IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 18, 2020, accepted April 25, 2020, date of publication April 29, 2020, date of current version May 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2991354

Semantic Segmentation of Crop and Weed using an Encoder-Decoder Network and Image Enhancement Method under Uncontrolled Outdoor Illumination

AICHEN WANG[1,3], YIFEI XU[2], XINHUA WEI[1,3], AND BINGBO CUI[1,3]

[1]Key Laboratory of Modern Agricultural Equipment and Technology, Jiangsu University, Ministry of Education, Zhenjiang 212013, China
[2]School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[3]School of Agricultural Equipment Engineering, Jiangsu University, Zhenjiang 212013, China

Corresponding author: Xinhua Wei (wei_xh@126.com)

This work was supported in part by the Natural Science Foundation of Jiangsu Province, China, under Grant BK20180861, and in part by the Synergistic Innovation Center of Jiangsu Modern Agricultural Equipment and Technology under Grant 4091600002.

ABSTRACT Weeds are among the major factors that could harm crop yield. Site-specific weed management has become an effective tool to control weed and machine vision combined with image processing is an effective approach for weed detection. In this work, an encoder-decoder deep learning network was investigated for pixel-wise semantic segmentation of crop and weed. Different input representations including different color space transformations and color indices were compared to optimize the input of the network. Three image enhancement methods were investigated to improve model robustness against different lighting conditions. The results show that for images without enhancement, color space transformation and vegetation indices without NIR (Near Infrared) information did not improve the segmentation results, while inclusion of NIR information significantly improved the segmentation accuracy, indicating the effectiveness of NIR information for precise segmentation under weak lighting condition. Image enhancement improved the image quality and consequently the robustness of segmentation models against different lighting conditions. The best MIoU value for pixel-wise segmentation was 88.91% and the best mean accuracy of object-wise segmentation was 96.12%. The deep network and image enhancement methods applied in this work provided promising segmentation results for weed detection and did not need large amount of data for model training, which is suitable for site-specific weed management.

INDEX TERMS Weed detection, semantic segmentation, deep learning, precision agriculture, image processing.

I. INTRODUCTION

Agriculture is facing tremendous challenges from weeds, which appear everywhere randomly in the field, and compete with crops for water, nutrients and sunlight, resulting in a detrimental impact on crop yields and quality if uncontrolled properly [1], [2]. Numerous studies have demonstrated a strong correlation between crop yield loss and weed competition. The production loss due to weeds can be up to 34% [3]–[6]. To control weeds, different operations have been adopted, among which chemical weeding has been the most widely used one since 1940s. However, conventional chemical weeding sprays herbicides uniformly to the whole field, resulting in the overuse of herbicides and further leading to catastrophic environmental pollution problems [3]. To counteract these issues, site-specific weed management (SSWM) was introduced. In SSWM, accurate weed identification is crucial, which provides necessary individual target information for spraying to the control system [7].

Machine vision is one of the most popular approaches and has been investigated extensively for weed identification [8]. Conventional procedures for weed detection with machine vision include image pre-processing, segmentation, feature extraction and classification [2], [9]. For the feature

The associate editor coordinating the review of this manuscript and approving it for publication was Liandong Zhu.

extraction procedure, handcrafted features are usually extracted and optimally selected, which are then used for classification. For images captured under ideal conditions and at specific plant growth stages, these conventional methods provide very promising classification results with high classification performances in the order of 80-95% in terms of accuracy [7]. However, for real applications in the field, the task becomes extremely challenging. Weed identification accuracy is influenced by weed density, weed distribution characteristics, varying lighting conditions in the field, occlusion or overlapping of the leaves of crops and weeds, and different growth stages of plants, etc[10]–[13]. Handcrafted features extracted from color, shape, texture and spectrum are not robust enough to the changes of these factors, leading to the poor robustness and low generalization capabilities of conventional crop-weed classification methods, and imposing difficulties to the practical applications of such methods in precision agriculture [7], [14].

Deep learning has been investigated extensively for image processing and has also been applied in agriculture including weed identification. Compared with conventional machine learning methods for identifying weeds from digital images, deep learning can automatically learn the hierarchical feature expression hidden deep into the images, avoiding the tedious procedures to extract and optimize handcrafted features [14]. In addition, semantic segmentation is one of the most effective approaches for alleviating the effect of occlusion and overlapping since pixel-wise segmentation can be achieved. Some deep learning algorithms have been investigated for weed detection. Dyrmann *et al.* [15] trained a fully CNN based on GoogLeNet architecture to detect weed locations in leaf occluded cereal crops, which yielded a recall of 46.3% and a precision of 86.6%. To cope with substantial environmental changes with respect to weed pressure, weed types, growth stages of the crop, visual appearance, and soil conditions, Lottes *et al.* [7] adopted a fully convolutional network (FCN) with an encoder-decoder structure and incorporated spatial information by considering image sequences. Both RGB and NIR (Near Infrared) images were used for model training. Results showed that the method substantially improved the accuracy of crop-weed classification. Similarly, Milioto *et al.* [16] constructed an end-to-end encoder-decoder semantic segmentation network, and fed the network with 14 different vegetation indices and alternate representations as input for semantic weed/crop/background segmentation. The proposed method could properly deal with heavily overlapping objects and a large variety of growth stages, yielding the best MIoU (mean intersection of union) value of 80.8% for pixel-wise segmentation. Though promising results can be obtained with these deep learning-based methods for weed identification, there are still room for improvement. The deep learning networks could learn effective features for weed detection, but are also affected by varying lighting conditions, which were not fully considered in the aforementioned studies. To further improve semantic segmentation accuracy, Chen *et al.* [17] proposed an encoder-decoder network with atrous separable convolution, for semantic image segmentation. The network could refine the segmentation results especially along object boundaries and yield state-of-art performance on PASCAL VOC 2012 and Cityscapes datasets. However, the network also did not consider varying lighting conditions.

Therefore, this work aimed at performing pixel-wise semantic segmentation of field images into soil, crop and weed. Specifically, (1) an encoder-decoder network with atrous separable convolution was investigated for semantic crop/weed/soil segmentation; (2) different input representations including different color space transformations and color indices, were compared to analyze the effect of input representations to the performance of the adopted network; and (3) model robustness with respect to lighting conditions was improved by image enhancement.

