



Few-leaf Learning: Weed Segmentation in Grasslands

Guldenring, Ronja; Boukas, Evangelos; Ravn, Ole; Nalpantidis, Lazaros

Published in:

Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems

Link to article, DOI:

[10.1109/IROS51168.2021.9636770](https://doi.org/10.1109/IROS51168.2021.9636770)

Publication date:

2021

Document Version

Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):

Guldenring, R., Boukas, E., Ravn, O., & Nalpantidis, L. (2021). Few-leaf Learning: Weed Segmentation in Grasslands. In *Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 3248-3254). IEEE. <https://doi.org/10.1109/IROS51168.2021.9636770>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Few-leaf Learning: Weed Segmentation in Grasslands

Ronja Gldenring, Evangelos Boukas, Ole Ravn and Lazaros Nalpantidis

Abstract—Autonomous robotic weeding in grasslands requires robust weed segmentation. Deep learning models can provide solutions to this problem, but they need to be trained on large amounts of images, which in the case of grasslands are notoriously difficult to obtain and manually annotate. In this work we introduce *Few-leaf Learning*, a concept that facilitates the training of accurate weed segmentation models and can lead to easier generation of weed segmentation datasets with minimal human annotation effort. Our approach builds upon the fact that each plant species within the same field has relatively uniform visual characteristics due to similar environmental influences. Thus, we can train a field-and-day-specific weed segmentation model on synthetic training data stemming from just a handful of annotated weed leaves. We demonstrate the efficacy of our approach for different fields and for two common grassland weeds: *Rumex obtusifolius* (broad-leaved dock) and *Cirsium vulgare* (spear thistle). Our code is publicly available at <https://github.com/RGring/WeedAnnotator>.

I. INTRODUCTION

Precision farming can pave the way towards more sustainable agricultural production. While different robots are being deployed in the fields to handle crops or control weeds, what the vast majority of those robots have in common is that they employ deep learning-based computer vision systems for plant detection. Deep learning constitutes a strong tool for detection, but comes at a significant cost: large amounts of annotated training data are required. In crop cultivation, plants are clearly separable from soil by means of simple color thresholding or by considering near-infrared (NIR) imagery that has higher reflectivity on chlorophyll content. However, this is not an option in grassland farming, as foreground vegetation (weeds) and background (grass) share similar color and chlorophyll content. Manual annotation of grassland data becomes an extremely laborious task and thus, such publicly available datasets are scarce. To the best of our knowledge, there is no real-world grassland dataset with pixel-level annotations publicly available.

This work introduces the concept of *Few-leaf Learning* which can generate field-and-day-specific segmentation models from just a handful of weed leaves. But how can training with just a few leaves, as shown in Fig. 1, lead to accurate weed segmentation? We build upon the observation that the

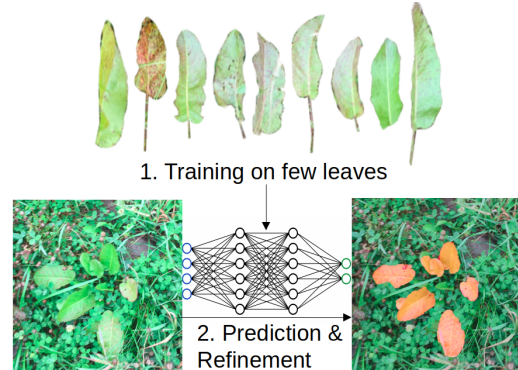


Fig. 1: Few-leaf Learning: A segmentation model can be trained on a very small number of manually annotated leaves.

visual appearance—phenotype—of plants belonging to a *specific species*—same genotype, captured within *one field and one day*, has relatively low variance. This is mainly, because plants within the same field have been exposed to the same environmental variations, which restrains the expression of phenotypic plasticity and thus results in similar leaf color, size, shape and thickness. As an example, Fig. 1 shows nine randomly drawn leaves from one of our data collection sessions, providing similar features in shape, texture and color. We train semantic segmentation models on synthetic data generated using standard image composition techniques. Hereby, a very small number of weed leaves needs to be manually annotated and then pasted on weed-free grassland background images, generating numerous combination of images. The model learns obviously relevant weed features, but also learns to separate those features from the variety of possible background features occurring in other plants, grass and mud. Despite our intended naming similarity, Few-leaf Learning is not related to the well-known *Few-shot Learning*, because in our case the network can and should be trained on a rather large number of synthetic training samples—though, just using only a few leaves to generate them.

