IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 August 2023, accepted 1 September 2023, date of publication 5 September 2023, date of current version 13 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3312191

RESEARCH ARTICLE

Weed Density Detection Method Based on a High Weed Pressure Dataset and Improved PSP Net

XIA LI, FANGTAO DUAN, MENGCHAO HU, JIAWEI HUA, AND XIWANG DU

Tianjin Key Laboratory for Advanced Mechatronic System Design and Intelligent Control, School of Mechanical Engineering, Tianjin University of Technology, Tianjin 300384, China

National Demonstration Center for Experimental Mechanical and Electrical Engineering Education, Tianjin University of Technology, Tianjin 300384, China

Corresponding author: Xia Li (lixia0415@email.tjut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 32171902 and Grant 32060417.

ABSTRACT Large-scale spraying on farmland is one of the most widely used weeding methods. Accurate weed density detection is of great significance for improving pesticide utilization and reducing environmental pollution. The purpose of this paper is to combine traditional image processing technology with deep learning technology to study semi-supervised annotation of the weed dataset and weed density detection method in a high-stress weed environment and provide a prescription map to guide variable spraying weeding operation. First, this paper uses a crop dataset to train U-Net to achieve crop segmentation and uses the color index and Otsu threshold segmentation algorithm to achieve vegetation segmentation. Then, weed segmentation is achieved by removing crop areas from vegetation segmentation results, and the segmentation results are made into a weed dataset. The improved PSP Net is trained using this dataset and weed segmentation is performed. The ratio of the number of weed pixels to the total number of pixels in the region is calculated for the obtained segmented image by region to measure the weed density. Finally, prescription maps representing different treatment intensities were generated based on the weed density threshold. Results indicate that EXG color index outperforms the other three indices for weed annotation. Compared with the original model, the MIoU, mPA, and Accuracy of the improved PSP Net model are increased by 2.15%, 0.92%, and 1.16%, respectively, and the model reasoning speed is increased by 6.9 times. The coefficient of determination between the predicted results of weed density and the manually labeled true values is 0.83, with a root mean square error of 0.17. The accuracy of the prescription map is 78%. The method proposed in this paper can effectively detect weed density in high-pressure weed environments and provide accurate prescription maps for variable spraying weed control.

INDEX TERMS Data annotation, image processing, prescription maps, semantic segmentation, weed density detection.

I. INTRODUCTION

Weed control is a critical aspect of agricultural production and plays a vital role in improving crop yield and quality [1]. How to remove weeds effectively has always been a challenging task for agricultural workers [2]. In modern agriculture, chemical herbicides are widely used to spray large areas of farmland to achieve weed control [3]. This indiscriminate treatment of the whole farmland does not conform to the characteristics of the patchy distribution of weeds in the actual large field environment. Spraying herbicides in weed-free areas results in the waste of herbicides and increases the cost of weeding [4]. Additionally, overuse of herbicides can hurt the farmland biodiversity and food safety [5]. Therefore, selective spraying equipment for reducing the use of herbicides has been widely studied [6]. This kind of equipment must be based on the accurate perception of target information to achieve unmanned precision weeding, and one of the key problems to be solved is to realize the accurate detection of weed density in the field [7], [8], [9].

In the latest study, convolutional neural networks show good performance in machine vision applications [10],

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

Among them, the semantic segmentation method based on supervised learning has been widely used in weed detection, and the commonly used structures include U-Net, Deeplabv3, PSP Net, Transformer, and so on [11]. This method extracts image features using deep convolutional neural network. It then separates weeds and backgrounds using various techniques to obtain images containing weed masks. The supervised deep learning method requires a large amount of image data and corresponding annotation images to ensure the accuracy of the model for weed segmentation [12]. However, when labeling pixel-level annotation images, annotators need to observe each pixel carefully and label it according to its category [13]. This labeling method requires a high degree of manual intervention and domain expertise, thus requiring a considerable amount of time and effort in the annotation process. Especially in a real field environment with high weed pressure, a single image may contain crops as well as a large number of weeds, with the leaves blocking each other. The boundary between the weeds and the background is very blurred. This makes the data annotation task more challenging [14]. How to reduce or avoid the labeling costs brought by the supervised learning method becomes a key issue in achieving weed density detection [15], [16].

Several studies have proposed viable solutions. Given the feature that crop planting is arranged according to the regular row spacing, some solutions use crop row detection as the basis for more precise segmentation of weeds within rows. Lottes and Stachniss assumed that the green vegetation between rows was weeds, and combined an image-based classifier and a geometric classifier in a probabilistic manner within crop rows to achieve crop and weed classification after training with a small amount of annotation data [17]. García-Santillán and Pajares utilized Mahalanobis distance to measure the image spectral similarity and combined it with the Bayesian classification method to identify and classify weeds in field crop rows [18]. Xu et al. improved the extraction stage of crop row lines and proposed a weed and crop corner extraction method based on absolute characteristic corner point (AFCP). By combining absolute corners to identify the position of weeds, this method avoided direct detection of crop rows, reduced the amount of calculation, and improved the processing speed [19]. The methods based on crop row detection rely heavily on regular seeding and continuous seedling rows. With the deterioration of these conditions, the detection accuracy of weed density decreased [20]. Given the characteristics of monoculture planting and easy labeling in the field, some schemes combine the deep learning method with traditional image processing technology. The study by Zou et al. adopted excessive green minus excessive red index to achieve vegetation segmentation for UAV images [21]. They adopted the modified U-Net to achieve crop segmentation in UAV image, removed crops from the vegetation segmentation result, and then the remaining vegetation was regarded as weeds. Jin et al. used the slightly modified excess green index to realize vegetation segmentation, used YOLO V3 to detect crop areas, and determined that all vegetation outside the crop areas was weeds [22]. This method avoids the direct annotation of weed data sets and shortens the annotation time. However, the segmentation of vegetation, crops, and weeds in three steps leads to long processing time and insufficient real-time performance of the model, which is not conducive to practical application. Based on the above studies, some methods are proposed to achieve vegetation segmentation by unsupervised means and further classify the results of vegetation segmentation to achieve weed segmentation. Shorewala et al. cut the vegetation segmentation images after removing the background into pixel blocks of equal size, and then used the fine-tuned CNN classifier to classify the blocks with plant pixels to identify weed areas and calculate weed density [20]. Mishra et al. trained the deep convolutional neural network (DCNN) model as training data after classifying the segmented pixel blocks to achieve direct weed classification [23]. This method simplifies the weed segmentation process. However, in the classification process, pixel blocks with fewer vegetation pixels are omitted, and the entire pixel block is classified into a certain category based on the proportion of weed pixels and crop pixels in the pixel block. This processing method leads to the pixel loss or misclassification of some weeds, which affects the final calculation result of weed density.

The aforementioned research has promoted the development of weed density detection methods. However, most studies have been conducted in environments with low weed pressure, and relatively fewer studies have been dedicated to the investigation of weed density detection methods in high-pressure weed environments. Furthermore, these studies often used indirect and multi-stage detection methods to avoid the difficulty of data labeling, which led to poor real-time performance of the model. To address the issue of weed density detection in high weed pressure environments, this paper proposes a weed density detection method based on a high weed pressure dataset and an improved PSP Net, combining traditional image processing methods with deep learning algorithms. The method starts from the labeling of the dataset and the direct detection of weeds.

