Computers and Electronics in Agriculture 212 (2023) 108114

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

From one field to another—Unsupervised domain adaptation for semantic segmentation in agricultural robotics

Federico Magistri [a,*], Jan Weyler [a], Dario Gogoll [a], Philipp Lottes [a], Jens Behley [a], Nik Petrinic [b], Cyrill Stachniss [a,c,d]

[a] University of Bonn, Photogrammetry and Robotics Lab, Nussallee 15, 53115 Bonn, Germany
[b] University of Oxford, Impact Engineering Lab, Oxford OX1 2JD, UK
[c] Lamarr Institute for Machine Learning and Artificial Intelligence, Germany
[d] University of Oxford, Department of Engineering Science, Oxford OX13PJ, UK

ARTICLE INFO

Dataset link: https://www.ipb.uni-bonn.de/html/projects/SemCUT_DomainAdaption_2022/data.zip, https://www.ipb.uni-bonn.de/html/projects/SemCUT_DomainAdaption_2022/code.zip

Keywords:
Domain adaptation
Semantic segmentation
Deep learning
Generative adversarial networks

ABSTRACT

In traditional arable crop fields, tractors treat the whole field uniformly applying large quantities of herbicides and pesticides for weed control and plant protection. Autonomous robots, instead, offer the potential to provide a per-plant treatment, thus turning weed control and plant protection environment-friendly. To this end, an autonomous robot has to reliably distinguish crops, weeds, and soil under a diverse range of environmental conditions using its onboard sensors. Such recognition ability forms the basis for targeted plant-specific treatments in the form of spot applications. Basically, all such perception systems used today rely on some form of machine learning technique. However, current learning-based solutions often show a performance decay when applied under new field conditions. This is a major bottleneck for real-world application and finally commercial adoption. In this paper, we propose a simple yet effective approach to unsupervised domain adaptation for semantic segmentation systems so that an existing segmentation pipeline can be adapted to different fields, different robots, and different crops. Our system yields a high segmentation performance in new target fields without the need for extra manual annotations. It exploits only annotations from the source domain, i.e., the original field used for training the robot's vision system. Our thorough evaluation shows that our approach achieves high accuracy when transferring an existing segmentation system to different environmental conditions, different plant species, and different robotic systems.

1. Introduction

Crops are a fundamental part of the production of food, feed, fuel, and fiber and thus a key pillar for our society. Current intensive crop production makes use of massive applications of agrochemicals, causing a negative impact on our ecosystem. Agricultural robots have the potential to revolutionize the standard practice (Asseng and Asche, 2019). Shifting from uniform agrochemicals application to a per-plant application, or even using alternative weeding tools to mechanically treat individual plants or use lasers, the use of agrochemicals can be reduced substantially (Khanna et al., 2022; Walter et al., 2017). Thus, robots may evolve to an effective and at the same time environment-friendly way to perform weed control (Pretto et al., 2020).

To achieve such a goal, autonomous robots must have a reliable vision system, preferably based on standard RGB cameras, and a suite of actuators like selective sprayers, lasers, or mechanical weeding tools to enable selective and targeted treatments. Typically, the vision system is responsible for a real-time classification system that distinguishes between crops, weeds, and soil. Today, such systems often use convolutional neural networks (CNN) (LeCun et al., 2015; Krizhevsky et al., 2017; LeCun et al., 1989) to predict a semantic mask where each pixel is assigned to a class.

In recent years, such CNN-based systems became the standard solution for robotic vision tasks in both crop fields (Zenkl et al., 2021; Lottes et al., 2018b; Mortensen et al., 2016; Barreto et al., 2021; Zhang et al., 2020; Milioto et al., 2018) and orchards environment (Kerkech et al., 2018; Zabawa et al., 2020; You et al., 2022), overcoming the requirement for handcrafted features (Haug et al., 2014; Roscher et al., 2014; Lottes et al., 2017; Jumpasut et al., 2008). These CNN-based classification systems typically achieve notable performances when they are trained and deployed in the same, or at least similar, field conditions (Hu et al., 2021).

The semantic segmentation task is a fully supervised one, meaning that it requires a large amount of image-label pairs to train the

* Corresponding author.
E-mail address: federico.magistri@igg.uni-bonn.de (F. Magistri).

https://doi.org/10.1016/j.compag.2023.108114
Received 16 December 2022; Received in revised form 22 June 2023; Accepted 24 July 2023
Available online 11 August 2023

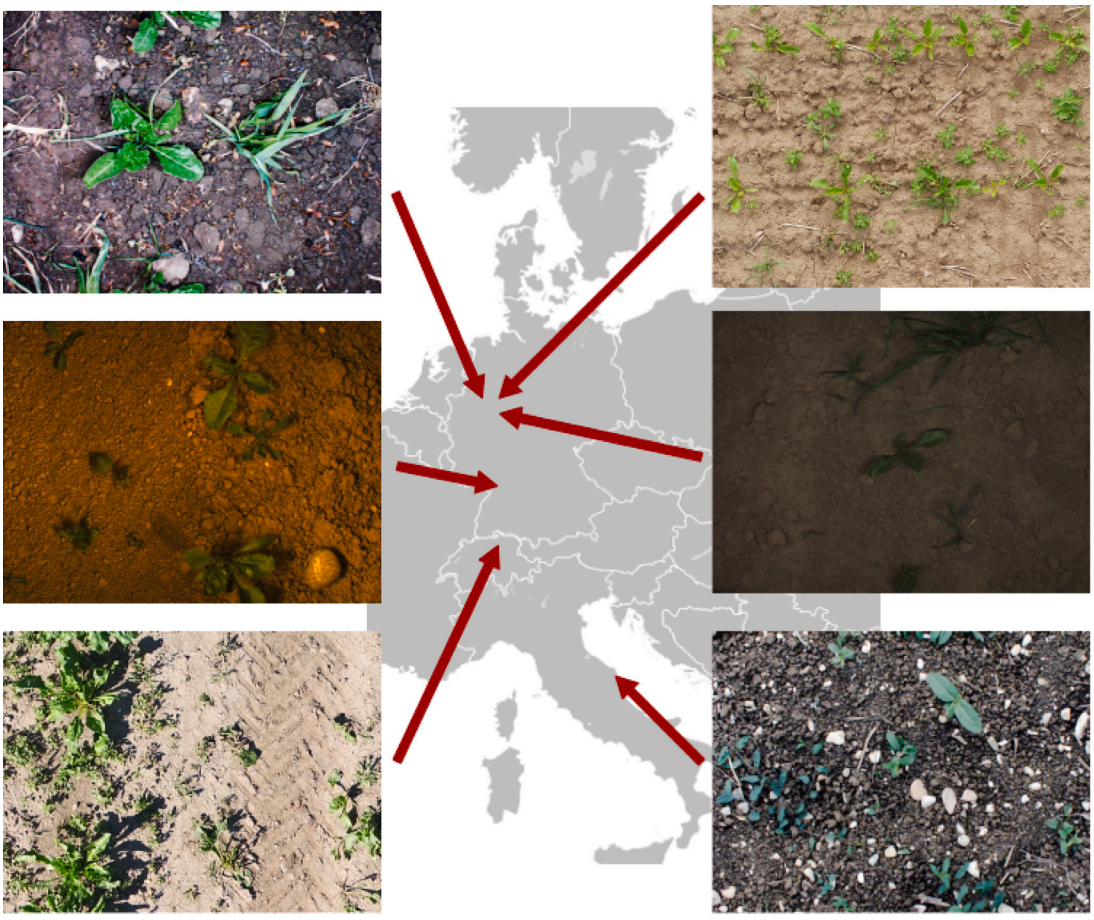


Fig. 1. Locations and examples of collected data. We use data collected on different agricultural fields spread through Europe. We show a single image from each of the dataset we used in this paper to appreciate the different environmental conditions that an autonomous robot has to face.

classification system. With labels being a major bottleneck, as they are typically time-consuming involving substantial manual labor.