Our work can significantly simplify the creation of annotated grassland datasets. We show that each time a data collection session has been performed on a field, Few-leaf Learning can be applied to predict the majority of weed occurrences within that session. Afterwards, the human annotator can additionally perform corrections or further annotations, as needed. We hope that our approach will encourage more publications of high-quality grassland datasets in the future, which is also our own ambition.

The contributions of this work are summarized as follows:

- We experimentally show, that nine manually annotated

All authors are with the Department of Electrical Engineering, DTU - Technical University of Denmark, Kgs. Lyngby, Denmark. {ronjag, evbou, or, lanalpa}@elektro.dtu.dk

The authors would like thank Dr. Frits K. van Evert (Wageningen UR) for sharing his valuable insights into plant phenotyping.

This work has been supported by the European Commission and European GNSS Agency through the project “Galileo-assisted robot to tackle the weed *Rumex obtusifolius* and increase the profitability and sustainability of dairy farming (GALIRUMI)”, H2020-SPACE-EGNSS-2019-870258.

leaves seem to be sufficient for the proposed Few-leaf Learning.

- We refine the mask predictions from the Few-leaf learning by applying GrabCut in an automatic manner (i.e. no user input is required). This leads to a notable performance improvement of $\approx 7\%$ for the *Intersection of Union* as well as the *Boundary F1 score*.
- We show generalizability of our final Few-leaf Learning pipeline by applying it to four different data collection sessions of different locations, days and plants, i.e. *Rumex obtusifolius* (broad-leaf dock) and *Cirsium vulgare* (spread thistle).
- We make our code publicly available at <https://github.com/RGring/WeedAnnotator>.

II. RELATED WORK

A. Agricultural Datasets

The automated detection of weeds in grassland has been investigated in agricultural robotics since the early 2000s, e.g. [1]–[5]. However, the proposed approaches are difficult to compare, because the research groups evaluated the approaches on their own unpublished datasets. In total three relevant published grassland/weed datasets were identified:

- *Rumex Ancenis dataset* [6] including monochrome images of *Rumex obtusifolius* with bounding box annotations. The images have been collected during one day and field under controlled lighting conditions.
- *GrassClover dataset* [7] including synthetic grass-clover images using image composition.
- *DeepWeeds dataset* [8] including 8 common Australian weeds in their natural environment. The annotations are on a classification level.

None of above provide pixel-level annotations on real-world images. This is a clear gap. In contrast, publicly available crop field datasets are of higher quality and most of them include annotations on a pixel-level which is relevant information for the weeding process, like e.g. cauliflower/broccoli crop dataset [9], sugar beet crop dataset [10] and carrot crop dataset [11].

B. Semi-automatic Annotation

The objective of semi-automatic annotation is to reduce the amount of human input, while maintaining the same annotation quality. This can especially save a significant amount of time on pixel-level annotations.

In interactive segmentation, masks are predicted based on negative and positive input provided by the annotator. Ideally, it requires only a few clicks to obtain high quality segmentation masks. GrabCut [12] is one of the pioneering works, formulated as an optimization problem. It requires bounding box input in order to generate a pixel-level annotation. The initial annotation proposal can be further refined with additional user input in form of scribble lines. More recently, interactive segmentation using deep learning techniques outperformed previous approaches, such as Deep Extreme Cut [13], f-BRS [14] and Inside Out Guidance [15].

Finally, another way of automating the annotation process is to automatically generate object proposals. The model generates them based on unlabeled or weakly-labeled data. Finally, the annotator corrects and complements the given object proposals to conclude the annotation process [16], [17].

C. Synthetic Image Generation

Synthetic images are artificial images, that have been generated programmatically. The great advantage of using synthetic data is the cost-efficiency of generating large amounts of data. There is no need of collecting data in the real world and more importantly there is no need of annotating the generated data.

Using image composition, new images are generated by pasting foreground objects to new background images. Dwibedi et. al. [18] pastes kitchen objects in new scenes. Their focus is patch-level realism, i.e. handling pasting artifacts. Tripathi et. al. [19] tackles global realism by extending the basic image composition with a GAN-network that learns where to position the foreground object in the scene. In agriculture, image composition has been successfully applied in the area of seed phenotyping [20], leaf segmentation and counting [21], [22] and clover-grass segmentation [7].

Another way of generating synthetic images is to render samples from a simulation environment, e.g. Cicco et. al. [23] simulate sugar beet leaves as kinematic chain, combining with real world textures. Building up such realistic simulations is complex and needs expert knowledge. Therefore, the costs might be shifted from manual annotation to creating appropriate simulation environments. Furthermore, the unrealistic appearance of rendered image is still a major issue. Domain Randomization [24], [25] as well as Domain adaptation (mainly GANs [26]–[28]) are popular approaches to close the the simulation-to-reality gap.