The contribution of this research can be summarized as follows:

1) Introducing a semi-supervised data labeling method. This method utilizes a U-Net model trained on crop datasets and the Otsu threshold segmentation algorithm based on color indices to convert direct annotations of weed areas into indirect annotations. This approach addresses the laborious and time-consuming task of manually annotating weed areas in images with high weed pressure.;

2) To modify the PSP Net semantic segmentation model to make it suitable for weed segmentation models in high weed pressure environments. To improve the real-time performance of the model, we introduced the lightweight MobilenetV3 as the backbone feature extraction network. To reduce the loss of high-frequency details in the reconstructed and enlarged

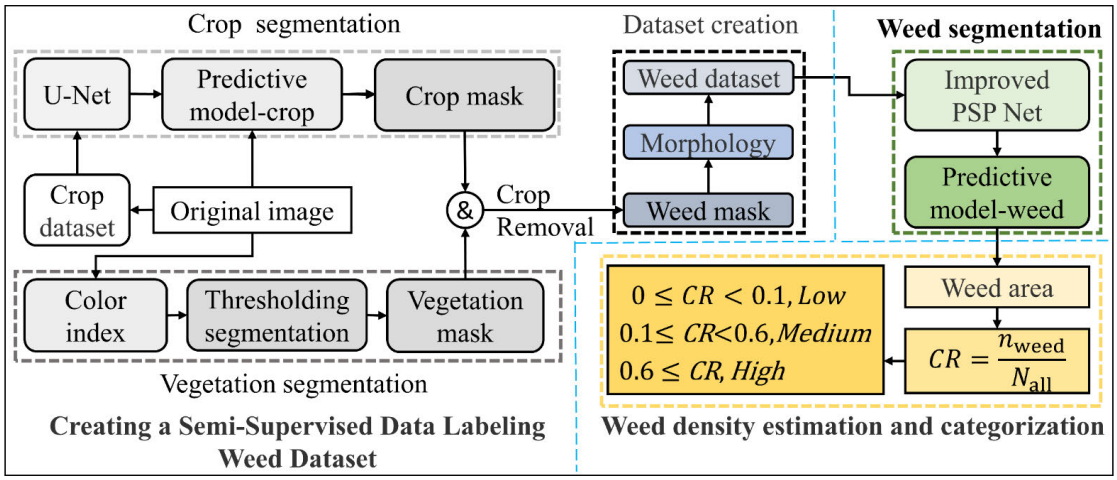

FIGURE 1. Flowchart of the proposed method.

feature maps after the pyramid feature extraction module, we used the bicubic interpolation method, which can provide smoother and more detailed interpolation results. After that, we introduced the Convolutional Block Attention Module to enhance the model's attention to key features and improve its robustness. Combined with a reasonable training strategy, PSP Net achieved significant performance improvement on the high weed pressure dataset.;

3) By applying the sliding window scanning algorithm and pixel counting algorithm, we achieved weed density calculation and grading by region, and generated a prescription map.

II. METHODOLOGY

As shown in Fig. 1, the weed density detection method proposed in this paper consists of three parts. The first part is the semi-supervised annotation of the high-pressure weed dataset. The second part is to use the improved PSP Net to achieve weed detection. The third part is the calculation of weed density and grading treatment. The main steps include: First, the U-Net semantic segmentation model is trained using the annotation crop dataset to achieve crop segmentation. Second, traditional image processing algorithms such as color index and threshold segmentation are used to achieve vegetation segmentation. Then, according to the crop segmentation results, the crop area is excluded from the obtained vegetation segmentation results and the remaining green vegetation is classified as weeds. After that, the weed mask images are processed by morphological operations such as dilation and erosion. The binary weed mask image is transformed into the annotation image of the weed data set in this paper. There are two categories of the dataset: crops and non-green parts (such as soil) in one, and weeds in the other. Then, the dataset is used to train the improved PSP Net semantic segmentation model to build a binary classification model for weed detection. Finally, a weed detection output image is divided into several equal-sized regions. The number of weed pixels and the total number of pixels are calculated in each region, and the ratio of the two is used as the index of weed density in the region. According to this index, the density level of the area is divided to determine the extent of weed treatment in the area.

A. SOURCE OF DATA

This paper adopts the open-source cauliflower field images (CWF-788) as the dataset, which were taken in two fields in Tongzhou, Beijing [24]. The dataset includes 788 images of cauliflower seedlings at 5 and 7 weeks after transplanting and the corresponding pixel-level labels. And the data set is divided into three parts: training set (400 images), test set (300 images), and verification set (88 images), which are independent of each other and have no duplicate image samples. Because of the low environment temperature, the plants grew slowly at the seedling stage, and the growers did not carry out weeding operations in the two fields. This provides a complex weed environment for later data collection and also facilitates the research of high-pressure weed identification. The sample dataset is illustrated in Fig. 2, which contains four images with different weed densities and the corresponding cauliflower annotation images.

B. CROP SEGMENTATION

A field image comprises crop plants, weed plants, and non-vegetation areas like soil. The purpose of this section is to segment the crop pixels in the image. Therefore, the pre-trained U-Net is trained by using the annotation crop dataset in this paper. U-Net [25] is a deep learning network structure commonly used for image segmentation. It adopts down-sampling and up-sampling mechanisms to improve the resolution and optimizes the transmission of feature signals through skip connection. The network uses low-level and high-level feature information and has efficient segmentation performance and feature extraction capability. Figure 3 shows the training loss curve, verification loss curve, and MIoU Variation curve in the training process. The trained U-Net has made predictions on the test set, and the obtained MIoU, mPA, and Accuracy are 97.86%, 98.91%, and 99.62%,

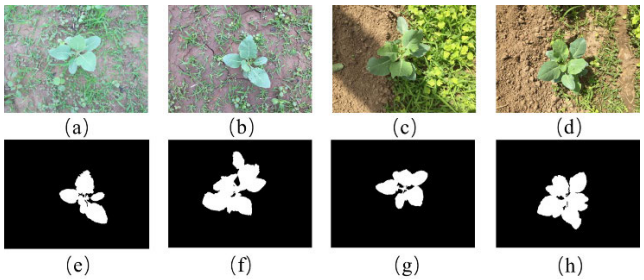

FIGURE 2. Sample dataset: (a-d) original image and (e-h) label image.

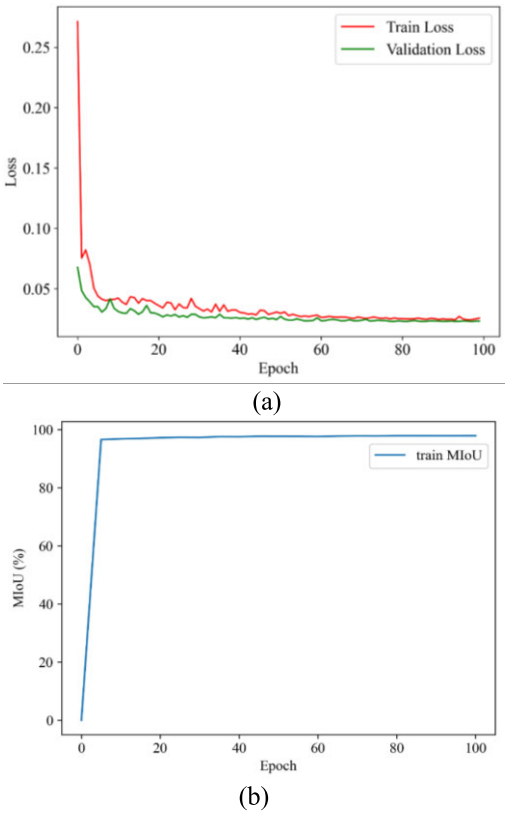

FIGURE 3. U-Net training: (a) training and validation loss curves and (b) MIoU variation curve.

respectively. By analyzing the images and data, it can be seen that the model has good convergence in the training process, and can accurately segment the crop region in the image in the test.

