Unsupervised Domain Adaptation for Weed Segmentation Using Greedy Pseudo-labelling

Yingchao Huang and Abdul Bais
University of Regina
3737 Wascana Pkwy, Regina, SK S4S 0A2, Canada
{abdul.bais, huang47y}@uregina.ca

Abstract

Automatic weed identification based on RGB images with convolutional neural networks (CNN) is a new frontier of precision agriculture. However, the CNN models expect a large volume of labelled data. Their performance deteriorates across different fields due to varied agricultural contexts. To address this, we propose an unsupervised domain adaptation (DA) framework leveraging pseudo-labelling. Our method involves co-training labelled source data with pseudo-labelled target data. We introduce a novel greedy pseudo-labelling strategy to optimize pseudo-label selection, maximizing gains while minimizing overfitting risks. Monitoring overfitting with covariance helps detect fluctuations in class pixel counts during co-training, enhancing target performance. The proposed framework has demonstrated superior performance by evaluation against literature approaches, including the input-level DA methods with Fourier Transform, feature-level with CycleGAN methods and AdaptSegNet, and output-level with self-training. It is tested with the ROSE challenge dataset from different cameras and years with diverse plant stages. Particularly in challenging conditions for plants across different years with varied plant stages, the proposed method outperforms existing literature that struggles to surpass the baseline.

1. Introduction

Automatic weed identification is recommended for weed control due to reduced labour costs and minimized herbicide usage [31]. It is conducted with machine vision systems and remote aerial imaging techniques [9, 19, 47] by processing the captured images and segmenting crops and weeds against background (soil, stones, crop residue, etc.). Weed segmentation with the captured images utilizes various computer vision methods, including image processing and machine learning (ML), to achieve precise and reliable weed identification [34]. The most successful ML technique in recent years is the use of convolutional neural networks (CNN) [14, 26, 51], which extract features automatically and classify images without any domain knowledge of the task they are dealing with [22].

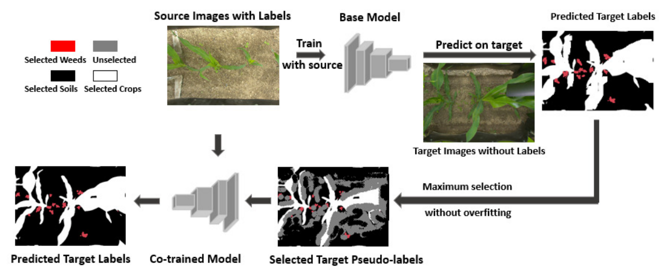

Figure 1. The base model is trained with source images, and the prediction of target images is optimized by selecting and incorporating a maximum number of pseudo-labels, ensuring effective co-training without overfitting alongside the source images.

However, due to the complexity of agricultural environments and the variability in weed species, appearance, lighting conditions, and growth stages, the performance of a CNN model trained on one specific set of images (i.e. source domain) deteriorates when it is deployed to new agricultural contexts (i.e. target domain) [1, 13, 41]. Maintaining comparable performance in such variable agricultural environments demands broad datasets containing samples of all conditions encountered in the field. Having enough labelled data for each newly studied field is costly and not feasible in real applications. Thus, domain adaptation (DA) techniques are proposed to adapt the existing models built on labelled source data to the unlabelled target data to achieve comparable performance and reduce the need for labelling.

DA is a technique to improve the performance of a model on a target domain containing insufficient annotated data by using the knowledge learned from another related source domain with adequate labelled data [42]. DA techniques can be classified based on the availability of labelled data in the source and target domains, including supervised DA, semi-supervised DA, unsupervised DA, self-supervised DA, and multi-source DA [39]. This study focuses on the unsupervised DA with the target labels unavailable.

DA techniques have been extended to agriculture in recent years, particularly in precision agriculture and automated agriculture systems, where adapting models across different agriculture contexts is crucial for accurate decision-making and resource optimization [37]. Examples include weed identification and herbicide application across various fields having varying weed species and growth patterns [13, 17, 45], crop disease identification based on environmental factors and specific crop varieties [3, 6, 46], and yield prediction and harvest planning with different soil types, weather patterns, and farming practices [29, 32, 44], etc.

In this work, we introduce an unsupervised DA framework leveraging pseudo-labelling called greedy pseudo-labelling weed segmentation tasks, which maximizes the utilization of pseudo-labelled target samples through a streamlined, one-stage co-training approach as shown in Figure 1. The base model is trained with the source images and makes predictions on the target. The maximum number of pseudo-labels determined by covariance analysis is selected from the predictions and included in the co-training process to improve the test performance on the target domain.

To mitigate overfitting of pseudo-labels during co-training, we employ a robust strategy to determine the maximum selection proportion based on covariance, considering class sizes. Overfitting is indicated by significant fluctuations in class sizes, as observed in previous studies [16, 54]. Covariance is effective in monitoring multiple variables simultaneously [25], allowing us to closely track fluctuations in class sizes pre- and post-DA to optimize selection proportions and enhance performance on the target domain.

We summarize the main contributions of this work as follows:

- We advocate for utilizing output-level approaches in weed identification, a novel technique not explored in existing literature for weed segmentation.
- Covariance in class sizes before and after DA is proposed and found effective in detecting overfitting and optimizing the pseudo-label selection proportion.
- We formulate a loss function incorporating the soft Intersection over Union (softIoU) of the labelled source and selected pseudo-labelled target pixels into the framework, enhancing the weed segmentation's effectiveness.
- The outlined method has demonstrated superior performance by evaluation against existing approaches at different adaptation levels, including the Fourier Transform (input-level) [41], AdaptSegNet (feature-level) [18], CycleGAN (feature-level) [13, 52] and CBST (output-level) [59] methods, using the ROSE challenge dataset [2].
- The proposed method effectively enhances adaptations in challenging conditions for plants from different years with varied stages, outperforming existing literature methods that fail to surpass the baseline.

2. Related work

DA at the input-level. DA for weed segmentation could be performed at different levels: input-level, feature-level, and output-level [39]. Adaptation at the input level achieves cross-domain uniformity of the visual appearance of the input images by statistical matching at the input level [40]. A rich line of work has been focusing on style transfer techniques, such as Fourier Transform [1, 21, 41, 53]. Utilizing the Fourier transform on an image enables the extraction of both its phase and amplitude components. By substituting the amplitude of a source image with that of a target image, the resulting image embodies the semantics of the source and the style of the target. In weed segmentation studies, the Fourier Transform has been explored for its efficacy [41], which showed that employing the Fourier Transform for style transfer yields superior performance compared to baseline methods in crop-weed segmentation tasks.

