Unsupervised Domain Adaptation for Transferring Plant Classification Systems to New Field Environments, Crops, and Robots

Dario Gogoll* Philipp Lottes* Jan Weyler Nik Petrinic Cyrill Stachniss

***Abstract*— Crops are an important source of food and other products. In conventional farming, tractors apply large amounts of agrochemicals uniformly across fields for weed control and plant protection. Autonomous farming robots have the potential to provide environment-friendly weed control on a per plant basis. A system that reliably distinguishes crops, weeds, and soil under varying environment conditions is the basis for plant-specific interventions such as spot applications. Such semantic segmentation systems, however, often show a performance decay when applied under new field conditions. In this paper, we therefore propose an effective approach to unsupervised domain adaptation for plant segmentation systems in agriculture and thus to adapt existing systems to new environments, different value crops, and other farm robots. Our system yields a high segmentation performance in the target domain by exploiting labels only from the source domain. It is based on CycleGANs and enforces a semantic consistency domain transfer by constraining the images to be pixel-wise classified in the same way before and after translation. We perform an extensive evaluation, which indicates that we can substantially improve the transfer of our semantic segmentation system to new field environments, different crops, and different sensors or robots.**

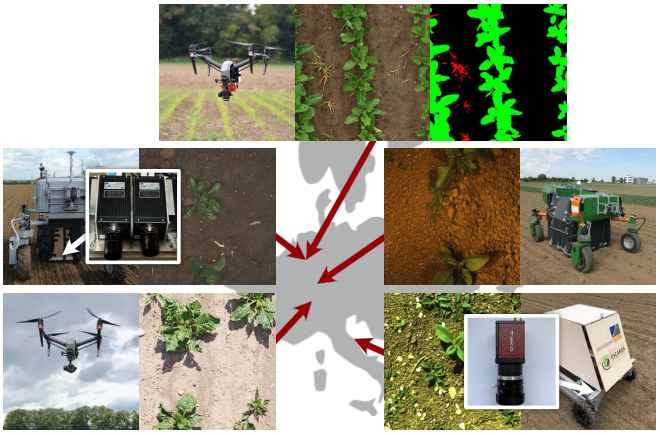

Fig. 1: Data acquisition under varying conditions, in different fields, and with different cameras and robots leads to highly distinctive image domains that challenge pre-trained plant classification systems to generalize well. We propose an effective approach for adapting existing systems to new environments, different crops, and other different field conditions.

I. Introduction

Crops make a substantial contribution to the production of food, feed, fuel, and fiber. Intensive crop production, however, has several negative impacts on our ecosystem, for example, through the massive application of agrochemicals. Farming robots have the potential to reduce negative impacts by a targeted, per-plant application of agrochemicals such as herbicides. Equipped with actuators, like selective sprayers, lasers, or mechanical tools, robots can enable selective and targeted treatments. Thus, robots may evolve to an effective and at the same time environment-friendly way to perform weed control.

Farming robots often rely on vision-based classification systems that distinguish between crops, weeds, and soil in real-time. They mostly use fully convolutional networks (FCNs) for semantic segmentation. These classification systems typically achieve a performance around 90+% when the trained classifiers are deployed in the same or at least similar field conditions [9], [13]. However, the performance of a classifier, which has been trained on a particular dataset, i.e., the *source domain*, suffers substantially when being deployed in new field environments or under changing conditions, i.e., in the *target domain*. This gap in performance between source and target domain is caused by the domain-shift between the domains. The domain-shift is affected by a different visual appearance, induced by different weed types, growth stages of plants, soil conditions, and illuminations.

Current approaches perform *supervised* domain adaptation of the classifiers to achieve a suitable performance on the target domain [9], [14]. Supervised retraining requires additional labels for new data from the target domain. In practice, however, we are often faced with scenarios where we solely have access to labeled images from the source domain and only *unlabeled* image data from the target domain, for example, when a robot enters a new field environment or is equipped with a new vision system. Thus, purely supervised approaches prevent the effective use of such classification systems at scale, due to the continuous label effort associated with domain changes.

In this paper, we aim at bridging the performance gap in visual crop and weed classification through transferring the visual classifier to the targeted domain without the need for an additional labeling effort. We target unsupervised domain adaptation towards an approach that enables us to train an FCN with suitable performance on the target domain while exploiting labels only from the source domain.

The main contribution of this work is an effective approach to unsupervised domain adaptation for plant segmentation systems in agriculture and thus adapt existing systems to new

*: authors with equal contribution.

D. Gogoll, J. Weyler, P. Lottes, and C. Stachniss are with the University of Bonn, Germany. P. Lottes and C. Stachniss are also with the Pheno-Inspect GmbH, Germany. N. Petrinic is with the University of Oxford, UK.

This work has partially been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy, EXC-2070 - 390732324 - PhenoRob.

environments, different value crops, and other farm robots. Our system yields a high segmentation performance in the target domain by solely using labeled RGB image data from the source domain and unlabeled RGB image data from the target domain. Our approach learns a mapping between the unpaired images from the source and target domain and is based on CycleGANs [20]. As CycleGANs give no guarantee to preserve the semantic information during the domain transfer, we enforce semantic consistency during domain transfer by constraining the images to be pixel-wise classified in the same way before and after translation.

