Weed Identification Using Deep Learning and Image Processing in Vegetable Plantation

XIAOJUN JIN[ID][1]**, JUN CHE**[2]**, AND YONG CHEN**[1]
[1]College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China
[2]Department of Artificial Intelligence Algorithm, Kedacom Inc., Shanghai 200030, China

Corresponding author: Yong Chen (chenyongjsnj@163.com)

ABSTRACT Weed identification in vegetable plantation is more challenging than crop weed identification due to their random plant spacing. So far, little work has been found on identifying weeds in vegetable plantation. Traditional methods of crop weed identification used to be mainly focused on identifying weed directly; however, there is a large variation in weed species. This paper proposes a new method in a contrary way, which combines deep learning and image processing technology. Firstly, a trained CenterNet model was used to detect vegetables and draw bounding boxes around them. Afterwards, the remaining green objects falling out of bounding boxes were considered as weeds. In this way, the model focuses on identifying only the vegetables and thus avoid handling various weed species. Furthermore, this strategy can largely reduce the size of training image dataset as well as the complexity of weed detection, thereby enhancing the weed identification performance and accuracy. To extract weeds from the background, a color index-based segmentation was performed utilizing image processing. The employed color index was determined and evaluated through Genetic Algorithms (GAs) according to Bayesian classification error. During the field test, the trained CenterNet model achieved a precision of 95.6%, a recall of 95.0%, and a F_1 score of 0.953, respectively. The proposed index $-19R + 24G - 2B \geq 862$ yields high segmentation quality with a much lower computational cost compared to the wildly used ExG index. These experiment results demonstrate the feasibility of using the proposed method for the ground-based weed identification in vegetable plantation.

INDEX TERMS Weed identification, deep learning, image processing, genetic algorithms, color index.

I. INTRODUCTION

Vegetable is considered one of the most nutrient-dense food all around the world due to its sufficient vitamins, minerals and antioxidants. Raising living standards boosts the consumption of green vegetables, which makes them a substantial part of our lives and possess great commercial value. Weeds compete with vegetables for water, sunlight and nutrients, leaving them prone to insect and disease infestation [1], [2]. The yield of vegetables decreased by 45%-95% in the case of weed-vegetable competition [3]. Excessive use of chemical herbicides results in over-application in areas of low or no weed infestation and causes environmental impacts including soil and ground water pollution [4]. Moreover, organic production of vegetables requires non-chemical weed control. Thus, hand weeding is still the primary option for weed control in vegetable plantation at present [5]. With the labor cost substantially increased, development of a visual

method of discriminating between vegetable and weed is an important and necessary step towards ecologically sustainable weed management.

A considerable amount of research has been conducted on various machine vision techniques for weed detection [2], [3], [6], [7]. Ahmed *et al.* [8] used Support Vector Machines (SVMs) to identify six weed species in a database of 224 images, they achieved 97.3% precision with the best combination of extractors. Herrera *et al.* [9] constructed a weed-crop classifier using shape descriptors and Fuzzy Decision-Making, and classification accuracy of 92.9% was obtained in a set of 66 images. Crop is usually much higher than weed at early growth stages. This height feature was used by Chen *et al.* [10] to establish a crop and weed discrimination method using a binocular stereo vision system. The discrimination between crop and weed was done by a height-based segmentation method using the depth dimension analysis. For the relative higher weeds, plant spacing information was utilized to distinguish weed from the crops.

The associate editor coordinating the review of this manuscript and approving it for publication was Hao Luo[ID].

In recent years, deep learning has demonstrated remarkable performance in extracting complex features from images automatically [11]–[13]. It has been widely employed as a promising method to image classification and object detection. There are two categories of methods found in deep learning for image detection [6]. The first is classifying the object and then draw bounding boxes around images to make the object classified. Second category is classifying object pixels, also known as semantic segmentation [14]. Olsen *et al.* [15] used the benchmark deep learning models to classify images of sixteen different types of weed, with an average classification accuracy of 95.1% and 95.7%, respectively. dos Santos Ferreira *et al.* [16] performed weed detection in soybean crop images and classified these weeds among grass and broadleaf by Convolutional Neural Networks (CNNs). Asad and Bais [17] made performance comparison of deep learning meta-architectures like SegNet and UNET and encoder blocks like VGG16 and ResNet-50 on high-resolution color images of canola fields. Textural feature analysis and morphological scanning were applied to sugar beet plant by Khurana and Bawa [18], and then a KNN classifier was used to classify weed plant from field crop. Veeranampalayam Sivakumar [19] evaluated and compared two object detection models, namely, Faster RCNN and the Single Shot Detector (SSD), over UAV imagery for weed detection in soybean fields. Faster RCNN was found to be the best model in terms of weed detection performance using mean Intersection over Union (IoU) and inference speed. Osorio *et al.* [20] presented three methods for weed estimation based on deep learning and image processing in lettuce crops. One method was based on support vector machines (SVM), the second method was based on YOLOV3 (you only look once V3) and the third one was based on Mask R-CNN. The performances were compared with the estimations of a set from weed experts and they found that these methods improved accuracy on weed coverage estimation and minimized subjectivity in human-estimated data.

There is no obvious row spacing and plant spacing in vegetable cultivation. Vegetables and weeds grow randomly, which makes weed identification in vegetable plantation more challenging than crop weed identification. Moreover, weeds in vegetables will also be mixed in vegetables during mechanized harvesting and need to be sorted manually. A variety of labor costs have pushed up sales prices. So far, little work has been found on identifying weeds in vegetable plantation, and previous crop weed identification used to be mainly focused on identifying weed directly, however, there is a large variation in weed species, but limited in vegetables. Therefore, we proposed methods to firstly identify and segment the vegetable using deep learning, in particular the architecture of Convolutional Neural Networks, then the remaining green objects in the segmented image were considered as weeds. This strategy can largely reduce the size of training image dataset as well as the complexity of weed detection, thereby enhancing the weed identification performance and accuracy.

The main objective of this research is to develop a weed identification algorithm based on deep learning and image processing for robotic weed removal in the vegetable plantation. The specific objectives were to 1) train a model using deep learning approach that capable of detecting the bounding boxes of vegetables. 2) extract and segment vegetation falling out of bounding boxes, in this case, weeds by image processing utilizing color feature.