While such systems provide accurate and robust prediction when deployed in environments similar to the environment used for training, the ability of CNNs to provide accurate prediction in the new environment is often unsatisfactory. Such behavior is mainly caused by the different characteristics between the training and testing environment. Fig. 1 showcases common examples of the different environmental conditions in which an agricultural robot should operate. In the following, we refer to the training environment as the source domain, while the target domain refers to the testing environment. This allows us to define as domain shift the difference between the source and target domain.

In the context of broad acre fields, several factors contribute to the domain shift. To name a few of these factors: different weed types can be present in different fields, the growth stage of plants may be different as well the soil conditions, and natural or artificial light conditions that vary between the fields. To close the gap between different domains, most contemporary methods use a fully supervised domain adaptation strategy to adapt the classifiers to attain a suitable performance in the targeted domain. However, supervised retraining or adaptation requires additional labels for novel data from the targeted domain. In practice, however, we often encounter scenarios where we have labeled images from the source domain and only have raw, unlabeled images from the target domain available that have been acquired during field operation of the robot in the new environment. Consequently, such a purely supervised approach to domain adaptation would prevent us from effectively use of such classification systems at scale due to the continuous labeling effort caused by domain changes.

We are not the first ones to suffer from such a domain shift and different approaches have been proposed to tackle this problem. In our previous conference paper (Lottes et al., 2018a), for example, we used a channel-wise combination of Gaussian smoothing and contrast stretch to improve generalization capabilities of our end-to-end trainable CNN to jointly estimate plant stem positions in the image plane and a pixel-wise semantic segmentation of crop and weeds. Building upon this, we exploit the structure of typical crop fields to improve the performance of our semantic segmentation CNN on unseen field (Lottes et al., 2020). In this paper, we propose a joint encoder for extracting image features that are then decoded by two task-specific decoders, one for semantic segmentation and one for domain transfer.

Other researchers take alternative paths, Potena et al. (2016) propose a pipeline based on RGB and near-infrared images. They first perform a binary segmentation to divide pixels into vegetation and soil, then they use a CNN to classify each vegetation pixel as either crop or weed. This approach, however, does not work with regular cameras due to the missing near-infrared channel that was used to identify vegetation pixels. Vasconcelos et al. (2021) propose to process images from both, source and target, domains with a contrast-limited adaptive histogram equalization, followed by a replacement between low-frequency amplitudes of source and target domains obtained with the fast Fourier transform. Cicco et al. (2017) propose to generate labeled dataset with a computer graphic engine. They can generate realistic images with the style of the target dataset. However, with this solution, there is still the need to generate one synthetic dataset for each target domain. Milioto et al. (2018) propose to use task-relevant information, namely vegetation indices, in addition to RGB information as input to the CNN to achieve better generalization capabilities on new field conditions. McCool et al. (2017) propose to use a mixture of lightweight CNNs for real-time crop/weed segmentation. Their

pipeline has three stages: they first train a large model, afterward they extract from it different lightweight CNNs using model compression techniques, and finally, they form a mixture model by combining the lightweight CNNs previously extracted, this lead to enhancement in performances while having low inference time. The work, however, does not address a transfer between domains. Blok et al. (2022) use an active learning strategy to select which images are better to annotate to maximize performances while reducing labeling efforts. In contrast to the aforementioned studies, our goal is to improve the generalization capabilities of semantic segmentation networks by explicitly addressing the domain shift present in images collected under different conditions without the need for labeling any single image of the target field. Wu et al. (2023) propose an unsupervised domain adaptation method for plant disease classification via uncertainty regularization, without the need for adversarial training. Kwak and Park (2022) suggest a multi-stage unsupervised adaptation method to classify crop types in satellite images. Our work is different as we need pixel-level adaptations with ground sampling distance around 1 $\frac{\text{mm}}{\text{px}}$ to accurately segments crops and weeds in images. In this article, we propose an approach that enables us to transfer any existing semantic segmentation CNN to new field conditions without the need for extra labeling efforts.

With the raise of generative adversarial networks (GANs) (Goodfellow et al., 2014), a diverse number of unsupervised domain adaptation methods exploit synthetically generated images for training semantic segmentation systems. Zhu et al. (2017) and Park et al. (2020) propose approaches to transfer the style of a source image into the style of a target image without changing the content of the source image without relying on matched image pairs. They transfer the style of images from the source towards the target domain such that the generated images are visually indistinguishable from real images of the target dataset. Experiments on different scenarios like translating summer conditions to winter conditions or real photos into the painting style of famous artists show outstanding qualitative results. Hoffman et al. (2018) propose CyCADA, an adaptation approach for semantic segmentation of urban scenes, building on top of CycleGAN (Zhu et al., 2017). They include a semantic consistency term in the loss function to help the CNN in retaining semantic content while transferring from synthetic to real images. Chen et al. (2019) propose CrDoCo a pixel-wise domain adaptation approach to have the same segmentation results on real and generated images.

Similarly to Gogoll et al. (2020), our goal is to tackle the domain shift in arable crop fields. To solve such problem, they first train a semantic segmentation network on the source domain in a fully supervised fashion. Afterward, they set up a cycle-consistent GAN (Zhu et al., 2017) where generated images from source to target domain are fed into a semantic segmentation network that is trained in parallel to the cycle-consistent GAN. The previously trained semantic segmentation network is additionally used to provide semantic predictions on generated images from the target to the source domain. In this way, is it possible to establish a cycle-consistent semantic loss. After convergence, the generated images are then used to train the semantic segmentation network for the target domain. In total, this leads to three different training stages where six networks have to work together. (two generators, two discriminators, and two semantic segmentation networks.) Bertoglio et al. (2023) extend such an approach by adding a constraint, in the form of an additional loss term, on the image phase to further improve semantic preservation under the assumption that phase component of an image contains information about its semantics, while the amplitude carries information about its style. Our proposed approach is different in two ways. First, we do not need to train a neural network in advance to successfully translate images from source to target domain. Second, our approach is able to translate images keeping semantic information without the need for a cycle-consistent loss term, making the translation simpler. Our proposed solution only exploits one generator, one discriminator, and one semantic segmentation network. We achieve a competitive adaptation performance on semantic segmentation tasks to different fields, different robots, and different crops.

In this paper, we bridge the performance gap in visual crop and weed segmentation between source and target without additional labeling effort. We aim at providing an unsupervised domain adaptation approach that enables us to train a CNN to attain suitable performance in the targeted domain, but we only use labels from the source domain. The main contribution of this work is an effective approach for unsupervised domain adaptation for plant segmentation in agriculture and thus we adapt existing systems to novel domains and environments, potentially with different value crops, but also acquired with different robots. Our proposed pipeline achieves a high segmentation performance in the targeted domain using labeled RGB images from the source domain and unlabeled RGB images from the target domain. In summary, we claim the following: our approach (i) attains a suitable performance for the semantic segmentation of crop, weed, and soil in the target domain without the need for extra labels from the target domain for the adaption of the segmentation approach, and (ii) allows to perform domain adaptation between different field environments, differences in the crops, and also robots used to acquire the data.