In sum, we make the following claims: our approach (i) provides a solid performance for the semantic segmentation of crop, weed, and soil in the target domain, while not requiring extra labels from the target domain for the adaption of the classifier, (ii) outperforms CycleGANs and other baselines on the target domain for all tested datasets, (iii) allows to perform domain adaptation between different field environment, different crops, and different robots and camera setups. All claims are experimentally validated on real-world data.

II. Related Work

Semantic segmentation of agricultural field scenes is a key step to interpret the sensor data for reliable robot-based weed control. They classify every pixel in an image and determine the semantic class it belongs to. Over the past few years, CNN-based approaches [2], [9], [11], [14], [15], [16], [17] became the standard solution for this task, overcoming the requirement of handcrafted features [4], [12], [7].

In our previous work [9], we present an end-to-end trainable FCN that simultaneously estimates plant stem locations and a pixel-wise semantic segmentation of plants. This work implements a joint encoder for extracting image features and two task-specific decoders for the task. Building upon this, we enhance the approach for improved performance in new field environments by exploiting crop arrangement information of the field [10]. Potena et al. [16] propose a fast classification pipeline for accurate crop-weed identification. Based on RGB and near-infrared images, they use CNNs for binary vegetation detection in a first step and sequentially perform a CNN-based crop-weed classification only on vegetation pixels. The work by Milioto et al. [14] proposes to extend the RGB input by task-relevant background knowledge such as vegetation indices to generalize better to new field conditions. McCool et al. [13] use a mixture of lightweight deep CNNs to segment weeds in real-time with high accuracy. For reasons of efficiency and improved accuracy, they introduce a three-stage approach: tune a pre-trained model, compute a lightweight deep CNN by employing model compression techniques, and combine several lightweight models to form a mixture model that enhances the performance. The work, however, does not address a transfer between domains. Our approach proposed in this paper enables us to enhance any existing segmentation system by generalization capabilities while keeping the same runtime for real-time deployment. In contrast to all these works, we aim at improving the generalization capabilities of the classifier by actively considering the domain shift of the image data in an unsupervised way.

A variety of unsupervised domain adaptation approaches use generative adversarial networks (GANs) [3] for training classification systems based on synthetic images. Zhu et al. [20] introduce CycleGAN, an image-to-image translation approach without relying on aligned image pairs. They translate images from the source towards the target domain such that the translated images are indistinguishable from real images of the target distribution. Experiments on different scenarios like translating horses to zebras or paintings to real photos show outstanding qualitative results. We also employ CycleGAN within our approach to translate source images towards the target domain of new field environments. Since CycleGAN fails in properly preserving semantics, we extend the approach and additionally enforce a semantic consistency in the domain translation. Hoffman et al. [6] proposes CyCADA, a domain adaptation approach for semantic segmentation of urban scenes, using CycleGAN [20]. They enforce semantic consistency between the real and synthetic images of urban scenes. Chen et al. [1] present CrDoCo for domain transfer that includes cross-domain consistency in the output space during training the target segmentation network. Similarly to our approach, they employ an image-to-image translation module for translating images between source and target domain. Inspired by these approaches, we perform domain adaptation for agriculture field scenes with additional semantic constraints. We achieve a strong adaptation performance so that we can translate our semantic segmentation system to new field environments, different crops, and other robots and sensors.

III. Unsupervised Domain Adaptation

We propose an unsupervised domain adaptation approach that can adapt existing segmentation systems for crop-weed classification to new domains, still yielding a high classification performance. We exploit unpaired image sets from a source domain X and target domain Y as well as labels only from X. Our domain adaptation approach is based on CycleGANs [20] and learns the mapping between the source and the target domain in an unsupervised manner. Our approach consists of two domain-specific fully convolutional neural networks (FCNs) for semantic segmentation, two generator networks for domain adaptation, and two discriminator networks.

In a first step, we train the Source-FCN, i.e., the FCN for domain X, in a supervised way using RGB images and labels from X. Second, we perform the actual domain adaptation to train a Target-FCN, i.e., the FCN for Y, without requiring labels from Y. We achieve this by jointly training two generator networks G and F along with the Target-FCN. The generator networks G and F perform the translation of images into the style of the opposite domain, i.e., $G : X \rightarrow Y$ and $F : Y \rightarrow X$. During the training of the GAN, their respective discriminators D_X and D_Y aim to distinguish between real and generated images. Technically,