C. VEGETATION SEGMENTATION

The leaves of plants usually appear green due to the presence of chlorophyll. This feature is often used in computer vision to distinguish between backgrounds such as plants and soil [26]. To achieve stable and efficient vegetation segmentation, this paper adopts the threshold segmentation method based on color index. The method first extracts useful color components from color images and calculates color indices based on these color components to detect and emphasize vegetation areas. The color index image is thresholded to obtain a binary image, in which vegetation and background areas are separated.

1) COLOR INDEX

Excess Green Index [27] (ExG) is a simple and fast index that can be applied to various types of images in the field of machine vision and agriculture. It is calculated by measuring the difference of reflectivity between green and red light wavelengths in the image, which can accurately detect and separate the green area and produce a near-binary image. The simplified equation of ExG is as follows:

$$ExG = \frac{2G - R - B}{G + R + B}, \tag{1}$$

where R, G, and B represent the pixel values of the three color channels (red, green, and blue) of the image.

Color Index of Vegetation Extraction (CIVE) mainly detects and extracts vegetation area through the linear combination of red and green channels [28]. It can effectively avoid the influence of uneven brightness and improve the detection accuracy of vegetation area. CIVE is calculated by the following formula:

$$CIVE = 0.441R - 0.811G + 0.385B + 18.78745. \tag{2}$$

Modified Excess Green Index (MExG) improves the adaptability to brightness and color by normalizing and optimizing the original EXG index [29]. It can effectively reduce the impact of brightness change and background noise on image segmentation and improve the quality of image segmentation. MExG is calculated by the following formula:

$$MExG = 1.262G - 0.884R - 0.311B. \tag{3}$$

Excess Green minus Excess Red Index [30] (ExGR) is mainly calculated by combining the red, green, and blue color channels, and the difference between the ExG and Excess Red [31] (ExR) indexes is also calculated. It can not only effectively detect the vegetation area, but also accurately extract the leaf information of vegetation, so as to obtain the growth status, type, and other information of vegetation, which greatly improves the accuracy of vegetation analysis. The simplified equation of ExGR is as follows:

$$ExR = \frac{1.4R - G}{G + R + B}, \tag{4}$$

$$ExGR = ExG - ExR. \tag{5}$$

2) THRESHOLD SEGMENTATION

Threshold segmentation is one of the fundamental tasks in image processing. This paper adopts the Otsu threshold segmentation method [32] to binarize the single-channel image obtained by color index calculation. The objective of this method is to select the optimal threshold so that the image can be divided into two parts, namely vegetation, and non-vegetation, and achieve the maximum variance between the pixel gray values of the two parts. Then, by enumerating different thresholds, the method compares the sum of the gray

value variances of the two parts after image segmentation to find the optimal threshold. By this method, the difference between vegetation pixel values and non-vegetation pixel values can be maximized so as to achieve effective binary segmentation. Finally, the obtained binary image can effectively reflect the vegetation distribution in the image.

D. DATASET CREATION

After the completion of crop segmentation and vegetation segmentation, the crop and vegetation mask images of the original image can be obtained. By performing the bitwise AND operation on these two mask images, a preliminary weed mask image can be obtained. To highlight weeds and improve the quality of annotation, it is necessary to use morphological operations to process weed mask images. Morphological operations are widely used in image processing for tasks like noise elimination, hole filling, and edge smoothing. Specifically, these tasks are accomplished through techniques such as dilation, erosion, opening, and closing. The morphologically processed weed mask image can be used as the annotation image to form the weed dataset together with the original image.

This section provides an overview of the process of creating the weed dataset and the associated image changes (Fig. 4).

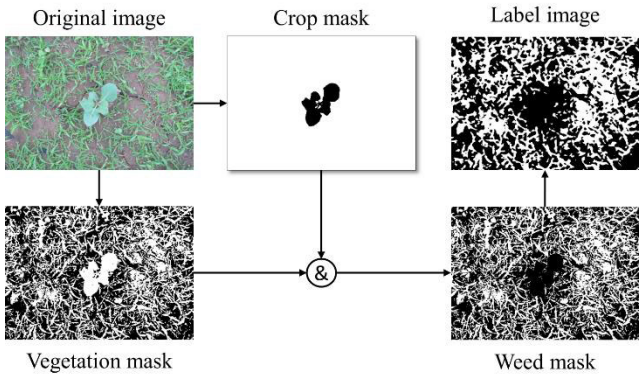

FIGURE 4. Image changes during the creation of the weed dataset.

E. WEED SEGMENTATION USING THE IMPROVED PSP NET

In this paper, the semi-supervised weed annotation process in a high-stress weed environment is proposed and a weed dataset is created by using this process. Based on this dataset, this section improves and trains the PSP Net model [33], and then obtains a weed segmentation model. The network structure of the improved PSP Net model is shown in Fig. 5. It is described in detail below.

1) MOBILENETV3-LARGE BACKBONE FEATURE EXTRACTION NETWORK

PSP Net uses the ResNet series as the backbone feature extraction network. In order to improve the real-time performance of the model, MobileNetV3-Large [34] is introduced in this paper to replace the original backbone feature extraction network. In the MobileNetV3-Large backbone feature extraction network, multiple layers of lightweight convolution and activation functions are adopted, such as Squeeze-and-Excitation (SE) module, bottleneck module, and depth-separable convolution. Among them, the Squeeze-and-Excitation module can learn the relationship between channels and further improve the representational ability of features; the bottleneck module can enhance the non-linear feature expression ability while maintaining low computational complexity; Depth-separable convolution can reduce the amount of calculation and improve the efficiency of the model. (MobileNetV3 will be used to refer to MobileNetV3-Large in the following description.)

2) PYRAMID SPATIAL POOLING MODULE

In traditional convolutional neural networks, pooling layers are often used to reduce the size of feature maps and achieve the scaling of spatial information. However, this can lead to the loss of some details and affects the performance of the model. To address this issue, PSP Net proposes the pyramid pooling module to pool the features from different scales and different sizes of receptive field and splice the pooling results of different scales to form a multi-scale feature pyramid. Nonlinear operation is added to the subsequent convolutional layers to further enhance the feature expression ability [33].

3) BICUBIC INTERPOLATION

In the pyramid feature extraction module of PSP Net, the Pyramid Spatial Pooling (PSP) is used to extract the feature image information of different scales. This paper uses the bicubic interpolation method to reconstruct and enlarge the feature map to avoid the loss of high-frequency details after the pooling of the last layer. Bicubic interpolation provides more precise and more detailed interpolation results than the traditional bilinear interpolation, has better performance in anti-aliasing and so on [35]. Through this method, the high-frequency details in the feature image can be better maintained in the process of high-quality image magnification and restoration, while better improving model accuracy and effects.

