Smart Agricultural Technology 4 (2023) 100188

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

A comparative study of Fourier transform and CycleGAN as domain adaptation techniques for weed segmentation

Riccardo Bertoglio *, Alessio Mazzucchelli, Nico Catalano, Matteo Matteucci

Department of Electronics, Information and Bioengineering, Politecnico di Milano, Via Ponzio 34/5, Milan, 20133, Italy

ARTICLE INFO

Editor: Spyros Fountas

Keywords:
Weed identification
Semantic segmentation
Domain adaptation
Fourier transform
Generative adversarial networks

ABSTRACT

Automatic weed identification is becoming increasingly important in the Precision Agriculture field as a fundamental capability for targeted spraying or mechanical weed destruction. Targeted weed elimination reduces herbicides' use and thus lowers the environmental impact of treatments. Convolutional Neural Networks are one of the most successful techniques to automatically detect weeds on RGB images. Such models require a high amount of labeled data to obtain satisfying detection performance. The agricultural context presents a high degree of variability, and it is thus unfeasible to expect a representative dataset for each specific condition that can appear in the fields. Domain Adaptation techniques are exploited to maintain high detection performance in different field conditions, lowering the need for labeled data. This study presents a comparison of the two main style transfer techniques for performing domain adaptation, that is, the Fourier Transform and the CycleGAN architecture. We used these techniques to reduce the domain gap in two use cases: one with images collected by different robots with different cameras and another with images collected by the same platform in different years. We show how, in the first case, the CycleGAN architecture attains satisfying performance and beats the simpler Fourier Transform. Instead, in the second case, all the tested DA techniques struggle to reach baseline performance. We also show how introducing a loss based on phase discrepancy in the CycleGAN architecture stabilizes the training and improves the performance. Moreover, we release a new dataset of labeled agricultural images and the code of our experiments for the reproducibility of the results and comparison with future works.

1. Introduction

Precision Agriculture is a farming management concept based on observing, measuring, and responding to inter and intra-field variability in crops. PA research aims to define decision support systems for whole-farm management, optimizing returns on inputs and thus preserving resources [1]. A typical agricultural task is weed destruction, which is usually tackled by spraying herbicides uniformly all over the field. The practice of uniform spraying poses environmental and economic concerns since part of the products are wasted because they are not targeted at weeds. The waste of herbicide products can contaminate the soil and decreases the crops' profits.

To reduce such waste, we need systems capable of targeted destruction to adapt the traditional weed removal practices to the concept of PA [2]. One option to implement such a targeted weed removal is to use smart implements capable of precise herbicide spraying limited to the weeds' area [3]. Alternatively, smart implements could perform mechanical destruction, thus eliminating the need for herbicides. Such smart implements need identification and distinction capabilities of weeds and crop plants, and RGB cameras can be used to capture weeds and crop plants [4]. Machine Learning (ML) algorithms then analyze images to detect the different plants and categorize them into crop or weed classes. The most successful ML technique in recent years is the use of Convolutional Neural Networks (CNN) [5–7], which, thanks to their ability to extract features automatically, can classify images without any domain knowledge of the task they are dealing with.

However, the performance of a classifier trained on a specific set of images (i.e., the source domain) decreases when deployed in new environments or under changing conditions (i.e., the target domain). Agricultural environments are intrinsically variable in time and spatial dimensions. Each plant is different from the others, and its appearance changes over time. Thus, the domain difference in the agricultural context is due to different weed types, growth stages of plants, soil conditions, and illumination. Maintaining good performance in such a variable environment requires broad datasets containing examples of all the possible conditions encountered in the field. Since the labeling

* Corresponding author.
E-mail address: riccardo.bertoglio@polimi.it (R. Bertoglio).

https://doi.org/10.1016/j.atech.2023.100188
Received 1 October 2022; Received in revised form 4 January 2023; Accepted 22 January 2023
Available online 26 January 2023

process is costly both in terms of time and money, it is not feasible to expect enough labeled data for each new field condition that needs to be classified. Thus, Domain Adaptation (DA) techniques have been proposed to reduce the need for labeled data.

DA is a concept under which we aim to learn a model on a source domain and attain good performance on a different (but related) target domain [8]. DA techniques are classified along different dimensions [9]. In this study, we address unsupervised DA, meaning that the target domain labels are not available. Indeed, to see weed segmentation models deployed in the real world, it is necessary to remove the dependence on labeled images for each new situation that can appear in the fields. The image labeling process takes too much time, while agricultural tasks must be performed in a timely manner. It is not feasible to postpone an agricultural task for weeks because the actual field conditions were not captured in the original labeled dataset. Beyond the unsupervised DA setting, our studied scenario is characterized by being a closed set task because all the possible classes (crop, weed, and background) appear both in the source and target domains. Finally, in our study, we address input-level DA [9] techniques to minimize the domain gap by aligning the visual appearance of source and target domain images. The downside of some of these input-level techniques is that they do not preserve semantic consistency when transferring the style. Indeed, pixels of a certain class can be mapped into a different class in the style-transformed images. This problem is typical of adversarial-based approaches, which require constraints to maintain semantic consistency.

Few pixel-level DA approaches for weed segmentation have been tested and presented in the literature. A method based on the CycleGAN architecture [10] has been proposed in [11]. The CycleGAN architecture is built on a pair of generative adversarial models, concurrently performing image translation between domains in both the source-to-target and target-to-source directions. The two adversarial modules are further tied by a cycle-consistency constraint, which is essential to preserve the geometrical properties of the input scene but provides no guarantees about the semantic consistency of the translations. In [11], the authors forced cross-domain image translations to preserve the semantic content by introducing a measure of semantic discrepancy between the original image and its translated counterpart. This measure is then minimized as an additional loss function in the learning process of the translation networks.

An alternative approach to perform style transfer and DA is via the use of the Fourier Transform [12]. When the Fourier transform is applied to an image, it is possible to extract its phase and amplitude components. The phase component carries the semantic information of the image, while the amplitude holds the style. Swapping the amplitude of a source image with the amplitude of a target image results in an image with the source semantics and the target style. In the weed segmentation field, the Fourier Transform has been tested in [13]. The authors showed how performing the style transfer through the Fourier Transform leads to higher performance compared to the baseline in the crop-weed segmentation task. They also demonstrated that preprocessing the images with the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique [14] further improves the performance.