DA at the feature-level. Adaptation at the feature level is to force the feature extractor to discover domain-invariant features by adjusting the distribution of latent representations from source and target domains [39]. The popular algorithms are adversarial-based [17, 18, 43] which introduces an additional domain classifier to distinguish features between the source and target as well as to confuse the domain discriminator by making the features from both domains indistinguishable to learn domain-invariant features, and generative-based [5, 10, 12, 27], involving generating synthetic data in the target domain or adapting existing source domain data to look like the target domain. For example, the adversarial-based approach AdaptSegNet was proposed for weed segmentation tasks across varied fields in [18] and found effective in handling changes in new field environments during real field inference. A considerable amount of research [1, 13, 24, 28, 52, 55] has been resorting to the generative-based with CycleGAN methods for weed segmentation tasks.

DA at the output-level.Self-training-based adaptation methods are widely used output-level approaches that co-train the model with labelled source and pseudo-labelled target samples. They have demonstrated success in diverse

$$pseudo\_softIoU = \frac{1}{c}\sum_{c=1}^{C}\frac{\sum_{i=1}^{n_s} y_{ic}\cdot y^*_{ic} + \sum_{j=1}^{n'_t} y_{jc}\cdot \hat{y}^*_{jc}}{\sum_{i=1}^{n_s} y_{ic} + y^*_{ic} - y_{ic}\cdot y^*_{ic} + \sum_{j=1}^{n'_t} y_{jc} + \hat{y}^*_{jc} - y_{jc}\cdot \hat{y}^*_{jc}} \tag{1}$$

domains, including city traffic scenes [11, 15, 56, 58] and medical image segmentation [30,48,57]. However, their application to weed segmentation is relatively scarce in the literature. Notably, recent work [24] suggests exploring self-training approaches as a promising future direction for weed segmentation.

The self-training-based adaptation methods involve gradually incorporating pseudo-labels by increasing their selection proportions while refining their quality through iterations [4, 49, 50]. Research has shown that iteratively refining these pseudo-labels during co-training improves the final performance. However, the quality of the initially selected pseudo-labels is pivotal. Incorrect labels can propagate errors through training, resulting in a less-than-optimal model [8]. Therefore, the success of co-training with pseudo-labels is tied to the initially selected pseudo-labels, but determining the optimal selection proportion presents a challenge [20, 36, 38].

3. Methodology

In this work, we propose using pseudo-labelling to segment weeds. Pseudo-labels of the target domain are used to co-train the model with the source domain based on $pseudo\_softIoU$, taking into account the softIoU of the source-labelled pixels and the selected pseudo-labelled target pixels, as shown in Eq (1). We seek to minimize the softIoU of both source-labelled and the selected target pseudo-labelled pixels by addition in Eq (1). C is the set of classes, y_{ic} and y_{jc} are the predicted label of source and target pixels. y^*_{ic} is the ground truth of a source label for pixel i, and $\hat{y}^*_{jc}$ is the target pseudo-label predicted by the base model for pixel j and treated as a true label in the co-training process. All the pixels from the source domain are counted in the loss function, annotated with n_s, whereas only the selected pseudo-labels are considered for the target domain, marked as n'_t.

The pseudo-labels y^*_{ic} included in the co-training are determined by the confidence thresholds $exp(-k_c)$, where c represents a class. The pixels with a confidence lower than the thresholds are filtered out using Eq (2) by setting the pseudo-label all zeros. As shown in Algorithm 1, the confidence threshold k_c is determined based on the selection proportion $\mathbf{p}_s$. This set stores all the pre-defined selection proportions, aiming to identify the optimal selection proportion. Subsequently, the optimal confidence threshold could be determined. To get the pseudo-labels of the target domain, we first need to predict the target images $\mathbf{X}_t$ with the neural network function by $f(\mathbf{w}, \mathbf{X}_t)$ with the prediction confidence of all pixels stored in $\mathbf{P}_{X_t}$. Then, we find the pseudo-labels $\hat{\mathbf{y}}_t$ based on the maximum prediction probabilities of each target pixel, which is stored in $\mathbf{MP}_{c,X_t}$ for each class c. $\mathbf{MP}_{c,X_t}$ is then used to match each prediction confidence by pseudo-labelled class and stored in $\mathbf{M}$.

The confidence threshold k_c is determined by ranking the prediction probabilities of all pixels assigned to class c. The ranked probabilities are stored in $\mathbf{m}_c$. Subsequently, the index of the threshold probability ind_c is calculated by $p_s \times length(\mathbf{m}_c)$, where p_s is the selection proportion. The k_c is then determined accordingly based on p_s and stored in $\mathbf{k}$. The thresholds for each selection proportion $\mathbf{k}$ are saved in $\mathbf{K}_{p_s}$.

$$\hat{y}_{j,c} = \begin{cases} 1 & \text{if } c \in C \text{ and } c = \arg\max \mathbf{p}_j, \\ & \mathbf{p}_j(c|\mathbf{w}, \mathbf{x}_{t,j}) \geq exp(-k_c) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Algorithm 1 Determination of $\mathbf{K}_{p_s}$

Input: Neural network $f(\mathbf{w})$, target images $\mathbf{X}_t$, selection portions $\mathbf{p}_s$
Output: $\mathbf{k}$

$\mathbf{P}_{X_t} = f(\mathbf{w}, \mathbf{X}_t)$
$\hat{\mathbf{y}}_t = argmax(\mathbf{P}_{X_t}, axis = 0)$
$\mathbf{MP}_{X_t} = max(\mathbf{P}_{X_t}, axis = 0)$
for $c = 1$ to C **do**
  $\mathbf{MP}_{c,X_t} = \mathbf{MP}_{X_t}(\hat{\mathbf{y}}_t == c)$
  $\mathbf{M} = [\mathbf{M}, matrix\_to\_vector(\mathbf{MP}_{c,X_t})]$
end for
for p_s in $\mathbf{p}_s$ **do**
  for $c = 1$ to C **do**
    $\mathbf{m}_c = sort(\mathbf{M}[c], order = descending)$
    $ind_c = length(\mathbf{m}_c) \times p_s$
    $k_c = -\log(\mathbf{m}_c[ind_c])$
    $\mathbf{k}.append(k_c)$
  end for
  $\mathbf{K}_{p_s}.append(\mathbf{k})$
end for
return $\mathbf{K}_{p_s}$

When the thresholds are determined for each pre-defined selection proportion, we select pseudo-labels from the predictions and include them in the co-training process. In the co-training process, we do not anticipate significant fluctuations in the allocation of pixels across classes due to the assumption of similarity and relation between source and target domains in DA [23]. To track these fluctuations, we propose the use of covariance in class distributions before and after the co-training. High covariance values indicate substantial fluctuations in pixel distribution among classes compared to the initial distribution. This is often attributed to a significant number of falsely assigned pseudo-labels resulting from a large selection proportion. Such mislabeled instances misguide the model, leading to inaccurate pixel labelling for respective classes. These falsely assigned labels are prone to overfitting, adversely affecting test performance. Hence, overfitting manifests itself in the covariance, providing a reliable indicator for its detection.