4) CONVOLUTIONAL BLOCK ATTENTION MODULE

Convolutional Block Attention Module (CBAM) is an attention module used to enhance the perception of Convolutional Neural Network (CNN) models [36]. By introducing channel attention mechanisms and spatial attention mechanisms, CBAM establishes a mapping relationship between channel features and spatial features, thus improving the representational capacity and performance of the network. In this paper, the CBAM attention module is added after the pyramid pooling module in order to enhance the representational capacity and improve the robustness of the model before global pooling.

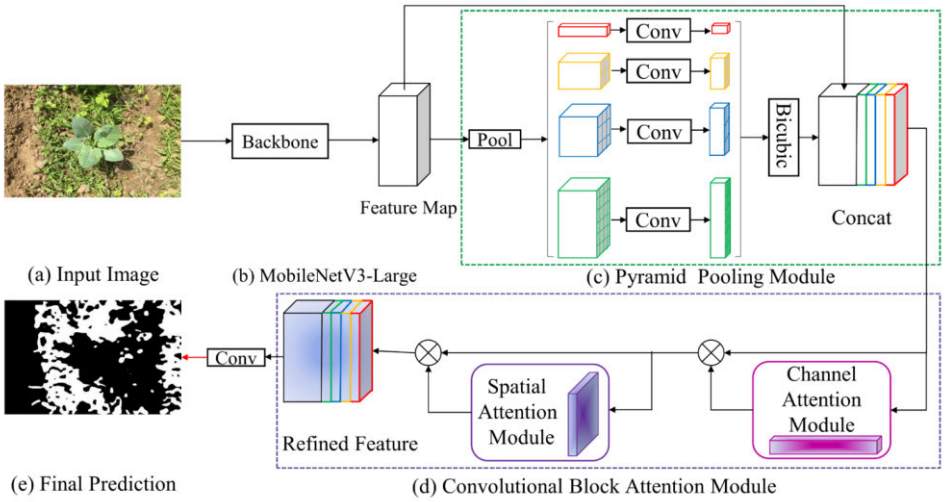

FIGURE 5. Overview of the improved PSP Net network architecture (a) Input image. (b) MobileNetV3-Large backbone feature extraction network to extract feature layers. (c) The feature pyramid pooling module obtains feature regions of different sizes and performs average pooling and bi-cubic linear interpolation operations to complete feature fusion. (d) CBAM attention mechanism enhanced feature layer. (e) The classifier classifies each pixel point and generates weed segmentation results.

5) EXPERIMENTAL PLATFORM

The main hardware configuration of the experimental platform is an Intel Core i5-8250U CPU with a main frequency of 1.6GHz, 16GB RAM, and NVIDIA GeForce MX150 GPU with 4G video memory. The running environment is Windows 10 operating system, Python version 3.7.12 language, Pytorch 1.8 deep learning framework with CUDA 11.1 parallel computing architecture, cuDNN 8.0.5 deep neural network GPU acceleration library, and OpenCV 4.5.4 computer vision library.

6) TRAINING STYATEGIES

The transfer learning method is used to train the model. The backbone feature extraction network is pre-trained by ImageNet large dataset to achieve a relatively optimal parameter space for the feature extraction network. Subsequently, the entire model was fine-tuned using the weed dataset created in Section II-D to improve the accuracy of its predicted segmentation in specific agricultural scenarios.

To avoid the issues of model overfitting and underfitting that may arise from a small sample size, this paper uses the online augmentation method to perform data augmentation operations on each epoch in the data input pipeline, including random cropping, resizing, and flipping. Taking into account computational resources and learning efficiency, this paper sets the batch size of training images to 2, the number of epochs to 100, and the number of images in the training set to 400. Therefore, during the training process, a maximum of 40,000 different images can be generated through transformations.

To reduce memory usage and improve computing efficiency, the Adam optimizer is adopted with an initial learning rate of 5e-4 and a momentum factor of 0.9, and CosineAnnealingLR is selected as the learning rate attenuation strategy. Focal-Dice Loss is used to reduce the sensitivity of the model to noise and background region, pay more attention to the division and accurate prediction of the target region, and improve the robustness and stability of image segmentation.

7) EVALUATION INDICATORS

In order to quantitatively analyze the performance of the model and test the effectiveness of the model improvement, this paper uses standardized evaluation indicators of semantic segmentation to measure the performance of the model in the verification set and test set. It mainly includes mean Pixel Accuracy (mPA), mean Intersection over Union (MIoU), and Accuracy.

MIoU is used to evaluate the segmentation accuracy of the overall target region of the model. It is the ratio of the intersection and union of the predicted area and the real area at the pixel level, reflecting the degree of overlap between them. It is defined as follows:

$$MIoU = \frac{1}{2}\left(\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FN + FP}\right), \quad (6)$$

where TP is the number of correctly classified weed pixels, TN is the number of correctly classified background pixels, FP is the number of background pixels incorrectly treated as weed pixels, and FN is the number of weed pixels incorrectly treated as background pixels.

MPA represents the mean pixel recognition accuracy of the two predicted categories, which is the main indicator for evaluating the pixel prediction accuracy of the model. It represents the average proportion of correctly predicted pixels in all categories and is defined as follows:

$$mPA = \frac{1}{2}\left(\frac{TP}{TP + FP} + \frac{TN}{TN + FN}\right). \quad (7)$$

Accuracy represents the proportion of the number of correctly predicted pixels to the total number of pixels in all categories and is defined as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}. \tag{8}$$

The calculation elements of the above evaluation indicators are calculated by the statistical results of the confusion matrix of all pixels in the test image dataset, which has statistical significance under the condition of ensuring the test sample size and can basically exclude the influence of random factors. In addition, the segmentation speed of the model is evaluated using the frame per second (FPS).

F. WEED DENSITY ESTIMATION AND CATEGORIZATION

1) CALCULATION OF WEED DENSITY

According to the output results of the semantic segmentation model of weeds, the weed-infested areas in the image can be identified. In this paper, the sliding window scanning algorithm is used to calculate the weed density in the image area covered by the window. In addition, the clustering rate (CR) is used to quantify weed density [19], which was defined as follows:

$$CR = \frac{n_{weed}}{N_{all}}, \tag{9}$$

where n_{weed} denotes the total number of weed plant pixels in the sliding window coverage area, while N_{all} represents the total number of pixels in the sliding window coverage area.

2) PRESCRIPTION MAP GENERATION

The prescription map is generated according to the calculation results of weed density, which can be used to guide the work of variable spraying or variable flame weeding. This paper sets two threshold values l_1 and l_2 based on the different weed densities [37]. When the weed density is less than l_1, the area will remain untreated. When the weed density is greater than l_1 but less than l_2, the area is treated with moderate intensity. High-intensity processing will be implemented when the weed density is greater than l_2. The accuracy of prescription graph generation can be calculated by the real prescription map and the predicted prescription map [38]. The accuracy is defined as follows:

$$accuracy = \frac{N_t}{N_a}, \tag{10}$$

where N_t represents the number of correctly predicted sliding window regions in the prescription map and N_a represents the total number of sliding window regions in the prescription map.

