World vegetable industry: production, breeding, trends. *Hortic Rev* **38**:299 (2011).
- Han J, Luo Y, Yang L, Liu X, Wu L and Xu J, Acidification and salinization of soils with different initial pH under greenhouse vegetable cultivation. *J Soil Sediment* **14**:1683–1692 (2014).
- Dai X, Xu Y, Zheng J and Song H, Analysis of the variability of pesticide concentration downstream of inline mixers for direct nozzle injection systems. *Biosyst Eng* **180**:59–69 (2019).
- Liu B and Bruch R, Weed detection for selective spraying: a review. *Curr Robot Rep* **1**:19–26 (2020).
- Hamuda E, Glavin M and Jones E, A survey of image processing techniques for plant extraction and segmentation in the field. *Comput Electron Agric* **125**:184–199 (2016).
- Slaughter DC, Giles DK and Downey D, Autonomous robotic weed control systems: a review. *Comput Electron Agric* **61**:63–78 (2008).
- Mennan H, Jabran K, Zandstra BH and Pala F, Non-chemical weed management in vegetables by using cover crops: a review. *Agronomy* **10**:257 (2020).
- Morales-Payan JP, Santos BM, Stall WM and Bewick TA, Effects of purple nutsedge (*Cyperus rotundus*) on tomato (*Lycopersicon esculentum*) and bell pepper (*Capsicum annuum*) vegetative growth and fruit yield. *Weed Technol* **672-676**:672–676 (1997).
- Laani W and Strange M, Low-input management of weeds in vegetable fields. *Calif Agric* **45**:11–13 (1991).
- Yu J, Sharpe SM, Schumann AW and Boyd NS, Deep learning for image-based weed detection in turfgrass. *Eur J Agron* **104**:78–84 (2019).
- Raja R, Slaughter DC, Fennimore SA, Nguyen TT, Vuong VL, Sinha N *et al.*, Crop signalling: a novel crop recognition technique for robotic weed control. *Biosyst Eng* **187**:278–291 (2019).
- Shaner DL and Beckie HJ, The future for weed control and technology. *Pest Manag Sci* **70**:1329–1339 (2014).
- Bakhshpour A, Jafari A, Nassiri SM and Zare D, Weed segmentation using texture features extracted from wavelet sub-images. *Biosyst Eng* **157**:1–12 (2017).
- Utstumo T, Urdal F, Brevik A, Dørum J, Netland J, Overskeid Ø *et al.*, Robotic in-row weed control in vegetables. *Comput Electron Agric* **154**:36–45 (2018).
- Siddiqi MH, Ahmad I and Sulaiman SB, Weed recognition based on erosion and dilation segmentation algorithm, in *2009 International Conference on Education Technology and Computer*. IEEE, Singapore, pp. 224–228 (2009).
- Sapkota B, Singh V, Neely C, Rajan N and Bagavathiannan M, Detection of Italian ryegrass in wheat and prediction of competitive interactions using remote-sensing and machine-learning techniques. *Remote Sens (Basel)* **12**:2977 (2020).
- Sabzi S, Abbaspour-Gilandeh Y and Arribas JI, An automatic visible-range video weed detection, segmentation and classification prototype in potato field. *Heliyon* **6**:e03685 (2020).
- Pérez-Ortiz M, Peña J, Gutiérrez PA, Torres-Sánchez J, Hervás-Martínez C and López-Granados F, A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. *Appl Soft Comput* **37**:533–544 (2015).
- Perez A, Lopez F, Benlloch J and Christensen S, Colour and shape analysis techniques for weed detection in cereal fields. *Comput Electron Agric* **25**:197–212 (2000).
- Wang A, Zhang W and Wei X, A review on weed detection using ground-based machine vision and image processing techniques. *Comput Electron Agric* **158**:226–240 (2019).
- Hasan AM, Soheli F, Diepeveen D, Laga H and Jones MG, A survey of deep learning techniques for weed detection from images. *Comput Electron Agric* **184**:106067 (2021).
- Kamilaris A and Prenafeta-Boldú FX, Deep learning in agriculture: a survey. *Comput Electron Agric* **147**:70–90 (2018).
- Jordan MI and Mitchell TM, Machine learning: trends, perspectives, and prospects. *Science* **349**:255–260 (2015).
- Liakos KG, Busato P, Moshou D, Pearson S and Bochtis D, Machine learning in agriculture: a review. *Sensors* **18**:2674 (2018).
- LeCun Y, Bengio Y and Hinton G, Deep learning. *Nature* **521**:436–444 (2015).
- Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B *et al.*, Recent advances in convolutional neural networks. *Pattern Recogn* **77**:354–377 (2018).
- Shi J, Li Z, Zhu T, Wang D and Ni C, Defect detection of industry wood veneer based on NAS and multi-channel mask R-CNN. *Sensors* **20**:4398 (2020).
- Zhou H, Zhuang Z, Liu Y, Liu Y and Zhang X, Defect classification of green plums based on deep learning. *Sensors* **20**:6993 (2020).
- Collobert R and Weston J, A unified architecture for natural language processing: Deep neural networks with multitask learning, in *Proceedings of the 25th International Conference on Machine Learning*. ICML, Helsinki Finland, pp. 160–167 (2008).
- Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K and Kuksa P, Natural language processing (almost) from scratch. *J Mach Learn Res* **12**:2493–2537 (2011).
- Hinton G, Deng L, Yu D, Dahl GE, Mohamed A, Jaitly N *et al.*, Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Sign Process Mag* **29**:82–97 (2012).

- 32 Schmidhuber J, Deep learning in neural networks: an overview. *Neural Netw* **61**:85–117 (2015).
- 33 Ghosal S, Blystone D, Singh AK, Ganapathysubramanian B, Singh A and Sarkar S, An explainable deep machine vision framework for plant stress phenotyping. *Proc Natl Acad Sci* **115**:4613–4618 (2018).
- 34 He T, Liu Y, Yu Y, Zhao Q and Hu Z, Application of deep convolutional neural network on feature extraction and detection of wood defects. *Measurement* **152**:107357 (2020).
- 35 Zhou D-X, Theory of deep convolutional neural networks: Downsampling. *Neural Netw* **124**:319–327 (2020).
- 36 Dhillon A and Verma GK, Convolutional neural network: a review of models, methodologies and applications to object detection. *Prog Artif Intell* **9**:85–112 (2020).
- 37 Raghu S, Sriaram N, Temel Y, Rao SV and Kubben PL, EEG based multi-class seizure type classification using convolutional neural network and transfer learning. *Neural Netw* **124**:202–212 (2020).
- 38 Ramos P, Prieto FA, Montoya E and Oliveros CE, Automatic fruit count on coffee branches using computer vision. *Comput Electron Agric* **137**:9–22 (2017).
- 39 Pantazi XE, Moshou D, Alexandridis T, Whetton RL and Mouazen AM, Wheat yield prediction using machine learning and advanced sensing techniques. *Comput Electron Agric* **121**:57–65 (2016).
- 40 Pantazi XE, Tamouridou AA, Alexandridis T, Lagopodi AL, Kontouris G and Moshou D, Detection of *Silybum marianum* infection with *Microbotryum silybum* using VNIR field spectroscopy. *Comput Electron Agric* **137**:130–137 (2017).
- 41 Ferentinos KP, Deep learning models for plant disease detection and diagnosis. *Comput Electron Agric* **145**:311–318 (2018).
- 42 Liu W, Wu G, Ren F and Kang X, DFF-ResNet: an insect pest recognition model based on residual networks. *Big Data Mining and Analytics* **3**:300–310 (2020).
- 43 Rustia DJA, Chao JJ, Chiu LY, Wu YF, Chung JY, Hsu JC *et al.*, Automatic greenhouse insect pest detection and recognition based on a cascaded deep learning classification method. *J Appl Entomol* **145**:206–222 (2021).
- 44 Pantazi X-E, Moshou D and Bravo C, Active learning system for weed species recognition based on hyperspectral sensing. *Biosyst Eng* **146**:193–202 (2016).
- 45 Binch A and Fox C, Controlled comparison of machine vision algorithms for Rumex and Urtica detection in grassland. *Comput Electron Agric* **140**:123–138 (2017).
- 46 Zhang M, Li C and Yang F, Classification of foreign matter embedded inside cotton lint using short wave infrared (SWIR) hyperspectral transmittance imaging. *Comput Electron Agric* **139**:75–90 (2017).
- 47 Maione C, Batista BL, Campiglia AD, Barbosa F Jr and Barbosa RM, Classification of geographic origin of rice by data mining and inductively coupled plasma mass spectrometry. *Comput Electron Agric* **121**:101–107 (2016).
- 48 Olsen A, Konovalov DA, Philippa B, Ridd P, Wood JC, Johns J *et al.*, DeepWeeds: a multiclass weed species image dataset for deep learning. *Sci Rep* **9**:2058 (2019).