II. MATERIALS AND METHODS

A. IMAGE DATASETS

Two image datasets were evaluated in this work. One is a publicly available sugar beet image dataset (http://www.ipb.uni-bonn.de/data/sugarbeets2016/) captured with a readily available agricultural robotic platform, BoniRob, on a sugar beet farm near Bonn in Germany over a period of three months in spring 2016 [18]. The other is an oilseed image dataset captured with a commercial RGB camera (Canon 60D, 50 mm lens, 5184 pixel × 3456 pixel) which was mounted on a gantry-type frame at a height of around 1.5 m above soil at our own test field on campus in early winter 2017. The sugar beet dataset consists of both RGB and corresponding NIR images, captured with a JAI AD-130GE multi-spectral camera at an image resolution of 1296 pixel × 966 pixel. The JAI AD-130GE camera was mounted to the bottom of the BoniRob robot chassis at a height of around 85 cm above soil, consisting of a RGB sensor and a NIR monochromatic sensor. The NIR monochromatic sensor collects signals within the spectral range of 750-1000 nm, with sensitivity peak at around 780 nm. The sugar beet was in early growth stage and weed density was relatively low, with slight overlapping of the leaves of sugar beet and weed. For the RGB images, it seems that the lighting condition is not well, as the brightness and contrast of the RGB images are low, as shown in Figure 1a. There are 283 images in the sugar beet dataset with ground-truth labeling provided by Chebrolu *et al.* [18] from which we randomly selected 200 images for training and 83 images for evaluation. For our oilseed dataset, the captured images with a resolution of 5184 pixel × 3456 pixel were cropped into more images with a resolution of 1550 pixel × 3456 pixel. The oilseed was also in the early growth stage, but with heavy weed pressure and overlapping. And the oilseed images (Figure 1b) were captured under the direct illumination of sunlight, with some shadow regions. These 68 RGB images were annotated by hand, with 50 images for training and 18 images for evaluation. To further verify the generalization capability of the

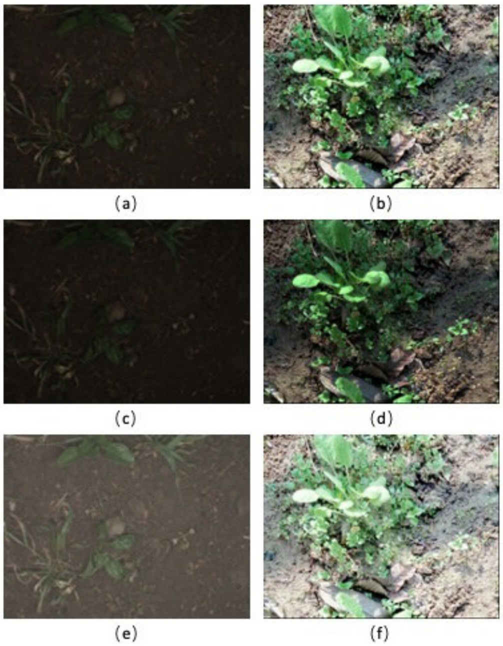


FIGURE 1. **(a) Original image from sugar beet dataset, (b) Original image from oilseed dataset, (c) Sugar beet image with *L* component subtracted by 40, (d) Oilseed image with *L* component subtracted by 100, (e) Sugar beet image with gamma value of 0.5, (f) Oilseed image with gamma value of 0.5.**

proposed approach, the datasets were augmented by gamma correction and changing the L component in HSL color space. For the sugar beet dataset, the gamma value was set as 0.5, 1.5 and 2, and the L component was added by 50, 100 and 150, and subtracted by 40, respectively. And for the oilseed dataset, the gamma value was set as 0.5, 1.5 and 2, and the L component was added by 50, and subtracted by 50 and 100, respectively. Examples of augmented images were shown in Figure 1.

B. IMAGE PREPROCESSING

Image preprocessing can help to improve the generalization capabilities of a classification model by aligning the training and test data distribution and improving the image quality [7]. As the lighting condition substantially affects the robustness of a classification model, three image enhancement methods were investigated in this work. For the input of deep network, Milioto *et al.* [16] deployed 14 different input representations including different vegetation indices and raw input in different color spaces to improve the performance of classification model and reduce the amount of images for training. Similarly, several different vegetation indices and color spaces were also evaluated as the input representations in this work.

1) IMAGE ENHANCEMENT

The two datasets involved in this work were acquired under totally different lighting conditions, as can be seen in Figure 1. For the sugar beet dataset, the images are with low brightness and contrast. By contrast, the images in our oilseed dataset were captured under the direct illumination of sunlight. The brightness and contrast of the oilseed images are high enough, with some regions close to saturation in the images. There are also some shadow regions in the oilseed images, imposing more difficulties for crop/weed classification. To alleviate the effect of different lighting conditions and improve the robustness of classification model, three image enhancement methods were evaluated.

a: HISTOGRAM EQUALIZATION

Histogram equalization (HE) is a powerful scheme for adjusting image intensities to enhance contrast. In grey scale histogram equalization, the method rearranges the grey values in such a way that the modified histogram resembles the histogram of uniform distribution [19]. The detailed principle and implementation procedures of HE can be reached in reference [20]. For color images with three channels, the same technique equalizing the image in three dimensional spaces causes unequal shift in the three components resulting in change of hue of the pixel [21]. The HE preprocessing for color image adopted in this work is to equalize only the intensity component in the color space of HSI, and then transform the equalized HSI image back to RGB color space [22].

b: AUTO CONTRAST

The process of contrast enhancement increases the perceptibility of the objects in the image. To enhance the contrast of the images involved in this research, the Auto Contrast algorithm used in the commercial software Adobe Photoshop CS6 (Adobe Systems Software Ireland Ltd.) was applied. The Auto Contrast operator does not adjust channels individually and does not introduce or remove color casts. It simply darkens the darkest pixels to pure black, lightens the lightest pixels to pure white, and redistributes all the other tonal values in between proportionally. This makes the highlights appear lighter and shadows appear darker. The Pseudocode demonstrating the process of Auto Contrast is as follows. The parameter ‘percent’ in the algorithm is the clipping percentage, and delta is a parameter to fine tune the enhanced image. For RGB images, the R, G and B channels are fed to the algorithm, while for NIR images, the input is just one channel.