The variance in pixel distribution among classes before and after adaptation is quantified through covariance, as depicted in Eq (3). For instance, we initially record the number of pixels assigned to each class predicted with the base model in an array $\mathbf{n}_0$. Subsequently, employing a selection proportion p_s to calculate the threshold k_c, we co-train the model using the selected pseudo-labels alongside the source domain, storing the resulting number of pixels for each class in an array $\mathbf{n}_{p_s}$. We then calculate the covariance between $\mathbf{n}_0$ and $\mathbf{n}_{p_s}$ for each selection proportion p_s within a set of the pre-defined selection proportions denoted by $\mathbf{p}_s$. Consequently, we have a covariance set including cov_{p_s} for each separate selection proportion p_s. The optimal selection proportion opt_{p_s} is the minimum of the covariance set, determined based on the selection proportion that generates the number of pixels for each class having the minimum covariance with $\mathbf{n}_0$.

$$opt_{p_s} = \min_{p_s \in \mathbf{p}_s} (cov_{p_s} | cov_{p_s} = cov(\mathbf{n}_0, \mathbf{n}_{p_s})) \tag{3}$$

The flow chart of the proposed framework is plotted in Figure 2. Our work uses the model designed in [1] to compare with the literature methods. The base model is built on a U-net segmentation network [33] with a VGG16 [35] pre-trained with ImageNet [7].

The base model, initially trained on source images, is tested on the target domain to generate target predictions, denoted as Step 1. Pseudo-labels are then selected in Step 2 based on confidence thresholds $\mathbf{K}_{p_s}$, calculated using pre-defined selection proportions $\mathbf{p}_s$ as outlined in Algorithm 1. The specified $\mathbf{p}_s$ values are pre-set as $[0.1, 0.2, 0.3, 0.4, 0.5]$ to have a maximum proportion 50% of the predictions included in the co-training as suggested in [59].

Subsequently, these selected pseudo-labels are employed to co-train the model with the source domain in Step 3. We prioritize the optimal selection proportion with minimal covariance, as shown in Eq (3), ensuring a robust co-training approach with covariance analysis in Step 4. The optimal selection proportion is determined in Step 5 by comparing it with the minimum predefined selection proportion of 0.1. If it is more than 0.1, it is used to select the consequent pseudo-labels to co-train the model and test it on the target domain to get the final predictions. Whereas equal to 0.1 indicates potential overfitting, often associated with a larger selection proportion and caused by numerous falsely labelled samples within the selected pseudo-labels. These falsely labelled samples manifest as substantial pixel fluctuations and heightened covariance. To prevent misleading the model and deteriorating test performance, we specifically integrate pseudo-labels with a minimal 0.1 selection proportion in Step 6. We then conduct co-training with the source domain in Step 7, followed by testing on the target domain to generate predictions in Step 8. When the target predictions are ready, this process repeats from Step 2, refining the pseudo-labels to determine an optimal selection proportion in subsequent rounds. Consequently, we optimize the optimal selection proportion more than the minimum predefined 0.1. Through this iterative refinement, we maximize pseudo-label selection proportions, leading to a significant enhancement in test performance.

4. Experiments

Datasets. The images we used in our experiments to test the proposed method are sourced from the ROSE Challenge, which involved four participating teams: BIPBIP, PEAD, ROSEAU, and WeedElec. These teams utilized various robots and camera systems during the ROSE field campaigns. The data for our experiments was gathered in 2019 and 2021 from an experimental field at the INRAE research center in Montoldre, France. The dataset comprises RGB images at various resolutions and semantic segmentation masks, capturing maize and bean plants and four weed types (Lolium perenne, Sinapis arvensis, Chenopodium album, Matricaria chamomilla) under natural daylight conditions. There are 1000 labelled images, including 125 per team and crop type for each year. The open-source dataset is available at [2].

In our study, we perform two types of adaptations: one involving different robots equipped with distinct cameras within the same field during the same period (May 2019) and the other involving the same robot used across different years and periods (May 2019 and September 2021) capturing varying plant growth stages. In the first type, we employ the images from BIPBIP and WeedElec with both crops (maize and beans). For the second type, we use images collected by BIPBIP with plants at different growth stages in the two years (2019 and 2021). Thus, we explore four combinations of source and target domains, each involving both crops.

Models and baselines. We assess our method against literature methods including the Fourier Transform [41], AdaptSegNet [18], CycleGAN methods [13,52], and CBST [59] method in both adaptations. To highlight the effectiveness of the proposed method, we set the baseline and upper