To the best of our knowledge, no study compares the current literature DA techniques for weed segmentation. This study aims to test and compare the Fourier Transform and CycleGAN techniques on the same dataset. We tested the original CycleGAN architecture with two extensions to force semantic consistency: the extension proposed by [11] by introducing IoU losses, and that proposed by [15], in which they minimize the phase difference between real and transformed images. Finally, we release a new labeled dataset of crop-weed images collected in bean and maize fields with four species of weeds. The dataset comprises eight sets of images collected by four robots with different RGB cameras and field of views in two different years, 2019 and 2021. We also release the code of our experiments to reproduce our results and for comparison with future works. Download the dataset and code at https://doi.org/10.17632/x8brgg2j28.2.

2. Domain adaptation techniques

In this study, we compare the main literature pixel-level unsupervised DA techniques for weed segmentation by testing them on the same dataset of agricultural images. The analyzed techniques are the Fourier Transform and the CycleGAN architecture. We test two modifications of the original CycleGAN architecture by introducing loss components to maintain semantic consistency in the transformed images. The results obtained from these techniques are compared to a baseline and an upper bound. In the following, we explain how we implemented the different tested techniques.

2.1. Baseline and upper bound

To compare the DA techniques with lower (baseline) and upper bounds, we designed a model based on a U-net segmentation network [16]. We implemented the encoder part of the network with a VGG16 [17] pre-trained on Imagenet [18] used as a backbone. In Fig. 1 we show a graphical representation of the segmentation network. We designed the network's output with a 1x1 convolutional softmax activation layer that predicts the probability of a pixel location to be of the class crop, weed, and background. We then produced the baseline results by training this model on the source domain and predicting on the target domain. Instead, we produced the upper bound results by directly training the model on the target domain to predict on the same domain.

We trained the baseline model using a loss function based on the softIoU computed as [19]:

$$softIoU = \frac{1}{|C|}\sum_{c}\frac{\sum_i p_{ic}\cdot p^*_{ic}}{\sum_i p_{ic} + p^*_{ic} - p_{ic}\cdot p^*_{ic}}$$

where C is the set of classes, p_{ic} is the prediction probability of a pixel i to be of class c, p^*_{ic} is the ground truth distribution, 1 if the pixel i belongs to class c, 0 otherwise.

Similarly to the learning scheduler used in [11], we set the initial learning rate of 0.0001 and we linearly decreased it towards zero, multiplying it by 0.1 every time the IoU loss did not improve after 4 epochs. We used the Adam optimizer [20] and 256x256 padded tiles (total padded size of 320x320) from the original images. We augmented the training set by performing horizontal and vertical flips with a probability of 0.5 in both directions. During testing, we did not use any padding.

2.2. Fourier transform

Starting from the well-known concept that an image amplitude contains information about its style while the phase carries its semantics, the authors of [12] proposed an unsupervised DA method based on amplitude swapping. Given an image in the source dataset and one in the target dataset, the authors propose to swap the low-frequency component of the amplitudes of their Fourier Transforms.

More formally, given $\beta \in [0,1]$ we define a mask M_β where the coordinate system origin (0,0) is in the center of the image as:

$$M_\beta(h,w) = \mathbb{1}_{(h,w)\in[-\beta H:\beta H,-\beta W:\beta W]}$$

We use the M_β mask to select the low-frequency part of the amplitude. Then, called $\mathcal{F}$ the Fourier Transform of an image and $\mathcal{F}^{-1}$ the inverse transform, sampled an image from the target domain $x^s \in D^s$, and an image from the target domain $x^t \in D^t$, we can perform the DA via amplitude swap by:

$$x^{s\to t} = \mathcal{F}^{-1}\left(\left[M_\beta \circ \mathcal{F}^A\left(x^t\right) + \left(1 - M_\beta\right)\circ \mathcal{F}^A\left(x^s\right), \mathcal{F}^P\left(x^s\right)\right]\right)$$

As noted by [12], setting β is a trade-off between transformation and introducing noise. In our experiments, we reached a good compromise setting $\beta = 0.1$.

To evaluate the Fourier Transform technique, we transformed the source images in the target style, and then we trained a U-Net model with the source labels. Then, we used this model to predict the unlabeled target domain images. For each source image, we randomly chose

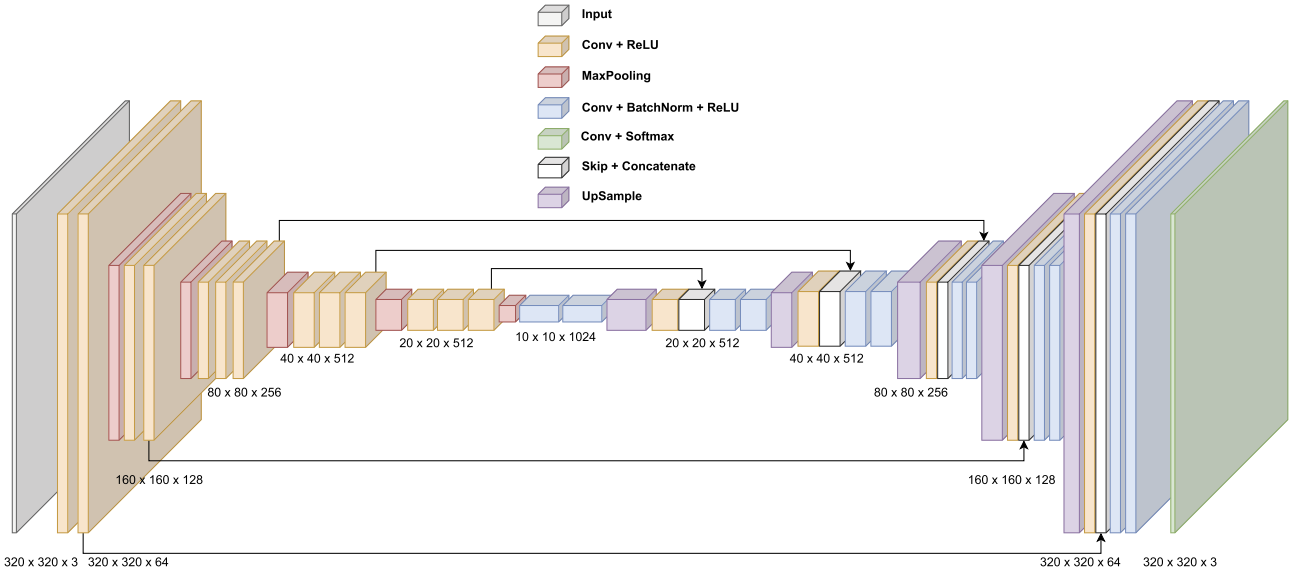

Fig. 1. Our U-net architecture for semantic segmentation.

a target image to perform the amplitude swap. By randomly choosing a different target image for each source image, we are sure to represent the distribution of target styles in the source-transformed images.