c: DEEP PHOTO ENHANCER

A Deep Photo Enhancer based on unpaired learning proposed by Chen *et al.* [23] was applied for image enhancement. As shown in Figure 2, the method is based on the framework of two-way generative adversarial networks (GANs) and U-Net was augmented with global features to act as a generator in the GAN model. Wasserstein GAN (WGAN) was improved with an adaptive weighting scheme, resulting in faster and better training converges. In addition, individual batch normalization layers for generators in the two-way GANs was used to better adapt to the characteristics of their own inputs. For enhancing the images in this work,

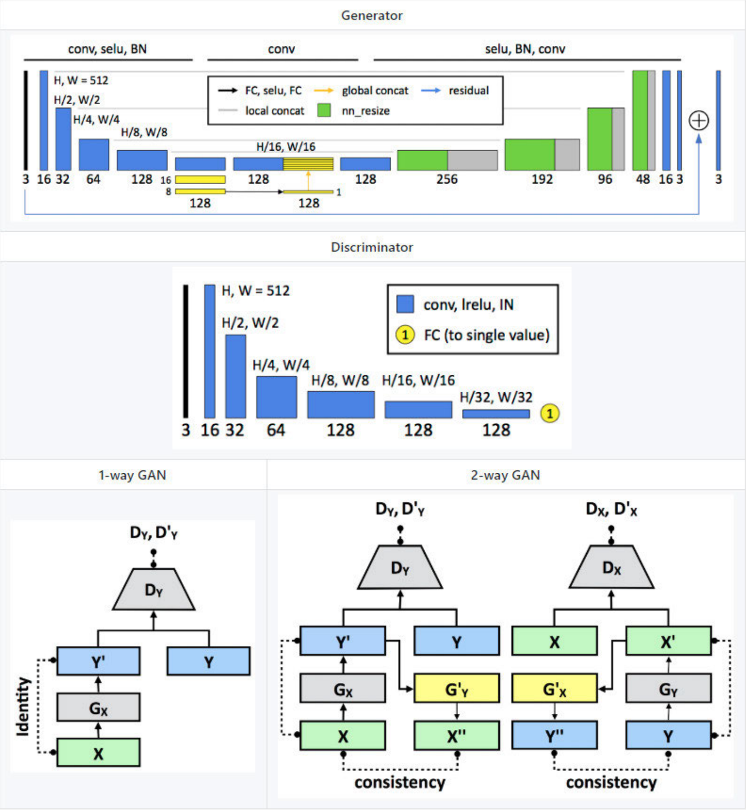


FIGURE 2. The network architectures of the proposed unpaired learning method for image enhancement [23].

the model provided by the authors was adopted, which was trained on photographer labels of the MIT-Adobe 5K dataset as well as an HDR dataset selected from Flickr images tagged with HDR.

2) INPUT REPRESENTATIONS

To facilitate greenness identification and plant classification, several frequently used color spaces and vegetation indices were involved to represent the input of model training. The color spaces of YCrCb and YCgCb have been proved to be effective for greenness segmentation by researchers [24], [25], therefore the raw images in these two color spaces were used as two input representations. The vegetation indices involved include NDI (Normalized Difference Index), NDVI (Normalized Difference Vegetation Index), ExG (Excess Green), ExR (Excess Red), ExGR (Excess Green minus Excess Red), CIVE (Color Index of Vegetation), VEG (Vegetative Index), and MExG (Modified Excess Green Index), COM1 (Combined Indices) and COM2, as calculated by Equations (1)-(10) [2], [16]. These indices were developed for vegetation extraction and are less sensitive to changes in field conditions.

$$\mathrm{NDI} = 128 * \left(\frac{G-R}{G+R}+1\right) \tag{1}$$

$$\mathrm{NDVI} = \frac{NIR+R}{NIR-R} \tag{2}$$

$$\mathrm{ExG} = \text{2G-R-B} \tag{3}$$

$$\mathrm{ExR} = \text{1.4R-G} \tag{4}$$

$$\mathrm{ExGR} = \text{ExG-ExR} \tag{5}$$

$$\mathrm{CIVE} = 0.441\mathrm{R} - 0.881\mathrm{G} + 0.385\mathrm{B} + 18.78745 \tag{6}$$

$$\mathrm{VEG} = \frac{G}{R^{0.667}B^{0.333}} \tag{7}$$

$$\mathrm{COM1} = \mathrm{ExG} + \mathrm{CIVE} + \mathrm{ExGR} + \mathrm{VEG} \tag{8}$$

$$\mathrm{MExG} = 1.262\mathrm{G} - 0.884\mathrm{R} - 0.311\mathrm{B} \tag{9}$$

$$\mathrm{COM2} = 0.36\mathrm{ExG} + 0.47\mathrm{CIVE} + 0.17\mathrm{VEG} \tag{10}$$

C. NETWORK ARCHITECTURE

An encoder-decoder network with atrous separable convolution, was investigated for semantic image segmentation in this work. As shown in Figure 3, the encoder module encodes multi-scale contextual information by applying depthwise atrous convolution at multiple scales. Atrous convolution is a powerful tool that allows to extract the features computed by deep convolutional neural networks at an arbitrary resolution. And depthwise separable convolution could drastically reduce computation complexity by factorizing a standard convolution into a depthwise convolution followed by a pointwise convolution. In the encoder-decoder network, the depthwise atrous convolution combines the atrous convolution and depthwise separable convolution to reduce computation complexity while maintaining similar (or better) performance.

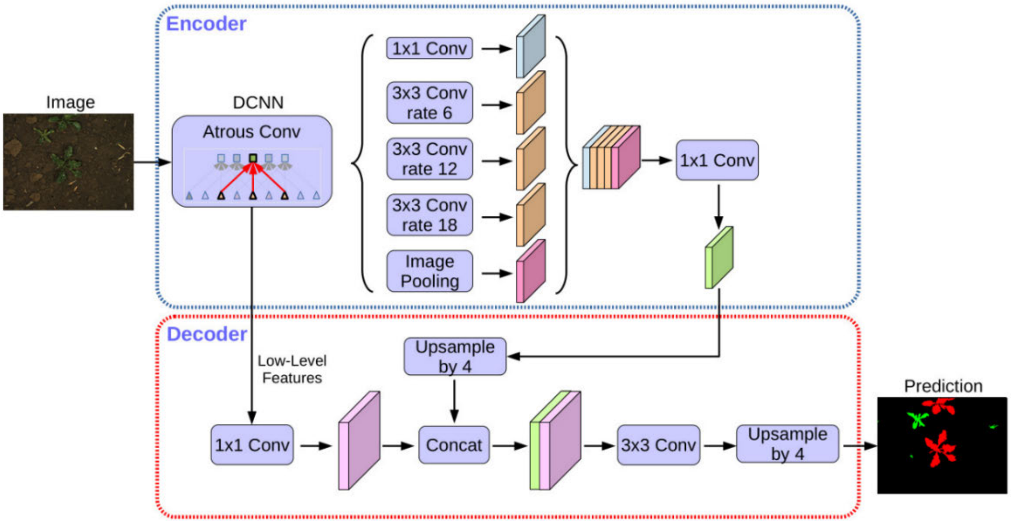


FIGURE 3. Encoder-decoder network with atrous separable convolution for semantic image segmentation [17].