III. PROPOSED PIPELINE

Our Few-leaf Learning pipeline consists of three main steps:

- Generation of synthetic data by pasting a small number of weed leaves on top of weed-free background images.
- Generation of weed mask predictions using a semantic segmentation model trained on the synthetic data only.
- Mask prediction refinement using GrabCut [12] in an automated manner.

A. Synthetic image generation

First, synthetic images are generated using image composition methodologies. As input, we provide a small number of precisely segmented leaves L and background field images B that do not contain the targeted plant. Additionally, plant parameters are required like the range of number of leaves $[p_l, p_u]$ and the maximum distance to the root center d_r . Finally, the synthetic images are composed according to the procedure described in Algorithm 1. We do not consider pure background images ($np > 0$) because the generated foreground images include already a significant amount of



Fig. 2: Examples of synthetic image compositions. Left two: Rumex obtusifolius, Right two: Cirsium vulgare

Algorithm 1 Image composition procedure

Input: Field background images B
Input: Upwards pointing leaves L
Input: range of number of leaves $[p_l, p_u]$, max root center dist d_r
Output: Composed image

$B_i \leftarrow \text{random_select}(B)$ \triangleright Background image
 $B_i \leftarrow \text{random_flip}(\text{random_scale}(\text{random_rotate}(B_i)))$
 $n_p \leftarrow \text{random_select}([1, 2, 3, 4])$ \triangleright Number of plants

for $i = 0$ to n_p **do**
 $x_i, y_i \leftarrow \text{random_select_position}(B_i)$
 $n_l \leftarrow \text{random_select}([p_l, \dots, p_u])$ \triangleright Number of leaves

for $j = 0$ to n_l **do**
 $L_j \leftarrow \text{random_select}(L)$
 $d_r \leftarrow \text{random_select}([0, \dots, d_r])$
 $\theta_r \leftarrow \text{random_select}([0, \dots, 360])$
 Paste L_j to B_i with angle θ_r and distance d_r to $[x_i, y_i]$
 Optional: Apply blending method
 end for
 Optional: Add distractor objects to B_i
end for

background pixels. In Fig. 2, example synthetic images are presented for both weeds – Rumex obtusifolius and Cirsium vulgare.

Previous studies showed, that local pasting artifacts can reduce the performance of detection methods [18], [19]. Gaussian filters smooth the sharp object boundaries. Furthermore, Poisson image blending [29] keeps the edges from the source patch, but adapts the colors from the target image (mode: *Normal*). An extension is to keep the stronger edges from either the source or the target (mode: *Mix*). We consider two variants of Poisson image blending. In variant *Poisson Normal*, we simply apply Poisson image blending in mode *normal* to the segmented leaf and the background image. In variant *Poisson Mix + Normal*, we first blend the leaf as a bounding box in mode *mix* and secondly add the segmented leaf in mode *normal*. This merges the grass structure of the target batch with the grass structure of the source image, while keeping the proper texture of the leaf. Figure 3 shows the global and local effects of both Poisson blending variants.

In order to avoid that blending artifacts are associated with the targeted weed, negative objects are pasted to the image [18], [19]. As a consequence, blending artifacts will be present both in background and foreground. We consider two types of distractor objects. The *grass distractor* is a

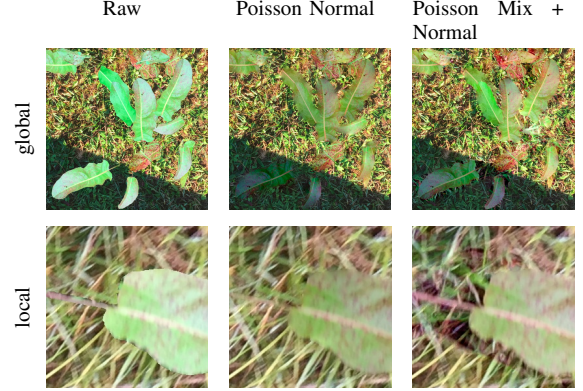


Fig. 3: Effects of Poisson blending [29]. Globally, both Poisson blending variants adapt to the color ambience of the background image. Locally, variant *Poisson Mix + Normal* merges the grass textures of the source patch and the target image.

randomly selected grass leaf object, that will be added to the image composition. The *shape distractor* [19] has the shape of one of the input leaves, but contains the texture and color of a random cutout in another randomly selected background image. A positive side-effect of adding distractor objects is the generation of partially occluded foreground objects.