2.3. CycleGAN

As pointed out in [11], the original CycleGAN architecture [10] cannot guarantee the preservation of the semantics in the transformed images. To enforce semantic consistency, the authors of [11] proposed constraining the labels of an image to be equal to the labels of its transformation. Thus, in the original CycleGAN implementation, they added two additional semantic losses involving both the source domain and the target domain. These losses force the source Fully Convolutional Network (FCN) to predict labels to be as similar as possible to the ground truth. At the same time, they force the target FCN to predict segmentations to be as close as possible to the prediction made by the source FCN on the translated images. The complete CycleGAN architecture is depicted in Fig. 2. In the figure, we highlighted in purple the components of the $\mathcal{L}_{semantic}$ loss as proposed in [11] and composed as:

$$\mathcal{L}_{semantic} = \mathcal{L}_{softIoU}(\bar{s}_{G(x)}, s_X) + \mathcal{L}_{softIoU}(\bar{s}_{F(G(x))}, s_X) + \mathcal{L}_{softIoU}(\bar{s}_Y, \bar{s}_{F(y)}) + \mathcal{L}_{softIoU}(\bar{s}_{G(F(y))}, \bar{s}_{F(y)})$$

where $\mathcal{L}_{softIoU}$ is a softIoU-based loss, s identifies the ground truth labels, and $\bar{s}$ the predicted labels. Moreover, $x \in X$ is an image from the source domain, and $y \in Y$ is an image from the target domain. Finally, G is the generator responsible for the source-to-target translation, and F is the generator for target-to-source translation.

We then added a constraint on the image phase to further improve semantic preservation. As explained earlier, an image's phase component contains information about its semantics, while the amplitude carries information about its style. Consequently, we penalized the phase discrepancy between real and transformed images through a loss function proposed in [15]:

$$\mathcal{L}_{phase}(T, x) = -\sum_j \frac{\langle \mathcal{F}(x)_j, \mathcal{F}(T(x))_j \rangle}{||\mathcal{F}(x)_j||_2 \cdot ||\mathcal{F}(T(x))_j||_2}$$

where $\langle , \rangle$ is the dot-product, $||\cdot||_2$ is the L_2 norm, $\mathcal{F}$ the Fourier Transform and T the transformation function of the generator being trained. We implemented the Fourier Transform $\mathcal{F}$ via the Fast Fourier Transform algorithm [21]. In Fig. 2 we depicted the new loss with orange rectangles and arrows. We applied this loss to both source-transformed and target-transformed images to enforce semantic consistency in both directions. At the implementation level, we summed the phase loss to the total loss of the CycleGAN architecture. Thus, the total loss adds up to:

$$\mathcal{L}_{total} = \mathcal{L}_{GAN} + \mathcal{L}_{cycle} + \mathcal{L}_{semantic} + (\mathcal{L}_{phase}(G(x), x) + \mathcal{L}_{phase}(F(y), y))/2$$

Where $\mathcal{L}_{GAN}$ is the adversarial loss as specified in [22], and $\mathcal{L}_{cycle}$ is used to enforce cycle consistency as proposed in [10]. We divided the sum of the phase losses by 2 to uniform its magnitude to that of the other losses.

We built the CycleGAN architecture as proposed in [11] but with U-net generators as represented in Fig. 1. The training has been performed with a learning rate of 0.0002 for the first 100 epochs, halving it after the 100th epoch and then halving it again every 20 epochs. To perform the predictions on the target domain, we then use the FCN f_y network trained in the CycleGAN model.

3. Dataset and experiments

The images we used in our experiments come from the ROSE Challenge [23], a benchmarking competition of agricultural robots focused on autonomous weed destruction. The ROSE Challenge has been organized by the National Laboratory of Metrology and Testing (LNE) and the National Research Institute for Agriculture, Food, and the Environment (INRAE). Four teams participated in the ROSE field campaigns with different robots and camera systems. The teams' names are BIPBIP, PEAD, ROSEAU, and WeedElec. The images used in our experiments were collected in the years 2019 and 2021 in an experimental field at the INRAE research center located in Montoldre, France. The teams scanned maize (*Zea mays*) and bean (*Phaseolus vulgaris*) plants with four kinds of weeds (*Lolium perenne, Sinapis arvensis, Chenopodium album, Matricaria chamomilla*) under natural daylight conditions. The dataset is composed of RGB images (with different resolutions) and semantic segmentation masks to distinguish the crop, weed, and background classes. There are 1000 labeled images in total (125 per team and crop type) per year.