II. MATERIALS AND METHODS
A. THE PROPOSED METHOD
The approach proposed in this research for the identification of weeds is composed of two stages. The first stage consists of the state-of-art CenterNet algorithm [21] for detecting bok choy in this study. Images of bok choy are collected and used as input data for training the neural network. The trained neural network is used for detecting bok choy and drawing bounding boxes around them, which generates bounding box coordinates and associate class probabilities as the output. In the second stage, a color index-based segmentation is performed on vegetation (pixels) outside the bounding boxes using color information, returning visual classification regarding the presence of weeds in the image. The employed color index was determined and evaluated through Genetic Algorithms (GAs) according to Bayesian classification error. Procedural steps of the proposed method are shown in Fig. 1. Rest of this section provides details of each step.

CenterNet model training and testing were performed in the PyTorch deep learning environment using a graphic processing unit (NVIDIA GeForce RTX 2080 SUPER, NVIDIA; Santa Clara, USA). Genetic algorithm was developed and implemented using the Python language with OpenCV library. Both algorithms were operated from a computer with an Intel(R) Core (TM) i9-9900K CPU @ 3.60GHz central processing unit (CPU).

B. IMAGE ACQUISITION
Images of bok choy or Chinese white cabbage (brassica rapa spp. chinensis) were acquired using a digital camera. The vegetable plantation where the images were acquired is located under the geographical coordinates of latitude 32°12′38.172″N and longitude 118°48′51.87″E, Nanjing, China. The original dimensions of the images were 3024 × 4032 pixels. Bok choy images were taken under various conditions, such as varied illumination conditions (Fig. 2a; Fig. 2b), complex backgrounds (Fig. 2c), various growth stages (Fig. 2d).

C. VEGETABLE DETECTION USING DEEP LEARNING
1) IMAGE AUGMENTATION
The training dataset contained 1150 images, these images were then expanded to 11500 images using data augmentation methods for the purpose of enhancing the richness of the experimental dataset. The collected images were pre-processed in terms of color, brightness, rotation, and
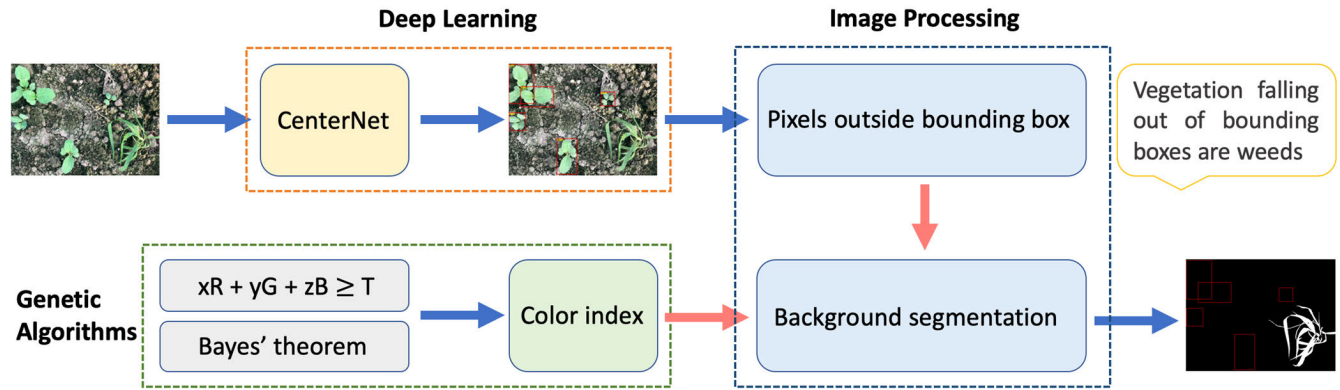
Deep Learning

Image Processing

FIGURE 1. Flow diagram for proposed weed detection methodology.

FIGURE 2. Bok choy images taken under various conditions: (a) low brightness, (b) high brightness, (c) complex background, (d) various growth stages.
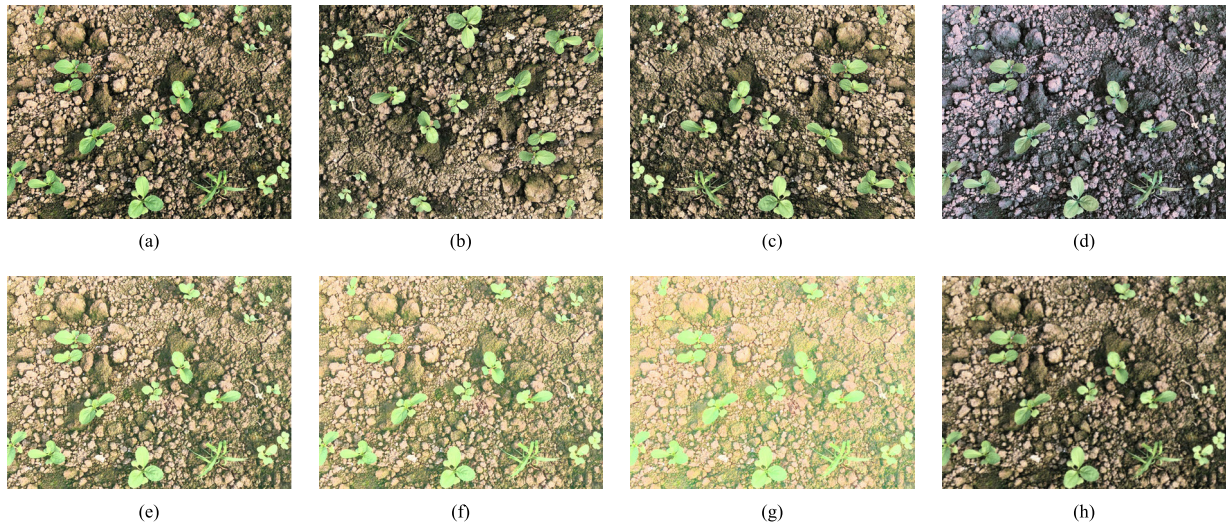

FIGURE 3. Image augmentation methods: (a) original image, (b) 180° clockwise rotation, (c) horizontal mirror, (d) color balance processing, (e-g) brightness transformation, and (h) blur processing.

image definition, and the dataset was augmented as shown in Fig. 3.