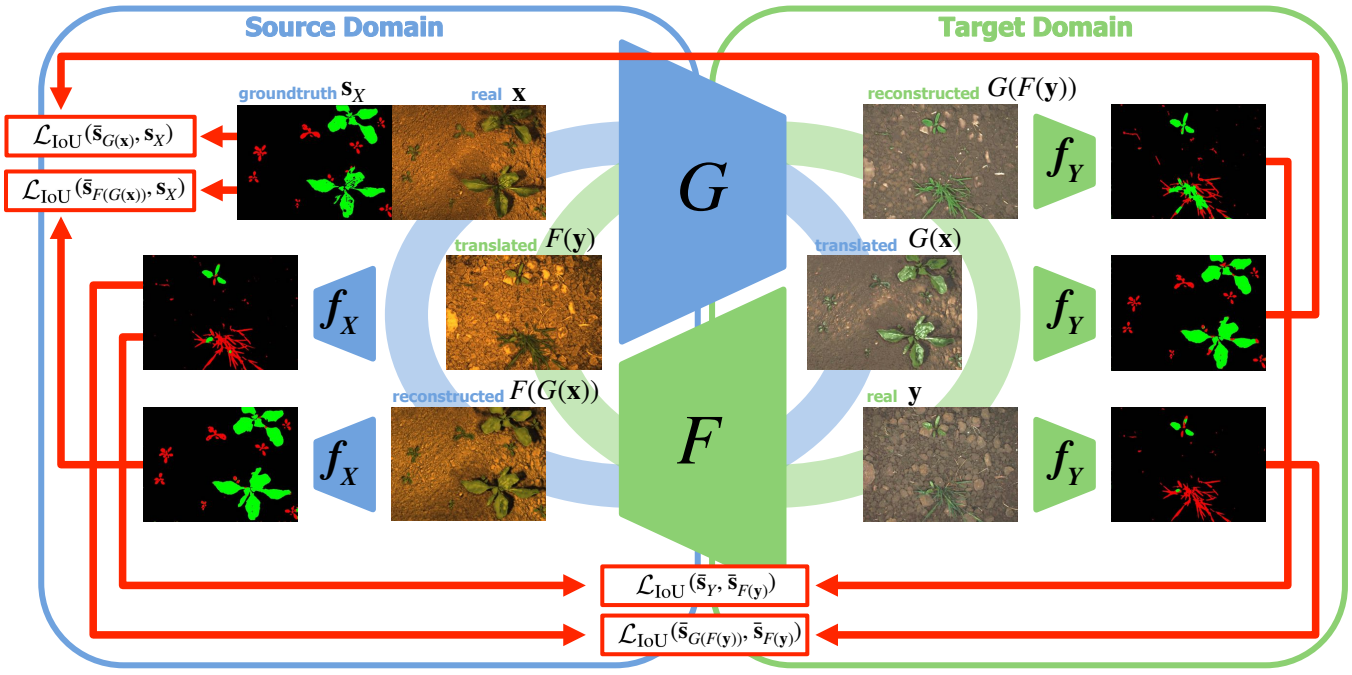

Fig. 2: Our approach consists of a Source-FCN f_X, Target-FCN f_Y and two generators G and F for mapping images in each direction. During training the semantic consistency domain adaptation we enforce the translated image $G(\mathbf{x})$ in the style of the target domain as well as the reconstructed image $F(G(\mathbf{x}))$ to be classified in the same way as it was before translation. Since we have no access to target labels, we use the prediction of the translated image $F(\mathbf{y})$ for enforcing semantic consistency in the opposite mapping direction.

the discriminator networks can be seen as dynamically-updated loss functions that train the generators to create images that are close to real images. At the same time, the generators try to deceive the discriminators for recognizing whether the image is real or generated [3].

We use the translated images $G(\mathbf{x})$ in the style of the target domain alongside with copied labels from the source domain to optimize the Target-FCN in a supervised manner. Thus, the generators must preserve semantic consistency such that the Target-FCN can be trained appropriately. Therefore, we propose a semantically consistent domain transfer by constraining the images to be classified in the same way before and after translation. As a result, our approach generates labeled images of the target domain that also enables us to retrain existing segmentation systems.

A. Domain-Specific FCNs

We differentiate two separate, domain-specific FCNs designed for the task of semantic segmentation. The Source-FCN classifies images that share the source domain distribution, whereas the Target-FCN works on images in the target domain. Both FCNs share the same network architecture. They take RGB images as input and output respective semantic segmentation maps, encoding a pixel-wise classification into crop, weed, and soil/background. The network architecture incorporates five fully convolutional building blocks based on the U-Net [19] architecture.

For training, we use a loss $\mathcal{L}_{\text{IoU}}$ approximating the intersection over union (IoU) metric. Plant pixels are typically under-represented concerning the amount of soil/background. The loss $\mathcal{L}_{\text{IoU}}$, however, is more stable with imbalanced

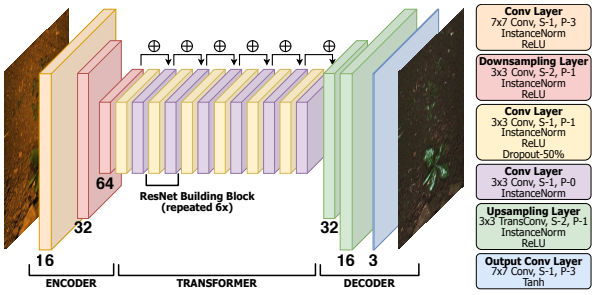

Fig. 3: The generators G and F work on RGB images and output translated RGB images in the opposite domain. The architecture consists of six ResNet building blocks [5], [20]. Numbers refer to the used filters in the layers.

class labels [18] and thus well-suited for our crop-weed classification. Additionally, we penalize errors for the weeds and crops by weighting $\mathcal{L}_{\text{IoU}}$ with a factor of four and two, respectively, as this weight factors led to the best segmentation performance within our hyperparameter search.