B. Model Training

We use the PSPNet network architecture [30] with Resnet18 [31] as backbone to train a semantic segmentation model on the synthetically generated images. We restrict the model size with a Resnet18 in order to avoid overfitting to the small number of different leaves.

C. Mask refinement

We investigate the usage of two classic interactive segmentation algorithms GrabCut [12] and marker-based watershed [32] to refine the model predictions. Originally, classic interactive segmentation algorithms expect the user to manually provide information about background and foreground regions. Instead, we use the prediction generated by the trained PSPNet to initialize background and foreground areas. This removes the necessity of user inputs and automates the methods. Concretely, our mask refinement consists of two steps as shown in Figure 4: (1) Mask shrinking by applying morphological erosion to the background and foreground

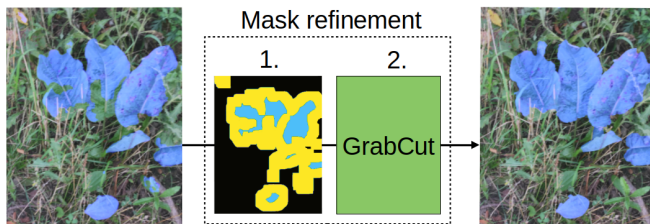


Fig. 4: The mask refinement consists of two steps: 1. The predicted mask will be shrunk by applying morphological erosion to background (black) and foreground (blue), resulting in a third unknown area (yellow). 2. The eroded mask initializes classical interactive segmentation algorithms like GrabCut [12] or watershed [32] to fill the unknown area.

area of the model prediction; (2) Applying classical interactive segmentation like GrabCut [12] or marker-based watershed [32] to the eroded mask and the corresponding image. Finally, we remove noise using morphological closing.

IV. EXPERIMENTAL SETUP

A. Data Acquisition

For this work, we have collected four sets of data. Each collection has been gathered on one field within one hour, so that we can expect relatively low variance of leaf features, which makes the proposed Few-leaf Learning applicable. The images were acquired using a 8-megapixel 1/3-inch camera with a resolution of 2448×3263 pixels while freely moving through the field. As a consequence, the collected images contain different levels of blur.

In data collection (dc) 1, 2 and 3 the weed *Rumex obtusifolius* has been manually annotated per leaf, i.e. for each leaf an individual instance has been created. In data collection 4, the weed *Cirsium vulgare* has been annotated per plant, as the plants are quite dense and it is most often infeasible for the annotator to identify all single leaves. Table I gives an overview of the data collections. For each gathering, the number of images containing instances of the considered weed (FG—foreground) and the number of weed-free field images (BG—background) is provided, as well as the average leaf/plant pixel size as a percentage of the whole image pixel size. In the case of dc4, the average object size is significantly bigger, as the annotation was performed per plant, rather than per leaf.

For the generation of synthetic training images, a pool of complete non-occluded leaves is extracted from the training split (see # *leaf pool* in table I). Hereby, blurry leaves are discarded. The synthetic training images are generated by randomly drawing n number of leaves from that pool and composing them with the background images from the training split. Validation and test results are reported on the real data, that has been annotated manually.

B. Metrics

Three standard metrics are considered to evaluate the semantic segmentation performance. The global accuracy (GAcc) computes the pixel accuracy over the dataset. Its

TABLE I: Summary of the four data collections (dc)

ID	dc1	dc2	dc3	dc4
Weed	Rumex obtusifolius			Cirsium vulgare
Date	Sep. 20, 2020	Sep. 03, 2020	Sep. 12, 2020	Sep. 20, 2020
Location	55°44'43.7"N 12°22'53.9"E	55°46'25.6"N 12°32'16.2"E	55°47'20.3"N 11°56'40.5"E	55°44'44.1"N 12°26'10.3"E
# FG imgs (train/val/test)	73 (36/18/19)	109 (54/27/28)	43 (7/18/18)	72 (6/33/33)
# BG imgs (train/val/test)	120 (60/30/30)	120 (60/30/30)	59 (27/16/16)	99 (50/24/25)
Annotation	leaf-based			plant-based
# Objects	589	619	329	125
# leaf pool	75	93	58	25
avg. object size	0.51 %	0.30 %	0.52 %	2.64 %

main disadvantage are misleading scores on imbalanced datasets. Also, we consider the widely known Intersection over Union (IoU) only for the foreground objects and call it therefore *Plant IoU*. While the IoU addresses the region similarity, the Boundary F1 score (BF score) [33] evaluates the contour similarity per image. Figure 5 shows example cases of the IoU and BF-score for different mask predictions.