2. Material and methods

A domain shift is a change in the data distribution between source and target dataset. Typically, CNN will perform well on the source data while failing on the target data if the domain shift is large. We propose an unsupervised domain adaptation approach yielding a high performance in crop-weed-soil segmentation to new field conditions. Our domain adaptation approach exploits a contrastive loss to replace the cycle consistency for the style transfer. Together with the style transfer, we learn to segment images from the source domain by sharing the weights among the two tasks, see Fig. 2. In this setting, our approach consists of a single GAN, namely a generator–discriminator pair, and one domain specific CNN for semantic segmentation on the target dataset. Thus, we are halving the complexity compared to the approach by Gogoll et al. (2020).

2.1. Generative adversarial network

A generative adversarial network or, in short, GAN, is a system of two neural networks competing against each other, where one network, called the generator, is trained to produce realistic looking images and the second network, the discriminator, is trained to recognize which images are real and which are generated. We use a GAN to generate images with the content of the source domain and the style of the target domain. Thus, the GAN is the first building block to close the gap between performances on source and target domain. In the domain adaption setting, a GAN learns to generate data from the source domain such that an adversarial discriminator is unable to distinguish the different domains. By mapping samples of the source domain into the target domain, we enable our model to learn on source data while still generalizing to target data. The generator G learns to map images from source to target dataset, $G : \mathcal{X}_s \rightarrow \mathcal{X}_t$, at the same time the discriminator D tries to distinguish between real images from the target domain $\mathcal{X}_t$ and generated images from the source domain $\hat{\mathcal{X}}_s$. This can be formalized with and adversarial loss:

$$\mathcal{L}_{\text{GAN}} = \min_G \max_D \mathbb{E}[\log D(\mathcal{X}_t)] + \mathbb{E}[1 - \log D(G(\mathcal{X}_s))] \tag{1}$$

Intuitively, the discriminator tries to maximize both terms, namely correctly classifying each image in real or fake, while the generator tries to fool the discriminator by minimizing the second term. In our approach, the generator is a CNN based on 9 ResNet (He et al., 2016) blocks. The discriminator is a simple sequence of 4 convolutional layers.

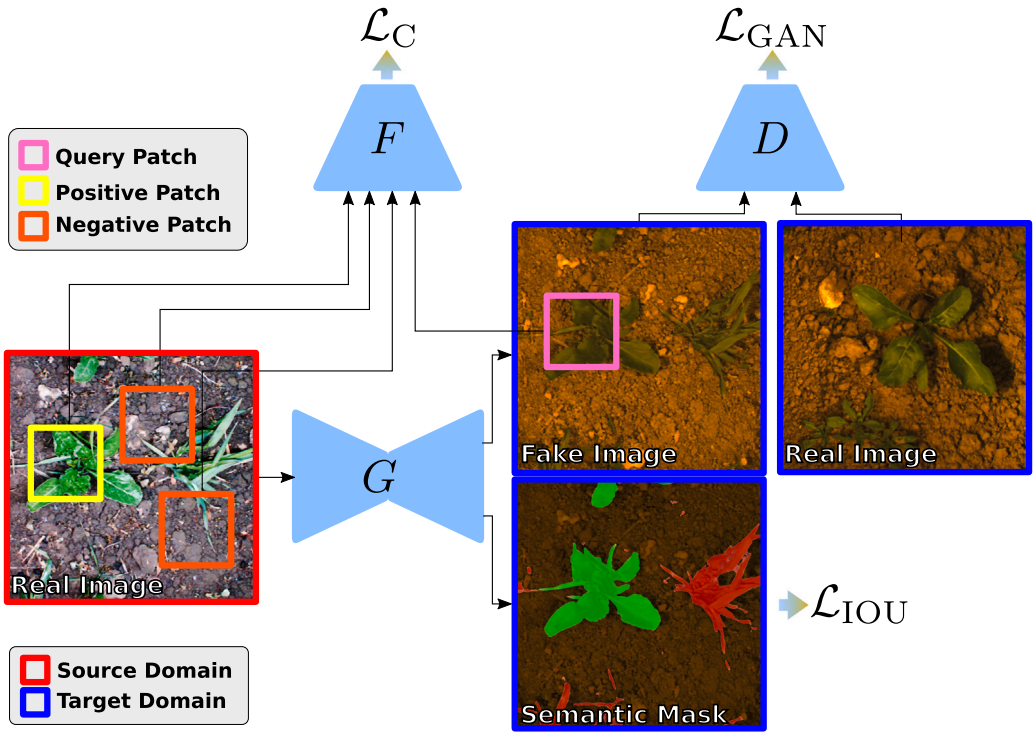


Fig. 2. Method overview. Overview of our domain adaptation approach. Our proposed solution is a single GAN, i.e. a generator–discriminator pair. Our Generator G takes as input images from the source domain and outputs a translated image and a semantic mask. We, then, use the translated images to compute the GAN and contrastive loss. In the first case, the discriminator, D, tries to differentiate generated images from the source domain and real images from the target domain. In the second case, we use patches extracted at different locations in real source images and generated source images. We feed these patches to a network, F, consisting of the encoder of the generator with the addition of a two-layer MLP. In this way, we can compute the contrastive loss at different layers. The semantic mask is used to define a semantic segmentation loss exploiting source labels.

2.2. Contrastive style transfer

The GAN loss previously defined is able alone to generate images that share high level characteristics, such as light conditions and soil type, with the target domain. However, such loss is not able to preserve pixel-wise details. In our case, this could lead to losing small plants completely or discriminative plant features such as plant veins or reflectance. To tackle such issue, we use a patch-wise contrastive strategy originally proposed by Park et al. (2020). During training, for each source image $\mathcal{X}_s$, we define a positive patch $\mathbf{p}^+$ and a set of N negative patches $\mathbf{p}_n^-$. After generating $\hat{\mathcal{X}}_s$, we extract the corresponding positive patch, namely the query patch $\mathbf{q}$. We can implement it, by extracting such negative patches $\mathbf{p}_n^-$ at random locations in the source image, $\mathcal{X}_s$. We, then, extract the query patch and the positive patch, $\mathbf{q}$ and $\mathbf{p}^+$, from the same pixel locations in the generated image, $\hat{\mathcal{X}}_s$, and in the source image, $\mathcal{X}_s$, respectively. We can now set up a noise contrastive estimation framework (Oord et al., 2018) to maximize mutual information between input and output. Each patch is mapped to a M-dimensional vector using the encoder of the generator network with the addition of a two-layer MLP. We denote such vectors as $\mathbf{v}^+$, $\mathbf{v}_n^-$, $\mathbf{v}^q$. In the embedding space define by the generator encoder and the MLP, we now want $\mathbf{v}^+$ and $\mathbf{v}^q$ to be closer and clearly separated from each negative patch $\mathbf{v}_n^-$. This can be formalized with a cross entropy:

$$\mathcal{L}_C = -\log\left[\frac{\exp(\mathbf{v}^q \cdot \mathbf{v}^+/\tau)}{\exp(\mathbf{v}^q \cdot \mathbf{v}^+/\tau) + \sum_{n=1}^{N} \exp(\mathbf{v}^q \cdot \mathbf{v}_n^-/\tau)}\right], \tag{2}$$

where the parameter τ scales the distances between the query vector and other vectors and all vectors are normalized. In such a loss, the dot product measures the similarity between vectors. As one would expect from a constrastive loss, Eq. (2) has a low value when the query vector $\mathbf{v}^q$ is similar to the positive vector $\mathbf{v}^+$ and it is not similar to $\mathbf{v}^-$. Additionally, the patch loss just define can computed also on features maps of intermediate layers and using positive patches at different spatial location.