B. Unsupervised Domain Adaptation Exploiting Semantics

For the training, we simultaneously use source images $\mathbf{x} \in X$, target images $\mathbf{y} \in Y$, and pixel-wise labels $\mathbf{s}_X \in X$. In a forward pass, we translate the images between the domains within two cycles, see Fig. 2. For the first cycle (blue), G translates $\mathbf{x}$ into the style of the target domain. Then, F translates $G(\mathbf{x})$ back into the style of the source domain. Both generators share the same network architecture, as shown in Fig. 3, consisting of an encoder, a decoder, and a transformer of six ResNet building blocks [5], [20] that are responsible for mapping features between

both domains. Following Zhu et al. [20], we enforce a cycle consistency constraint $F(G(\mathbf{x})) \approx \mathbf{x}$ through an L1-loss $\mathcal{L}_{\text{cycle}} = ||F(G(\mathbf{x})) - \mathbf{x}||_1$ and apply an adversarial loss $\mathcal{L}_{\text{GAN}}(G, D_Y)$ as proposed by Goodfellow et al. [3]. Here, G tries to generate $G(\mathbf{x})$ that look similar to $\mathbf{y}$, while D_Y tries to distinguish between $G(\mathbf{x})$ and $\mathbf{y}$. The same procedure is performed for the second cycle (green in Fig. 3), but with the respective images and networks of the other domain. The two loss terms add up to equal shares from both cycles. We refer to the CycleGAN paper by Zhu et al. [20] for more details.

The problem with CycleGANs is that they give no guarantee to preserve the semantic information of $\mathbf{s}_X$ for the translated image $G(\mathbf{x})$. Semantic consistency between $G(\mathbf{x})$ and $\mathbf{s}_X$, however, is key to train the Target-FCN. Therefore, we propose additional semantic consistency constraints for training the domain adaption. The key idea is that all images of one cycle, i.e., $\mathbf{x}, G(\mathbf{x}), F(G(\mathbf{x}))$ (blue in Fig. 3) or $\mathbf{y}, F(\mathbf{y}), G(F(\mathbf{y}))$ (green in Fig. 3), should share the same semantic information. We include the two domain-specific FCNs into the training process and compute additional semantic losses that finally add to the overall training objective. We freeze the weights of the pre-trained Source-FCN f_X to keep stable predictions on images in the style of the source domain and initialize the Target-FCN f_Y according to the weights of the Source-FCN. During the training procedure, we jointly optimize the generators along with the Target-FCN towards a high classification performance in the target domain.

In our approach we propose to use two additional semantic losses for each cycle. Fig. 2 illustrates the additional loss terms for the first cycle (blue). First, we compute $\mathcal{L}_{\text{IoU}}(\bar{\mathbf{s}}_{G(\mathbf{x})}, \mathbf{s}_X)$ between the source labels $\mathbf{s}_X$ and the prediction of the Target-FCN on the translated image, i.e. $\bar{\mathbf{s}}_{G(\mathbf{x})} = f_Y(G(\mathbf{x}))$. This forces the generator to produce semantically consistent images in the target domain and serves as the training objective for the Target-FCN. Second, we force also the reconstructed images to be semantically aligned with the ground truth by $\mathcal{L}_{\text{IoU}}(\bar{\mathbf{s}}_{F(G(\mathbf{x}))}, \mathbf{s}_X)$, with $\bar{\mathbf{s}}_{F(G(\mathbf{x}))} = f_X(F(G(\mathbf{x})))$. This adds to the cycle consistency constraint, thus, supports the stability and convergence of the generator F. In case of the second cycle (green), we have no access to labels in the target domain. Thus, we constrain the predictions $\bar{\mathbf{s}}_Y = f_Y(\mathbf{y})$ as well as $\bar{\mathbf{s}}_{G(F(\mathbf{y}))} = f_Y(G(F(\mathbf{y})))$ to match with the prediction of the Source-FCN for the translated image, i.e. $\bar{\mathbf{s}}_{F(\mathbf{y})} = f_X(F(\mathbf{y}))$. We argue, that the Source-FCN provides stable predictions on images in the style of the source domain.

Thus, our proposed full semantic loss adds up to:

$$\begin{aligned}\mathcal{L}_{\text{semantic}}&(G, F, f_X, f_Y) \\ &= \mathcal{L}_{\text{IoU}}(\bar{\mathbf{s}}_{G(\mathbf{x})}, \mathbf{s}_X) + \mathcal{L}_{\text{IoU}}(\bar{\mathbf{s}}_{F(G(\mathbf{x}))}, \mathbf{s}_X) \\ &+ \mathcal{L}_{\text{IoU}}(\bar{\mathbf{s}}_Y, \bar{\mathbf{s}}_{F(\mathbf{y})}) + \mathcal{L}_{\text{IoU}}(\bar{\mathbf{s}}_{G(F(\mathbf{y}))}, \bar{\mathbf{s}}_{F(\mathbf{y})}).\end{aligned} \tag{1}$$

Consequently, the full objective composes to:

$$\begin{aligned}\mathcal{L}\,(G, F, D_X, D_Y, f_X, f_Y) &= \mathcal{L}_{\text{GAN}}\,(G, D_Y, F, D_X) \\ &+ \mathcal{L}_{\text{cycle}}(G, F) + \mathcal{L}_{\text{semantic}}(G, F, f_X, f_Y).\end{aligned} \tag{2}$$