C. Training configuration

For all experiments the same training configuration is applied. The Resnet18 backbone of the PSP-Network is initialized with weights, pre-trained on ImageNet [34]. The network is trained on 50 synthetically generated images (downscaled to 612×816 pixels) for 300 epochs. It is updated using the Adam Optimizer [35] with a learning rate = 0.0001 and the Lovasz Loss [36]. The following image-based augmentations were applied: random flip, random scale, random rotate, random crop and motion blur. Note, that during validation the last model state (at epoch 300) will be considered. Likewise, during application of Few-leaf Learning, there will be no ground-truth validation split available to perform early stopping to prevent overfitting.

V. EXPERIMENTAL EVALUATION

We have conducted four experiments to evaluate the proposed Few-leaf Learning. In the first three, we are considering just the training- and validation- split of data collection 1 to define the best set of hyperparameters. This collection provides leaf-based annotations, allowing in-depth evaluations. Then, in the the fourth experiment, we are using these outcomes and apply our pipeline to the test splits of all four data collections to evaluate the generalization capabilities over different fields, days and plants.

A. Blending Methods: Ablation Study

In this experiment, the effect of different blending methods is investigated. We are considering the methods discussed in Sec. III-A, as well as a number of combinations of those, where each leaf has been blended with a certain probability:

- *Blending Mix1*: [R : 0.5, GB : 0.5, P_N : 0.0, P_{M+N} : 0.0],
- *Blending Mix2*: [R : 0.33, GB : 0.33, P_N : 0.0, P_{M+N} : 0.34],
- *Blending Mix3*: [R : 0.25, GB : 0.25, P_N : 0.25, P_{M+N} : 0.25].

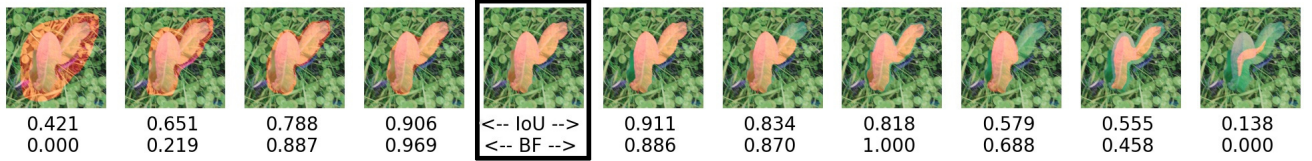


Fig. 5: IoU and BF-score for predictions of different mask qualities. Note, that the framed mask represents the ground truth (IoU=BF=1) with over-segmentations to the left and under-segmentations to the right.

We have used 9 segmented leaves to generate 50 image compositions with the different considered blending methods and additional distractor objects. Table II shows our considered metrics (GAcc, Plant IoU and BF) obtained for the different blending methods. Most blending methods do not have a significant influence on the model performance. Considered in isolation, *Poisson Mix + Normal* performs best but mixing it with raw pasting and Gaussian edge smoothing leads to even better performance. Thus, in all further experiments, *Blending Mix2* will be used.

TABLE II: Results on the validation split of data collection 1 for models trained on image compositions with different blending methods.

Blending method	GAcc	Plant IoU	BF
Raw (R)	0.984	0.598	0.651
Gaussian Blur (GB)	0.984	0.585	0.627
Poisson Normal (P_N)	0.979	0.467	0.432
Poisson Mix + Normal (P_{M+N})	0.985	0.617	0.670
Blending Mix1	0.984	0.594	0.646
Blending Mix2	0.985	0.624	0.684
Blending Mix3	0.985	0.620	0.657
Blending Mix2 + shape distractor	0.984	0.610	0.666
Blending Mix2 + grass distractor	0.985	0.619	0.686

B. How few leaves are required?

In this experiment, the minimum required number of leaves is determined in order to predict the majority of the foreground objects. From the leaf pool of intact leaves (see table I), we draw randomly $n = 1, 3, 9, 15, 21$ leaves to generate synthetic image compositions. For each number of n leaves, we repeat the experiment 10 times. Figure 6 shows the mean and variance of Plant IoU and BF-score over the drawn number of leaves. The Plant IoU and BF-score both increase with increasing number of leaves and saturate at some point—this appears to be 9 leaves.