III. RESULTS AND DISCUSSION

A. COMPARISON AND ANALYSIS OF SEMI-SUPERVISED DATA ANNOTATION RESULTS

This paper introduces a semi-supervised data annotation method that removes crop pixels from vegetation pixels and considers the remaining vegetation pixels as weed pixels to create a weed dataset. By training with a crop dataset, U-Net achieved an accuracy of 99.62% on the test set, indicating that its segmentation results are almost indistinguishable from manual annotations. In this case, where crop areas can be accurately segmented, the accuracy of vegetation segmentation directly affects the annotation of weeds in the weed dataset. A dataset with complete and accurate annotation and comprehensive information is beneficial for improving the performance of the trained model. As a result, this paper applied the color indices mentioned in Section II-C-I) to the data annotation work, comparing the four annotation results obtained to the manually annotated results to select the data annotation method that is closest to the manual annotation results. This was done to ensure the quality of data annotation. In this study, 26 cauliflower images were randomly selected, and the vegetation in the images was annotated at the pixel level using the labelme data labeling software, and the obtained labels were then converted into binary mask images in PNG format. Due to the high density of weed distribution, the labeling time for each image was between 1.5 and 2 hours. To reduce the deviation caused by sparse data, this paper enhances the data of these 26 images and their corresponding mask images, and finally expands the dataset to 52 images after several random rotations, horizontal and vertical flipping and random cropping operations.

Subsequently, four different color indexes were used in this study to perform semi-supervised data annotation on the 52 original images. According to the manual labeling results, the accuracy of the four groups of semi-supervised labeling results was calculated as shown in Table 1. Additionally, three original images and corresponding semi-supervised annotation results were randomly selected to verify whether the actual annotation results corresponded to the accuracy calculation results, as shown in Fig. 6.

TABLE 1. Accuracy of semi-supervised annotation based on four different color indexes.

Evaluation index	Color index			
	ExG	CIVE	ExGR	MExG
Accuracy/%	85	84.93	84.65	80.56

The results indicate that the accuracy of the dataset annotation results obtained by the ExG method in the high-pressure weed environment is superior to the other three color indices. By comparing the images mixed with the annotation results and the original images, it can be found that ExG color index can detect more weed areas and has better processing ability for shadow parts. Weed density detection is often used in the scenarios such as variable spraying and variable flaming weeding, requiring accurate detection of all weeds as much as possible to prevent an excessive application or less application. Therefore, this paper selects ExG as the formula for calculating the color index to obtain vegetation information in the image.

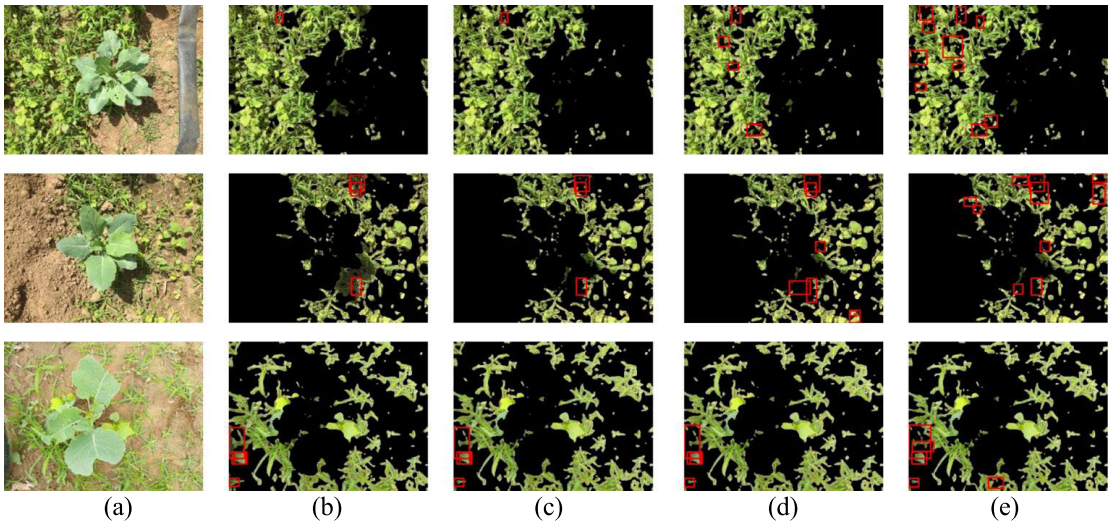

FIGURE 6. Comparison of semi-supervised annotation results based on four different color indices (Mixed with original images): (a) original image, (b) ExG annotation results, (c) CIVE annotation results, (d) ExGR annotation results, (e) MExG annotation results. The red boxes are areas with significant differences.

TABLE 2. Ablation results of the improved PSP Net model.

Networks	Backbone	MIoU/%	mPA/%	Accuracy/%	FPS
PSP Net	ResNet50	88.08	93.99	94.02	1.92
PSP Net	MobileNetV3	87.76	93.61	93.87	13.67
PSP Net+Bicubic	MobileNetV3	89.08	94.24	94.58	13.32
PSP Net+Bicubic+CBAM	MobileNetV3	90.23	94.91	95.18	13.26

B. PERFORMANCE EVALUATION OF THE IMPROVED PSP NET

Based on the weed dataset generated by the semi-supervised annotation method in this paper, the improved PSP Net model was trained according to the training strategy in Section II-E-6) to identify and segment weed regions in complex field images of cauliflower. An ablation experiment was conducted on the improved PSP Net model to verify the effect of model optimization. The results are shown in Table 2.

Table 2 show that the introduction of the MobileNetV3 backbone CNN for feature extraction has improved the speed of model inference. Despite its strengths, MobileNetV3's performance is somewhat limited in detecting smaller objects, and it is sensitive to information loss in certain convolutional kernels, which results in a slight drop in model accuracy. Some optimization measures, such as replacing bicubic linear interpolation and adding the CBAM attention mechanism, supplement the decreased accuracy and achieve some improvement. Although the model speed is reduced, this speed loss can be ignored compared to the speed increase brought by the MobileNetV3 backbone feature extraction network. Finally, through a series of optimizations, compared with the original PSP Net model with ResNet50 as the backbone feature extraction network, the MIoU, mPA and Accuracy of the improved model are increased by 2.15%, 0.92% and 1.16% respectively, and the reasoning speed of the model is increased by 6.9 times.

As illustrated in Fig. 7, it can be observed that after 100 epochs of training, the training and validation losses of the improved PSP Net become stable, with no significant gap between the two curves. This indicates that the improved PSP Net does not experience overfitting while being trained on the high weed pressure dataset.

In order to verify the effectiveness of the improved PSP Net model, this section compares its performance with that of three other models: FCN, DeeplabV3+, U-Net and Swin-UNet. The evaluation results of these models are presented in Table 3.

According to the data presented in the analysis table, the MIoU, mPA, Accuracy and FPS of the improved PSP Net model in the test set are 90.23%, 94.91%, 95.18%, and 13.26, respectively, outperforming both FCN and DeeplabV3+ models. For U-Net, the MIoU, mPA, Accuracy and FPS in the test set are 91.44%, 95.57%, 95.86% and 3.37, respectively. Although the model accuracy of U-Net is better than that of the improved PSP Net, the model reasoning speed of the improved PSP Net is four times that of U-Net. Moreover, the size of the improved PSP Net model is only 8.87% of that of the U-Net model. To balance the real-time performance and accuracy of Swin-Unet, we set the image_size of Swin-Unet

TABLE 3. The improved PSP Net model compared to other models.