2.3. Sharing semantic features

While the constrastive loss can preserve details from source images $\mathcal{X}_s$ to generated images $\hat{\mathcal{X}}_s$, there is no guarantee that the generated images share semantic information with the source images. This happens more frequently when there is a dominant class in both datasets. In our application, the soil represents a large part of the images while crops and weeds being less frequent, especially the latter. Such a class unbalance typically leads to a wrong adaptation, meaning that in the generated images we can have regions with the appearance of soil while being annotated as crop/weed or vice versa. Such failure cases are a crucial aspect for domain adaptation approaches, given that we use the source labels $\mathcal{Y}_s$ paired with the transformed images $\hat{\mathcal{X}}_s$. The resulting inconsistency between visual appearance and semantic annotations leads to inaccurate segmentation results on the target domain $\mathcal{X}_t$. To enforce a semantic consistency between source and generated images, we note that pixel belonging to different classes should undergo different transformations. To this end, we can exploit the annotations of the source domain $\mathcal{Y}_s$. Our idea is that we can generate images that share the semantic content of the source domain while being visually similar to the target domain by jointly learning: (i) to semantically segment images from the source domain and (ii) to translate image from source to target domain. Formally speaking, our generator G takes as input an image from the source domain and outputs a translated image $\hat{\mathcal{X}}_s$ together with a semantic mask $\mathcal{P}$. The translated image contributes to the losses previously defined, $\mathcal{L}_{GAN}$ and $\mathcal{L}_C$. We, then, use the semantic mask to define a loss between the source labels $\mathcal{Y}_s$ and the semantic masks:

$$\mathcal{L}_{IOU} = 1 - \frac{i(\mathcal{P}, \mathcal{X}_s)}{u(\mathcal{P}, \mathcal{X}_s)}, \tag{3}$$

with:

$$i(\mathcal{P}, \mathcal{X}_s) = \sum_{c \in C} \mathcal{P}^c * \mathcal{X}_s^c, \tag{4}$$

and

$$u(\mathcal{P}, \mathcal{X}_s) = \sum_{c \in C} \mathcal{P}^c + \mathcal{X}_s^c - \mathcal{P}^c * \mathcal{X}_s^c. \quad (5)$$

In Eqs. (4) and (5), the variable c refers to a class belonging to a set of classes C, where $C = \{\text{soil, crop, weed}\}$ in our use case, and the symbol $*$ represents the element-wise multiplication. This loss approximates and maximizes the intersection over union (IOU) and it is particularly suited for unbalanced classes (Rahman and Wang, 2016), fitting well our datasets given that the soil is over-represented in all of them.

2.4. Overall loss function

We optimize the weighted sum of the previously defined terms:

$$\mathcal{L} = w_{gan}\mathcal{L}_{\text{GAN}} + w_c\mathcal{L}_{\text{C}} + w_{iou}\mathcal{L}_{\text{IOU}}. \quad (6)$$

Thus, during training, the GAN loss, $\mathcal{L}_{\text{GAN}}$, is responsible for the high-level style-transfer, soil type, plants color, light conditions and so on. The contrastive loss, $\mathcal{L}_{\text{C}}$, refines low-level details such as plants boundaries or cracks or rock in the terrain, while the semantic loss, $\mathcal{L}_{\text{IOU}}$, preserves semantic consistency between source and generated images.

2.5. Domain-specific CNN

Once we generate the transformed images $\hat{\mathcal{X}}_s$, we pair them with the source label $\mathcal{Y}_s$ to train the semantic segmentation model that we will deploy on the target domain $\mathcal{X}_t$. For this task, we use ERFNet (Romera et al., 2018). It takes RGB images as input and output respective semantic segmentation mask, encoding a pixel-wise classification into crop, weed, and soil. For training, we use the loss defined in Eq. (3) approximating the IOU. This loss is more stable with imbalanced class labels and thus well-suited for our crop-weed segmentation problem where plant pixels (crop or weed) are typically under-represented with respect to the amount of soil pixels. We, additionally, use a class weighting scheme to tackle the heavy class imbalance in our datasets.

3. Results

We consider the problem of unsupervised adaptation, where we are provided source data $\mathcal{X}_s$, source labels $\mathcal{Y}_s$, and target data $\mathcal{X}_t$, but no target labels. The objective is to generate a set of images, $\hat{\mathcal{X}}_s$, which share the content of $\mathcal{X}_s$ while having similar appearances of $\mathcal{X}_t$. Thus, using $\hat{\mathcal{X}}_s$ and $\mathcal{Y}_s$, being able to learn a model f that can correctly predict the label for the target data $\mathcal{X}_t$ without the need for target labels. We provide experiments to support our claim the following: our approach (i) attains a suitable performance for the semantic segmentation of crop, weed, and soil in the target domain without the need for extra labels from the target domain for the adaption of the segmentation approach, and (ii) allows to perform domain adaptation between different field environments, differences in the crops, and also robots used to acquire the data. All claims are experimentally validated on real-world data.

3.1. Experimental setup

In our experiments, we use dataset collected under different degrees of domain shift to show the performance of our approach in typical real-worlds scenarios for agricultural robots. In total, we perform our experiments on six different real-world datasets, which we collected with ground robots and UAVs. We acquired all datasets such that the ground sampling distance is around 1 $\frac{\text{mm}}{\text{px}}$. In total, we evaluate our approach on 3916 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.

Table 1
Dataset overview. Key characteristics of the dataset used in this article.

Name	# Images	Crop	Leaf stage	Camera	Robot
UGV-Bonn	2148	Sugar Beet	4–8	JAI	BoniRob
UGV-Stuttgart	665	Sugar Beet	2–8	JAI	BoniRob
UAV-Bonn	379	Sugar Beet	4–12	ZX5s	Inspire-II
UAV-Zurich	336	Sugar Beet	4–12	ZX5s	Inspire-II
Sunflower	83	Sugar Beet	4–6	JAI	Self-built
Sugarbeet	305	Sunflower	4–6	JAI	BoniRob

Table 1 summarizes the key properties of the used datasets in our experiments. Additionally, we provide few sample images from each dataset to show the variations of the used datasets in Fig. 3.

We evaluate our domain adaptation strategy by computing semantic segmentation metrics on the target domain when using the generated images as training set. Specifically, we report: (i) the mean intersection over union (IoU) over the three considered semantic classes (crop, weed, soil) and (ii) the per-class precision and recall. We additionally compute the Frechet inception distance (FID) introduced by Heusel et al. (2017) to capture the similarity between generated and real images.

For each experiment, we report the semantic segmentation results on the target domain when using only the source domain as training set. We refer to this approach as vanilla from now on. Furthermore, we show the results on the target domain when using only the target domain as training set. This approach can be seen as an upper boundary as it corresponds to the fully supervised setting. Furthermore, to better evaluate our model, called SemCUT, we use different baselines. CycleGAN (Zhu et al., 2017) and CUT (Park et al., 2020) that do not use labels from the source domain but also Sem-Cycle-GAN (Gogoll et al., 2020), which exploits source labels during the training of the adaptation network. We additionally compare our approach to DUA by Mirza et al. (2022), a non-generative approach that updates the statistics of the batch normalization layers to adapt a model trained on the source domain to the target domain. We describe training details together with qualitative results for generated images and semantic segmentation masks in the supplementary material.

3.2. Translating between locations (Bonn vs. Stuttgart) using UGV data

In the first experiment, we use two datasets of sugar beet fields, the firs was collected near Bonn, Germany, and the second was collected near Stuttgart, Germany. Besides the different visual aspect of the field due to different environmental conditions, the dataset collected in Stuttgart uses a different artificial lighting source that was not used when collecting the dataset in Bonn, which leads to a different visual appearance as shown in Fig. 3. In the Bonn dataset, the sugar beet show a number of leaves between 4 and 8, while the Stuttgart dataset has a number of leaves between 2 and 8. We collected both datasets using the BoniRob robot from DeepFields Robotics (Ruckelshausen et al., 2009). The robot is equipped with a AD-130GE camera produced by JAI.