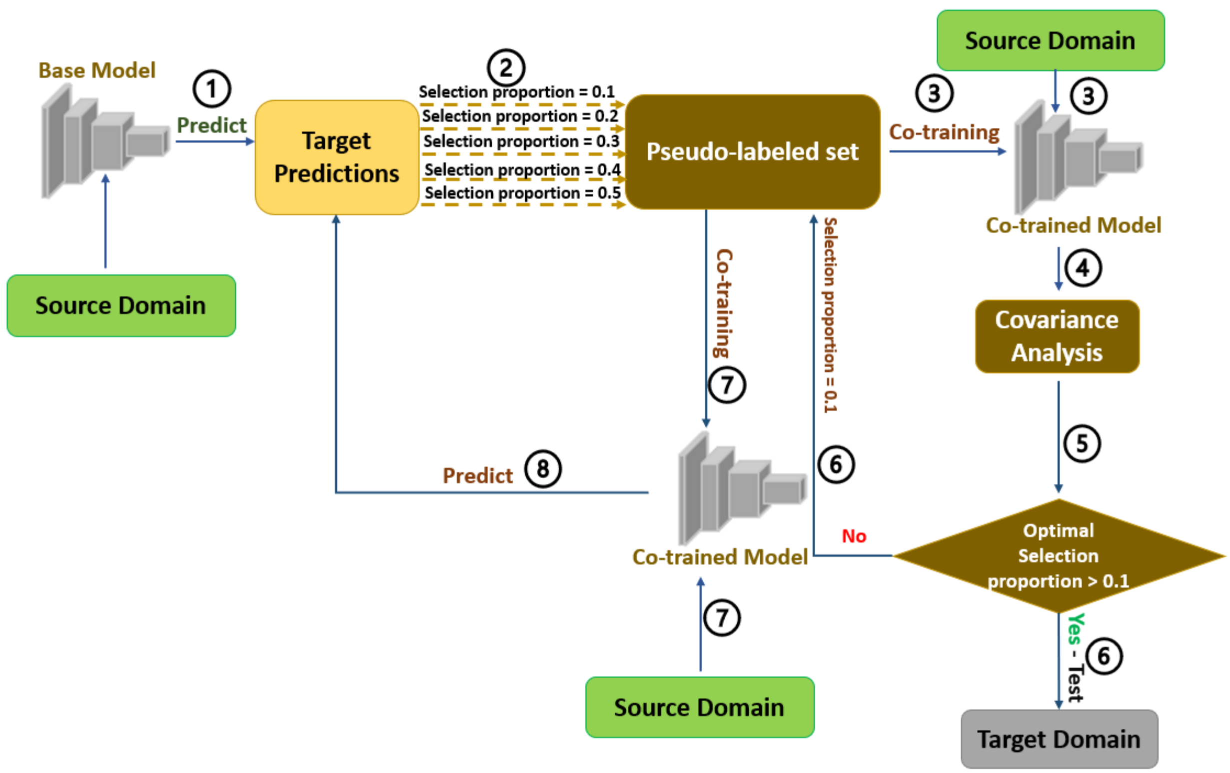

Figure 2. The scheme of the proposed greedy pseudo-labelling method, which shows the optimization of selection proportions and refinement of pseudo-labels.

bound for all adaptation scenarios. The baseline comprises predictions by the base model trained on the source domain with softIoU and tested on the target domain without any DA. On the other hand, the upper bound consists of predictions generated by the model trained directly on the target domain.

All the methods use the same U-net segmentation network [33] with a VGG16 [35] pre-trained with ImageNet [7]. Like the learning scheduler in [13], we set the initial learning rate of 0.0001 and linearly reduce it to zero, multiplied by 0.1 every time the IoU loss does not improve after four epochs. Adam optimizer and padded size of 320×320 from the original images are used in this study. Horizontal and vertical flips are performed with a probability of 0.5 in both directions. The test set does not use any padding during testing. Among the CycleGAN methods, we consider the architecture outlined in [13] (referred to as CGAN L_semantic) and the same architecture augmented with an additional phase loss, as proposed in [52] (referred to as CGAN L_phase). This evaluation allows us to compare our method's effectiveness against established techniques.

4.1. Results

The performance comparisons are detailed in Table 1. Our proposed method exhibits outstanding results, surpassing the performance of existing methods with an average mean IoU of 0.805. This achievement beats the Fourier Transform approach by 0.029, the AdaptSegNet by 0.036, and the CGAN L_semantic, CGAN L_phase, CBST by a substantial margin of 0.148, 0.156, and 0.392, respectively. Although the Fourier Transform shows superior test performance over the proposed method for adapting beans from WeedElec to BIPBIP, the proposed method obtains comparable results.

The test performance of the proposed method for the adaptation of maize from WeedElec to BIPBIP is constrained by the quality of pseudo-labels generated with the base model, which has a poor mean IoU of 0.465. The proposed method selects pseudo-labels from the predictions with the base model, and if the performance of the base model is inadequate, the quality of the pseudo-labels cannot be guaranteed. This may lead to numerous false labels, restricting the enhancement of test performance by co-training the model with these selected pseudo-labels.

Importantly, our method significantly enhances target

Table 1. Mean IoU (crop, weed and background classes) for the eight combinations of adaptations across different robots and years.

Crop	Source Domain	Traget Domain	Baseline	Fourier Transform [41]	CGAN L_semantic [13]	CGAN L_phase [52]	AdaptSegNet [18]	CBST [59]	Ours	Upper Bound
bean	BIPBIP	WeedElec	0.801	0.763	0.800	0.814	0.809	0.310	**0.823**	0.858
	WeedElec	BIPBIP	0.684	0.805	0.817	**0.830**	0.798	0.355	0.828	0.848
maize	BIPBIP	WeedElec	0.844	0.826	0.807	0.840	0.818	0.422	**0.861**	0.866
	WeedElec	BIPBIP	0.465	**0.806**	0.743	0.732	0.755	0.355	0.734	0.885
bean	2019	2021	0.635	0.690	0.631	0.552	0.672	0.391	**0.748**	0.810
	2021	2019	0.807	0.772	0.352	0.541	0.772	0.489	**0.821**	0.848
maize	2019	2021	0.769	0.739	0.362	0.310	0.710	0.400	**0.779**	0.805
	2021	2019	0.843	0.806	0.741	0.569	0.814	0.580	**0.849**	0.885
Average			0.731	0.776	0.657	0.649	0.769	0.413	**0.805**	0.851

performance in different adaptation scenarios across various years for plants with different growth stages. This achievement stands out as the existing literature methods have struggled to surpass the baselines. The study by Bertoglio et al. [2] acknowledged challenges related to distinct class proportions in the two domains and a domain gap arising from plants with varying shapes. In contrast, our method effectively tackles these challenges, resulting in a notable enhancement of target performance.

For a clear visual comparison of segmentation performance, we provide examples in Figure 3, featuring predictions from the proposed method alongside literature methods such as Fourier Transform and AdaptSegNet, which closely follow as the second-best alternatives in the overall adaptation results. The examples include diverse DA scenarios involving various robots and plant growth stages. The visuals highlight variations in soil colour among different robots and differences in plant growth stages across different years. Predictions from various methods closely align with the ground truth across different robots. However, the proposed method excels in predicting weed edges and capturing minor weeds during adaptations across different years as shown in Figure 3 (c) and (d).

4.2. Discussion

It is noted that the test performance of CBST is incredibly lower than the other methods, as shown in Table 1. Despite this, it is useful and efficient in segmenting traffic scenes in [59]. This discrepancy in performance could be attributed to the class imbalance in the weed dataset and the potential overfitting of the pseudo-labels. The overfitting is evident in Figure 4, where the mean IoU initially peaks but significantly reduces as the selection proportions increase during later stages. The error propagates with iterations when more pseudo-labels are integrated into the co-training process due to inevitable false labels. Thus, we oppose iterative co-training and propose a streamlined one-stage co-training in this work. However, it is challenging to determine the optimal selection proportion to mitigate overfitting and achieve optimal adaptation performance [20, 36, 38].