2) IMAGE ANNOTATION

Manual annotation was applied by drawing bounding boxes onto the vegetable (in this paper, bok choy) in the input images using a custom software LabelImg. Corresponding XML format label files were generated to train the CenterNet. 80% and 20% of the dataset were used for training and testing, respectively.

3) TRAINING AND TESTING

The CenterNet [21] model is a cutting edge and brand-new object detector, which is anchor-free and depends on the key points estimation. In CenterNet, objects are represented as a single point, and heatmap is used to predict the centers of objects. Heatmap is created using a Gaussian kernel and an FCN, estimated centers are derived from the peak values in the heatmap [22]. Based on the center localization, object properties such as size and

TABLE 1. Individual loss definition.

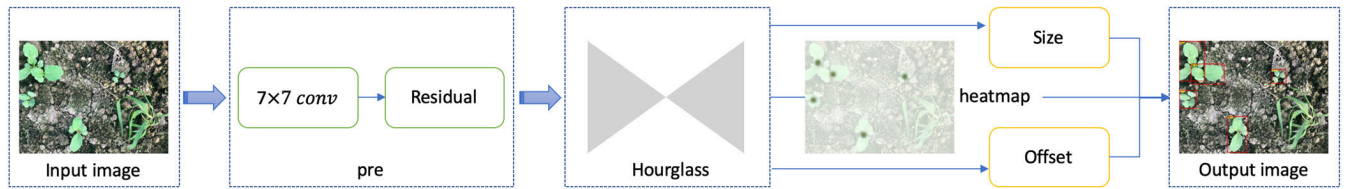| Loss	Equation	Parameters	Meaning
L_k	$\dfrac{-1}{N}\sum \begin{cases} (1-\hat{Y}_{xyc})^{\alpha}\log(\hat{Y}_{xyc}), & if\ Y_{xyc}=1 \\ (1-Y_{xyc})^{\beta}(\hat{Y}_{xyc})^{\alpha}\log(1-\hat{Y}_{xyc}), & otherwise \end{cases}$	N	Number of key-points
		α, β	Hyper-parameters of the focal loss
L_{size}	$\dfrac{-1}{N}\sum_{k=1}^{N}\left|\hat{S}_{p_k}-S_k\right|$	S_k	Object size for object k
		p_k	Center point for object k
L_{off}	$\dfrac{-1}{N}\sum_{P}\left|\hat{O}_{\tilde{P}}-(\dfrac{p}{R}-\tilde{p})\right|$	$\tilde{p}$	Key-points location
		$\hat{O}$	Predicted local offset

FIGURE 4. Detection model of CenterNet.

dimension can be regressed directly without any prior anchor [23].

CenterNet is a single-stage detection model, and it does not require the non-maximum suppression (NMS) as the post-processing step, thus it provides much faster detection. For the feature extraction, Hourglass was selected as backbone architecture in this study. To train the network, each ground truth key point is transformed to the lower size of key-point heatmap using a Gaussian kernel, with focal loss L_k. In addition, CenterNet also predicts the local offset to reduce the error caused by the size resample from input image to key-points heatmap. The offset is trained with an L_{off} loss. Finally, the object size is regressed from the center points with loss L_{size}. Hence, the loss function (L_{det}) is made up of three components: key-point loss (L_k), offset loss (L_{off}) and object size loss (L_{size}):

$$L_{det} = L_k + \lambda_{size}L_{size} + \lambda_{off}L_{off} \qquad (1)$$

where λ_{size} and λ_{off} are constants for loss weighting. In this paper, we used $\lambda_{size} = 0.1$ and $\lambda_{off} = 1$ as suggested by the author. The equations of each individual loss and its meaning were presented in Table 1.

At inference time, each key point location is given by an integer coordinate (x_i, y_i). Afterwards, CenterNet uses key-point values as a measure of its detection confidence, and produces a bounding box at location

$$(\hat{x}_i + \delta\hat{x}_i - \frac{\hat{w}_i}{2}, \hat{y}_i + \delta\hat{y}_i - \frac{\hat{h}_i}{2},$$
$$\hat{x}_i + \delta\hat{x}_i + \frac{\hat{w}_i}{2}, \hat{y}_i + \delta\hat{y}_i + \frac{\hat{h}_i}{2}) \qquad (2)$$

where $(\delta\hat{x}_i, \delta\hat{y}_i)$ is the offset prediction and $(\hat{w}_i, \hat{h}_i)$ is the size prediction.

The state-of-art CenterNet was proved to be simpler, faster, and more accurate than traditional bounding-box based detectors [21]. In this study, we adopted and implemented the CenterNet model as the vegetable detector.

The CenterNet detection model is shown in Fig. 4. All images in the training set are resampled to a fixed size. While detecting the object, it predicts bounding boxes as well as their confidence scores.

During the detection process, objects are represented as a single point - the center point of its bounding box, which is then obtained from key-point estimation. This anchor-free strategy used by CenterNet to only detect the object from the center point and regress the object size enables it to work more accurate and efficiently, and thereby making it outperforms most of the detection methods.