In our experiments, we performed two kinds of style transfer. In the first set of experiments, we used images collected by different robots with different cameras but of the same field and in the same period (May 2019). Specifically, we used the images of the teams BIPBIP and WeedElec, with both maize and bean crops, being the most similar in terms of shooting conditions. In the second set of experiments, we used images collected by the same robot and camera but in different years and periods (May 2019 and September 2021). The images were collected by the BIPBIP team, and plants were at different growth stages

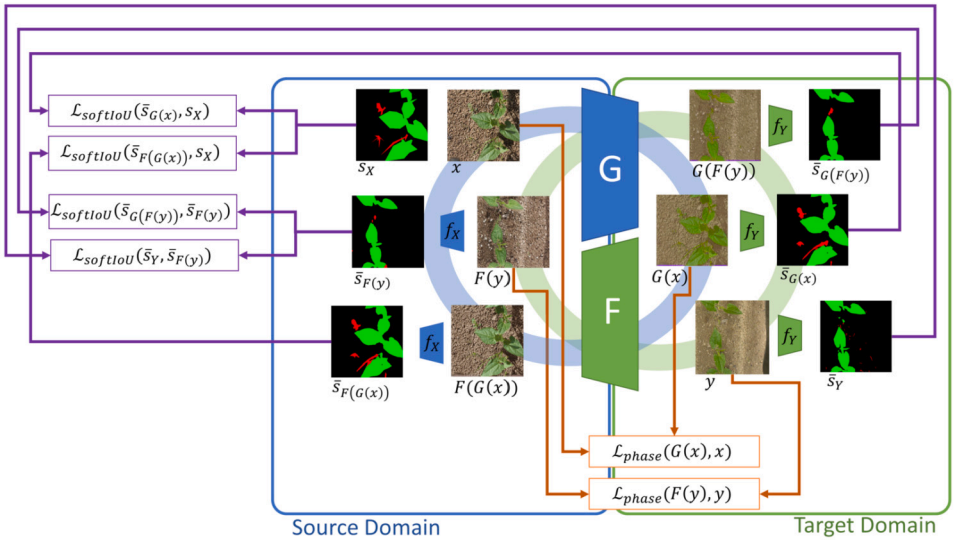

Fig. 2. The model proposed in [11] consists of two generators responsible for translate source-to-target (generator G) and from target-to-source (generator F). The model also includes two segmentation networks: f_X to segment in the source domain, f_Y to segment in the target domain. The $\mathcal{L}_{softIoU}$ loss developed by [11] (purple arrows) enforces semantic consistency using the mask predicted by the segmentation modules and the ground truth. The other additional loss (orange arrows) consists of a phase loss $\mathcal{L}_{phase}$ that encourages the generators to minimize the phase discrepancy between original and translated images. In the image, s identifies the ground truth labels, and $\bar{s}$ the predicted labels.

Table 1
Mean IoU (crop, weed, and background classes) results for the four combinations of the BIPBIP and WeedElec teams' images.

Crop	Source Domain	Target Domain	Baseline	Fourier Transform	CGAN L_semantic	CGAN L_phase	Upper Bound
Bean	BIPBIP	WeedElec	0.741	0.787	**0.827**	0.808	0.870
	WeedElec	BIPBIP	0.726	0.835	0.827	**0.844**	0.847
Maize	BIPBIP	WeedElec	**0.843**	0.838	0.587	0.645	0.893
	WeedElec	BIPBIP	0.614	0.799	0.770	**0.841**	0.883

Table 2
Mean IoU (crop, weed, and background classes) results for the four combinations of the 2019 and 2021 BIPBIP images.

Crop	Source year	Target year	Baseline	Fourier Transform	CGAN L_semantic	CGAN L_phase	Upper Bound
Bean	2019	2021	0.726	**0.738**	0.539	0.539	0.805
	2021	2019	**0.807**	0.787	0.426	0.628	0.847
Maize	2019	2021	**0.766**	0.741	0.239	0.276	0.808
	2021	2019	**0.854**	0.837	0.705	0.460	0.883

in the two years. Thus, we came up with four combinations of source and target domains and crops for each set of experiments. The dataset [24] and code of the experiments are available at this URL https://doi.org/10.17632/x8brgg2j28.2.

In both sets of experiments, we compared three unsupervised DA approaches against a baseline and an upper bound. We summarized the results in Tables 1 and 2. Table 1 shows the comparison results for the two image sets of BIPBIP and WeedElec teams collected in 2019, while Table 2 shows the comparison results for the image sets of the BIPBIP team collected in the years 2019 and 2021. For each combination of crop-source-target, we tested the Fourier Transform technique, the CycleGAN architecture as proposed in [11] (named "CGAN L_semantic"), and the same CycleGAN architecture but with the additional phase loss as proposed in [15] (named "CGAN L_phase"). We measured the algorithms' segmentation performance with the IoU over the classes crop, weed, and background.

The first set of experiments was characterized by a domain shift caused by different camera models and acquisition parameters. In this setting, the generative models outperformed the baseline and the Fourier Transform in three out of four comparisons. The baseline overperformed all the other methods in the experiment with BIPBIP as the source domain, WeedElec as the target domain, and the maize crop. We hypothesize that a higher proportion of background pixels gave the FCN f_y a poor crop-weed discrimination capability. Indeed, the WeedElec maize images were taken at a larger distance from the ground than the BIPBIP images, which turned into a higher portion of the image being just soil (that is, background class). Also, since the CycleGAN models are trained with image tiles, a significant portion of them consisted of just soil. In Fig. 3 we show the transformation of the same source image into the target domain with the three tested methods. We can observe that in the case of the BIPBIP to WeedElec (bean crop) transformation, the target image consists of only ground. The CGAN L_semantic method completely removed the plant from the source image, while the CGAN L_phase retained the crop plant.

The Fourier Transform method achieved better results than the baseline in three comparisons and comparable performance in one. Finally, the introduction of the phase loss in the CycleGAN architecture (CGAN L_phase) led to overperforming the CGAN L_semantic approach in three comparisons. In Fig. 4, for each of the four comparisons, we show the predictions of the three tested methods, plus the baseline and the ground truth. In general, we noticed that the CycleGAN models found more weeds than those in the ground truth. By inspecting the images closely, we verified that the CycleGAN models produced a better predic-