For both adaptation directions, our proposed solution provides the highest mean IoU reaching above 72% in both directions while CUT and Sem-Cycle-GAN yield competitive performances only in one direction, see Table 3. In both directions, the higher mean IoU obtained by our approach is due to the higher values in the precision metrics, while the recall is on pair with the other baselines, Fig. 4. Interestingly, CUT obtains lowest FID values in both datasets, see Table 2, but this does not translate to better segmentation results.

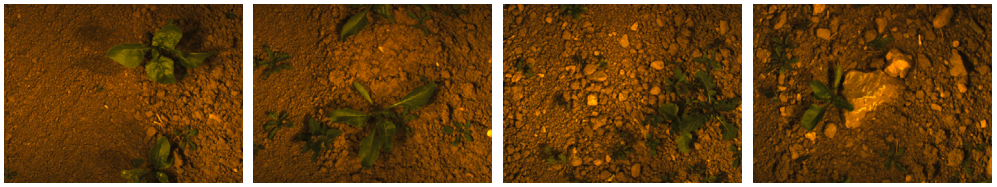
3.3. Translating between locations (Bonn vs. Zurich) using UAV data

Similar to the first experiment, we collected two datasets flying over sugar beet fields with the same UAV. We collect one dataset again in Bonn the other in Zurich, Switzerland. We collected both datasets with

Table 2
FID. Comparison of data similarity between target doman and generated images. We also report the FID score between source and target domain. Interestingly, the FID score does not correlate with better segmentation performances on the target domain.

Datasets		Approach				
Source	Target	Cycle-GAN Zhu et al. (2017)	CUT Park et al. (2020)	Sem-Cycle-GAN Gogoll et al. (2020)	SemCUT (ours)	
UGV-Bonn	UGV-Stuttgart	244.33	189.11	**67.44**	96.88	91.15
UGV-Stuttgart	UGV-Bonn	244.33	90.65	**26.85**	76.91	80.94
UAV-Bonn	UAV-Zurich	171.36	286.90	**71.03**	93.93	75.04
UAV-Zurich	UAV-Bonn	171.36	169.46	80.49	**76.27**	94.63
Sunflower	Sugarbeet	421.51	364.56	**84.13**	222.2	108.21
Sugarbeet	Sunflower	421.51	**117.13**	125.81	123.13	126.90

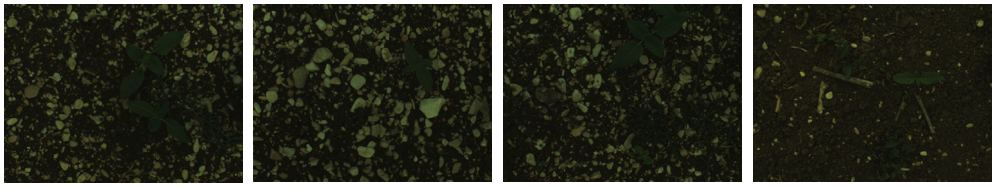
(a) UGV-Bonn



(b) UGV-Stuttgart

(c) UAV-Bonn

(d) UAV-Zurich



(e) Sunflower

(f) Sugarbeet

Fig. 3. Example images of the different datasets. We show a few sample for each dataset used in this paper.

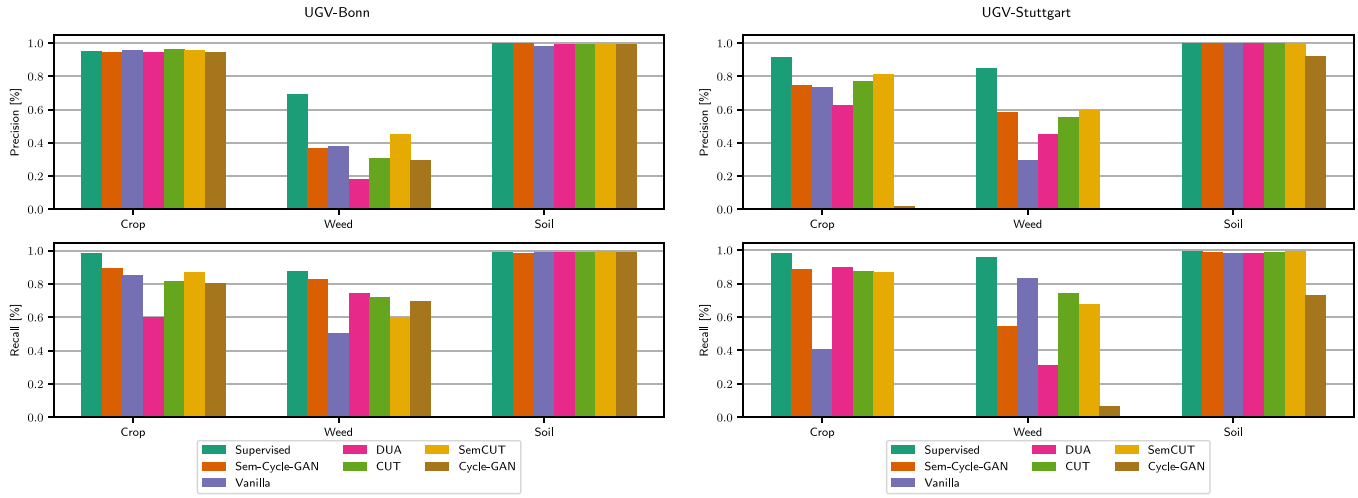


Fig. 4. UGV-Bonn vs UGV-Stuttgart. Per-class precision and recall.

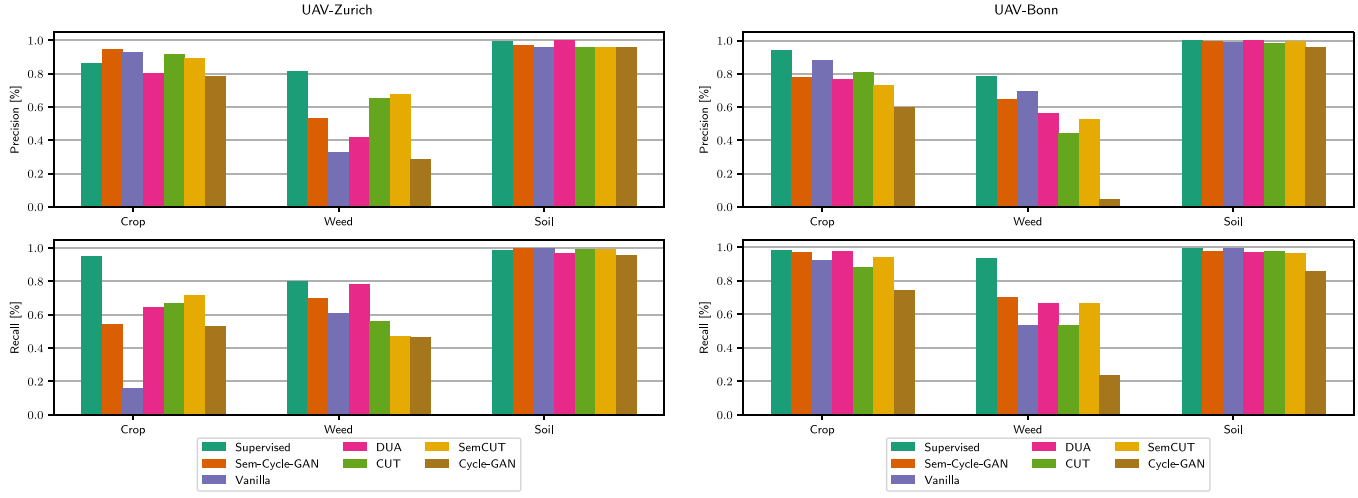


Fig. 5. UAV-Bonn vs UAV-Zurich. Per-class precision and recall.