D. WEED IDENTIFICATION UTILIZING IMAGE PROCESSING

Once the vegetable was found, the remaining green objects falling out of the bounding boxes were marked as weeds. To extract weeds from other elements of the scene (i.e. soil, straws, stones, and other residues), color index-based segmentation using a binary-coded genetic algorithms (GAs) identifying weed in RGB color space for the outdoor field conditions was studied and implemented. Evaluation was then carried out by comparing the output of the segmentation result with the widely used excess green (ExG) index [24].
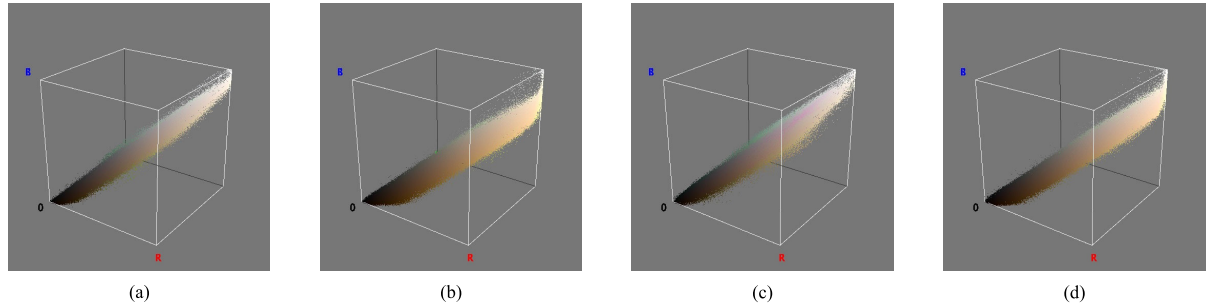
FIGURE 5. Image pixel distribution in RGB color space: (a) image pixel distribution of Fig. 2a, (b) image pixel distribution of Fig. 2b, (c) image pixel distribution of Fig. 2c, (d) image pixel distribution of Fig. 2d.
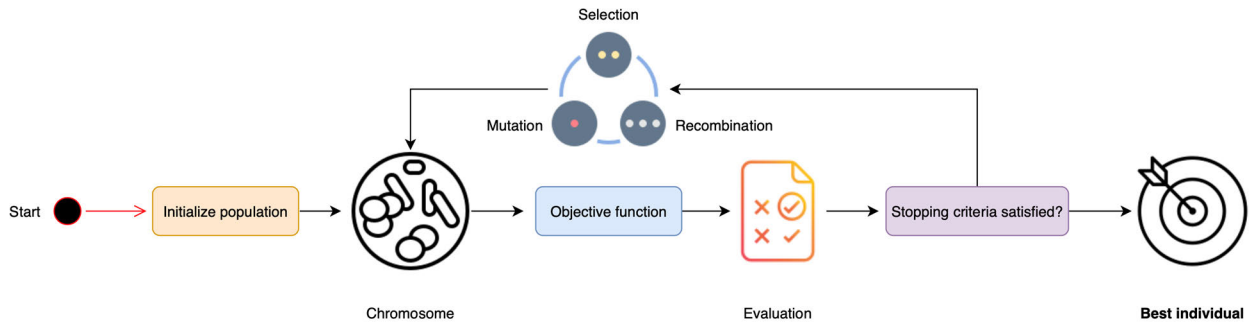

FIGURE 6. Procedures of a genetic algorithm.

1) COLOR INDEX-BASED SEGMENTATION

Image pixel distribution in RGB color space is shown in Fig. 5. Segmentation refers to the process of finding a plane that intersect the RGB color cube, thereby classifying image into vegetation and non-vegetation pixels. The plane is defined by the equation:

$$xR + yG + zB = T \tag{3}$$

In order to separate vegetation from the background, the value of x, y, z and T need to be determined.

2) GENETIC ALGORITHMS

Genetic Algorithms (GAs) are a family of adaptive search methods based on the mechanics of natural selection and natural genetic evolutionary system, which are particularly efficient in dealing with difficult combinatorial search problems without being trapped by local optima through their parallel exploration of the search space [25].

The three parameters x, y and z are in the range $(-255, 255)$, while T is limited to $[0, 1024)$ [26]. This will lead to $510 \times 510 \times 510 \times 1024$ possible combinations. Thus, an efficient searching algorithm is necessary to solve this problem. GAs behave well in exploiting accumulated information of an initially unknown domain in a highly efficient way. Therefore, GA was selected to design a search engine in this work. Procedures of a genetic algorithm is shown in Fig. 6. To start the algorithm, an initial population was generated randomly.

3) CHROMOSOME

An 88-bit binary string, or a chromosome, was used to represent parameters in Eq. 3 that encoded by permutation method. The relative locations of the Bytes in chromosome are crucial because of how GAs choose better combinations of parameters. Based on the length and range of parameters, the string was organized with the sign as the first byte in the chromosome, where 0 means negative, while 1 indicates positive. The next bytes were binary values of the parameters. Table 2 shows the structure of chromosome string.

4) POPULATION SIZE

The basic element of a GA is called individual, which is characterized by a set of parameters (variables) known as Genes. Genes are then joined in a string to form a Chromosome. A set of individuals is referred to as a population. In this work, a population size of 200 was used to generate color index parameters.

5) SELECTION

The probability of an individual is selected for reproduction is (directly or inversely) proportional to its fitness relative to the rest of the population. Roulette wheel selection, which selects the individual with the highest fitness randomly picked individuals, was chosen and implemented in GA.

6) CROSSOVER AND MUTATION

Crossover exchanges information between the selected chromosomes and generates a new offspring population for the

Param	Chromosome											Value
	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7	Byte 8	Byte 9	Byte 10	Byte 11	
x	0	0	0	0	0	1	0	0	1	1	-	-19
y	1	0	0	0	0	1	1	0	0	0	-	24
z	0	0	0	0	0	0	0	0	1	0	-	-2
T	1	1	1	0	1	0	1	1	1	1	0	862

next population, which is implemented by cutting individually the two parent bit strings into two or more-bit string segments. Offspring bit strings are created from combining the two-bit string segments undergoing crossing over. A crossover probability of 0.8 was selected. Mutation provides random alternation on bit strings in the crossover operation by shifting, inverting or rotating one or more genetic elements during reproduction. The mutation rate of 0.2 was used in this application.

7) EVALUATION

Bayesian classification error (BCE) was applied for function evaluation. For each given color, BCE (r, g, b) was defined as:

$$BCE(r, g, b) = \begin{cases} 0.5 \times p2(r, g, b), & xR + yG + zB \geq T \\ 0.5 \times p1(r, g, b), & xR + yG + zB < T \end{cases} \tag{4}$$

where p1 (r, g, b) and p2 (r, g, b) are the distribution probabilities of weed and background in the color space, respectively

$$p1(r, g, b) = \frac{C_w}{C_t} \tag{5}$$

$$p2(r, g, b) = \frac{C_b}{C_t} \tag{6}$$

where C_w is the number of occurrences of color (r, g, b) in weed pixels. C_b is the number of occurrences of color (r, g, b) in background pixels. C_t is the total number of pixels in the reference images.

Obviously, the theoretical minimum value of BCE (r, g, b) is:

$$BCE(r, g, b)_{min} = \min(0.5 \times p1(r, g, b), 0.5 \times p2(r, g, b)) \tag{7}$$

Thereby, the definitions of BCE and its corresponding theoretical minimum value BCE_{min} with respect to the reference images are:

$$BCE = \sum_{i=0}^{I_u} BCE(r_i, g_i, b_i) \tag{8}$$

$$BCE_{min} = \sum_{i=0}^{I_u} BCE(r_i, g_i, b_i)_{min} \tag{9}$$

where I_u is the number of unique pixels in the reference images.