C. Effectiveness of mask refinement

In this experiment, we want to show the impact of the mask refinement step, where we apply GrabCut as well as watershed to the mask predictions generated by PSPNet. A model trained with 9 different leaves with an average performance (Plant IoU = 0.632) is used here. The predicted masks are refined, as described in Sec. III-C. The results of Table III show that GrabCut improves the performance significantly. Both, the Plant IoU as well as the BF-score

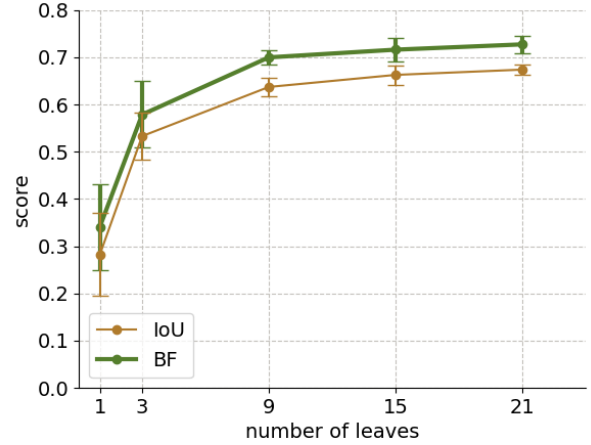


Fig. 6: Mean Plant IoU and BF score (each averaged over 10 iterations) for models trained on synthetic images generated with various numbers of randomly drawn leaves.

TABLE III: Performance results for the mask refinement step using GrabCut or watershed, compared to using no refinement.

Post-processing method	GAcc	Plant IoU	BF
None	0.985	0.632	0.685
GrabCut	0.988	0.703	0.751
Watershed	0.978	0.559	0.567

have improved by approximately 7%. On the contrary, the marker-based watershed is shown to have a negative effect on the overall pipeline performance.

In Fig. 7, we illustrate qualitative results using GrabCut in the mask refinement. Even if the model predictions tend to under-segment the detected leaves (second column), those under-segmentations are often corrected by GrabCut (third column), with the trade-off of losing small-scale predictions, like e.g. in the second row.

D. Generalizability over different fields, days and plants

In this last experiment, the final pipeline is applied to the test splits of all four data collections (dc) to evaluate the generalization capabilities. As an outcome of the three previous experiments, our final pipeline (i) applies Mix2 image blending, (ii) randomly draws 9 segmented leaves,

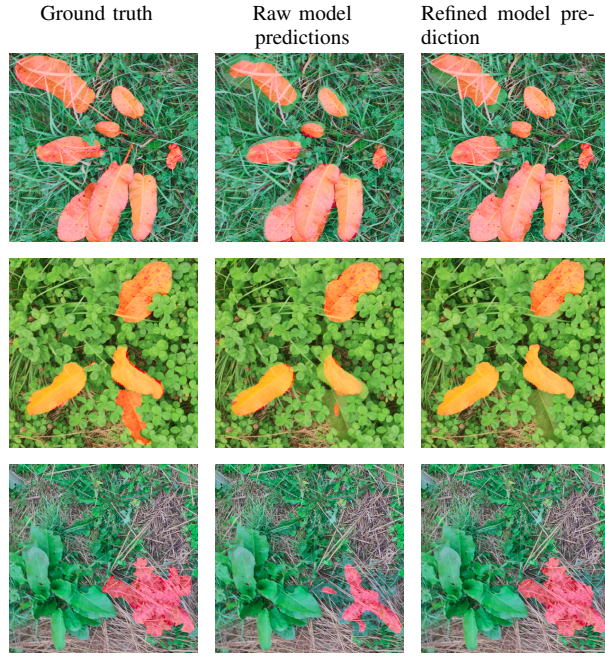


Fig. 7: Qualitative examples of model predictions without and with GrabCut refinement. GrabCut generally improves large detections but suppresses small ones.

TABLE IV: Performance of the Few-leaf training pipeline on the four different data collections, specified in Table I.

Data Collection	GAcc	Plant IoU	BF
1	0.986	0.701	0.704
2	0.991	0.585	0.592
3	0.989	0.745	0.726
4	0.989	0.723	0.729

and (iii) uses GrabCut for post-processing.

Table IV contains the obtained experimental results and shows clearly that the Few-leaf training pipeline generalizes over different days (dc1 vs. dc2 vs. dc3), different fields (dc1 vs. dc2 vs. dc3) and different plants (dc1/dc2/dc3 vs. dc4). Though, the performance on data collection 2 is noticeably worse. This is due to the fact that it contains smaller plants, with an average leaf size significantly smaller compared to the average leaf size of data collection 1 and 3 (see again Table I).