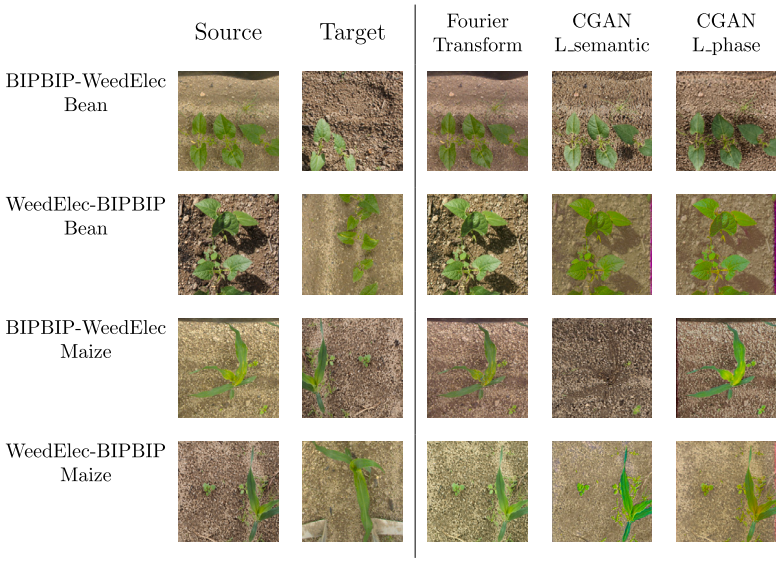

Fig. 3. Comparison of the results of the DA techniques used. In the first two columns, we display the source and target images taken by the two teams BIPBIP and WeedElec. We display the results of the transformation of the source images into the target style in the remaining three columns. The third column shows the transformation via the Fourier Transform technique, then the transformation via the CGAN L_semantic method, and finally, the CGAN L_phase method.

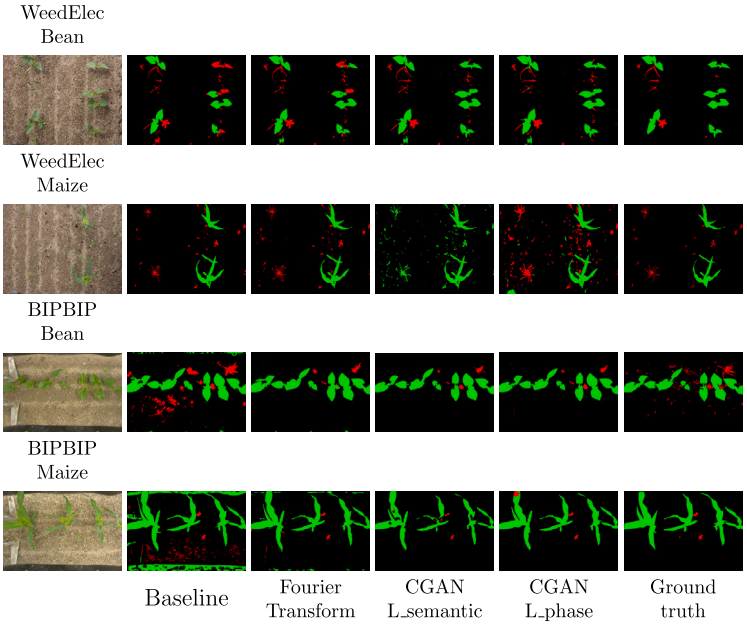

Fig. 4. Comparison of segmentation predictions performed with four models. In the first column is displayed the target image being segmented. The baseline column displays the segmentation performed by training the model with source labels. In the first column, we show the Fourier Transform predictions from a model trained on the translated dataset. The CGAN L_semantic and CGAN L_phase columns show the segmentation predictions produced with the CycleGAN models. In the last column, we show the ground truth.

tion than the ground truth in many cases. Indeed, we noticed that the human annotators had missed the smallest and merely visible weeds.

The second set of experiments was characterized by a domain shift due to a different growth stage of the plants. In this setting, the generative models presented performance under the baseline, as shown in Table 2. The style-transfer methods exploited in this study were thus helpful in the case of a domain gap caused by different camera models, but they struggled when the domain gap was because of plants with different shapes. In this setting, features-level DA methods could perform better than pixel-level ones. Also, the Fourier Transform did not improve the results compared to the baseline but in one case, although with a slight improvement.

In Fig. 5, we show the transformation of the same source image into the target domain with the three tested methods. We notice that the Fourier transform produced some visible artifacts. Moreover, even the CycleGAN transformed images present alterations which lead to low performance. Fig. 6 compares the mask predictions of the three tested DA methods. The CycleGAN architectures produced many wrongly predicted pixels.

In the Appendix, we reported tables with the experiment results specifying the IoU of the three classes (background, crop, weed) (Tables 3–7). Since the images are heavily unbalanced toward soil pixels, the background class showed higher performance, followed by the crop and weed classes.

Finally, in Fig. 7 and 8, we show the plots of the loss and mean IoU, respectively, for each of the four comparisons in the first set of experiments, that is, with the BIPBIP and WeedElec 2019 images

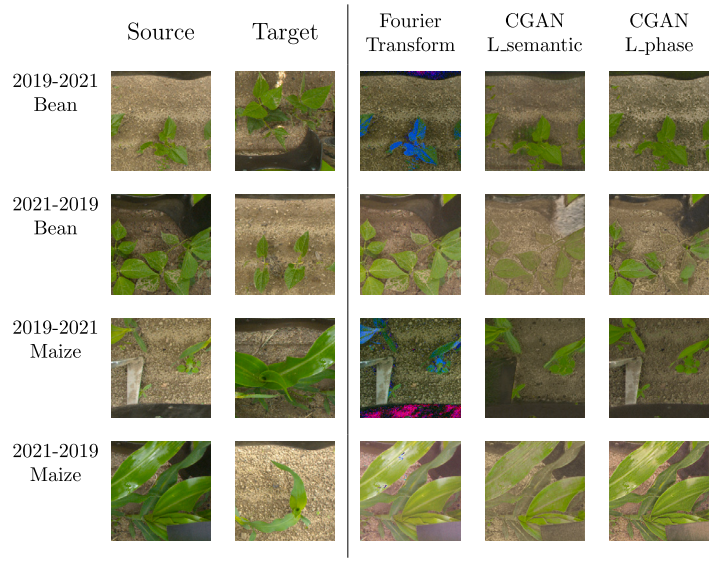

Fig. 5. Comparison of the results of the DA techniques used. In the first two columns, we display the source and target images taken by the team BIPBIP in the years 2019 and 2021. We display the results of the transformation of the source images into the target style in the remaining three columns. The third column shows the transformation via the Fourier Transform technique, then the transformation via the CGAN L_semantic method, and finally, the CGAN L_phase method.