8) STOPPING CRITERION

The algorithm terminates whenever one of the two conditions was satisfied:

(a) if the iterations number reached 2000.

(b) if the best fitness value which is equal to the predefined threshold (theoretical minimum error) of acceptance was located.

When any one of the above conditions was met, the best-fit chromosome string according to Table 2 was decoded as the parameters of the color index.

III. RESULTS AND DISCUSSION
A. PERFORMANCE OF THE VEGETABLE DETECTION

In order to adapt the input required for the Hourglass framework, the input images were resized to 512 × 512 pixels. Batch size was set to 4 and 24 maximal number of epochs were used for the purpose of better analyzing the training process. Other parameters (momentum, initial learning rate, weight decay regularization, etc.) referred to the default parameters in the CenterNet model. In the training stage, we followed the training process of the original paper on CenterNet and the model was trained after setting the training parameters. To optimize the training loss, the optimization algorithm Adam was used to update network weights iterative based in training data. The initial values of the training parameters are shown in Table 3.

For object detection task, the results of testing can be divided into four types, including true positive (TP), true negative (TN), false positive (FP), and false negative (FN). In this context, TP represents the bounding boxes containing target vegetables that are correctly identified; FP means the bounding boxes without target vegetables that are incorrectly identified as target vegetables; and FN indicates target vegetables not identified as vegetables and no bounding boxes were drew. The precision, recall and F_1 score [27] are used as the performance indices of predictive ability. Precision and recall are defined as follows:

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

TABLE 3. Initialization parameters of CenterNet.

Size of input images	Batch	Momentum	Initial learning rate	Decay	Training steps
512×512	4	0.9	1.25e-4	0.0001	55,200

TABLE 4. Evaluation of different confidence scores.

Confidence score	True positive	False positive	Precision	Recall	F_1 score
0.9	1327	12	0.991	0.84	0.909
0.8	1422	33	0.977	0.90	0.937
0.7	1469	44	0.971	0.93	0.950
0.6	1501	69	0.956	0.95	0.953
0.5	1533	111	0.932	0.97	0.951

TABLE 5. GA parameters and performance results.

	GA parameters and performance results	Values
	Population size	200
	Selection	Roulette wheel selection
Parameters	Crossover probability	0.8
	Mutation rate	0.2
	Number of Iterations	2000
	Evaluation times	400000
	Generation with optima	423
	BCE_{min}	0.113%
Results	BCE	0.21%
	x	-19
	y	24
	z	-2
	T	862
Color index	$-19R + 24G - 2B \geq 862$	

The F_1 score is also one of the significant measures to evaluate the model. It is a harmonic means of the precision and recall defined as follows:

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (12)$$

Precision-recall curve (PRC) is made up of precision (vertical axis) and recall (horizontal axis). It is a more objective judgment criterion for the evaluation of the performance of the model. Comparison results of the different Intersection-over-Union (IoU) thresholding values on the integrated test set measured by PRCs are shown in Figure 7. In this study, 0.5 was adopted as the IoU thresholding value.

In order to determine the thresholding value of the confidence scores, 1580 ground truths bounding boxes from the test set were used to evaluate the model (Table 4). Five threshold values (0.5, 0.6, 0.7, 0.8 and 0.9) were used and tested. 0.6 was found as the optimal value as it yielded a precision of 95.6% and a recall of 95%.
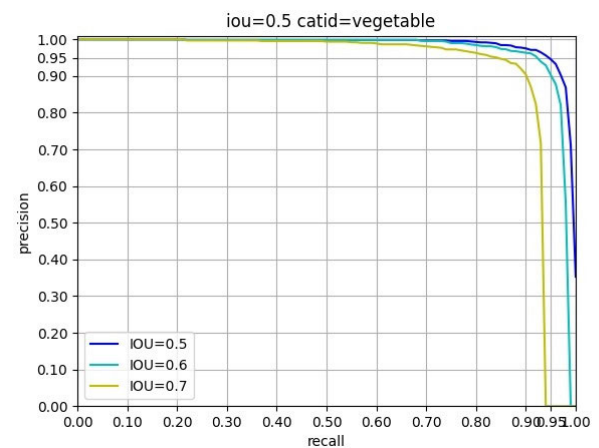

FIGURE 7. Precision-recall curve with different IoU thresholding values.

Fig. 8 shows the trained model worked on images acquired under various conditions (Fig. 2). The result images
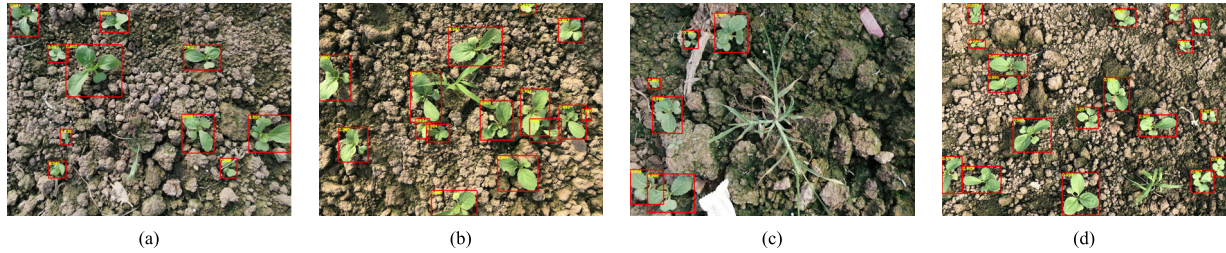
FIGURE 8. Detection of vegetables under various conditions using the CenterNet detection model. (a) detection result of Fig 2a, (b) detection result of Fig 2b, (c) detection result of Fig 2c, (d) detection result of Fig 2d.

demonstrated that the CenterNet detection model can provide high classification accuracy. Hence, it is feasible to identify the vegetable using the proposed deep learning model.

Detection of vegetables in the presence of broadleaf weed is given in Fig 9. It can be seen from the result that the trained CenterNet model can also distinguish between vegetables and broadleaf weed. It is also worth mentioning that there is a large variation in weed species. The traditional way is to identify weed directly. Therefore, deep learning models need to be trained with different types of weed dataset for classification. If there is an unknown type of weed that has not appeared in the training dataset, then the detection is most likely to fail. In contrast, the proposed method trains the model to detect only vegetables. In this way, there is no need to handle various weed species. Even if we encounter unknown weeds, it is unlikely to result in misidentification.