- 49 dos Santos FA, Matte Freitas D, Gonçalves da Silva G, Pistori H and Theophilo Folhes M, Weed detection in soybean crops using ConvNets. *Comput Electron Agric* **143**:314–324 (2017).
- 50 Yu J, Sharpe SM, Schumann AW and Boyd NS, Detection of broadleaf weeds growing in turfgrass with convolutional neural networks. *Pest Manag Sci* **75**:2211–2218 (2019).
- 51 Yu J, Schumann AW, Sharpe SM, Li X and Boyd NS, Detection of grassy weeds in bermudagrass with deep convolutional neural networks. *Weed Sci* **68**:545–552 (2020).
- 52 Veeranampalayam Sivakumar ANV, Li J, Scott S, Psota E, Jhala A, Luck JD *et al.*, Comparison of object detection and patch-based classification deep learning models on mid-to late-season weed detection in UAV imagery. *Remote Sens (Basel)* **12**:2136 (2020).
- 53 Le VNT, Ahderom S and Alameh K, Performances of the lbp based algorithm over cnn models for detecting crops and weeds with similar morphologies. *Sensors* **20**:2193 (2020).
- 54 Quan L, Feng H, Lv Y, Wang Q, Zhang C, Liu J *et al.*, Maize seedling detection under different growth stages and complex field environments based on an improved faster R-CNN. *Biosyst Eng* **184**:1–23 (2019).
- 55 Potena C, Nardi D and Pretto A, Fast and accurate crop and weed identification with summarized train sets for precision agriculture, in *International Conference on Intelligent Autonomous Systems*. Springer, Singapore, pp. 105–121 (2016).
- 56 Redmon J and Farhadi A, Yolov3: An incremental improvement. *arXiv preprint arXiv:180402767* (2018).
- 57 Zhou X, Wang D and Krähenbühl P, Objects as points. *arXiv preprint arXiv:190407850* (2019).
- 58 Ren S, He K, Girshick R and Sun J, Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv preprint arXiv:150601497* (2015).
- 59 Redmon J, Divvala S, Girshick R and Farhadi A, You only look once: unified, real-time object detection, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Las Vegas, NV, pp. 779–788 (2016).
- 60 Redmon J and Farhadi A, YOLO9000: better, faster, stronger, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, San Francisco, CA, pp. 7263–7271 (2017).
- 61 Algabri M, Mathkour H, Bencherif MA, Alsulaiman M and Mekhtiche MA, Towards deep object detection techniques for phoneme recognition. *IEEE Access* **8**:54663–54680 (2020).
- 62 Gong B, Ergu D, Cai Y and Ma B, Real-time detection for wheat head applying deep neural network. *Sensors* **21**:191 (2021).
- 63 Lu J, Behbood V, Hao P, Zuo H, Xue S and Zhang G, Transfer learning using computational intelligence: a survey. *Knowl-Based Syst* **80**:14–23 (2015).
- 64 Deng J, Dong W, Socher R, Li L-J, Li K and Fei-Fei L, Imagenet: a large-scale hierarchical image database, in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Miami, FL, pp. 248–255 (2009).
- 65 Ghosal S, Zheng B, Chapman SC, Potgieter AB, Jordan DR, Wang X *et al.*, A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenomics* **2019**:1525874–1525814 (2019).
- 66 Everingham M, Eslami SA, Van Gool L, Williams CK, Winn J and Zisserman A, The pascal visual object classes challenge: a retrospective. *Int J Comput Vis* **111**:98–136 (2015).
- 67 Woebbecke DM, Meyer GE, Von Bargen K and Mortensen DA, Color indices for weed identification under various soil, residue, and lighting conditions. *Trans ASAE* **38**:259–269 (1995).
- 68 Gée C, Bossu J, Jones G and Truchetet F, Crop/weed discrimination in perspective agronomic images. *Comput Electron Agric* **60**:49–59 (2008).
- 69 Otsu N, A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* **9**:62–66 (1979).
- 70 Carranza-García M, Torres-Mateo J, Lara-Benítez P and García-Gutiérrez J, On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data. *Remote Sens (Basel)* **13**:89 (2021).
- 71 Alkentar SM, Alsahwa B, Assalem A and Karakolla D, Practical comparison of the accuracy and speed of YOLO, SSD and faster RCNN for drone detection. *J Eng* **27**:19–31 (2021).
- 72 Sharma C, Singh S, Poornalatha G and Ajitha Shenoy KB, Performance analysis of object detection algorithms on YouTube video object dataset. *Eng Lett* **29**:813–817 (2021).