Table 3
Mean IoU in [%]. Semantic segmentation results after traning the ERFNet using as training set images generated with different adaptation approaches.

Datasets		Approach						
Source	Target	Vanilla	DUA Mirza et al. (2022)	Cycle-GAN Zhu et al. (2017)	CUT Park et al. (2020)	Sem-Cycle-GAN Gogoll et al. (2020)	SemCUT (ours)	Supervised
UGV-Bonn	UGV-Stuttgart	53.94	59.95	23.19	71.59	68.69	**72.71**	90.36
UGV-Stuttgart	UGV-Bonn	69.04	58.01	67.04	68.42	72.53	**72.59**	85.46
UAV-Bonn	UAV-Zurich	46.01	63.10	45.45	66.97	**74.44**	69.05	88.51
UAV-Zurich	UAV-Bonn	**74.42**	71.83	53.12	67.05	64.11	66.47	82.54
Sunflower	Sugarbeet	53.31	49.46	51.96	54.06	**56.85**	52.23	88.74
Sugarbeet	Sunflower	32.01	46.23	59.27	**70.43**	64.87	63.26	81.12

the DJI Inspire 2 drone equipped with a DJI Zenmuse X5s camera. Both datasets have a number of leaves between 4 and 12, thus presenting a large diversity regarding the aspects of the crops.

As reported in the previous experiment, we can see the same uncorrelation between FID values and semantic segmentation results. In fact, while CUT has the lowest FID value from Bonn to Zurich (71.03%), the approach proposed by Gogoll et al. (2020) obtains a higher mean IoU, reaching 74.44%. Considering the adaptation from Zurich to Bonn, the Sem-Cycle-Gan has the lowest FID value 76.27%, while we report the best segmentation performances when no adaptation is involved, reaching a 74.42% of mean IoU with the vanilla approach, with particularly high precision values as can be seen in Fig. 5. We believe this behavior can be explained by the small amount of variation presents in the Bonn dataset, especially considering light and soil conditions. We can appreciate this by look at the low mean IoU (46.01%) while adapting from Bonn to Zurich using the vanilla approach. We also point out that the pair of datasets considered here have a total of 715 images, in contrast, in the first experiment we use 2813 images. We believe that this is a crucial point for the adaptation capability of GAN-based approaches.

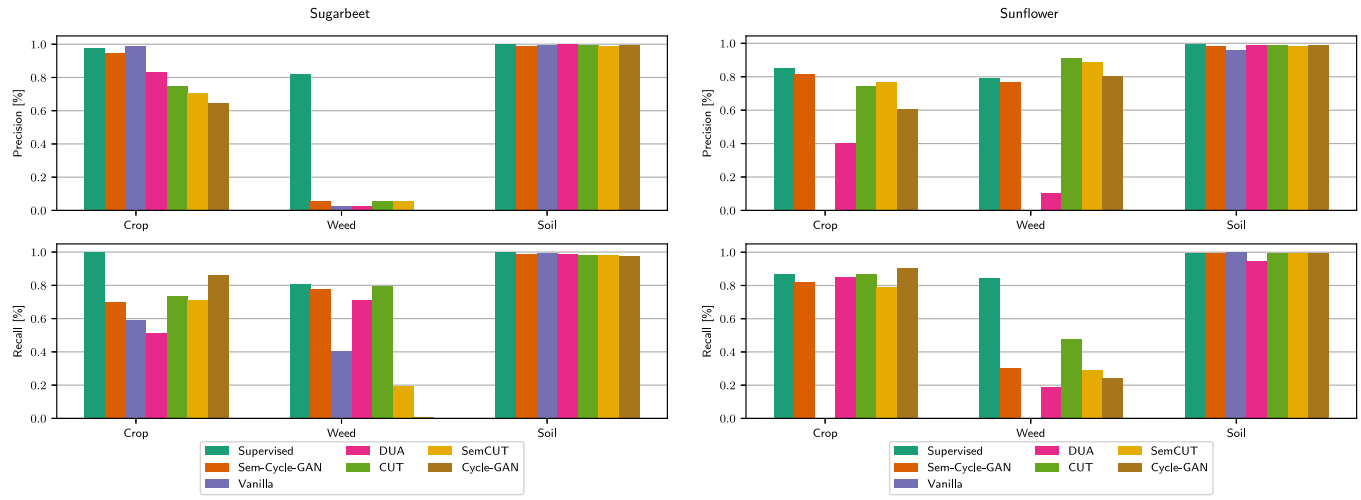


Fig. 6. Sugarbeet vs Sunflower. Per-class precision and recall.

3.4. Translating between species (Sunflower vs. Sugarbeet)

In the third experiment, we use one dataset collect on a sunflower field in Ancona, Italy, and one on a sugar beet field collected in Bonn. Thus, the main difference in this pair is the crop type but also the varieties of weed present in the field due to the different conditions between Italy and Germany. The first dataset was collected with a small self-made robot, while the second using BoniRob. Both robots are equipped with the JAI AD-130GE camera.

We notice, in average, unsatisfactory results in the adaptation from the Sunflower dataset to the Sugarbeet dataset. As can be seen in Fig. 6 all the approaches have weed precision below 10%. This is most likely due to the small size of the Sunflower dataset. While adapting from sugarbeet to sunflower, CUT yields better performance in terms of mean IoU reaching a 70.43% while other approaches stay below 65%. We notice that, in this direction, adding semantic labels while training the adaptation network is not helping the segmentation results on the target domain. This can be explained by the different crop and weed shape. In fact, as already mentioned, the weed types are substantially different in the two datasets and the represented crops, sugar beets and sunflower, have different shapes. Again the experimental evaluation shows no correlation between mean IoU and FID values.

3.5. Ablation study

To validate our design choice, we run a series of ablation studies by changing the loss formulation in (6). Specifically, we train our adaptation network without one different loss term for each run. See Table 4. We use only the pair of largest datasets, corresponding to the experiments presented in Section 3.2 for this ablation for computational limits. The main outcome is that without the semantic loss, $\mathcal{L}_{\text{IOU}}$, the adaptation to the target domain fails. While providing a small improvement over the vanilla approach using the GAN loss, $\mathcal{L}_{\text{GAN}}$, and the semantic loss, $\mathcal{L}_{\text{IOU}}$, when transferring from UGV-Bonn to UGV-Stuttgart and using the contrastive loss, $\mathcal{L}_{\text{C}}$, and the semantic loss, $\mathcal{L}_{\text{IOU}}$, when transferring from UGV-Stuttgart to UGV-Bonn. In both cases, the best adaptation performances can be reached using our proposed loss function defined in (6).

Table 4
Mean IoU in [%]. Ablation Study. Semantic segmentation results after training ERFNet using as training set images generated with different loss terms.