To mitigate the overfitting and determine the optimal selection proportions, we propose using covariance of the pixel distributions across classes before and after DA in this study. To highlight the effectiveness of this approach, we present visualizations of the covariance alongside the target mean IoU in Figure 5, using examples of bean plants adapted from BIPBIP to WeedElec and BIPBIP maize from 2021 to 2019. The covariance values are normalized to $[0, 1]$ for enhanced comparability. Despite some fluctuations in the curves, there is a consistent overall trend between the covariance and the mean IoU. As covariance increases, the target mean IoU decreases, and the optimal selection proportion aligns with the minimum covariance and the maximum target mean IoU.

Overfitting becomes pronounced when a substantial portion of pseudo-labels is incorporated into the co-training process. Illustrated in Figure 5, the adaptation of bean plants from BIPBIP to WeedElec reveals a reduction in the target mean IoU from 0.823 to 0.819 as the selection proportion increases to 50%. This decline primarily stems from including a considerable number of false pseudo-labels during co-training, misleading the model and adversely affecting its test performance. The magnitude of the selection proportion leading to overfitting varies across different adaptation scenarios, depending on the baseline. Taking the example of adapting from WeedElec to BIPBIP for maize with a baseline of 0.465, even a modest selection proportion of 20%, as depicted in Table 6, results in a reduction in mean IoU from 0.646 to 0.522. This discrepancy underscores the poor quality of initially selected pseudo-labels, contributing to the overfitting of false labels with a relatively large selection proportion and consequent degradation in the model's performance.

To enhance the performance of the target domain, especially when dealing with a baseline, we employ an iterative refinement strategy for pseudo-labels. This involves incorporating a modest selection proportion and performing co-training, then optimizing the selection proportion on the co-trained model instead of the initial base model. This choice is deliberate, as a conservative selection proportion not only guards against overfitting but also has the potential to enhance the overall quality of the pseudo-labels. By refining the selection proportion based on the co-trained model, we aim to elevate the performance of the target domain further.

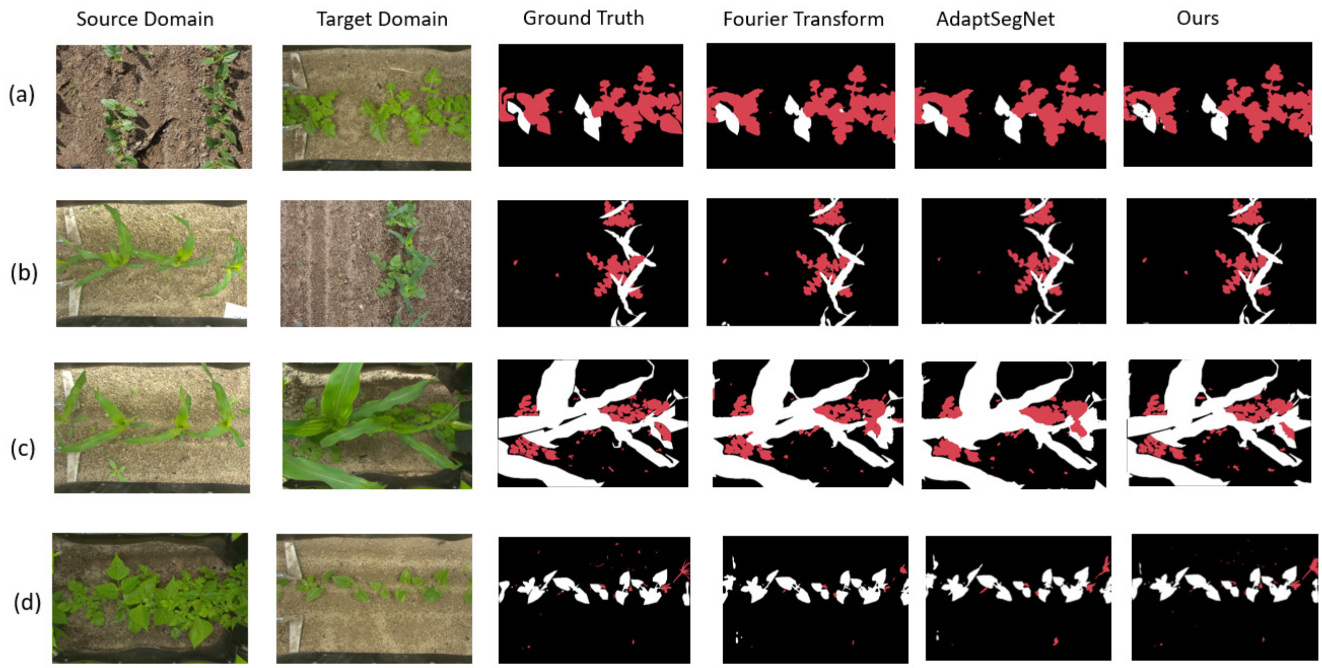

Figure 3. Comparisons of segmentation predictions performed with Fourier Transform, AdaptSegNet, and the proposed method across different robots and growth stages: (a) adaptation of beans from WeedElec to BIPBIP, (b) maize from BIPBIP to WeedElec, (c) maize collected with BIPBIP from 2019 to 2021, (d) beans collected with BIPBIP from 2021 to 2019.

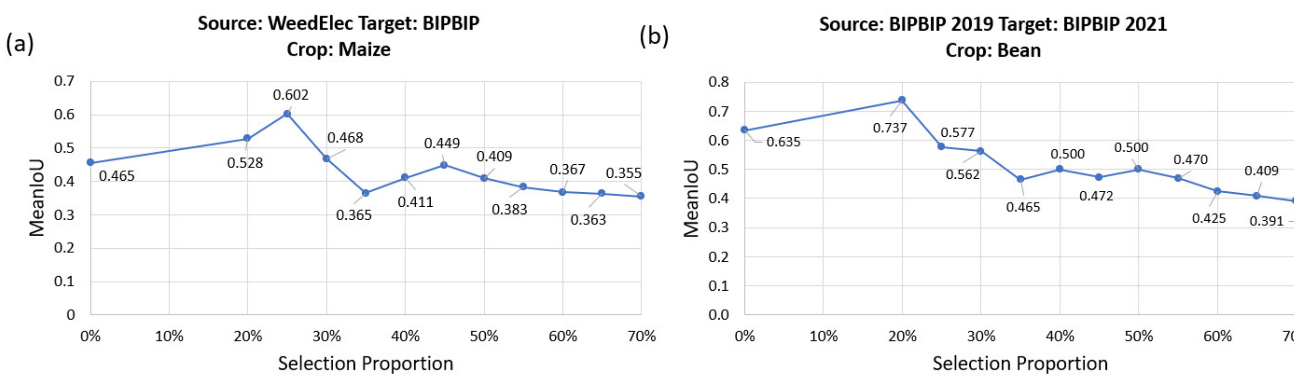

Figure 4. The covariance of the pixel count assigned to each class and the target mean IoU alongside the selection proportions for adaptations across different robots.