Networks	Backbone	MIoU/%	mPA/%	Accuracy/%	FPS	Model size（MB）
FCN	ResNet50	82.7	91.0	91.9	0.5	269
DeeplabV3+	MobileNetv2	89.64	94.65	94.87	10.07	22.4
U-Net	ResNet50	91.44	95.57	95.86	3.37	94.9
Swin-UNet	Swin Transformer	90.34	95.13	95.22	2.33	105
Our	MobileNetV3	90.23	94.91	95.18	13.26	8.42

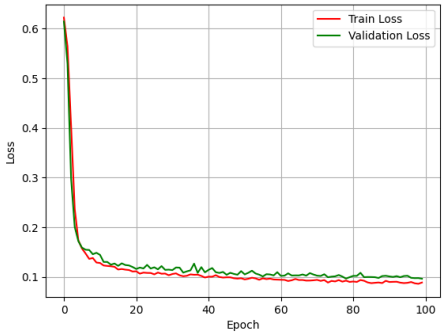

FIGURE 7. Training and validation curves of the improved PSP Net.

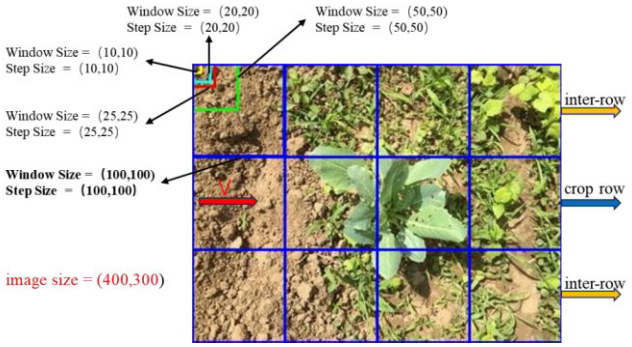

FIGURE 8. Schematic diagram of sliding windows with different sizes.

to 224 × 224 during model training, while for other models, the image_size was uniformly set to 512 × 512. On the test set, Swin-Unet achieved an mIoU, mPA, Accuracy, and FPS of 90.34%, 95.13%, 95.22%, and 2.33, respectively. The model accuracy of Swin-Unet is slightly higher than the improved PSP Net, but even with a smaller image_size, the inference speed of Swin-Unet is still lower than the improved PSP Net. In addition to comparing with the above models, this paper also attempted to use the Segment Anything Model for segmenting high weed pressure images. However, only some crop areas with obvious features appeared in the segmentation results. Moreover, the model size was about 42.46 times larger than the models used in this paper. As an excellent segmentation model, further optimization is needed for the application of the Segment Anything Model in weed segmentation.

Finally, considering the model accuracy, inference speed, and size, it can be concluded that the improved PSP Net is more suitable for weed density detection in the field. This proves the effectiveness of the improved PSP Net proposed in this paper. The excellent segmentation accuracy and relatively small model size demonstrated by U-Net also prove the correctness of applying the U-Net model in crop segmentation stages with low real-time requirements.

C. COMPARISON OF DENSITY CALCULATION RESULTS WITH REAL VALUES

This paper adopts the sliding scanning algorithm to segment an image into multiple pixel regions and quantify the number of weed pixels in each region. The weed density in each

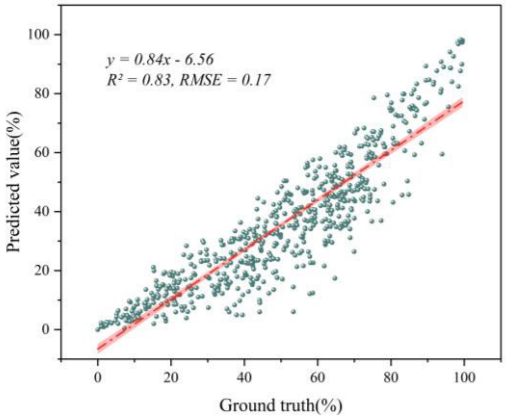

FIGURE 9. Regression analysis results of manual annotation and model prediction.

region is then calculated by the formula of weed density calculation. To determine the main parameters of the sliding window scanning algorithm, this paper set the window shape as a square based on the spraying shape of the sprayer. Subsequently, based on the original image size of 400 × 300, five sliding window sizes were selected for comparison, including 100×100, 50×50, 25×25, 20×20, and 10×10. Then, 300 test images were batched and inputted to the sliding window scanning algorithm. The scanning time for each image was recorded, and the average scanning time was calculated.

According to Table 4, the average scanning time gradually decreases as the sliding window size increases. The shortest average scanning time was observed when the sliding

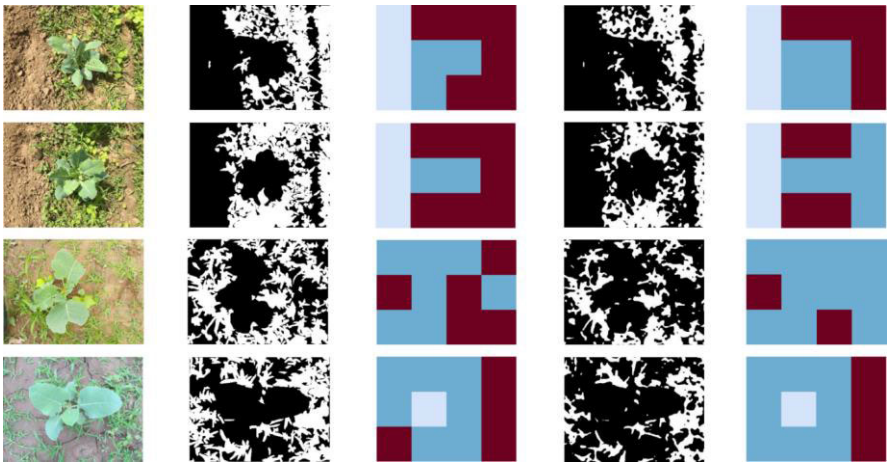

FIGURE 10. Prescription map generation: (a) original image, (b) and (c) manual labeling results and their corresponding prescription maps. (d) and (e) model prediction results and their corresponding prescription maps. Representing low weed density and requiring no treatment. Representing medium weed density and requiring moderate intensity treatment. Representing high weed density and requiring high-intensity treatment.

TABLE 4. Average scanning time of a single image under different sliding window sizes and step sizes.

Window Size(px)	Step Size(px)	average scanning time(s)
(10,10)	(10,10)	0.1747
(20,20)	(20,20)	0.0387
(25,25)	(25,25)	0.0251
(50,50)	(50,50)	0.0067
(100,100)	(100,100)	0.0019

window size was 100 × 100. Figure 8 shows that selecting a sliding window size of 100 × 100 can divide the image into 12 regions, with four regions in the middle row belonging to crop rows and eight regions in the upper and lower rows belonging to inter-row areas. This satisfies the need for this paper to calculate weed density by dividing the image into regions. Moreover, this also meets the needs of some weed control equipment that adopts different treatment methods in intra-row and inter-row areas. To avoid scanning the same region repeatedly during scanning, it is necessary to ensure that the step size of the sliding window scanning algorithm is consistent with the sliding window size. Therefore, the step size of the sliding window scanning algorithm selected in this paper is 100 × 100 to match the sliding window size.