The proposed Few-leaf training becomes especially relevant when it comes to the creation of high-quality grassland datasets. We picture it as a tool, that can be applied after each data collection session to generate labels for the majority of plant occurrences. The intention is to reduce the manual labeling labour, leaving the human annotator with only a few additional annotations and corrections. We want to get an intuition of how big is the proportion of sufficiently predicted weeds, respectively of how much work remains for the human annotator. We determine for each predicted mask m_p^i corresponding to image i the subset of its ground

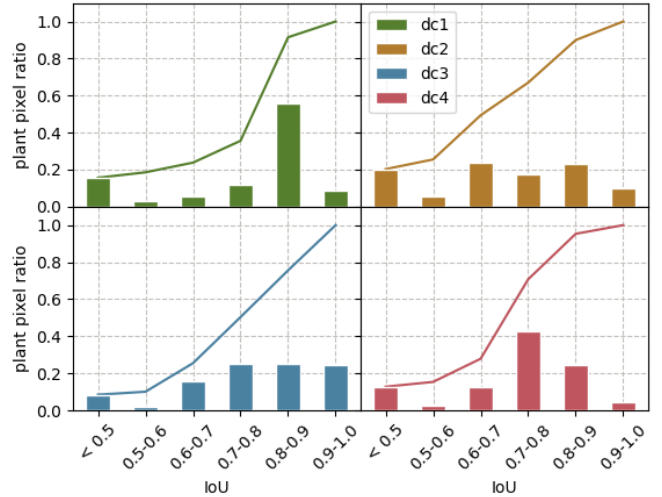


Fig. 8: *Plant pixel ratio* over IoU for the test sets of the four different data collections (dc1 - dc4). The *plant pixel ratio* represents the ratio of ground truth pixels that has been predicted within the pre-defined IoU ranges, shown by the bars. The continuous line represents the cummulative *plant pixel ratio*.

truth instances S_i that results in the maximum IoU between m_p^i and S_i . We refer to this maximum value as $IoU_{max,i}$. As a consequence, we know for each ground truth instance the IoU it has been predicted with; it is $IoU = IoU_{max,i}$. Finally, we can determine the overall percentage of ground truth weed pixels in the whole dataset that are predicted with a certain IoU. In Fig. 8, we show the ratio of ground truth pixel (named “*plant pixel ratio*”) that has been predicted within given ranges of IoUs. The predominant portion of pixels is predicted with an $IoU > 0.6$ and depending on which label precision is targeted, one can decide if Few-leaf Learning could support during the process of annotating.

VI. CONCLUSION

Our work manifests how domain knowledge—in our case plant phenotyping observations—can lead to meaningful simplifications of demanding computer vision problems. More precisely, we were able to use just 9 random leaves to synthetically generate multiple images and ultimately train accurate weed segmentation models. The mask predictions are further refined by applying GrabCut in an automated way. The proposed pipeline has been evaluated on different days, different grassland fields and for two different weeds (*Rumex obtusifolius*, *Cirsium vulgare*) exhibiting good generalization capabilities. The presented results show that a significant amount of effort can be saved during annotation using Few-leaf Learning. Hopefully, this will allow for the creation of new and more detailed grassland weed datasets that will enable autonomous robotic weeding. Nevertheless, we believe that our pipeline can be also applied to other domains by properly adapting the domain-specific composition procedure.