By adopting this approach, we maximize the potential of pseudo-labels and mitigate overfitting due to the base model's initially low performance, thereby improving the test performance. Illustrated in Figure 6 for adaptation from WeedElec to BIPBIP for maize, we initiate the co-training process with a modest selection proportion of 0.1, effectively preventing overfitting and enhancing the target performance from the baseline 0.465 to 0.646. Subsequently, the predicted labels generated by the co-trained model become the basis for subsequent iterations. Through this iterative refinement, we identify an optimal selection proportion of 0.3, marking a substantial boost in target performance from 0.646 to 0.734 for the mean softIoU.

5. Conclusions and future work

The major contribution of this paper is that we propose a novel DA method for weed segmentation based on pseudo-labelling. Our method seeks to optimize the selection pro-

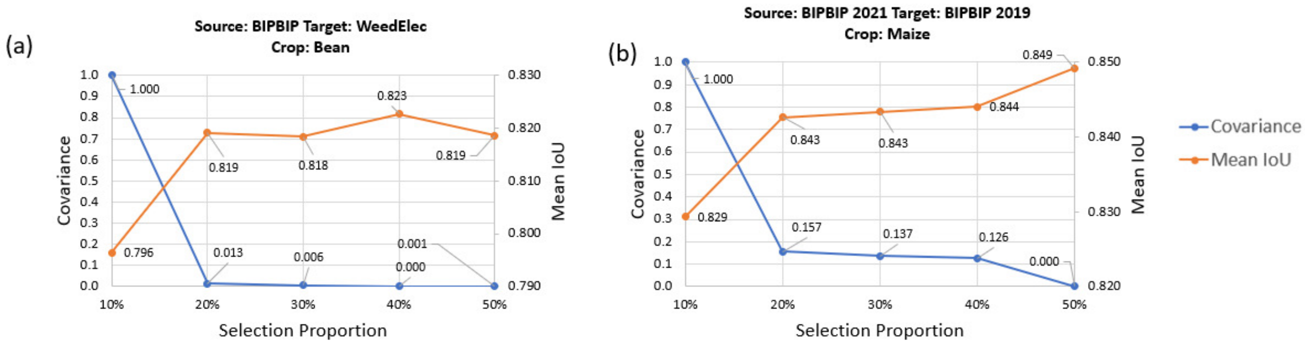

Figure 5. The covariance of the pixel count assigned to each class and the target mean IoU alongside the selection proportions for adaptations across different robots.

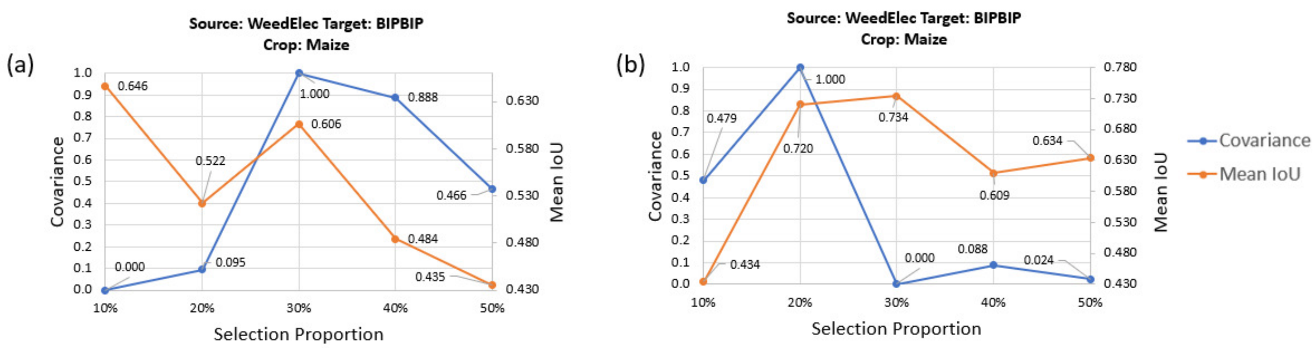

Figure 6. The comparison of covariance and target mean IoU for adaptation from WeedElec to BIPBIP for maize plants: (a) without pseudo-labels refinement, (b) with pseudo-labels refinement.

portions to maximize the gains of pseudo-labels with a one-stage co-training instead of iteratively increasing the selection proportions in the co-training process. Covariance is used to track the number of pixels assigned to each class to monitor the overfitting and optimize the selection proportions. To demonstrate the effectiveness of our method, we evaluate it using the ROSE challenge dataset, comparing its performance against the input-level adaptation method with the Fourier Transform, feature-level methods with CycleGAN and AdaptSegNet, and the popular output-level adaptation method with CBST. The results are highly promising, with the superior adaptability of our proposed method in the challenging conditions of plants from different years and varied stages, outperforming existing literature methods that fail to surpass the baseline.

However, the effectiveness of the proposed method is constrained by the quality of pseudo-labels derived from the base model, called baseline, trained with the source domain and performing prediction directly on the target domain without any adaptation. When the base model performs poorly, the test performance of the proposed method is constrained. For instance, in the adaptation of maize across different robots from WeedElec to BIPBIP, the base model yields a mean IoU of 0.465, limiting the test performance of the proposed method to 0.734. This performance is notably lower than other adaptations with higher base model performances.

We suggest considering the class imbalance in the co-training process to improve the adaptation scenarios with poor baselines. The images for weed segmentation are heavily imbalanced toward soil pixels and present a class-biased challenge. Using the same selection proportion for all classes may lead to model bias towards these large classes. In future work, we plan to customize the class selections in the co-training process to address this issue and enhance our model's performance. We will consider the large classes and class transfer differences, which may improve the accuracy and robustness of the segmentation results.

Acknowledgements. This research is supported by an NSERC Discovery grant (NSERC Ref: RGPIN-2021-04171) titled "Crop Stress Management using Multisource Data Fusion."