A simple yet effective decoder first bilinearly upsamples the encoder features by a factor of 4 and then these features are concatenated with corresponding low-level features from the network backbone. After the concatenation, several 3×3 convolutions are applied to refine the features followed by another simple bilinear upsampling by a factor of 4. With these operations, the decoder module could refine the segmentation results along object boundaries. More details regarding the network can be found at reference [17].

The effectiveness of this encoder-decoder network has been approved on the benchmarks of PASCAL VOC 2012 and Cityscapes datasets, achieving the test set performance of 89.0% and 82.1% without any post-processing. The network was implemented relying on the Google TensorFlow library with the programming language Python 3.5.

D. TRAINING DEEP NETWORK

As we know, training a deep model from scratch is computationally expensive and requires large mounts of labelled data. However, in this work we only have 200 and 50 images for training for the sugar beet and oilseed, respectively, making it impossible to train the models from scratch. Hence, we utilized knowledge in other segmentation domain to solve our problem via transfer learning [26] in a low-cost way. Transfer learning for a convolutional neural network that consists of convolution base and fully-connected layers at the end, means to retrain the final layers of the network with new traing data based on a previously trained network. This process will slightly adjust the weights for final layers of the pre-trained model according to the input images. Therefore, in this work we trained the encoder-decoder network based on a pretrained model on PASCAL VOC 2012 dataset from VOC challenges with 11530 images. This leads to much less computation load and training data, while remaining comparable segmentation accuracy.

E. EVALUATION

Three classes were considered (soil, weed and crop) in this work. The performance of segmentation was firstly measured in terms of pixel intersection over union (IoU) averaged across the 3 classes. The mean intersection over union (MIoU) can be calculated using Equation (11).

$$\mathrm{IoU} = \frac{TP}{FP + TP + FN} \tag{11}$$

For automated weeding, it is more important to recognize the targeted object accurately, since the weeding actuator cannot perform pixel-wise operation. Therefore, an object-wise metric was also calculated to indicate the model's performance. We analyzed all objects with area bigger than 320 pixels, which was calculated by dividing the desired minimum object detection size of 1 cm^2 by the spatial resolution of 2 mm^2/px in the 1296×966 images.

III. RESULTS AND DISCUSSION

A. IMAGE ENHANCEMENT

Figure 4 compares the visual effect of different image enhancement methods. For raw RGB images (Figure 4a1) from sugar beet dataset, the brightness and contrast were low. After enhancement, the brightness and contrast of the images were improved significantly. The images enhanced by HE method (Figure 4a2) were with the highest brightness and contrast, but the color of the images was distorted and seemed vary unnatural. That may be caused by the irreducible singularities of the transformation between RGB and HSI spaces and the fact that HE is only performed on the intensity component [22]. For RGB images enhanced by Photoshop Auto Contrast (Figure 4a3), they looked bright and sharp, with visually appealing color. And for RGB images processed by Deep Photo Enhancer (Figure 4a4), the brightness and contrast were also improved, but the change was not that significant compared with the images processed by HE and

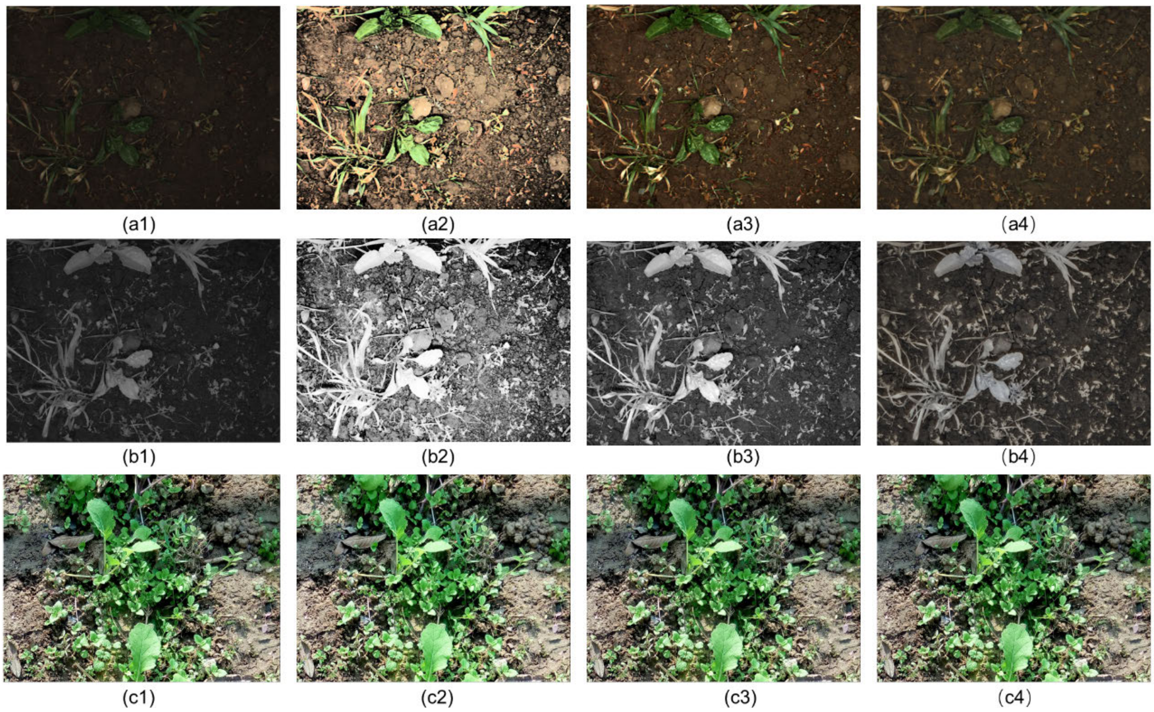


FIGURE 4. Raw and enhanced images. (a1-a4) Raw and enhanced RGB images of sugar beet by HE, Photoshop Auto Contrast and Deep Photo Enhancer; (b1-b4) Raw and enhanced NIR images of sugar beet by HE, Photoshop Auto Contrast and Deep Photo Enhancer; (c1-c2) Raw and enhanced RGB images of oilseed by HE, Photoshop Auto Contrast and Deep Photo Enhancer.