REFERENCES

- [1] G. Polder, F. van Evert, A. Lamaker, A. de Jong, G. van der Heijden, L. Lotz, A. van der Zalm, and C. Kempenaar, "Weed detection using textural image analysis," in *EFITA/WCCA Conference*, 2007.
- [2] F. K. van Evert, G. Polder, G. van der Heijden, C. Kempenaar, and L. A. P. Lotz, "Real-time vision-based detection of rumex obtusifolius in grassland," *Weed Research*, vol. 49, no. 2, pp. 164–174, 2009.
- [3] F. Ahmed, M. H. Kabir, S. Bhuyan, H. Bari, and E. Hossain, "Automated weed classification with local pattern-based texture descriptors," *Int. Arab J. Inf. Technol.*, vol. 11, pp. 87–94, 2014.
- [4] T. Kounalakis, G. Triantafyllidis, and L. Nalpantidis, "Weed recognition framework for robotic precision farming," in *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, pp. 466–471, 2016.
- [5] T. Kounalakis, G. Triantafyllidis, and L. Nalpantidis, "Image-based recognition framework for robotic weed control systems," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 9567–9594, 2018.
- [6] T. Kounalakis, G. Triantafyllidis, and L. Nalpantidis, "Deep learning-based visual recognition of rumex for robotic precision farming," *Computers and Electronics in Agriculture*, vol. 165, Oct. 2019.
- [7] S. Skovsen, M. Dyrmann, A. K. Mortensen, M. S. Laursen, R. Gislum, J. Eriksen, S. Farkhani, H. Karstoft, and R. N. Jorgensen, "The grassclover image dataset for semantic and hierarchical species understanding in agriculture," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2019-June, pp. 2676–2684, 2019.
- [8] A. Olsen, D. A. Kononov, B. Philippa, P. Ridd, J. C. Wood, J. Johns, W. Banks, B. Girgenti, O. Kenny, J. Whinney, B. Calvert, M. Rahimi Azghadi, and R. D. White, "DeepWeeds: A multiclass weed species image dataset for deep learning," *Scientific Reports*, vol. 9, 2 2019.
- [9] A. Bender, B. Whelan, and S. Sukkarieh, "A high-resolution, multimodal data set for agricultural robotics: A ladybird's-eye view of brassica," *Journal of Field Robotics*, vol. 37, 07 2019.
- [10] N. Chebrolu, P. Lottes, A. Schaefer, W. Winterhalter, W. Burgard, and C. Stachniss, "Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields," *The International Journal of Robotics Research*, 2017.
- [11] S. Haug and J. Ostermann, "A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks," in *Computer Vision - ECCV 2014 Workshops*, pp. 105–116, 2015.
- [12] C. Rother, V. Kolmogorov, and A. Blake, "'GrabCut': Interactive foreground extraction using iterated graph cuts," in *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, (New York, NY, USA), p. 309–314, Association for Computing Machinery, 2004.
- [13] K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool, "Deep Extreme Cut: From extreme points to object segmentation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 616–625, 2018.
- [14] O. B. A. K. Konstantin Sofiuk, Ilia Petrov, "f-BRS: Rethinking back-propagating refinement for interactive segmentation," *arXiv preprint arXiv:2001.10331*, 2020.
- [15] S. Zhang, J. H. Liew, Y. Wei, S. Wei, and Y. Zhao, "Interactive object segmentation with Inside-Outside Guidance," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [16] D. P. Papadopoulos, J. Uijlings, F. Keller, and V. Ferrari, "We don't need no bounding-boxes: Training object class detectors using only human verification," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 854–863, 2016.
- [17] O. Russakovsky, L. Li, and L. Fei-Fei, "Best of both worlds: Human-machine collaboration for object annotation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2121–2131, 2015.
- [18] D. Dwibedi, I. Misra, and M. Hebert, "Cut, Paste and Learn: Surprisingly easy synthesis for instance detection," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-October, pp. 1310–1319, 2017.
- [19] S. Tripathi, S. Chandra, A. Agrawal, A. Tyagi, J. M. Rehg, and V. Chari, "Learning to generate synthetic data via compositing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 461–470, 2019.
- [20] Y. Toda, F. Okura, J. Ito, S. Okada, T. Kinoshita, H. Tsuji, and D. Saisho, "Training instance segmentation neural network with synthetic datasets for crop seed phenotyping," *Communications Biology*, vol. 3, p. 173, 04 2020.
- [21] D. Kuznichov, A. Zvirin, Y. Honen, and R. Kimmel, "Data augmentation for leaf segmentation and counting tasks in rosette plants," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2019-June, pp. 2580–2589, 2019.
- [22] D. Ward, P. Moghadam, and N. Hudson, "Deep leaf segmentation using synthetic data," in *BMVC*, 2018.
- [23] M. Di Cicco, C. Potena, G. Grisetti, and A. Pretto, "Automatic model based dataset generation for fast and accurate crop and weeds detection," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.
- [24] M. Andrychowicz, B. Baker, M. Chociej, R. Józefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, J. Schneider, S. Sidor, J. Tobin, P. Welinder, L. Weng, and W. Zaremba, "Learning dexterous in-hand manipulation," *The International Journal of Robotics Research*, vol. 39, no. 1, pp. 3–20, 2020.
- [25] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 23–30, 2017.
- [26] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2242–2251, 2017.
- [27] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.
- [28] Y.-H. Tsai, W.-C. Hung, S. Schuster, K. Sohn, M.-H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7472–7481, 2018.
- [29] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," in *ACM SIGGRAPH 2003 Papers*, SIGGRAPH '03, (New York, NY, USA), p. 313–318, Association for Computing Machinery, 2003.
- [30] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6230–6239, 2017.
- [31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [32] F. Meyer, "Color image segmentation," in *1992 International Conference on Image Processing and its Applications*, pp. 303–306, 1992.
- [33] G. Csúszka and D. Larlus, "What is a good evaluation measure for semantic segmentation?," in *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, 01 2013.
- [34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.
- [35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2015.
- [36] M. Berman and M. B. Blaschko, "Optimization of the jaccard index for image segmentation with the lovász hinge," *CoRR*, vol. abs/1705.08790, 2017.