Examination of the detection cases also revealed that a vegetable may be missed due to occlusion (Fig. 10). In Fig. 10, several vegetables were planted too close and were completely occluded. Encountering such cases in the field would result in missing identification. However, this case can be further improved or solved by adding more occluded images into the training dataset. Moreover, better result should be obtained with a more advanced growth stage due to the fact that at this stage vegetables tend to have smaller canopies and their leaves do not extend very much [10].
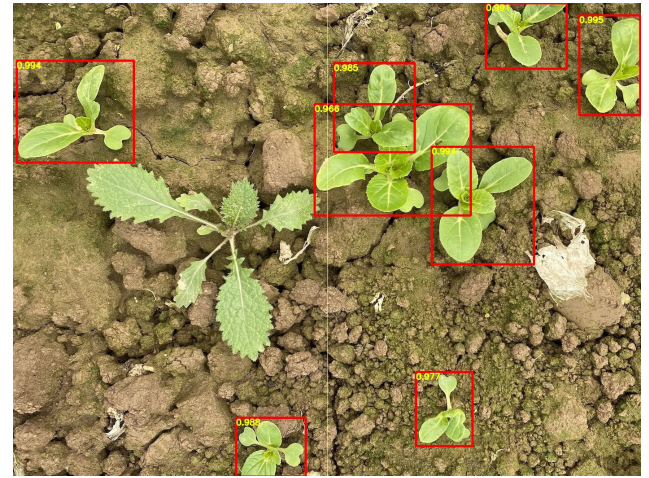
B. WEED SEGMENTATION PERFORMANCE

During the GA experiments, 20 images under various conditions were selected as reference images. Meanwhile, the same set of four selected unsegmented-images in Fig 8 was used as test images to evaluate GA color index segmentation results. The theoretical minimum value BCE_{min} of the reference images was 0.113% calculated by Eq. 9. GA parameters and performance results are shown in Table 5.

Fitness of 0.21% was the best value achieved with BCE function for all population sizes under 2000 generations. There was no substantial improvement of the best fitness when increasing the population size or iterations number. Thus, the corresponding result of the color index is:

$$\begin{cases} -19R + 24G - 2B \geq 862, & \text{weed} \\ -19R + 24G - 2B < 862, & \text{background} \end{cases} \quad (13)$$

FIGURE 9. Detection of vegetables in the presence of broadleaf weed.
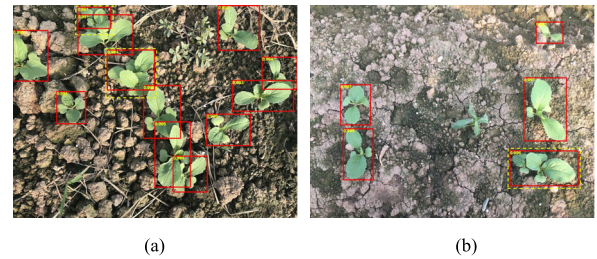
FIGURE 10. The yellow dashed boxes enclose missed vegetable cases. (a-b) occlusion.
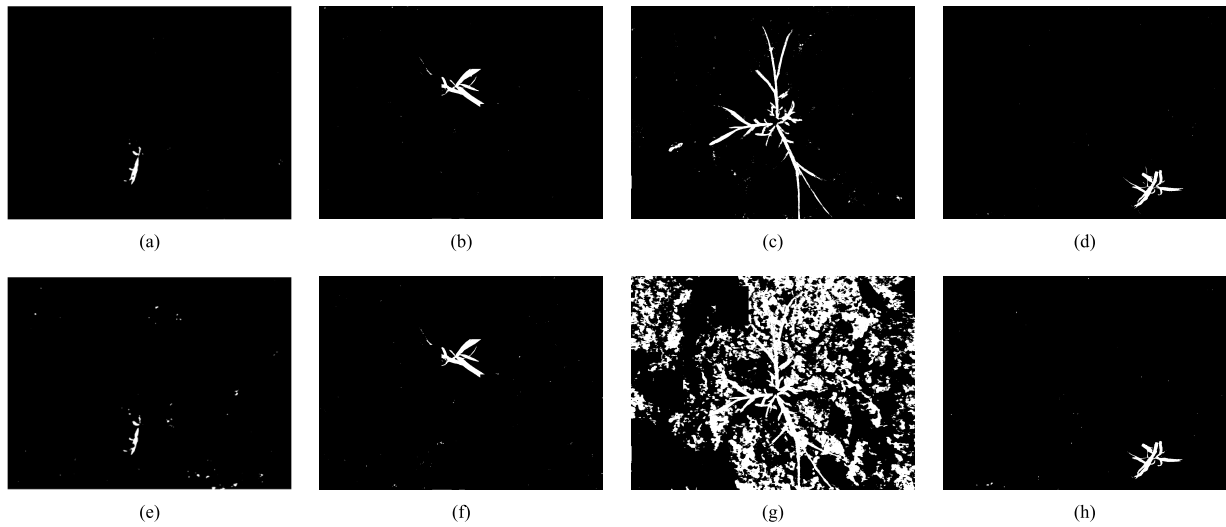
Figure 11 represents the results when the proposed index was applied to the images in Figure 8, which shows that weeds were effectively segmented from backgrounds for images taken under natural conditions. To further verify segmentation performance, evaluation was carried out by comparing the segmentation result with the excess green index (ExG) index.

The ExG index has been widely used and has performed well in separating plants from background [2], [24]. ExG index converted color image into greyscale image, which was easy to transformed into a black-and-white image to derive a binary image using the method of Otsu [28].

Examination of Fig. 11 indicated that the result obtained from ExG + Otsu method was corrupted by more noises, and

TABLE 6. Comparison between ExG + Otsu and the proposed index.