References

[1] Riccardo Bertoglio, Alessio Mazzucchelli, Nico Catalano, and Matteo Matteucci. A comparative study of fourier transform and cyclegan as domain adaptation techniques for weed segmentation. *Smart agricultural technology*, 4:100188, 2023. 1, 2, 4

[2] R. Bertoglio, A. Mazzucchelli, N. Catalano, and M. Matteucci. A comparative study of fourier transform and cyclegan as domain adaptation techniques for weed segmentation - code and data. 2023. 2, 4, 6

[3] Lei Chen, Jiaxian Zou, Yuan Yuan, and Haiyan He. Improved domain adaptive rice disease image recognition based on a novel attention mechanism. *Computers and Electronics in Agriculture*, 208:107806, 2023. 2

[4] Sentao Chen, Mehrtash Harandi, Xiaona Jin, and Xiaowei Yang. Domain adaptation by joint distribution invariant projections. *IEEE Transactions on Image Processing*, 29:8264–8277, 2020. 3

[5] Yue Linn Chong, Jan Weyler, Philipp Lottes, Jens Behley, and Cyrill Stachniss. Unsupervised generation of labeled training images for crop-weed segmentation in new fields and on different robotic platforms. *IEEE Robotics and Automation Letters*, 8(8):5259–5266, 2023. 2

[6] Zhelin Cui, Kanglong Li, Chunyan Kang, Yi Wu, Tao Li, and Mingyang Li. Plant and disease recognition based on pmf pipeline domain adaptation method: Using bark images as meta-dataset. *Plants*, 12(18), 2023. 2

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. 4, 5

[8] Emilio Dorigatti, Jann Goschenhofer, Benjamin Schubert, Mina Rezaei, and Bernd Bischl. Positive-unlabeled learning with uncertainty-aware pseudo-label selection, 2022. 3

[9] Lydia Elstone, Kin Yau How, Samuel Brodie, Muhammad Zulfahmi Ghazali, William P. Heath, and Bruce Grieve. High speed crop and weed identification in lettuce fields for precision weeding. *Sensors*, 20(2), 2020. 1

[10] Borja Espejo-Garcia, Nikos Mylonas, Loukas Athanasakos, Eleanna Vali, and Spyros Fountas. Combining generative adversarial networks and agricultural transfer learning for weeds identification. *Biosystems Engineering*, 204:79–89, 2021. 2

[11] J. Fan, B. Gao, H. Jin, and L. Jiang. Ucc: Uncertainty guided cross-head cotraining for semi-supervised semantic segmentation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9937–9946, Los Alamitos, CA, USA, jun 2022. IEEE Computer Society. 3

[12] Mulham Fawakherji, Ciro Potena, Ibis Prevedello, Alberto Pretto, Domenico D. Bloisi, and Daniele Nardi. Data augmentation using gans for crop/weed segmentation in precision farming. In *2020 IEEE Conference on Control Technology and Applications (CCTA)*, pages 279–284, 2020. 2

[13] Dario Gogoll, Philipp Lottes, Jan Weyler, Nik Petrinic, and Cyrill Stachniss. Unsupervised domain adaptation for transferring plant classification systems to new field environments, crops, and robots. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2636–2642, 2020. 1, 2, 4, 5, 6

[14] A S M Mahmudul Hasan, Ferdous Sohel, Dean Diepeveen, Hamid Laga, and Michael G.K. Jones. A survey of deep learning techniques for weed detection from images. *Computers and Electronics in Agriculture*, 184:106067, 2021. 1

[15] L. Hoyer, D. Dai, Y. Chen, A. Koring, S. Saha, and L. Van Gool. Three ways to improve semantic segmentation with self-supervised depth estimation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11125–11135, Los Alamitos, CA, USA, jun 2021. IEEE Computer Society. 3

[16] Yingchao Huang, Abdul Bais, and Amina E. Hussein. Domain adaptation using class-balanced self-paced learning for soil classification with libs. *IEEE Transactions on Plasma Science*, 51(9):2742–2755, 2023. 2

[17] Talha Ilyas, Jonghoon Lee, Okjae Won, Yongchae Jeong, and Hyongsuk Kim. Overcoming field variability: unsupervised domain adaptation for enhanced crop-weed recognition in diverse farmlands. *Frontiers in Plant Science*, 14, 2023. 2

[18] T Ilyas, J Lee, O Won, Y Jeong, and H Kim. Overcoming field variability: unsupervised domain adaptation for enhanced crop-weed recognition in diverse farmlands. *Frontiers in Plant Science*, 14:1234616, 2023. 2, 4, 6

[19] Shubham Innani, Prasad Dutande, Bhakti Baheti, Sanjay Talbar, and Ujjwal Baid. Fuse-pn: A novel architecture for anomaly pattern segmentation in aerial agricultural images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2960–2968, June 2021.

[20] Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondrej Chum. Label propagation for deep semi-supervised learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5065–5074, 2019. 3, 6

[21] Arnesh Kumar Issar, Aryan Mehta, Karan Uppal, Saurabh Mishra, and Kirtan Mali. [RE] FDA: Fourier domain adaptation for semantic segmentation, 2021. 2

[22] Manjunath Jogin, Mohana, M S Madhulika, G D Divya, R K Meghana, and S Apoorva. Feature extraction using convolution neural networks (cnn) and deep learning. In *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pages 2319–2323, 2018. 1

[23] Xiaofeng Liu, Chaehwa Yoo, Fangxu Xing, Hyejin Oh, Georges El Fakhri, Je-Won Kang, and Jonghye Woo. Deep unsupervised domain adaptation: A review of recent advances and perspectives. *APSIPA Transactions on Signal and Information Processing*, 11(1):–, 2022. 4

[24] Federico Magistri, Jan Weyler, Dario Gogoll, Philipp Lottes, Jens Behley, Nik Petrinic, and Cyrill Stachniss. From one field to another—unsupervised domain adaptation for semantic segmentation in agricultural robotics. *Computers and Electronics in Agriculture*, 212:108114, 2023. 2, 3

[25] Dennis K. J. Lin Mohsen Ebadi, Shojaeddin Chenouri and Stefan H. Steiner. Statistical monitoring of the covariance matrix in multivariate processes: A literature review. *Journal of Quality Technology*, 54(3):269–289, 2022. 2

[26] Amin Nasiri, Mahmoud Omid, Amin Taheri-Garavand, and Abdolabbas Jafari. Deep learning-based precision agriculture through weed recognition in sugar beet fields. *Sustainable Computing: Informatics and Systems*, 35:100759, 2022. 1