Photoshop Auto Contrast. Different from raw RGB images, the NIR images (Figure 4b1) from sugar beet dataset were with relatively high contrast, thanks to the ability of NIR camera to capture information in low illumination environment. After enhancement, the brightness and contrast of the NIR images were improved significantly. However, the images processed by Deep Photo Enhancer (Figure 4b4) seems to be with color and not grayscale image. After analysis it was found that the images consisted of three channels, which was caused by the three channels output of Deep Photo Enhancer. With respect to the oilseed dataset, the raw RGB images were with high brightness and contrast, but with the direct illumination of sunlight and some shadow regions. The difference between raw and enhanced RGB images was visually marginal.

Algorithm 1 Auto Contrast

```
Input: RGB or NIR images with a dimension of row*col
Output: Enhanced images
I←R*Parameter1+G*Parameter2+B*Parameter3
I←I/(max(I))
I_sort←sort(I)
I_out←I
I_min←I_sort(row*col*percent)
I_max←I_sort(row*col*percent)
for i←1 to row do
    for j←1 to col do
        if I(I, j) < I_min then
            I_out(I, j) = I_min
        else if I(I, j) < I_max then
            I_out(I, j) = 1
        else
            I_out(i, j) = (I(i, j)-I_min)*(1-I_min)/(I_max-
                I_min)+I_min
        end if
    end for
end for
k←(I_out + delta)/(I+delta)
I_out[][][1]←R*k
I_out[][][2]←G*k
I_out[][][3]←B*k
```

B. PERFORMANCE OF SEMANTIC SEGMENTATION

Table 1 illustrates the pixel-wise segmentation results with different input representations and image enhancement methods. Corresponding segmentation results are shown in Figure 5 and Figure 6. For the sugar beet dataset, transformation of color space did not improve the segmentation results, with RGB space yielding the highest MIoU value of 72.01%. Compared with color images in RGB, YCrCb and YCgCb spaces, the model trained with NIR images yielded much better result, with MIoU value of 79.28%. This is consistent with the fact shown in Figure 4 that NIR images are with higher brightness and contrast, facilitating the discrimination between sugar beet from weeds. With respect to the performance of vegetation indices, different indices and their combinations were compared. It should be noted that since the deep network adopted in this work

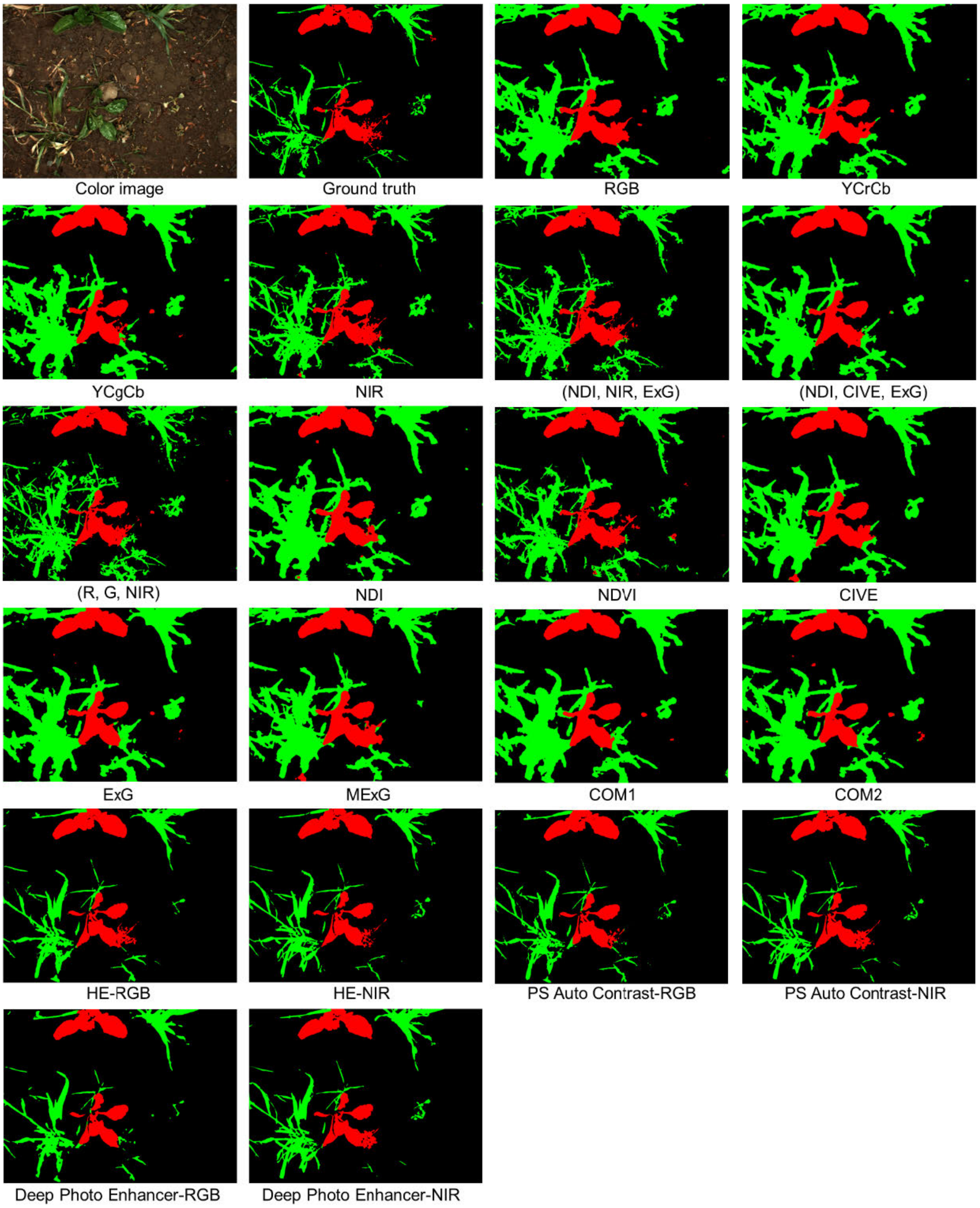


FIGURE 5. Visualization of pixel-wise segmentation results for sugar beet dataset. The caption of each sub-image, except 'Color image' and 'Ground truth', denotes the input for the deep network, with which the segmentation results were obtained, corresponding to Table 1.

only allows three channels as input, all representations listed in Table 1 are set as input with the format of (channel1, channel2, channel3). For grayscale images that only have one channel like NIR image, the three input channels are identical. From the MIoU values it can be observed that the input representations including NIR information (No. 5, 7 and 9) provided much better performance than those without NIR information. By contrast, other vegetation indices did not benefit the segmentation, with some even deteriorating the performance. This again confirms the effectiveness of NIR information for precise segmentation under weak lighting condition.