Test image	BCE		Running time (ms)	
	The proposed index	ExG + Otsu	The proposed index	ExG + Otsu
Fig 8a	0.024%	0.098%	6.26	8.35
Fig 8b	0.008%	0.020%	6.14	9.43
Fig 8c	0.256%	14.39%	7.73	11.3
Fig 8d	0.015%	0.032%	7.05	9.98

FIGURE 11. Results when the proposed index and ExG + Otsu were applied to the images in Figure 8. (a) the proposed index applied to Fig. 8a, (b) the proposed index applied to Fig. 8b, (c) the proposed index applied to Fig. 8c, (d) the proposed index applied to Fig. 8d, (e) ExG + Otsu applied to Fig. 8a, (f) ExG + Otsu applied to Fig. 8b, (g) ExG + Otsu applied to Fig. 8c, (h) ExG + Otsu applied to Fig 8d.
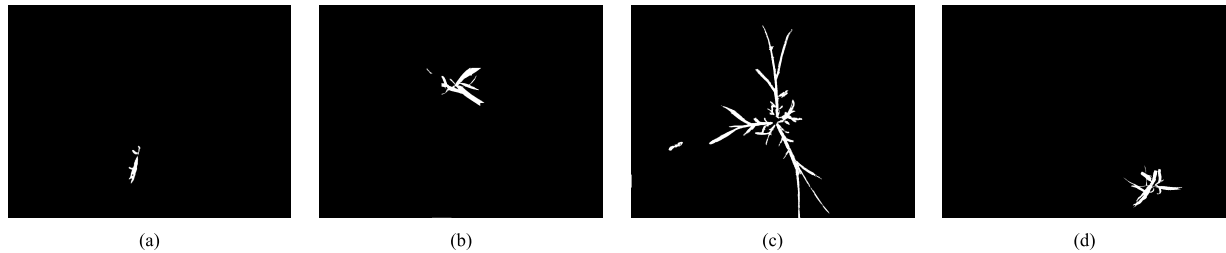
FIGURE 12. Results after area filter (a) area filter applied to Fig. 11a, (b) area filter applied to Fig. 11b, (c) area filter applied to Fig. 11c, (d) area filter applied to Fig 11d.

was not able to separate weeds from complex background (Fig. 11g). The performance of ExG + Otsu method was poorer than the proposed color index in this study as ExG with Otsu threshold tends to result in under-segmentation. On the other hand, ExG segmentation increases in computation time since it requires two steps.

Table 6 compares the proposed index with ExG + Otsu and shows their BCEs and image processing times. Based on BCE metric, it is evident that the proposed index yields high segmentation quality with a much lower computational cost compared to the ExG + Otsu method, which render it suitable

to work in vegetable plantation under natural conditions for intelligent robotic weed control.

In order to obtain a color index that capable of dealing with different lighting, several images under various lighting conditions were selected as reference images during the GA searching. Fig. 11a, and Fig. 11b represent the segmentation results when the proposed index was applied to the images in Figure 8a (low brightness) and Figure 8b (high brightness), respectively, which shows that weeds were effectively segmented from background for images taken under different lighting conditions.
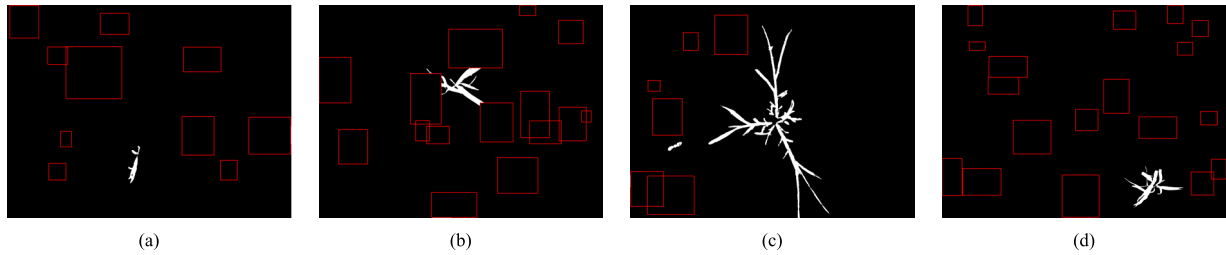
FIGURE 13. Final segmentation with vegetable regions marked with red box (a) vegetable regions marked to Fig. 12a, (b) vegetable regions marked to Fig. 12b, (c) vegetable regions marked to Fig. 12c, (d) vegetable regions marked to Fig. 12d.

Due to the color similarities between weeds and background, some background pixels were misclassified as weeds (noises). These noises generally were spread disjointedly within the image. An area filter using a thresholding technique was employed to eliminate the relatively small noise regions in the binary image. The area of each connected regions was calculated. Objects smaller than a preset threshold (derived through trial and error) were considered as noise and filtered (Fig. 12a, Fig. 12b, Fig. 12c, Fig. 12d). Final segmentation results with vegetable regions marked with red box are shown in Fig. 13a, Fig. 13b, Fig. 13c and Fig. 13d.

The proposed GA method with fitness function of Bayesian classification error can also be applied to calculate color index for the segmentation of other vegetations in agriculture area. Likewise, this process can be easily reproduced by simply replacing the reference images and counting pixel distribution probabilities of the corresponding targets.

IV. CONCLUSION

In this study, we proposed an approach to identify weeds in vegetable plantation using deep learning and image processing. The algorithm was depicted in two steps. A CenterNet model was trained to detect vegetables. The trained CenterNet achieved a precision of 95.6%, a recall of 95.0% and a F_1 score of 0.953. Then the remaining green objects in the color image were considered as weeds. To extract weeds from the background, a color index was determined and evaluated through Genetic algorithms (GAs) according to Bayesian classification error. In this way, the model focuses on identifying only the vegetables and thus avoid handling various weed species.

The contributions of this paper were to: 1) study and present an approach to identify weeds in vegetable plantation using deep learning and image processing 2) develop a visual method of discriminating between vegetable and weed in a novel and indirect way 3) propose a color index to extract weeds from the background under natural conditions.

The proposed algorithm in this paper can be used for robotic weeding, e.g. chemical weeding or mechanical weeding, however, for organic vegetables, only mechanical weeding can be applied. Given the high-level performance in this work, it was demonstrated that the proposed method is suitable for the ground-based weed identification in vegetable plantation under various conditions, including varied illumination, complex backgrounds as well as various growth stages and has application value for the sustainable development of the vegetable industry. Future work will be conducted to identify weeds in in-situ videos. Meanwhile, it would also be interesting to evaluate the accuracy reached in the detection of vegetables by optimizing the deep learning model.