Datasets		Approach				
Source	Target	Vanilla	No $\mathcal{L}_{\text{GAN}}$	No $\mathcal{L}_{\text{C}}$	No $\mathcal{L}_{\text{IOU}}$	All terms
UGV-Bonn	UGV-Stuttgart	53.94	30.79	60.44	57.39	72.71
UGV-Stuttgart	UGV-Bonn	69.04	70.61	52.65	52.67	72.59

4. Discussion

For a successful deployment of robotic platforms in changing field conditions, we have to provide it with a robust perception system. Most machine learning approaches show degrading performance when applied on data showing a domain shift, which is exemplarily shown by our experimental results for the vanilla approach which uses a model trained solely on the source domain data. The semantic segmentation performance compared to the same approach trained with labels from the target domain drops consistently in our different experiments. In this work, we target a perception approach that can be adapted to novel field conditions in a target domain $\mathcal{X}_t$ by exploiting labeling effort that went into the annotation of a dataset in a source domain $\mathcal{X}_s$. To this end, we transfer images from the source domain into the target domain, while preserving the semantics of the source domain. Given the translated images, we can then re-train a semantic segmentation approach on the translated target images using the existing source labels.

In our experiments, we study different levels of domain gaps in terms of field conditions, robotic platforms, and targeted crops. From the presented results, we can conclude that our method provides advantages over other approaches for style transfer, when the source data and target data is large, as shown in our first experiment. However, in situations with only few data points, our proposed methods does not show clear advantages over competing methods as shown in the second experiment. In the third experiment, where we have a difference in the targeted crop, we can see that the source and target domain can have an influence on the performance. While the transfer from sugarbeet to sunflowers improves the performance of the semantic segmentation approach consistently, we see no considerable improvement for the transfer from sunflowers to sugarbeets. We suspect that this difference in performance is mainly caused by the size of the dataset, see Table 1, but also might be attributed to the larger difference in appearance of the crops.

5. Conclusions

Our experiments show that the usage of semantics for style transfer improves the results in the target domain. Furthermore, our results indicate that the FID score measuring the appearance is unrelated to the final performance of the semantic segmentation in the target

domain indicating that visual appearance of the generated images is not sufficient to guarantee good semantic segmentation performance.

Despite the promising results, there are several avenues for future research to close the generalization gap between a source and target domain. Recently, diffusion models (Sohl-Dickstein et al., 2015) showed convincing performance in several image synthesis tasks (Dhariwal and Nichol, 2021; Ramesh et al., 2022; Nichol et al., 2021) surpassing GANs in terms of quality of the generated results. Thus, replacing the generative model for the style transfer could be a promising direction to further improve the performance. Furthermore, the recent success of self-supervised representation learning (Chen et al., 2020; He et al., 2020; Grill et al., 2020) could be leveraged to learn good initial representations in the target domain that could be the starting point for fine-tuning using transferred images.

CRediT authorship contribution statement

Federico Magistri: Conceptualization, Methodology, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Jan Weyler:** Conceptualization, Methodology, Investigation, Visualization, Writing – review & editing. **Dario Gogoll:** Methodology, Visualization. **Philipp Lottes:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Jens Behley:** Conceptualization, Methodology, Visualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing. **Nik Petrinic:** Conceptualization, Writing – review & editing. **Cyrill Stachniss:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used in this paper can be downloaded at: https://www.ipb.uni-bonn.de/html/projects/SemCUT_DomainAdaption_2022/data.zip. The code used in this paper can be downloaded at: https://www.ipb.uni-bonn.de/html/projects/SemCUT_DomainAdaption_2022/code.zip.

Acknowledgments

This work has partially been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Germany's Excellence Strategy, EXC-2070 – 390732324 – PhenoRob, by the Deutsche Forschungsgemeinschaft, Germany under STA 1051/5-1 within the FOR 5351 (AID4Crops), and by the Federal Ministry of Food and Agriculture (BMEL), Germany based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE), Germany under the innovation support programme under funding no 28DK108B20 (RegisTer).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compag.2023.108114.

References

Asseng, S., Asche, F., 2019. Future farms without farmers. Science Robotics 4 (27), eaaw1875.

Barreto, A., Lottes, P., Ispizua, F., Baumgarten, S., Wolf, N., Stachniss, C., Mahlein, A.-K., Paulus, S., 2021. Automatic UAV-based counting of seedlings in sugar-beet field and extension to maize and strawberry. Comput. Electron. Agric..

Bertoglio, R., Mazzucchelli, A., Catalano, N., Matteucci, M., 2023. A comparative study of Fourier transform and CycleGAN as domain adaptation techniques for weed segmentation. Smart Agric. Technol. 4, 100188.

Blok, P.M., Kootstra, G., Elghor, H.E., Diallo, B., van Evert, F.K., van Henten, E.J., 2022. Active learning with MaskAL reduces annotation effort for training Mask R-CNN on a broccoli dataset with visually similar classes. Comput. Electron. Agric. 197, 106917.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A simple framework for contrastive learning of visual representations. In: Proc. of the Intl. Conf. on Machine Learning. ICML, URL https://arxiv.org/pdf/2002.05709.pdf.

Chen, Y., Lin, Y., Yang, M., Huang, J., 2019. CrDoCo: Pixel-level domain transfer with cross-domain consistency. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. CVPR.

Cicco, M., Potena, C., Grisetti, G., Pretto, A., 2017. Automatic model based dataset generation for fast and accurate crop and weeds detection. In: Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems. IROS, URL.

Dhariwal, P., Nichol, A., 2021. Diffusion models beat GANs on image synthesis. In: Proc. of the Conf. on Neural Information Processing Systems. NeurIPS.

Gogoll, D., Lottes, P., Weyler, J., Petrinic, N., Stachniss, C., 2020. Unsupervised domain adaptation for transferring plant classification systems to new field environments, crops, and robots. In: Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems. IROS, URL http://www.ipb.uni-bonn.de/pdfs/gogoll2020iros.pdf.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. In: Proc. of the Advances in Neural Information Processing Systems. NIPS, pp. 2672–2680, URL https://arxiv.org/pdf/1406.2661.pdf.

Grill, J.-B., Strub, F., Altché, F., Tallec, C., ichemond, P.H.R., Buchatskaya, E., Doersch, C., Pires, B.A., Guo, Z.D., mad Gheshlaghi Azar, M., Piot, B., Kavukcuoglu, K., Munos, R., Valko, M., 2020. Bootstrap your own latent: A new approach to self-supervised learning. In: Proc. of the Conf. on Neural Information Processing Systems. NeurIPS.

Haug, S., Michaels, A., Biber, P., Ostermann, J., 2014. Plant classification system for crop / weed discrimination without segmentation. In: Proc. of the IEEE Winter Conf. on Applications of Computer Vision. WACV, pp. 1142–1149.

He, K., Fan, H., Wu, Y., Xie, S., Girshick, R., 2020. Momentum contrast for unsupervised visual representation learning. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. CVPR, URL proceedings:he2020cvpr-mcfu.pdf.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. CVPR, URL https://arxiv.org/pdf/1512.03385.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S., 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Proc. of the Advances in Neural Information Processing Systems. NIPS.

Hoffman, J., Tzeng, E., Park, T., Zhu, T., Isola, P., Saenko, K., Efros, A.A., Darrell, T., 2018. CyCADA: Cycle-consistent adversarial domain adaptation. In: Proc. of the Intl. Conf. on Machine Learning. ICML, URL http://proceedings.mlr.press/v80/hoffman18a/hoffman18a.pdf.

Hu, K., Wang, Z., Coleman, G., Bender, A., Yao, T., Zeng, S., Song, D., Schumann, A., Walsh, M., 2021. Deep learning techniques for in-crop weed identification: A review. arXiv preprint arXiv:2103.14872.

Jumpasut, A., Petrinic, N., Elliott, B., Siviour, C., Arthington, M., 2008. An error analysis into the use of regular targets and target detection in image analysis for impact engineering. J. Appl. Mech. Mater. 13–14, 203–210. http://dx.doi.org/10.4028/www.scientific.net/AMM.13-14.203, URL https://www.scientific.net/AMM.13-14.203.pdf.