The real weed mask images and predicted mask images corresponding to 52 original images obtained in Section III-A are used to calculate weed density in this paper. Finally, the density calculation results of two sets of different weed mask images are obtained, and each set contains 624 data.

As illustrated in Fig. 9, this paper conducts regression analysis on the two sets of density calculation results obtained. The analysis results show that the relationship between the weed density predicted by the model and the real weed density manual annotation is $y = 0.84x - 6.56$. The coefficient of determination (R^2) is 0.83 and the root mean square error (RMSE) is 0.17. These results indicate there is a strong correlation between the real value and the predicted value, and the predicted result is accurate and reliable. The established prediction model of weed density can predict the real weed density well.

D. PRESCRIPTION MAP GENERATION

In this paper, the threshold $l_1 = 10\%$ and $l_2 = 60\%$ are set to divide the weed density into three levels [39]. Through the threshold screening of the two groups of weed density data obtained in Section III-C, weed areas within different threshold intervals are divided into corresponding levels to generate a visual prescription map. By comparing the real prescription map with the model prediction prescription map, 486 of the 624 weed areas are accurately classified, and 138 areas are treated at a lower level than the real one.

From the above data, the accuracy of the prescription map predicted by the model is calculated to be 78%. And even if it is predicted incorrectly, the processing intensity in this region will not differ much from the real situation. This is because some small weeds are missed during the weed detection process, causing the predicted weed density to be lower than the real weed density. When the real weed density in this area is close to the classification threshold, the predicted result is lower than the threshold, while the real result is higher than the threshold, resulting in a classification difference but without a big grade leap..

At the same time, the classification according to the threshold weakens the influence of small errors to some extent, making the model errors more tolerable. Moreover, by adopting a threshold-based classification approach, the impact of minor errors is weakened to some extent, resulting in a higher

tolerance for error in the model. This threshold classification method can enhance the robustness and stability of the model to a certain degree. Figure 10 shows visual prescription maps generated based on the manual annotation results and model prediction results

IV. CONCLUSION

In this paper, we focus on the problem of weed density detection in environments with high weed pressure. We study the fast labeling of high weed pressure images and propose an improved convolutional neural network. Firstly, we introduce a simple semi-supervised labeling method that is adapted for images under high weed pressure. This labeling method enables an efficient annotation of high weed pressure images. Secondly, we propose a modified PSP Net model that can accurately segment areas under high weed stress.

Among them, semi-supervised data annotation is carried out in three steps. First, the U-net semantic segmentation model is trained by using crop data with supervision annotation to detect crops. Second, the ExG color index and Otsu threshold algorithm are used for vegetation segmentation of the original image. Finally, the crop pixels contained in the vegetation segmentation result are removed, and the remaining vegetation is regarded as weeds. And the final weed annotation is obtained by morphological processing of the image, which forms a complete weed dataset together with the original image.

Based on this dataset, this paper improves and trains the PSP Net model. The MIoU, mPA, and Accuracy of the improved PSP Net model are increased by 2.15%, 0.92%, and 1.16%, respectively, and the model detection speed is increased by 6.9 times. Compared with the original model, the improved model has significantly improved the accuracy and real-time performance.

Finally, based on the improved model, this paper realizes weed detection and weed density calculation. The coefficient of determination (R^2) between the weed density results obtained by this method and the real value is 0.83, and the root mean square error (RMSE) is 0.17. The accuracy of the prescription map generated by dividing the weed density threshold is 78%. The above results show that there is a strong correlation between the detection results of the method proposed in this paper and the real results, and the prediction results are accurate and reliable.

In farmland with high weed pressure and patchy weed distributions, manual supervision and annotation of weeds is a costly task. Given the singleness and regularity of crop planting, the supervised annotation method of crops in this paper is more beneficial to reduce the time cost of dataset annotation and accelerate the application of the supervised convolutional neural network in weed density detection tasks.

The method proposed in this paper can be applied to some weeding methods based on the change in weed density, such as variable spraying and variable flame weeding, which is beneficial to reduce weeding costs and environmental pollution. The future work will optimize the vegetation segmentation method to enhance the annotation quality of the weed dataset, and integrate it with variable spraying weeding robots for weed density detection and application in the actual field scene.

ACKNOWLEDGMENT

The authors would like to thank Prof. Li and his team for their generous contribution of the dataset that was crucial in facilitating our research. The provision of their dataset enabled them to progress with greater speed and efficiency, ultimately enriching the results of our study. They are truly grateful for their indispensable support.

REFERENCES

[1] Z. Wu, Y. Chen, B. Zhao, X. Kang, and Y. Ding, "Review of weed detection methods based on computer vision," *Sensors*, vol. 21, no. 11, p. 3647, May 2021.

[2] H. Fang, G. Xu, X. Xue, M. Niu, and L. Qiao, "Study of mechanical-chemical synergistic weeding on characterization of weed–soil complex and weed control efficacy," *Sustainability*, vol. 15, no. 1, p. 665, Dec. 2022.

[3] Y. Liu, "(Retracted) field weed recognition algorithm based on machine learning," *J. Electron. Imag.*, vol. 31, no. 5, May 2022, Art. no. 051413.

[4] J. Liu, I. Abbas, and R. S. Noor, "Development of deep learning-based variable rate agrochemical spraying system for targeted weeds control in strawberry crop," *Agronomy*, vol. 11, no. 8, p. 1480, Jul. 2021.

[5] M. Ataei and M. Abdollahi, "A systematic review of mechanistic studies on the relationship between pesticide exposure and cancer induction," *Toxicol. Appl. Pharmacol.*, vol. 456, Dec. 2022, Art. no. 116280.

[6] S. Fountas, N. Mylonas, I. Malounas, E. Rodias, C. Hellmann Santos, and E. Pekkeriet, "Agricultural robotics for field operations," *Sensors*, vol. 20, no. 9, p. 2672, May 2020.

[7] R. Raja, D. C. Slaughter, S. A. Fennimore, and M. C. Siemens, "Real-time control of high-resolution micro-jet sprayer integrated with machine vision for precision weed control," *Biosyst. Eng.*, vol. 228, pp. 31–48, Apr. 2023.

[8] M. S. Alam, M. Alam, M. Tufail, M. U. Khan, A. Güneş, B. Salah, F. E. Nasir, W. Saleem, and M. T. Khan, "TobSet: A new tobacco crop and weeds image dataset and its utilization for vision-based spraying by agricultural robots," *Appl. Sci.*, vol. 12, no. 3, p. 1308, Jan. 2022.

[9] H. Peng, Z. Li, Z. Zhou, and Y. Shao, "Weed detection in paddy field using an improved RetinaNet network," *Comput. Electron. Agricult.*, vol. 199, Aug. 2022, Art. no. 107179.

[10] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.

[11] A. Milioto and C. Stachniss, "BonNet: An open-source training and deployment framework for semantic segmentation in robotics using CNNs," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 7094–7100.

[12] Y. Liu, Y.-H. Wu, P. Wen, Y. Shi, Y. Qiu, and M.-M. Cheng, "Leveraging instance-, image- and dataset-level information for weakly supervised instance segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1415–1428, Mar. 2022.