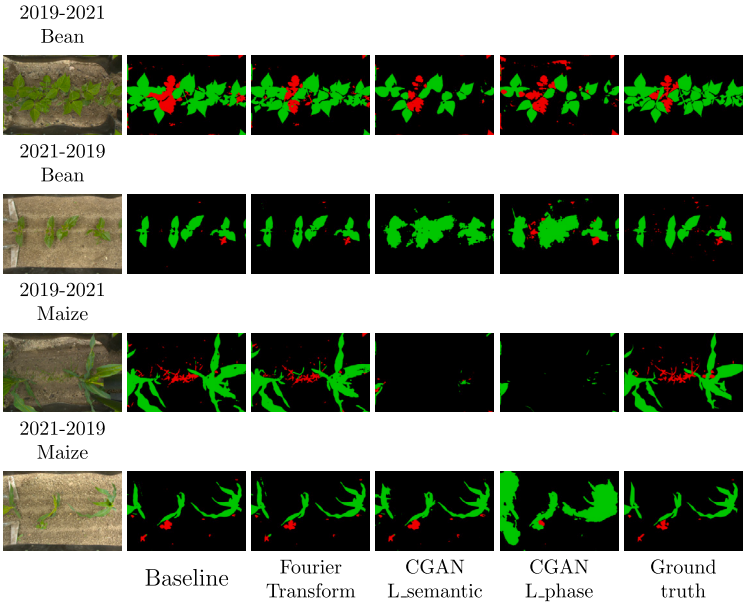

Fig. 6. Comparison of segmentation predictions performed with four models. In the first column is displayed the target image being segmented. The baseline column displays the segmentation performed by training the model with source labels. In the first column, we show the Fourier Transform predictions from a model trained on the translated dataset. The CGAN L_semantic and CGAN L_phase columns show the segmentation predictions produced with the CycleGAN models. In the last column, we show the ground truth.

as domains. The plots show the validation mean IoU and $\mathcal{L}_{softIoU}$ loss obtained during the training of the CGAN L_semantic and CGAN L_phase models. We noticed that CGAN L_phase has a more stable loss (and even lower in some cases). With loss stability, we mean that the width of the loss curve oscillations is smaller compared to a more unstable loss curve. This is evident in the BIPBIP (source), WeedElec (target), and maize crop experiment. Moreover, we noticed a more stable loss between the 100th and 175th epochs of the BIPBIP, WeedElec, and bean experiment, and after the 125th epoch of WeedElec, BIPBIP, and maize experiment. Similarly, the mean IoU is higher and more stable with the CGAN L_phase model. The effectiveness of the additional loss is particularly evident in the maize experiments.

4. Conclusions

In this study, we compared two unsupervised DA techniques for weed segmentation. At the time of writing, few approaches have been proposed, namely, the Fourier Transform and the CycleGAN architecture. We tested the original CycleGAN architecture with two modifications by adding losses based on the IoU and phase discrepancy. We tested the DA methods with two kinds of domain gaps. The first domain gap comes from images collected by different robots with different cameras and the other from images collected by the same platform in different years. Regarding the first domain gap, the Fourier Transform improved the baseline with a lower training procedure. The CycleGAN architecture, together with the IoU-based loss, showed a higher DA power at the cost of an intensive training process. We demonstrated

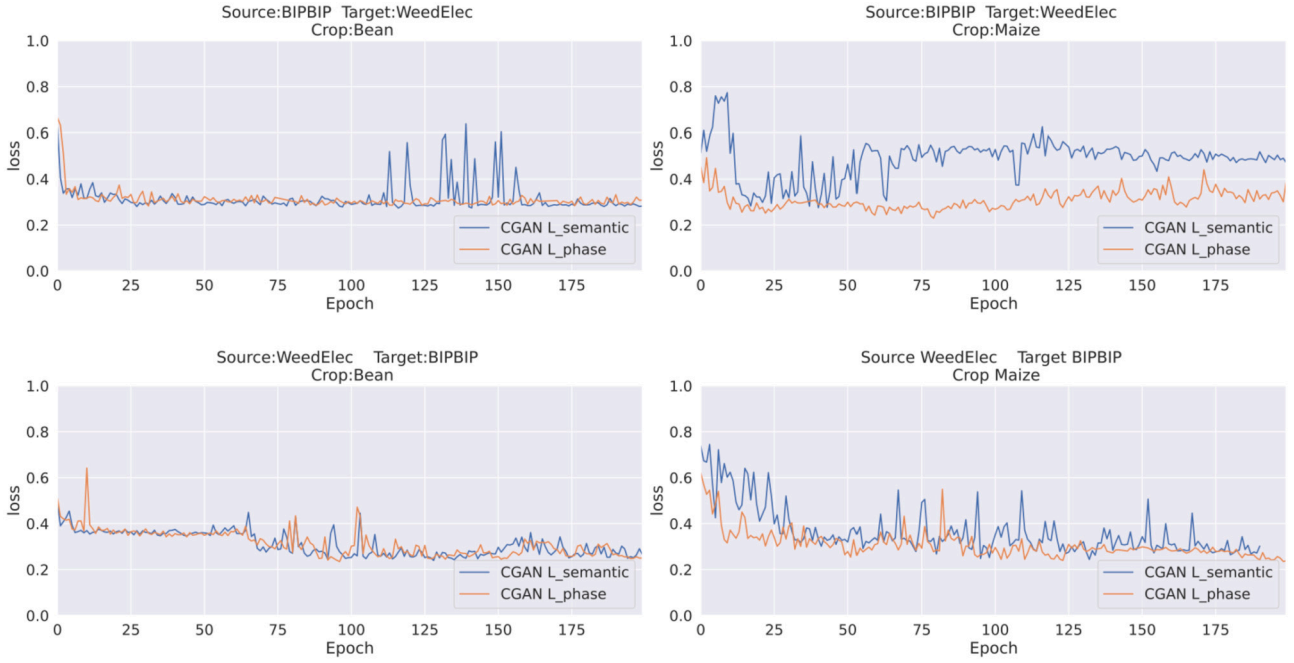

Fig. 7. Comparison plots of the losses for the training of the four experiments. CGAN L_semantic is the loss as proposed by [11], while CGAN L_phase is the loss that includes the additional phase discrepancy term.