Kerkech, M., Hafiane, A., Canals, R., 2018. Deep leaning approach with colorimetric spaces and vegetation indices for vine diseases detection in UAV images. Comput. Electron. Agric. 155, 237–243.

Khanna, M., Atallah, S.S., Kar, S., Sharma, B., Wu, L., Yu, C., Chowdhary, G., Soman, C., Guan, K., 2022. Digital transformation for a sustainable agriculture in the United States: Opportunities and challenges. Agricult. Econ.

Krizhevsky, A., Sutskever, I., Hinton, G., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60 (6), 84–90, URL https://dl.acm.org/doi/pdf/10.1145/3065386.

Kwak, G.-H., Park, N.-W., 2022. Unsupervised domain adaptation with adversarial self-training for crop classification using remote sensing images. Remote Sens. 14 (18), 4639.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444, URL https://www.nature.com/articles/nature14539.pdf.

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1 (4), 541–551.

Lottes, P., Behley, J., Chebrolu, N., Milioto, A., Stachniss, C., 2018a. 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.

Lottes, P., Behley, J., Chebrolu, N., Milioto, A., Stachniss, C., 2020. Robust joint stem detection and crop-weed classification using image sequences for plant-specific treatment in precision farming. J. Field Robotics (JFR) 37, 20–34. http://dx.doi.org/10.1002/rob.21901, URL http://www.ipb.uni-bonn.de/pdfs/lottes2019jfr.pdf.

Lottes, P., Behley, J., Milioto, A., Stachniss, C., 2018b. Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. IEEE Robotics Autom. Lett. (RA-L) 3, 3097–3104. http://dx.doi.org/10.1109/LRA.2018.2846289, URL https://arxiv.org/abs/1806.03412.

Lottes, P., Höferlin, M., Sander, S., Stachniss, C., 2017. Effective vision-based classification for separating sugar beets and weeds for precision farming. J. Field Robotics (JFR) 34, 1160–1178. http://dx.doi.org/10.1002/rob.21675, URL http://www.ipb.uni-bonn.de/wp-content/papercite-data/pdf/lottes16jfr.pdf.

McCool, C., Perez, T., Upcroft, B., 2017. Mixtures of lightweight deep convolutional neural networks: Applied to agricultural robotics. In: Proc. of the IEEE Intl. Conf. on Robotics & Automation. ICRA.

Milioto, A., Lottes, P., Stachniss, C., 2018. 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, URL https://arxiv.org/pdf/1709.06764.

Mirza, M.J., Micorek, J., Possegger, H., Bischof, H., 2022. The norm must go on: Dynamic unsupervised domain adaptation by normalization. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. CVPR, URL http://proceedings:mirza2022cvpr.pdf.

Mortensen, A., Dyrmann, M., Karstoft, H., Jörgensen, R.N., Gislum, R., 2016. Semantic segmentation of mixed crops using deep convolutional neural network. In: Proc. of the International Conf. of Agricultural Engineering. CIGR.

Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., Chen, M., 2021. GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models. arXiv preprint 2112.10741.

Oord, A.v.d., Li, Y., Vinyals, O., 2018. Representation learning with contrastive predictive coding. arXiv preprint 1807.03748.

Park, T., Efros, A.A., Zhang, R., Zhu, J.-Y., 2020. Contrastive learning for unpaired image-to-image translation. In: Proc. of the Europ. Conf. on Computer Vision. ECCV, pp. 319–345.

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

Pretto, A., Aravecchia, S., Burgard, W., Chebrolu, N., Dornhege, C., Falck, T., Fleckenstein, F., Fontenla, A., Imperoli, M., Khanna, R., Liebisch, F., Lottes, P., Milioto, A., Nardi, D., Nardi, S., Pfeifer, J., Popović, M., Potena, C., Pradalier, C., Rothacker-Feder, E., Sa, I., Schaefer, A., Siegwart, R., Stachniss, C., Walter, A., Winterhalter, W., Wu, X., Nieto, J., 2020. Building an aerial-ground robotics system for precision farming. IEEE Robot. Autom. Mag. 28 (3), 29–49.

Rahman, M.A., Wang, Y., 2016. Optimizing intersection-over-union in deep neural networks for image segmentation. In: Intl. Symp. on Visual Computing. pp. 234–244.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M., 2022. Hierarchical text-conditional image generation with CLIP latents. arXiv preprint 2204.06125.

Romera, E., Alvarez, J.M., Bergasa, L.M., Arroyo, R., 2018. Erfnet: Efficient residual factorized ConvNet for real-time semantic segmentation. IEEE Trans. Intell. Transp. Syst. (ITS) 19 (1), 263–272. http://dx.doi.org/10.1109/TITS.2017.2750080, URL http://www.robesafe.uah.es/personal/eduardo.romera/pdfs/Romera17tits.pdf.

Roscher, R., Herzog, K., Kunkel, A., Kicherer, A., Töpfer, R., Förstner, W., 2014. Automated image analysis framework for high-throughput determination of grapevine berry sizes using conditional random fields. Comput. Electron. Agric. 100, 148–158. http://dx.doi.org/10.1016/j.compag.2013.11.008.

Ruckelshausen, A., Biber, P., Dorna, M., Gremmes, H., Klose, R., Linz, A., Rahe, F., Resch, R., Thiel, M., Trautz, D., et al., 2009. BoniRob–an autonomous field robot platform for individual plant phenotyping. Precis. Agric. 9 (841), 1.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S., 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In: Proc. of the Intl. Conf. on Machine Learning. ICML.

Vasconcelos, G.J., Spina, T.V., Pedrini, H., 2021. Low-cost domain adaptation for crop and weed segmentation. In: Proc. of the Iberoamerican Congress on Pattern Recognition. CIARP.

Walter, A., Finger, R., Huber, R., Buchmann, N., 2017. Opinion: Smart farming is key to developing sustainable agriculture. In: Proc. of the National Academy of Sciences, Vol. 114, No. 24. pp. 6148–6150. http://dx.doi.org/10.1073/pnas.1707462114, arXiv:https://www.pnas.org/content/114/24/6148.full.pdf.

Wu, X., Fan, X., Luo, P., Choudhury, S.D., Tjahjadi, T., Hu, C., 2023. From laboratory to field: Unsupervised domain adaptation for plant disease recognition in the wild. Plant Phenomics 5, 0038.

You, A., Kolano, H., Parayil, N., Grimm, C., Davidson, J.R., 2022. Precision fruit tree pruning using a learned hybrid vision/interaction controller. In: Proc. of the IEEE Intl. Conf. on Robotics & Automation. ICRA.

Zabawa, L., Kicherer, A., Klingbeil, L., Töpfer, R., Kuhlmann, H., Roscher, R., 2020. Counting of grapevine berries in images via semantic segmentation using convolutional neural networks. ISPRS J. Photogramm. Remote Sens. 164, 73–83, URL https://arxiv.org/pdf/2004.14010.pdf.

Zenkl, R., Timofte, R., Kirchgessner, N., Roth, L., Hund, A., Van Gool, L., Walter, A., Aasen, H., 2021. Outdoor plant segmentation with deep learning for high-throughput field phenotyping on a diverse wheat dataset. Front. Plant Sci. 12.

Zhang, Z., Kayacan, E., Thompson, B., Chowdhary, G., 2020. High precision control and deep learning-based corn stand counting algorithms for agricultural robot. Auton. Robots 44 (7), 1289–1302.

Zhu, J., Park, T., Isola, P., Efros, A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc. of the IEEE/CVF Intl. Conf. on Computer Vision. ICCV, pp. 2223–2232, URL https://arxiv.org/pdf/1703.10593.pdf.