As stated previously, three image enhancement methods were applied to improve the image quality. And the enhanced images were then used for model training. After enhancement, the brightness and contrast of the images were

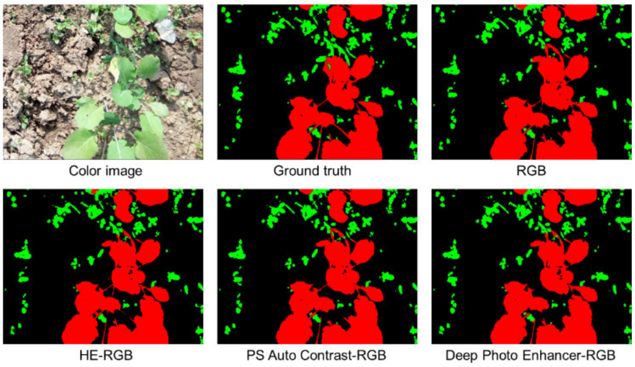


FIGURE 6. **Visualization of pixel-wise segmentation results for oilseed dataset. The caption of each sub-image, except 'Color image' and 'Ground truth', denotes the input for the deep network, with which the segmentation results were obtained, corresponding to Table 1.**

improved significantly, and the performance of segmentation models was boosted correspondingly, which were all superior than the result obtained by Milioto *et al.* [16] using 14 channels as input. For enhanced RGB images, comparison of MIoU values showed that the three image enhancement methods performed similarly, with HE being slightly inferior. For enhanced NIR images, model trained with images enhanced by PS Auto Contrast yielded the best results, followed by model trained with images enhanced by Deep Photo Enhancer. It can be also seen that models trained with enhanced images all yielded superior segmentation results than those trained without enhancement. This could be attributed to the low brightness and contrast of the RGB images that would result in missing boundary information of the objects. From Figure 5 it can be seen that the segmented objects (sugar beet and weed) by models trained with NIR information contain more abundant details along the object boundaries. By contrast, the boundaries of objects segmented by models trained without NIR information are much smoother, which seems like they are processed by dilation operation.

TABLE 1. **Pixel-wise test performance.**

DATASET	NO.	ENHANCEMENT	INPUT REPRESENTATION	MIOU [%]
SUGAR BEET	1	--	RGB	72.01
	2	--	YCRCB	71.25
	3	--	YCGCB	68.63
	4	--	NIR	79.28
	5	--	(NDI, NIR, EXG)	78.87
	6	--	(NDI, CIVE, EXG)	70.68
	7	--	(R, G, NIR)	78.82
	8	--	NDI	69.39
	9	--	NDVI	75.60
	10	--	CIVE	71.03
	11	--	EXG	69.35
	12	--	MEXG	71.38
	13	--	COM1	68.81
	14	--	COM2	69.13
	15	HE	RGB	84.53
	16	HE	NIR	85.80
	17	PS-AC	RGB	85.37
	18	PS-AC	NIR	87.13
	19	DPE	RGB	85.18
	20	DPE	NIR	86.64
	21	MILIOTO'S CNN (14 CHANNELS)		80.80
OILSEED	22	--	RGB	88.54
	23	HE	RGB	88.27
	24	PS-AC	RGB	88.62
	25	DPE	RGB	88.91

*HE: Histogram Equalization, PS-AC: PS Auto Contrast, DPE: Deep Photo Enhancer.

TABLE 2. **Pixel-wise test performance on augmented dataset.**

DATASET	NO.	ENHANCEMENT	INPUT	MIOU [%]
Augmented sugar beet dataset by changing *L* component	1	--	RGB	75.82
	2	HE	RGB	84.29
	3	PS-AC	RGB	85.36
	4	DPE	RGB	83.12
Augmented sugar beet dataset by gamma correction	5	--	RGB	76.30
	6	HE	RGB	82.93
	7	PS-AC	RGB	84.74
	8	DPE	RGB	82.05
Augmented oilseed dataset by changing *L* component	9	--	RGB	71.52
	10	HE	RGB	87.23
	11	PS-AC	RGB	86.08
	12	DPE	RGB	86.05
Augmented oilseed dataset by gamma correction	13	--	RGB	73.76
	14	HE	RGB	87.65
	15	PS-AC	RGB	86.44
	16	DPE	RGB	85.78

*HE: Histogram Equalization, PS-AC: PS Auto Contrast, DPE: Deep Photo Enhancer, CL: change L component of the image, GC: Gamma correction.

For our oilseed dataset captured under better lighting condition, the difference between raw and enhanced RGB images is visually marginal, and pixel-wise segmentation results also demonstrate that image enhancement did not change the segmentation performance (Figure 6). The deep network is capable of learning effective features hidden deep into the images with high brightness and contrast, regardless of shadow regions. And the three image enhancement methods did not cause any negative effect on segmentation. Comparing the segmentation results of the two datasets shows that the MIoU values of the oilseed dataset are all greater than those of the sugar beet dataset. This may be mainly caused by some mistaken labels in the sugar beet dataset (Figure 7). For those mistaken labelled objects shown in Figure 7, the deep model correctly segmented most of them. However, when calculating MIoU, they were not counted as true positives since they were different from the labels in ground truth images provided. In addition, the poor illumination in the sugar beet dataset may also has some effect. The proposed method could alleviate but not totally eliminate the negative effect of poor illumination. This can be also confirmed by the semantic segmentation results of the augmented datasets (Table 2), from which we can see that after changing the brightness and contrast of the images by altering *L* component and gamma values, the MIoU values for segmentation without image enhancement (No. 1, 5, 9 and 13) decreased significantly to less than 77%, while after enhancement, the MIoU values all increased to over 82%, comparable but slightly less than those shown in Table 1 (No. 15-20, and 23-25).