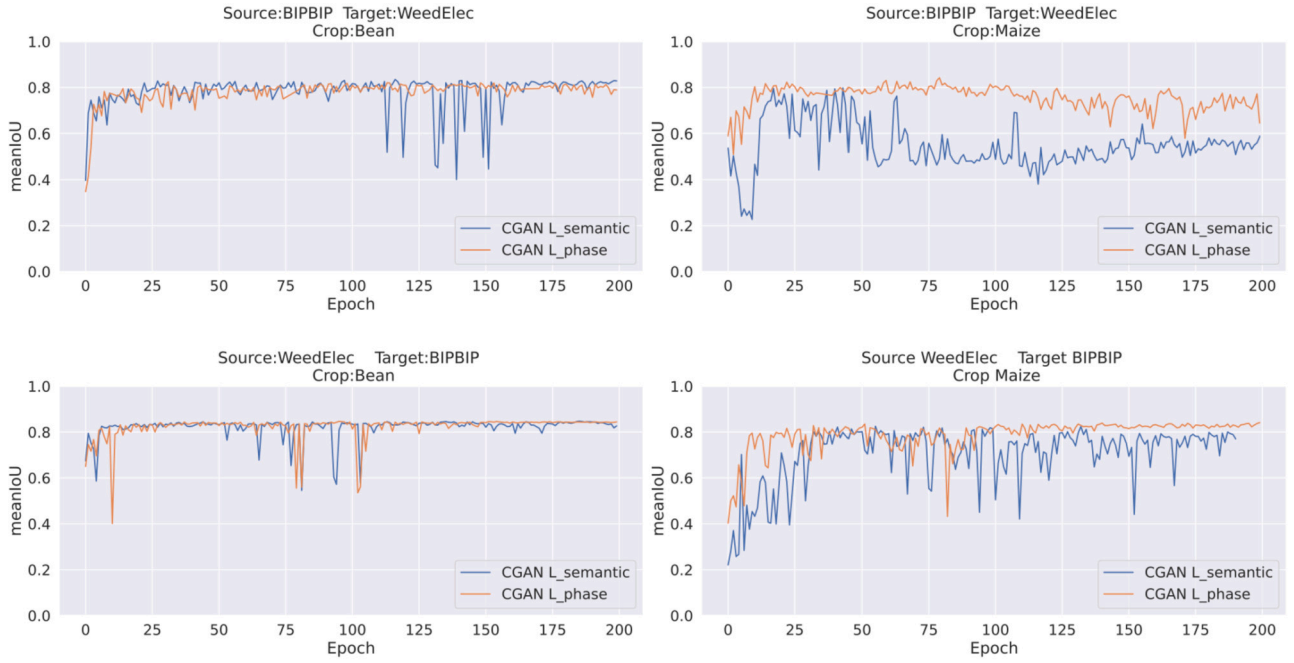

Fig. 8. Comparison plots of meanIoU for the model proposed by [11] (CGAN L_semantic) and the same model with the addition of the phase loss (CGAN L_phase). The four plots represent the accuracy over the four experiments.

how using an additional loss function based on phase discrepancy further improved the performance and stabilized the training. One of the four combinations of experiments presented DA results worse than the baseline. Indeed, different class proportions of the two domains lead to worse performance. In this case, we suggest using a weighted IoU loss to weigh more the minority classes. The second domain gap was more challenging, and none of the DA techniques could consistently beat the baseline. Both the Fourier Transform and CycleGAN models introduced artifacts in the transformed images. Other models (like features-level methods) should be tested to overcome this problem. Together with this study, we released the entire dataset and code for the reproduction of the results (including the implementation of literature techniques for which no code was available) and comparison with future works.

Data availability

The data and code used in the experiments can be downloaded here https://doi.org/10.17632/x8brgg2j28.2.

Acknowledgements

We thank the ROSE challenge and the METRICS project for providing us the data. This work has been partially funded by the H2020 METRICS project under grant agreement No. 871252.

Appendix A

Table 3
IoU results of the baseline. The segmentation model is trained on the source domain and tested on the target without any form of DA.

Crop	Source Domain	Target Domain	Baseline			
			bgIoU	cropIoU	weedIoU	meanIoU
Maize	BIPBIP 2019	WeedElec 2019	0.987	0.835	0.708	0.843
	WeedElec 2019	BIPBIP 2019	0.861	0.425	0.557	0.614
	BIPBIP 2019	BIPBIP 2021	0.912	0.836	0.549	0.766
	BIPBIP 2021	BIPBIP 2019	0.976	0.845	0.741	0.854
Bean	BIPBIP 2019	WeedElec 2019	0.976	0.697	0.551	0.741
	WeedElec 2019	BIPBIP 2019	0.944	0.724	0.511	0.726
	BIPBIP 2019	BIPBIP 2021	0.939	0.781	0.456	0.726
	BIPBIP 2021	BIPBIP 2019	0.980	0.802	0.640	0.807

Table 4
IoU results of the Fourier Transform DA. The segmentation model is trained on the source-transformed images and tested on the target domain.

Crop	Source Domain	Target Domain	Fourier Transform			
			bgIoU	cropIoU	weedIoU	meanIoU
Maize	BIPBIP 2019	WeedElec 2019	0.986	0.839	0.690	0.838
	WeedElec 2019	BIPBIP 2019	0.963	0.740	0.693	0.799
	BIPBIP 2019	BIPBIP 2021	0.912	0.834	0.477	0.741
	BIPBIP 2021	BIPBIP 2019	0.971	0.805	0.735	0.837
Bean	BIPBIP 2019	WeedElec 2019	0.974	0.792	0.597	0.787
	WeedElec 2019	BIPBIP 2019	0.979	0.822	0.704	0.835
	BIPBIP 2019	BIPBIP 2021	0.935	0.794	0.484	0.738
	BIPBIP 2021	BIPBIP 2019	0.978	0.772	0.612	0.787

Table 5
IoU results of the CGAN L_semantic DA. The segmentation model is trained on the source-transformed images and tested on the target domain.