C. Training

The goal of the training procedure is to obtain a Target-FCN, which performs well on the target domain. At this point, we assume to have already a trained expert for the source domain, i.e., the Source-FCN. We train the generators and discriminators along with the Target-FCN within our domain adaption procedure. In this paper, we train all models, including the domain-specific FCNs, for 400 epochs with an initial learning rate of 0.0002. We linearly decrease the learning rate towards zero after 100 epochs. We use the Adam [8] optimizer and randomly initialize the weights using the normal distribution $\theta_{\text{init}} \sim \mathcal{N}(0, 0.0002)$. For all semantic losses, we use the IoU-based loss as described in Sec. III-A. For regularization, we augment image data during the training. From each training image, we randomly sample a patch of size 240×240 px and rotate it randomly, keeping the original scale on which the Source-FCN was trained for. During test time, we do not conduct data augmentation but perform the inference image translation and classification on the original image size.

IV. Experimental Evaluation

We design the experiments to support our three claims: our approach (i) provides solid performance for crop-weed classification in new domains without requiring extra labels of the target domain for its adaption, (ii) outperforms CycleGANs and other baselines on the target domain, (iii) can transfer a classifier between different field environments, different crops, and different camera setups.

We carry out the experiments on eight different real-world datasets, which we collected with farm robots as well as UAVs. On the field robots, we used the cameras AD-130GE from JAI or the MAKO G-158 from Allied Vision. With UAVs, we used the Zenmuse X5s from DJI. All cameras provide RGB images at different image resolutions. We acquired all datasets such that the ground sampling distance is around $1\frac{\text{mm}}{\text{px}}$. Thus, scaling plays a minor role for the domain adaption. In total, we evolute our approach on 6.221 images containing sugar beets, sunflowers, different weed types, different growth stages, and different soil conditions. The datasets were collected under natural or artificial lighting conditions. Tab. I summarizes the key properties of the used datasets in our experiments.

We selected challenging scenarios for crop-weed classification in practice, where the performance typically suffers if no adaptation of the classifier is performed. First, we transfer a classifier for detecting sugar beets and weeds between different fields. Second, we transfer a classifier between the task of detecting sugar beets and sunflowers. Third, we transfer a classifier between two sugar beet fields, where we acquired the data with different camera systems.

For the assessment of the classifier as well as the domain adaption performance, we measure the average IoU across the classes crop, weed, and soil, i.e.,

$$\text{mIoU} = \frac{1}{C}\sum_{c=1}^{C} \text{IoU}_c \text{ with } c = \{\text{crop, weed, soil}\}, \tag{3}$$

TABLE I: Key properties of the datasets and pixel-wise classification results. We report the average IoU across crop, weed, and soil.

Source Domain							Target Domain					
Data	#Images	Crop	Leaf Stage	Camera	Robot	Upper Bound	Data	Vanilla	DiAda	CGAN	Ours	Upper Bound
Transfer between different fields.												
BONN	2148	Beet	4-8	JAI	BoniRob	81	STUTTGART	24	65	61	**72**	84
STUTTGART	665	Beet	2-8	JAI	BoniRob	84	BONN	11	57	65	**74**	81
UAV-BONN	380	Beet	4-12	ZX5s	Inspire-II	85	UAV-ZURICH	39	45	52	**61**	66
UAV-ZURICH	336	Beet	4-12	ZX5s	Inspire-II	66	UAV-BONN	42	54	61	**85**	85
Transfer between different crop types.												
SUGARBEET	305	Beet	4-6	JAI	BoniRob	76	SUNFLOWER	29	38	61	**67**	75
SUNFLOWER	97	Sunflower	4-6	JAI	Self-built	75	SUGARBEET	38	58	43	**70**	76
Transfer between different cameras / robots.												
MAKO	920	Beet	2	MAKO	Self-built	70	JAI	13	**49**	38	**49**	62
JAI	1370	Beet	2-4	JAI	BoniRob	62	MAKO	39	45	33	**62**	70

where the data association between the samples in the prediction and the ground truth is given by the pixel-wise alignment of the image data.

For comparison, we also evaluate different baselines. First, we naively deploy the trained Source-FCN on the target domain to estimate the expected classifier performance for a new domain without adaptation. We refer to this baseline with the term "vanilla". Next, we evaluate the original implementation of "CycleGAN" (CGAN). In this case, we perform the domain adaptation using CycleGAN and train the Target-FCN on the translated images along with the source labels. Finally, we perform "direct adaption" (DiAda). Here, we preprocess the source and target domain images considering the same channel-wise color correction. We treat each image and each channel c independently. We standardize the image by subtracting the channel mean μ_c and dividing by the channel-wise standard deviation σ_c:

$$\bar{\mathbf{x}}_{ij}^{c} = \frac{\mathbf{x}_{ij}^{c} - \mu_c}{\sigma_c}, \text{ with } c = \{\text{red, green, blue}\}. \quad (4)$$

Subsequently, we perform contrast stretching of the entire image, i.e.,

$$\tilde{\mathbf{x}}_{ij} = \frac{\bar{\mathbf{x}}_{ij} - \min(\bar{\mathbf{x}})}{\max(\bar{\mathbf{x}}) - \min(\bar{\mathbf{x}})}, \quad (5)$$

After color correction, we train the Source-FCN using the preprocessed images from the source domain and deploy the classifier on the preprocessed images from the target domain (analog to the vanilla approach). The direct adaption entails a shift of the intensity distribution towards a common intermediate distribution for both domains. Thus, the pixel intensities of the RGB images are distributed more similarly which may reduce the domain-shift. In our previous publication [11], we showed that the direct adaptation of both domains leads to a substantial increase in the performance of the Target-FCN.