[13] M. H. Saleem, J. Potgieter, and K. M. Arif, "Automation in agriculture by machine and deep learning techniques: A review of recent developments," *Precis. Agricult.*, vol. 22, no. 6, pp. 2053–2091, Dec. 2021.

[14] I. Sa, Z. Chen, M. Popovic, R. Khanna, F. Liebisch, J. Nieto, and R. Siegwart, "WeedNet: Dense semantic weed classification using multispectral images and MAV for smart farming," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 588–595, Jan. 2018.

[15] Y. Hua, D. Marcos, L. Mou, X. X. Zhu, and D. Tuia, "Semantic segmentation of remote sensing images with sparse annotations," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[16] A. Abdalla, H. Cen, L. Wan, R. Rashid, H. Weng, W. Zhou, and Y. He, "Fine-tuning convolutional neural network with transfer learning for semantic segmentation of ground-level oilseed rape images in a field with high weed pressure," *Comput. Electron. Agricult.*, vol. 167, Dec. 2019, Art. no. 105091.

[17] P. Lottes and C. Stachniss, "Semi-supervised online visual crop and weed classification in precision farming exploiting plant arrangement," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 5155–5161.
[18] I. D. García-Santillán and G. Pajares, "On-line crop/weed discrimination through the Mahalanobis distance from images in maize fields," *Biosyst. Eng.*, vol. 166, pp. 28–43, Feb. 2018.
[19] Y. Xu, R. He, Z. Gao, C. Li, Y. Zhai, and Y. Jiao, "Weed density detection method based on absolute feature corner points in field," *Agronomy*, vol. 10, no. 1, p. 113, Jan. 2020.
[20] S. Shorewala, A. Ashfaque, R. Sidharth, and U. Verma, "Weed density and distribution estimation for precision agriculture using semi-supervised learning," *IEEE Access*, vol. 9, pp. 27971–27986, 2021.
[21] K. Zou, X. Chen, F. Zhang, H. Zhou, and C. Zhang, "A field weed density evaluation method based on UAV imaging and modified U-Net," *Remote Sens.*, vol. 13, no. 2, p. 310, Jan. 2021.
[22] X. Jin, J. Che, and Y. Chen, "Weed identification using deep learning and image processing in vegetable plantation," *IEEE Access*, vol. 9, pp. 10940–10950, 2021.
[23] A. M. Mishra, S. Harnal, V. Gautam, R. Tiwari, and S. Upadhyay, "Weed density estimation in soya bean crop using deep convolutional neural networks in smart agriculture," *J. Plant Diseases Protection*, vol. 129, no. 3, pp. 593–604, Mar. 2022.
[24] N. Li, X. Zhang, C. Zhang, H. Guo, Z. Sun, and X. Wu, "Real-time crop recognition in transplanted fields with prominent weed growth: A visual-attention-based approach," *IEEE Access*, vol. 7, pp. 185310–185321, 2019.
[25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Munich, Germany: Springer, 2015, pp. 234–241.
[26] M.-K. Lee, M. R. Golzarian, and I. Kim, "A new color index for vegetation segmentation and classification," *Precis. Agricult.*, vol. 22, no. 1, pp. 179–204, Feb. 2021.
[27] D. Woebbecke, G. Meyer, K. Von Bargen, and D. Mortensen, "Plant species identification, size, and enumeration using machine vision techniques on near-binary images," *Proc. SPIE*, vol. 1836, pp. 208–219, May 1993.
[28] T. Kataoka, T. Kaneko, H. Okamoto, and S. Hata, "Crop growth estimation system using machine vision," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics (AIM)*, Jul. 2003, pp. b1079–b1083.
[29] X. P. Burgos-Artizzu, A. Ribeiro, M. Guijarro, and G. Pajares, "Real-time image processing for crop/weed discrimination in maize fields," *Comput. Electron. Agricult.*, vol. 75, no. 2, pp. 337–346, Feb. 2011.
[30] G. E. Meyer and J. C. Neto, "Verification of color vegetation indices for automated crop imaging applications," *Comput. Electron. Agricult.*, vol. 63, no. 2, pp. 282–293, Oct. 2008.
[31] A.-G. Manh, G. Rabatel, L. Assemat, and M.-J. Aldon, "AE-Automation and emerging technologies: Weed leaf image segmentation by deformable templates," *J. Agric. Eng. Res.*, vol. 80, no. 2, pp. 139–146, 2001.
[32] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
[33] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.
[34] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324.
[35] Z. Huang and L. Cao, "Bicubic interpolation and extrapolation iteration method for high resolution digital holographic reconstruction," *Opt. Lasers Eng.*, vol. 130, Jul. 2020, Art. no. 106090.
[36] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, V. Ferrari, Ed., Oct. 2018, pp. 3–19.
[37] P. Gonzalez-de-Santos, A. Ribeiro, C. Fernandez-Quintanilla, F. Lopez-Granados, M. Brandstoetter, S. Tomic, S. Pedrazzi, A. Peruzzi, G. Pajares, G. Kaplanis, M. Perez-Ruiz, C. Valero, J. del Cerro, M. Vieri, G. Rabatel, and B. Debilde, "Fleets of robots for environmentally-safe pest control in agriculture," *Precis. Agricult.*, vol. 18, no. 4, pp. 574–614, Aug. 2017.
[38] H. Huang, J. Deng, Y. Lan, A. Yang, X. Deng, S. Wen, H. Zhang, and Y. Zhang, "Accurate weed mapping and prescription map generation based on fully convolutional networks using UAV imagery," *Sensors*, vol. 18, no. 10, p. 3299, Oct. 2018.
[39] F. López-Granados, J. Torres-Sánchez, A. Serrano-Pérez, A. I. de Castro, F.-J. Mesas-Carrascosa, and J.-M. Peña, "Early season weed mapping in sunflower using UAV technology: Variability of herbicide treatment maps against weed thresholds," *Precis. Agricult.*, vol. 17, no. 2, pp. 183–199, Apr. 2016.

XIA LI received the Ph.D. degree in agricultural mechanization engineering from China Agricultural University, Beijing, China, in 2013. She is currently a Professor with the School of Mechanical Engineering, Tianjin University of Technology. Her research interests include intelligent pneumatically powered deep loosening, agricultural robots, and machine vision.

FANGTAO DUAN received the bachelor's degree in mechanical engineering from Shandong Agriculture and Engineering University, Shandong, China, in 2021. He is currently pursuing the master's degree in mechanical engineering with the Tianjin University of Technology. His current research interests include agricultural image processing, machine vision, and deep learning.

MENGCHAO HU received the bachelor's degree in mechanical design, manufacturing and automation from the Shaanxi University of Technology, Shaanxi, China, in 2020. He is currently pursuing the master's degree in mechanical engineering with the Tianjin University of Technology. His current research interests include agricultural robotics and vision detection.

JIAWEI HUA received the bachelor's degree in mechanical design, manufacturing and automation from the Qing Gong College, North China University of Science and Technology, Hebei, China, in 2018. He is currently pursuing the master's degree in mechanical engineering with the Tianjin University of Technology. His current research interests include object detection and deep learning.

XIWANG DU received the bachelor's degree in mechanical and electronic engineering from the Tianjin Sino-German University of Applied Sciences, Tianjin, China, in 2022. He is currently pursuing the master's degree in mechanical engineering with the Tianjin University of Technology. His current research interests include agricultural image processing and the structural design of weeding robots.