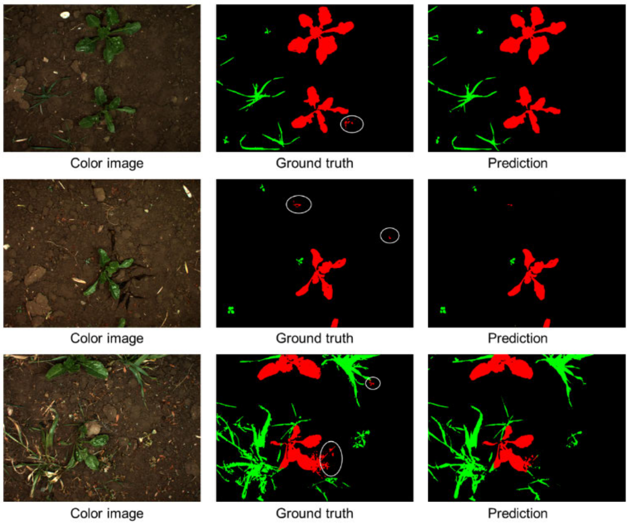


FIGURE 7. Examples of mistaken labels in ground truth images (white circles) for sugar beet dataset.

Generally it can be concluded that for the sugar beet dataset, image enhancement improved the image quality and thus the robustness of segmentation models in terms of different lighting conditions, and for the oil seed dataset, image enhancement did not degrade the performance of segmentation models. Another point is that the deep network applied in this work does not need large amount of data for model training thanks to the advantage of transfer learning. The segmentation model for the oilseed dataset trained 50 images and yielded MIoU values around 88%.

C. PERFORMANCE OF OBJECT-WISE SEGMENTATION

Table 3 illustrates the object-wise segmentation results with different input representations and image enhancement methods. The connected areas bigger than 320 pixels in ground truth and prediction images, which were treated as objects, were counted and the mean accuracy of the connected areas were calculated. For the sugar beet dataset, the mean accuracy of different input representations did not differ from each other obviously, with the mean accuracy ranging from 93.55% to 96.06%. However, after image enhancement, the mean accuracy all decreased, which was counter-intuitive since image enhancement improved the performance of pixel-wise segmentations. Analysis found that two reasons may lead to the result. The first is that some objects were wrongly labelled in the ground truth images, as shown in Figure 7. The second is that the mean accuracy was calculated as the ratio of true positives and all objects, which tended to be larger for coarser segmentations, since the objects in the coarser segmentations were larger than the objects in ground truth images and covered the latter ones more easily. For segmentations No. 4, 5, 7 and 21 in Table 1 and Table 3 whose pixel-wise accuracies were close, their object-wise accuracies were also very close to each, indicating that the image segmentation did not reduce the segmentation accuracy. This can be further confirmed by the object-wise segmentation results for the oilseed dataset.

TABLE 3. Object-wise test performance (HE: Histogram Equalization, PS-AC: PS Auto Contrast, DPE: Deep Photo Enhancer)

Dataset	No.	Enhancement	Input Representation	Mean Accuracy [%]
Sugar beet	1	--	RGB	96.06
	2	--	YCrCb	95.31
	3	--	YCgCb	94.36
	4	--	NIR	95.04
	5	--	(NDI, NIR, ExG)	95.31
	6	--	(NDI, CIVE, ExG)	94.22
	7	--	(R, G, NIR)	95.87
	8	--	NDI	94.06
	9	--	NDVI	94.48
	10	--	CIVE	94.86
	11	--	ExG	93.55
	12	--	MExG	93.86
	13	--	COM1	94.12
	14	--	COM2	94.39
	15	HE	RGB	92.75
	16	HE	NIR	89.69
	17	PS-AC	RGB	94.29
	18	PS-AC	NIR	91.01
	19	DPE	RGB	93.50
	20	DPE	NIR	91.37
	21	Milioto's CNN (14 channels)		94.74
Oilseed	22	--	RGB	95.76
	23	HE	RGB	94.80
	24	PS-AC	RGB	95.80
	25	DPE	RGB	96.12

*HE: Histogram Equalization, PS-AC: PS Auto Contrast, DPE: Deep Photo Enhancer.

TABLE 4. Runtime of classifiers.

Inputs	Preprocessing	Network	Total
Sugar beet	--	60 ms	60 ms
	HE: 10 ms		70 ms
	PS-AC: 6 ms		66 ms
	DPE: 12 ms		72 ms
Oilseed	--	190 ms	190 ms
	HE: 38 ms		228 ms
	PS-AC: 20 ms		210 ms
	DPE: 44 ms		234 ms

D. RUNTIME

The training time for 200 sugar beet images for 40k iterations is about 8 hours, and 50 oilseed images for 40000 iterations is about 3 hours, on a workstation with an Intel i7 CPU (256 GB RAM) and NVIDIA GTX1080Ti GPU (88 GB GPU memory). For implementing the classifier on our workstation, the runtime is shown in Table 4. We can see that the total inference time is less than 100 ms for a camera with a resolution of 1296 × 966 pixel, which meets the requirement of real-time processing. For a higher resolution image, it takes longer time to process.

IV. CONCLUSION

In this work, an encoder-decoder deep learning network was investigated for pixel-wise semantic segmentation of crop and weed. Different input representations including different color space transformations and color indices were compared to optimize the input of the network. Three image enhancement methods were investigated to improve model robustness against different lighting conditions. Results shows that color space transformation and vegetation indices without NIR information did not improve the segmentation results, while inclusion of NIR information significantly improved the segmentation accuracy, indicating the effectiveness of NIR information for precise segmentation under weak lighting condition. Image enhancement improved the image quality and thus the robustness of segmentation models against different lighting conditions. Another point is that the deep network applied in this work does not need large amount of data for model training. Future work will be focused on model compression, through which the trained model can be applied on mobile platforms with less computing capability.

REFERENCES

[1] W. S. Lee, V. Alchanatis, C. Yang, M. Hirafuji, D. Moshou, and C. Li, "Sensing technologies for precision specialty crop production," *Comput. Electron. Agricult.*, vol. 74, no. 1, pp. 2–33, Oct. 2010.

[2] 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.