To better understand the performance of the transferred classifiers on the target domain, we additionally evaluate the performance of the domain-specific FCNs on the domain, on which we trained them explicitly. This gives us information about the theoretically possible performance if the classifier had access to the labels of the target domain during the training phase. In other words this performance reflects an upper boundary for the expected generalization performance of the transferred classifiers.

A. Transfer Between Different Fields

The first set of experiments is designed to show that our approach can transfer a crop-weed classifier between different field environments and provides a reliable performance in the target domain without requiring additional labeled data for its adaption. We consider four datasets recorded in different field environments. We collected the datasets BONN and STUTTGART with the DeepField Robotics BoniRob platform as depicted in Fig. 1 (left, middle), and the datasets UAV-BONN and UAV-ZURICH with the DJI Inspire II UAV. All datasets contain sugar beets at 2-12 leaf growth stages and serve a substantial amount of different weeds types. The STUTTGART dataset has roughly 15% of the crops and weeds overlapping in the images. Fig. 4 depicts an example image for the respective datasets. The RGB images reveal the different domains that we face in this experiment.

We perform one classifier transfer between the BoniRob datasets BONN and STUTTGART and one between the UAV datasets UAV-BONN and UAV-ZURICH. We perform the transfers in both directions. Tab. I summarizes the segmentation performance obtained by our approach and the considered baselines. First, we see that the vanilla baseline fails on all target domains indicating the need for any kind of domain adaptation methods. Our approach substantially outperforms all other baselines. As a noteworthy result, our transferred Target-FCN between the UAV datasets UAV-ZURICH to UAV-BONN achieves the same mIoU concerning the upper boundary of around 85% mIoU. For the reverse experiment as well as the transfer between the BONN and STUTTGART datasets, our approach obtains around 8% less mIoU compared to the upper boundary. Compared to the CycleGAN baseline, our approach gains around 13% and 18% concerning the direct adaptation baseline.

Fig. 4 reveals excellent crop-weed classification results for our approach, whereas the Target-FCN trained with the CycleGAN approach produces more errors between all