REFERENCES

[1] T. W. Berge, A. H. Aastveit, and H. Fykse, "Evaluation of an algorithm for automatic detection of broad-leaved weeds in spring cereals," *Precis. Agricult.*, vol. 9, no. 6, pp. 391–405, Dec. 2008.

[2] E. Hamuda, M. Glavin, and E. Jones, "A survey of image processing techniques for plant extraction and segmentation in the field," *Comput. Electron. Agricult.*, vol. 125, pp. 184–199, Jul. 2016.

[3] H. Mennan, K. Jabran, B. H. Zandstra, and F. Pala, "Non-chemical weed management in vegetables by using cover crops: A review," *Agronomy*, vol. 10, no. 2, p. 257, Feb. 2020.

[4] X. Dai, Y. Xu, J. Zheng, and H. Song, "Analysis of the variability of pesticide concentration downstream of inline mixers for direct nozzle injection systems," *Biosyst. Eng.*, vol. 180, pp. 59–69, Apr. 2019.

[5] D. C. Slaughter, D. K. Giles, and D. Downey, "Autonomous robotic weed control systems: A review," *Comput. Electron. Agricult.*, vol. 61, no. 1, pp. 63–78, Apr. 2008.

[6] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 8, p. 2674, Aug. 2018.

[7] A. Wang, W. Zhang, and X. Wei, "A review on weed detection using ground-based machine vision and image processing techniques," *Comput. Electron. Agricult.*, vol. 158, pp. 226–240, Mar. 2019.

[8] F. Ahmed, H. A. Al-Mamun, A. S. M. H. Bari, E. Hossain, and P. Kwan, "Classification of crops and weeds from digital images: A support vector machine approach," *Crop Protection*, vol. 40, pp. 98–104, Oct. 2012.

[9] P. Herrera, J. Dorado, and Á. Ribeiro, "A novel approach for weed type classification based on shape descriptors and a fuzzy decision-making method," *Sensors*, vol. 14, no. 8, pp. 15304–15324, Aug. 2014.

[10] Y. Chen, X. Jin, L. Tang, J. Che, Y. Sun, and J. Chen, "Intra-row weed recognition using plant spacing information in stereo images," presented at the Kansas City, Missouri, St. Joseph, MI, USA, 2013. [Online]. Available: http://elibrary.asabe.org/abstract.asp?aid=43357&t=5

[11] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[13] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.

[15] A. Olsen, D. A. Konovalov, B. Philippa, P. Ridd, J. C. Wood, J. Johns, W. Banks, B. Girgenti, O. Kenny, J. Whinney, B. Calvert, M. R. Azghadi, and R. D. White, "DeepWeeds: A multiclass weed species image dataset for deep learning," *Sci. Rep.*, vol. 9, no. 1, p. 2058, Feb. 2019.

[16] A. D. S. Ferreira, D. M. Freitas, G. G. D. Silva, H. Pistori, and M. T. Folhes, "Weed detection in soybean crops using ConvNets," *Comput. Electron. Agricult.*, vol. 143, pp. 314–324, Dec. 2017.

[17] M. H. Asad and A. Bais, "Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network," *Inf. Process. Agricult.*, vol. 7, no. 4, pp. 535–545, Dec. 2020.

[18] G. Khurana and N. K. Bawa, "Weed detection approach using feature extraction and KNN Classification," in *Advances in Electromechanical Technologies*. Singapore: Springer, Sep. 2020, pp. 671–679.

[19] A. N. V. Sivakumar, J. Li, S. Scott, E. Psota, A. J. Jhala, J. D. Luck, and Y. Shi, "Comparison of object detection and patch-based classification deep learning models on mid-to late-season weed detection in UAV imagery," *Remote Sens.*, vol. 12, no. 13, p. 2136, Jul. 2020.

[20] K. Osorio, A. Puerto, C. Pedraza, D. Jamaica, and L. Rodríguez, "A deep learning approach for weed detection in lettuce crops using multispectral images," *AgriEngineering*, vol. 2, no. 3, pp. 471–488, Aug. 2020.

[21] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, *arXiv:1904.07850*. [Online]. Available: http://arxiv.org/abs/1904.07850

[22] A. Karami, M. Crawford, and E. J. Delp, "Automatic plant counting and location based on a few-shot learning technique," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5872–5886, 2020.

[23] M. Algabri, H. Mathkour, M. A. Bencherif, M. Alsulaiman, and M. A. Mekhtiche, "Towards deep object detection techniques for phoneme recognition," *IEEE Access*, vol. 8, pp. 54663–54680, 2020.

[24] D. M. Woebbecke, G. E. Meyer, K. Von Bargen, and D. A. Mortensen, "Color indices for weed identification under various soil, residue, and lighting conditions," *Trans. ASAE*, vol. 38, no. 1, pp. 259–269, 1995.

[25] L. Tang, L. Tian, and B. L. Steward, "Color image segmentation with genetic algorithm for in-field weed sensing," *Trans. ASAE*, vol. 43, no. 4, pp. 1019–1027, 2000.

[26] H. Mao, B. Hu, Y. Zhang, D. Qian, and S. Chen, "Optimization of color index and in weed recognition threshold segmentation," *Trans. CSAE*, vol. 23, no. 9, pp. 154–158, 2007.

[27] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.

[28] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. TSMC-9, no. 1, pp. 62–66, Jan. 1979.

XIAOJUN JIN was born in 1987. He received the B.S. and M.S. degrees in mechanical engineering from Nanjing Forestry University, Nanjing, China, in 2009 and 2012, respectively, where he is currently pursuing the Ph.D. degree in mechanical engineering. His research interests include computer vision and image processing, machine learning, and software development.

JUN CHE was born in 1989. He received the B.S. and M.S. degrees in mechanical engineering from Nanjing Forestry University, Nanjing, China, in 2011 and 2014, respectively. He is currently working with the Artificial Intelligence Algorithm Department, Kedacom Inc., Shanghai, China. His research interests include machine vision and the application of artificial intelligence algorithms on edge devices.

YONG CHEN received the B.S. degree in mechanical engineering from the Nanjing Institute of Technology, Nanjing, China, in 1987, the M.S. degree in power engineering from Southeast University, Nanjing, in 1999, and the Ph.D. degree in mechanical engineering from Nanjing Forestry University, Nanjing, in 2005. He is currently a Professor with Nanjing Forestry University. His research interests include robotics and mechatronics.

• • •