[3] J. Gao, W. Liao, D. Nuyttens, P. Lootens, J. Vangeyte, A. Pižurica, Y. He, and J. G. Pieters, "Fusion of pixel and object-based features for weed mapping using unmanned aerial vehicle imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 67, pp. 43–53, May 2018.

[4] 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.

[5] C. McCarthy, S. Rees, and C. Baillie, "Machine vision-based weed spot spraying: A review and where next for sugarcane?" in *Proc. 32nd Annu. Conf. Austral. Soc. Sugar Cane Technol. (ASSCT)*, 2010, pp. 424–432.

[6] 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.

[7] P. Lottes, J. Behley, A. Milioto, and C. Stachniss, "Fully convolutional networks with sequential information for robust crop and weed detection in precision farming," 2018, *arXiv:1806.03412*. [Online]. Available: http://arxiv.org/abs/1806.03412

[8] A. H. Saeedeh Taghadomi-Saberi, "Improving field management by machine vision—A review," in *Agricult. Eng. Int., CIGR J.*, vol. 17, no. 3, pp. 1–20, 2015.

[9] M. Weis and M. Sökefeld, "Detection and identification of weeds," in *Precision Crop Protection—The Challenge and Use of Heterogeneity*. Dordrecht, The Netherlands: Springer, 2010, pp. 119–134.

[10] C. Lin, "A support vector machine embedded weed identification system," M.S. thesis, Univ. Illinois Urbana-Champaign, Champaign, IL, USA, 2009.

[11] F. López-Granados, "Weed detection for site-specific weed management: Mapping and real-time approaches," *Weed Res.*, vol. 51, no. 1, pp. 1–11, Feb. 2011.

[12] J. Romeo, G. Pajares, M. Montalvo, J. M. Guerrero, M. Guijarro, and J. M. de la Cruz, "A new expert system for greenness identification in agricultural images," *Expert Syst. Appl.*, vol. 40, no. 6, pp. 2275–2286, May 2013.

[13] D. L. Shaner and H. J. Beckie, "The future for weed control and technology," *Pest Manage. Sci.*, vol. 70, no. 9, pp. 1329–1339, Sep. 2014.

[14] J. Tang, D. Wang, Z. Zhang, L. He, J. Xin, and Y. Xu, "Weed identification based on K-means feature learning combined with convolutional neural network," *Comput. Electron. Agricult.*, vol. 135, pp. 63–70, Apr. 2017.

[15] M. Dyrmann, R. N. Jørgensen, and H. S. Midtiby, "RoboWeedSupport-Detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network," *Adv. Animal Biosci.*, vol. 8, no. 2, pp. 842–847, Jul. 2017.

[16] A. Milioto, P. Lottes, and C. Stachniss, "Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 2229–2235.

[17] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.

[18] N. Chebrolu, P. Lottes, A. Schaefer, W. Winterhalter, W. Burgard, and C. Stachniss, "Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields," *Int. J. Robot. Res.*, vol. 36, no. 10, pp. 1045–1052, Sep. 2017.

[19] S. K. Naik and C. A. Murthy, "Hue-preserving color image enhancement without gamut problem," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1591–1598, Dec. 2003.

[20] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2008.

[21] M.-S. Shyu and J.-J. Leou, "A genetic algorithm approach to color image enhancement," *Pattern Recognit.*, vol. 31, no. 7, pp. 871–880, Jul. 1998.

[22] N. Bassiou and C. Kotropoulos, "Color image histogram equalization by absolute discounting back-off," *Comput. Vis. Image Understand.*, vol. 107, nos. 1–2, pp. 108–122, Jul. 2007.

[23] Y.-S. Chen, Y.-C. Wang, M.-H. Kao, and Y.-Y. Chuang, "Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6306–6314.

[24] S. Sabzi, Y. Abbaspour-Gilandeh, and G. García-Mateos, "A fast and accurate expert system for weed identification in potato crops using Metaheuristic algorithms," *Comput. Ind.*, vol. 98, pp. 80–89, Jun. 2018.

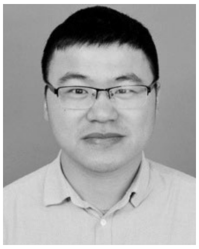
[25] J.-L. Tang, X.-Q. Chen, R.-H. Miao, and D. Wang, "Weed detection using image processing under different illumination for site-specific areas spraying," *Comput. Electron. Agricult.*, vol. 122, pp. 103–111, Mar. 2016.

[26] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, p. 9, Dec. 2016.

AICHEN WANG received the B.S. degree from the College of Engineering, Huazhong Agriculture University (HZAU), Wuhan, China, in 2011, and the Ph.D. degree in agricultural engineering from Zhejiang University, in 2017. He joined Jiangsu University, in 2017, where he is currently an Assistant Researcher. His current research interests include agricultural information intelligent perception, agricultural robot, and unmanned agricultural machinery.

YIFEI XU received the B.S. degree in computer science and technology from the South China University of Technology (SCUT), Guangzhou, China, in 2011, and the Ph.D. degree from the School of Computer Science, Zhejiang University, in 2017. From 2017, he was with the School of Software, Xi'an Jiaotong University as an Assistant Professor. His current research interests include deep learning, hyperspectral image processing, and image semantic segmentation.

XINHUA WEI received the B.S. degree from the School of Mechanical and Electrical Engineering, Shandong Agricultural University (SAU), Tai'an, China, in 1994, the M.S. degree from the College of Engineering, China Agricultural University, Beijing, China, in 1997, and the Ph.D. degree in measurement technology and automatic equipment from Southeast University, in 2008. From March 1997 to March 2004, he was with the School of Mechanical and Electrical Engineering, SAU as an Associate Professor. He joined Jiangsu University, in 2008, where he is currently a Professor. His current research interests include agricultural information intelligent perception, agricultural robot, and unmanned agricultural machinery.



BINGBO CUI received the B.S. and M.S. degrees in measurement, control, and instrument from the Chongqing University of technology, Chongqing, China, in 2005 and 2012, respectively, and the Ph.D. degree in precision instrument and machinery from Southeast University, Nanjing, China, in 2017. Since November 2017, he has been an Assistant Professor with the School of Agricultural Equipment Engineering, Jiangsu University, Zhenjiang, China. He was a recipient of the National Excellent Doctoral Dissertation Award Nomination in measurement, control and instrument in 2019. His research interests include inertial navigation, nonlinear filtering, integrated navigation, and their application in autonomous vehicle control.