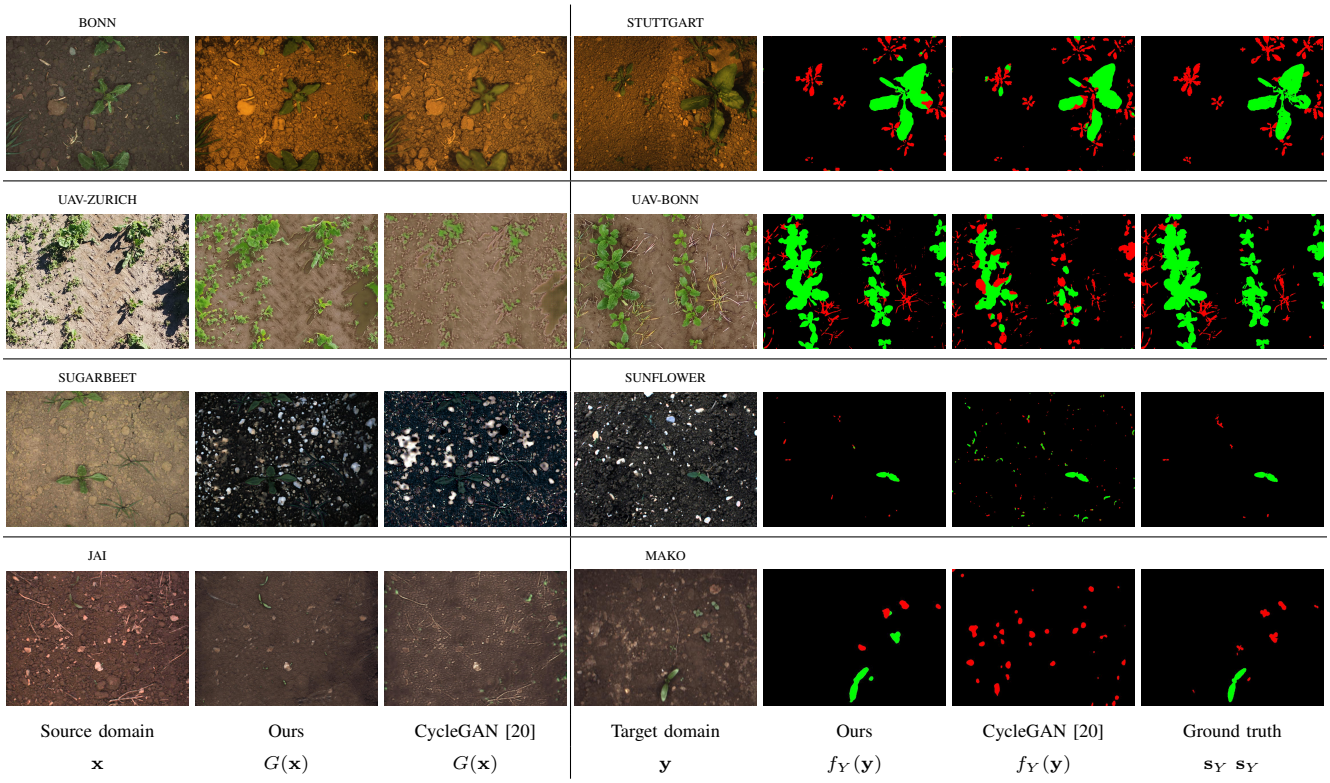

Fig. 4: Qualitative results. Our approach exploiting semantics provides a better translation into the target domain compared to CycleGANs. Our approach preserves fine structures and properly transfers the semantic information in a pixel-wise manner. The CycleGAN approach suffers from missing semantic information. It wrongly translates pixels that belong to small vegetation objects or fine structures. For the sake of brevity, we show only one direction and neglect to visualize the results based on the direct adaption baseline.

considered classes. The inspection of the translated images $G(\mathbf{x})$ for the BONN and UAV-BONN datasets illustrates a solid domain adaption to the target domain. Our visual inspection of the entire test datasets shows that our approach reliably generates images in the style of the target domain while keeping the semantics of the source domain images. In case of the CycleGAN baseline, the generators miss details in the texture as well as the semantic information for small vegetation objects and fine structures like grass-weeds. This in turn leads to a weaker crop-weed classification in the target domain, since the target FCN is trained with lower quality image material. The direct adaption baseline mostly suffers from a confusion between the classes crop and weed by the Target-FCN.

These results demonstrate the ability of our approach for transferring an FCN for crop-weed classification to different field environments without the need of extra labels from the target domain. The different growth stages of plants and weeds as well as the exposure conditions seem to have little influence on performance. Furthermore, these results convey that using our proposed semantic consistency constraints allow our approach to adequately preserve the semantic information between the source and target domain.

B. Transfer Between Different Crop Types

The second set of experiments is designed to show that our approach can transfer a classifier between sugar beets and sunflowers. Please note that the challenge in this transfer is less in the appearance of the crop plants themselves but rather in the substantially different appearance of the overall fields. We consider the SUGARBEET dataset, which we recorded with the BoniRob on a field near Bonn, Germany, and the SUNFLOWER dataset, which we recorded with a self-built robot near Ancona, Italy. Both datasets contain crop plants at 4-6 leaf growth stage and 14 weed species in total. Compared to the previous experiment, the soil conditions differ more and both datasets contain small weeds of size 0.5-4 cm^2.

Tab. I summarizes the obtained segmentation performance. Our approach achieves the best generalization capabilities to a new crop and outperforms all baselines. Our approach achieves 67% mIoU for the SUNFLOWER and 70% mIoU for the SUGARBEET dataset, which is around 7% less mIoU compared to the achievable upper boundary. Neither the CycleGAN nor the direct adaptation baseline perform on a comparable level. The direct adaptation fails on the SUNFLOWER dataset due to a wrong segmentation of the vegetation. The white gravel is often classified as plant or weed. Compared to that, the CycleGAN performs better. It can generate the basic style of the target domains, thus, making the Target-FCN aware of the new soil conditions. This underlines the need for active domain adaptation considering specific characteristics of both domains. However, the CycleGAN approach is not able to properly translate small weeds in a sufficient quality. As a result, the Target-FCN predicts most of the vegetation in the respective target domains as crop plants.

The qualitative results in Fig. 4 illustrate that our approach

performs better for those small vegetation objects. It correctly preserves the semantic information during translation into the target domain. Most of the remaining error between our approach and the theoretical upper bound is caused by wrongly classified plant and weed contours.

C. Transfer Between Different Cameras

The last set of experiments is designed to show that our approach is able to transfer a classifier between two different robots that employ different cameras. The MAKO dataset is recorded with the MAKO-G 158 camera under natural lighting conditions on a field near Ulm, Germany, whereas the JAI dataset is recorded with the BoniRob using the JAI-AD-130GE camera on a field near Zurich, Switzerland. Most of the vegetation in these datasets is of size 0.2-4.0 cm^2. These datasets are the most challenging ones as the obtained upper bound performance already suggests, see Tab. I. Our inspection of the upper bound performance reveals that the FCN classifiers have problems with distinguishing very small sugar beets and weeds at a size of 0.2 cm^2-2.0 cm^2.

Also in this experiment our approach outperforms the other baselines approaches, except for the direct adaptation. Both our approach and DiAda obtain 49% mIoU for the transfer of the MAKO to the JAI data. For the reverse run, our approach achieves 62% mIoU on the MAKO data outperforming vanilla, direct adaptation, and CycleGAN. Concerning all experiments, the good performance for the direct adaptation on the JAI data can be treated as an outlier. On average, our approach obtains 11% less mIoU compared to the upper boundary. The performance loss is mainly caused by wrong prediction of small plants and weeds.