Crop	Source Domain	Target Domain	CGAN L_semantic			
			bgIoU	cropIoU	weedIoU	meanIoU
Maize	BIPBIP 2019	WeedElec 2019	0.981	0.529	0.252	0.587
	WeedElec 2019	BIPBIP 2019	0.960	0.732	0.619	0.770
	BIPBIP 2019	BIPBIP 2021	0.672	0.032	0.011	0.239
	BIPBIP 2021	BIPBIP 2019	0.966	0.665	0.484	0.705
Bean	BIPBIP 2019	WeedElec 2019	0.976	0.826	0.679	0.827
	WeedElec 2019	BIPBIP 2019	0.980	0.822	0.679	0.827
	BIPBIP 2019	BIPBIP 2021	0.829	0.443	0.347	0.540
	BIPBIP 2021	BIPBIP 2019	0.877	0.353	0.048	0.426

Table 6
IoU results of the CGAN L_phase DA. The segmentation model is trained on the source-transformed images and tested on the target domain.

Crop	Source Domain	Target Domain	CGAN L_phase			
			bgIoU	cropIoU	weedIoU	meanIoU
Maize	BIPBIP 2019	WeedElec 2019	0.946	0.487	0.502	0.645
	WeedElec 2019	BIPBIP 2019	0.973	0.829	0.723	0.841
	BIPBIP 2019	BIPBIP 2021	0.683	0.145	0.001	0.276
	BIPBIP 2021	BIPBIP 2019	0.827	0.340	0.211	0.460
Bean	BIPBIP 2019	WeedElec 2019	0.975	0.811	0.637	0.808
	WeedElec 2019	BIPBIP 2019	0.980	0.839	0.714	0.844
	BIPBIP 2019	BIPBIP 2021	0.855	0.834	0.277	0.539
	BIPBIP 2021	BIPBIP 2019	0.923	0.549	0.413	0.628

Table 7
IoU results of the upper bound. The segmentation model is trained and tested on the target domain.

Crop	Source Domain	Target Domain	Upper Bound			
			bgIoU	cropIoU	weedIoU	meanIoU
Maize	BIPBIP 2019	WeedElec 2019	0.993	0.892	0.794	0.893
	WeedElec 2019	BIPBIP 2019	0.979	0.880	0.789	0.883
	BIPBIP 2019	BIPBIP 2021	0.920	0.864	0.641	0.809
	BIPBIP 2021	BIPBIP 2019	0.979	0.880	0.789	0.883
Bean	BIPBIP 2019	WeedElec 2019	0.981	0.970	0.759	0.870
	WeedElec 2019	BIPBIP 2019	0.984	0.862	0.697	0.847
	BIPBIP 2019	BIPBIP 2021	0.955	0.869	0.592	0.805
	BIPBIP 2021	BIPBIP 2019	0.984	0.862	0.697	0.847

References

[1] R. Bertoglio, C. Corbo, F.M. Renga, M. Matteucci, The digital agricultural revolution: a bibliometric analysis literature review, IEEE Access (2021).

[2] K.G. Liakos, P. Busato, D. Moshou, S. Pearson, D. Bochtis, Machine learning in agriculture: a review, Sensors 18 (8) (2018) 2674.

[3] B. Liu, R. Bruch, Weed detection for selective spraying: a review, Curr. Robot. Rep. 1 (1) (2020) 19–26.

[4] X. Wu, S. Aravecchia, P. Lottes, C. Stachniss, C. Pradalier, Robotic weed control using automated weed and crop classification, J. Field Robot. 37 (2) (2020) 322–340.

[5] A.M. Hasan, F. Sohel, D. Diepeveen, H. Laga, M.G. Jones, A survey of deep learning techniques for weed detection from images, Comput. Electron. Agric. 184 (2021) 106067.

[6] A. Milioto, P. Lottes, C. Stachniss, Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2018, pp. 2229–2235.

[7] P. Lottes, J. Behley, A. Milioto, C. Stachniss, Fully convolutional networks with sequential information for robust crop and weed detection in precision farming, IEEE Robot. Autom. Lett. 3 (4) (2018) 2870–2877.

[8] G. Csurka, Domain adaptation for visual applications: a comprehensive survey, arXiv preprint, arXiv:1702.05374, 2017.

[9] M. Toldo, A. Maracani, U. Michieli, P. Zanuttigh, Unsupervised domain adaptation in semantic segmentation: a review, Technologies 8 (2) (2020).

[10] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.

[11] D. Gogoll, P. Lottes, J. Weyler, N. Petrinic, C. Stachniss, Unsupervised domain adaptation for transferring plant classification systems to new field environments, crops, and robots, in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2020, pp. 2636–2642.

[12] Y. Yang, S. Soatto, FDA: Fourier domain adaptation for semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 4085–4095.

[13] G.J. Vasconcelos, T.V. Spina, H. Pedrini, Low-cost domain adaptation for crop and weed segmentation, in: Iberoamerican Congress on Pattern Recognition, Springer, 2021, pp. 141–150.

[14] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J.B. Zimmerman, K. Zuiderveld, Adaptive histogram equalization and its variations, Comput. Vis. Graph. Image Process. 39 (3) (1987) 355–368.

[15] Y. Yang, D. Lao, G. Sundaramoorthi, S. Soatto, Phase consistent ecological domain adaptation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9011–9020.

[16] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[17] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint, arXiv:1409.1556, 2014.

[18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: a large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 248–255.

[19] M.A. Rahman, Y. Wang, Optimizing intersection-over-union in deep neural networks for image segmentation, in: International Symposium on Visual Computing, Springer, 2016, pp. 234–244.

[20] D.P. Kingma, J. Ba Adam, A method for stochastic optimization, arXiv preprint, arXiv:1412.6980, 2014.

[21] J.W. Cooley, J.W. Tukey, An algorithm for the machine calculation of complex Fourier series, Math. Comput. 19 (90) (1965) 297–301.

[22] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144.

[23] Rose challenge, http://challenge-rose.fr/en/home/. (Accessed 3 January 2023).

[24] R. Bertoglio, A. Mazzucchelli, N. Catalano, M. Matteucci, A comparative study of Fourier transform and CycleGAN as domain adaptation techniques for weed segmentation - code and data, https://doi.org/10.17632/x8brgg2j28.2, 2023.