[27] Cheng Nong, Xiaohui Fan, and Jia Wang. Semi-supervised learning for weed and crop segmentation using uav imagery. *Frontiers in Plant Science*, 13:927368, Jul 2022. 2

[28] Chunshi Nong, Xijian Fan, and Junling Wang. Semi-supervised learning for weed and crop segmentation using uav imagery. *Frontiers in Plant Science*, 13, 2022. 2

[29] Isaac Kofi Nti, Adib Zaman, Owusu Nyarko-Boateng, Adebayo Felix Adekoya, and Frimpong Keyeremeh. A predictive analytics model for crop suitability and productivity with tree-based ensemble learning. *Decision Analytics Journal*, 8:100311, 2023. 2

[30] Cheng Ouyang, Carlo Biffi, Chen Chen, Turkay Kart, Huaqi Qiu, and Daniel Rueckert. Self-supervised learning for few-shot medical image segmentation. *IEEE Transactions on Medical Imaging*, 41(7):1837–1848, 2022. 3

[31] Lorena Parra, Jose Marin, Salima Yousfi, Gregorio Rincón, Pedro Vicente Mauri, and Jaime Lloret. Edge detection for weed recognition in lawns. *Computers and Electronics in Agriculture*, 176:105684, 2020. 1

[32] Rhorom Priyatikanto, Yang Lu, Jadu Dash, and Justin Sheffield. Improving generalisability and transferability of machine-learning-based maize yield prediction model through domain adaptation. *Agricultural and Forest Meteorology*, 341:109652, 2023. 2

[33] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. 4, 5

[34] Halil Mertkan Sahin, Tajul Miftahushudur, Bruce Grieve, and Hujun Yin. Segmentation of weeds and crops using multispectral imaging and crf-enhanced u-net. *Computers and Electronics in Agriculture*, 211:107956, 2023. 1

[35] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 4, 5

[36] Yuanshuang Sun, Yinghua Wang, Hongwei Liu, Liping Hu, Chen Zhang, and Siyuan Wang. Gradual domain adaptation with pseudo-label denoising for sar target recognition when using only synthetic data for training. *Remote Sensing*, 15(3), 2023. 3, 6

[37] Tanha Talaviya, Dhara Shah, Nivedita Patel, Hiteshri Yagnik, and Manan Shah. Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. *Artificial Intelligence in Agriculture*, 4:58–73, 2020. 2

[38] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. 3, 6

[39] Marco Toldo, Andrea Maracani, Umberto Michieli, and Pietro Zanuttigh. Unsupervised domain adaptation in semantic segmentation: A review. *Technologies*, 8(2), 2020. 2

[40] Marco Toldo, Umberto Michieli, and Pietro Zanuttigh. Unsupervised domain adaptation in semantic segmentation via orthogonal and clustered embeddings. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1357–1367, 2021. 2

[41] Gustavo J. Q. Vasconcelos, Thiago V. Spina, and Helio Pedrini. Low-cost domain adaptation for crop and weed segmentation. In João Manuel R. S. Tavares, João Paulo Papa, and Manuel González Hidalgo, editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 141–150, Cham, 2021. Springer International Publishing. 1, 2, 4, 6

[42] Hemanth Venkateswara and Sethuraman Panchanathan. *Introduction to Domain Adaptation*, pages 3–21. Springer International Publishing, Cham, 2020. 1

[43] Qi Wang, Junyu Gao, and Xuelong Li. Weakly supervised adversarial domain adaptation for semantic segmentation in urban scenes. *IEEE Transactions on Image Processing*, 28(9):4376–4386, 2019. 2

[44] Shengzhe Wang, Ling Wang, Zhihao Lin, and Xi Zheng. Temporal convolution domain adaptation learning for crops growth prediction, 2022. 2

[45] Jan Weyler, Thomas Läbe, Federico Magistri, Jens Behley, and Cyrill Stachniss. Towards domain generalization in crop and weed segmentation for precision farming robots. *IEEE robotics and automation letters*, 8(6):3310–3317, 2023. 2

[46] Xing Wu, Xiaohui Fan, Peng Luo, Sudipto Das Choudhury, Tardi Tjahjadi, and Cheng Hu. From laboratory to field: Unsupervised domain adaptation for plant disease recognition in the wild. *Plant Phenomics*, 5:0038, Mar 2023. 2

[47] Zhangnan Wu, Yajun Chen, Bo Zhao, Xiaobing Kang, and Yuanyuan Ding. Review of weed detection methods based on computer vision. *Sensors*, 21(11), 2021. 1

[48] Xiangyi Yan, Junayed Naushad, Shanlin Sun, Kun Han, Hao Tang, Deying Kong, Haoyu Ma, Chenyu You, and Xiaohui Xie. Representation recovering for self-supervised pre-training on medical images. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2684–2694, 2023. 3

[49] Lihe Yang, Lei Qi, Litong Feng, Wayne Zhang, and Yinghuan Shi. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In *CVPR*, 2023. 3

[50] Lihe Yang, Wei Zhuo, Lei Qi, Yinghuan Shi, and Yang Gao. St++: Make self-training work better for semi-supervised semantic segmentation. In *CVPR*, 2022. 3

[51] Qiangli Yang, Yong Ye, Lichuan Gu, and Yuting Wu. Msfcanet: A multi-scale feature convolutional attention network for segmenting crops and weeds in the field. *Agriculture*, 13(6), 2023. 1

[52] Y. Yang, D. Lao, G. Sundaramoorthi, and S. Soatto. Phase consistent ecological domain adaptation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*

(CVPR), pages 9008–9017, Los Alamitos, CA, USA, jun 2020. IEEE Computer Society. 2, 4, 5, 6

[53] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4084–4094, 2020. 2

[54] Yunlong Zhang, Changxing Jing, Huangxing Lin, Chaoqi Chen, Yue Huang, Xinghao Ding, and Yang Zou. Hard class rectification for domain adaptation. *Knowledge-Based Systems*, 222:107011, 2021. 2

[55] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 2

[56] Yi Zhu, Zhongyue Zhang, Chongruo Wu, Zhi Zhang, Tong He, Hang Zhang, R Manmatha, Mu Li, and Alexander J Smola. Improving semantic segmentation via efficient self-training. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2021. 3

[57] Adrian Ziegler and Yuki M. Asano. Self-supervised learning of object parts for semantic segmentation, 2022. 3

[58] Yang Zou, Zhiding Yu, B.V.K. Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. 3

[59] Yang Zou, Zhiding Yu, B. V. K. Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 297–313, Cham, 2018. Springer International Publishing. 2, 4, 6