Noteworthy is the low performance obtained by CycleGAN. CycleGANs can not preserve the semantic information for small plants during the image translation. The wrong translation prevents a correct adjustment of the Target-FCN to the MAKO data. The qualitative results in Fig. 4 illustrate for the JAI to MAKO experiment, that the CycleGAN maps pixels from the actual soil class to vegetation pixels in the target domain, whereas our approach translates the semantic information correctly. This result confirms the beneficial properties for our domain adaptation exploiting the semantic information induced by the source labels.

V. Conclusion

In this paper, we presented an unsupervised domain adaptation approach to the problem of pixel-wise crop-weed classification. Our approach enables to train an FCN that achieves a solid performance in changing domains such as new field environments, different crops, and different sensors or robots, while exploiting labeled data only from a source domain. Our approach learns a mapping between unpaired images from the source and target domain by exploiting cycle as well as semantic consistency constraints. Our extensive evaluation demonstrates that we outperform CycleGANs and other baselines in changing domains and substantially improve the generalization capabilities of crop-weed classification systems. We believe that this approach is an important step towards the real-world deployment of agricultural robots at scale.

References

[1] Y. Chen, Y. Lin, M. Yang, and J. Huang. CrDoCo: Pixel-level Domain Transfer with Cross-Domain Consistency. *arXiv preprint*, abs/2001.03182v1, 2020.

[2] M. Di Cicco, C. Potena, G. Grisetti, and A. Pretto. Automatic Model Based Dataset Generation for Fast and Accurate Crop and Weeds Detection. *arXiv preprint*, abs/1612.03019v3, 2016.

[3] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. *arXiv preprint*, abs/1406.2661, 2014.

[4] S. Haug, A. Michaels, P. Biber, and J. Ostermann. Plant classification system for crop / weed discrimination without segmentation. In *Proc. of the IEEE Winter Conf. on Applications of Computer Vision (WACV)*, pages 1142–1149, 2014.

[5] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[6] J. Hoffman, E. Tzeng, T. Park, J. Zhu, P. Isola, K. Saenko, A.A. Efros, and T. Darrell. CyCADA: Cycle-Consistent Adversarial Domain Adaptation. *arXiv preprint*, abs/1711.03213v3, 2017.

[7] A. Jumpasut, N. Petrinic, B. Elliott, C. Siviour, and M. Matthew. An error analysis into the use of regular targets and target detection in image analysis for impact engineering. *Journal Applied Mechanics and Materials*, 13-14:203–210, 01 2008.

[8] D.P. Kingma and J.Ba. Adam: A method for stochastic optimization. *arXiv preprint*, abs/1412.6980, 2014.

[9] P. Lottes, J. Behley, N. Chebrolu, A. Milioto, and C. Stachniss. Joint Stem Detection and Crop-Weed Classification for Plant-specific Treatment in Precision Farming. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2018.

[10] P. Lottes, J. Behley, N. Chebrolu, A. Milioto, and C. Stachniss. Robust joint crop-weed classification and stem detection using image sequences. *Journal of Field Robotics (JFR)*, 37(1):20–34, 2020.

[11] P. Lottes, J. Behley, A. Milioto, and C. Stachniss. Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. *IEEE Robotics and Automation Letters (RA-L)*, 3:3097–3104, 2018.

[12] P. Lottes, M. Höferlin, S. Sander, and C. Stachniss. Effective Vision-based Classification for Separating Sugar Beets and Weeds for Precision Farming. *Journal of Field Robotics (JFR)*, 34:1160–1178, 2017.

[13] C.S. McCool, T. Perez, and B. Upcroft. Mixtures of Lightweight Deep Convolutional Neural Networks: Applied to Agricultural Robotics. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2017.

[14] A. Milioto, P. Lottes, and C. Stachniss. Real-time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2018.

[15] A.K. Mortensen, M. Dyrmann, H. Karstoft, R. N. Jörgensen, and R. Gislum. Semantic Segmentation of Mixed Crops using Deep Convolutional Neural Network. In *Proc. of the Intl. Conf. of Agricultural Engineering (CIGR)*, 2016.

[16] C. Potena, D. Nardi, and A. Pretto. Fast and accurate crop and weed identification with summarized train sets for precision agriculture. In *Proc. of Int. Conf. on Intelligent Autonomous Systems (IAS)*, 2016.

[17] A. Pretto, S. Aravecchia, W. Burgard, N. Chebrolu, C. Dornhege, T. Falck, F. Fleckenstein, A. Fontenla, M. Imperoli, R. Khanna, F. Liebisch, P. Lottes, A. Milioto, D. Nardi, S. Nardi, J. Pfeifer, M. Popovic, C. Potena, C. Pradalier, E. Rothacker-Feder, I. Sa, A. Schaefer an R. Siegwart, C. Stachniss, A. Walter, V. Winterhalter, X. Wu, and J. Nieto. Building an Aerial-Ground Robotics Systemfor Precision Farming. *IEEE Robotics & Automation Magazine*, 2020.

[18] M. A. Rahman and Y. Wang. Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation. In *Int. Symp. on Visual Computing*, 2016.

[19] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *arXiv preprint*, abs/1505.04597, 2015.

[20] J. Zhu, T. Park, P. Isola, and A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. of the IEEE Intl. Conf. on Computer Vision (ICCV)*, pages 2223–2232, 2017.