Evaluating synthetic vs. real data generation for AI-based selective weeding

Naeem Iqbal[1], Justus Bracke[1], Anton Elmiger[1], Hunaid Hameed[1] and Kai von Szadkowski[1]

Abstract: Synthetic data has the potential to reduce the cost for ML training in agriculture but poses its own set of problems compared to real data acquisition. In this work, we present two methods of training data acquisition for the application of machine vision algorithms in the use case of selective weeding. Results from ML experiments suggest that current methods for generating synthetic data in the field of agriculture cannot fully replace real data but may greatly reduce the quantity of real data required for model training.

Keywords: synthetic images, plant detection, phenotyping, deep learning, agriculture

1 Introduction

As a response to the ever-rising demand for training data in Machine Learning (ML) applications, much research has been conducted on simplifying the process of data recording and labelling. This is also true for the agricultural domain, where e.g., various high-throughput field phenotyping systems for recording plant data have been developed in recent years [Bo21]. Different approaches have also been proposed to reduce the effort for data annotation, such as the deep interactive object selection method presented in [Xu16]. Given the laborious and costly nature of real data acquisition, it is no wonder that the use of synthetic data (SD) has become a popular alternative in recent years [Ni21].

However, synthetic data comes with its own pitfalls, first and foremost the inherent and sometime subtle difference between reality and simulation, commonly referred to as the (simulation) reality gap. This special case of a domain gap between data from various sources is the deciding factor in how useful SD can be for a particular ML use case, its impact depending on the characteristics of the used data and algorithms. It is thus difficult to predict how well SD can be used for any particular use case or algorithm, as the relevance of different aspects of the data may vary between contexts and inference targets.

While synthetic data has been used in numerous applications, as reviewed by Nikolenko [Ni21], its application for machine vision tasks are of special interest in the context of agriculture, where current and future smart farming applications depend on the ability of

[1] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Forschungsbereich Planbasierte Robotersteuerung, Berghoffstraße 11, Osnabrück, 49090; Contact: kai.von_szadkowski@dfki.de; naeem.iqbal@dfki.de; justus_felix.bracke@dfki.de; anton.elmiger@dfki.de; hunaid.hamid@dfki.de

agricultural machinery and robots to make sense of camera data, e.g., for crop row detection or selective weeding. Aggravating the laborious nature of real data acquisition, agricultural contexts pose additional problems such as uncontrollable external factors (e.g., weather conditions) or seasonality of the objects of interest. Due to the fractal nature and thus complex shape of plants, data annotation can also become more costly or error-prone in agricultural contexts, a point where ML models trained on SD could provide useful assistance even if their performance is not fully on par with models trained on real data.

It is thus of little surprise that a number of recent studies have used synthetic data for different machine vision applications in agriculture. Ubbens et al. used synthetic plant models created via a Lindenmayer system to generate relatively simple image data for a leaf counting task, discussing how various aspects of the data impacted model performance [Ub18]. Barth et al. created synthetic data of pepper plants in a harvesting scenario, using an elaborate mix of procedural generation and 3D-scanned parts and training a segmentation model to facilitate autonomous harvesting with promising results [Ba18]. Carbone et al. even produced both RGB and infrared using the Unity game engine[2], finding that the inclusion of synthetic near-infrared (NIR) data can improve segmentation results [CPN20]. Similarly using a game engine (Unreal Engine 4[3]), Di Cicco et al. explored the use of synthetic data for crop and weed detection in sugar beets [Ci17]. These (and other) studies show the potential benefit of using synthetic data for various applications in the agricultural domain, often finding that the best results can be achieved when synthetic and real data are combined.

In this study, we present our approaches for generating real and synthetic data for a simple selective weeding use case that requires differentiating between crop and weed plants in a maize field. Using state-of-the-art machine vision algorithms, we evaluate the benefits of synthetic data for future applications in this research area.

2 Data acquisition

2.1 Synthetic data

To create a synthetic data set of a maize field, we used *Syclops*, a modular software pipeline allowing to generate synthetic data for agricultural environments. Syclops is currently under development in the Agri-Gaia[4] project and is intended to be made available open-source over the course of the project. For the data used in this work, we utilized Syclops' module interfacing it with the open-source 3D software Blender[5]. This allowed rendering photorealistic images using Blender's raytracing engine *cycles*. Achieving a

[2] https://unity.com
[3] https://www.unrealengine.com
[4] https://www.agri-gaia.de/
[5] https://www.blender.org/

quality of synthetic data close to photorealism helps to minimize the simulation-reality-gap in the RGB data and realistically depict e.g., shadows or subsurface scattering (scattering of light in organic and other translucent materials such as plant leaves).

A combination of commercially available and free assets was used in the 3D environments generated with Syclops. Models from the Maxtree [Ma22] plant library were utilized for the maize plants, adapting them to match the desired growth stages of their counterparts in the real data (Fig. 1). The weed models stem from the Graswald [Ha22] library and were similarly adjusted to fit the scene. Using Syclops' tooling for generating agricultural scenes, the plant models were distributed on a virtual field environment in specific distribution patterns (rows for maize, scattering for weeds), using random transformations such as scaling and rotation to increase variability.

Fig. 1: Rendering of 3D-Models of the virtual maize plants

To simulate changing lighting conditions in real outdoor environments, lighting of the scene was set up using a large number of HDR (high dynamic range) images, randomly chosen for each output image. For this, HDR imagery was used that is freely available from Polyhaven [Za22]. Similarly, ground materials were varied for each image, also using Polyhaven assets, thus sampling from a collection of soil and dirt textures to reproduce realistic backgrounds for the plants. Additionally, an elevation map was used to create micro-displacements of the ground to better simulate ground structures such as pebbles, tire tracks, or clods of soil. Finally, a shadow caster object was created in the virtual scene representing the recording setup of the real data set (Section 2.2), further facilitating a close match of real and synthetic data.

Overall, the combination of these elements results in a data set with variable composition and illumination, as shown in Figure 2. In total, 6500 images were rendered (matching the resolution of the real data set, see section 2.2) with their associated bounding boxes for the two classes *Maize* and *Weed*, totaling 45997 and 119316 instances respectively.

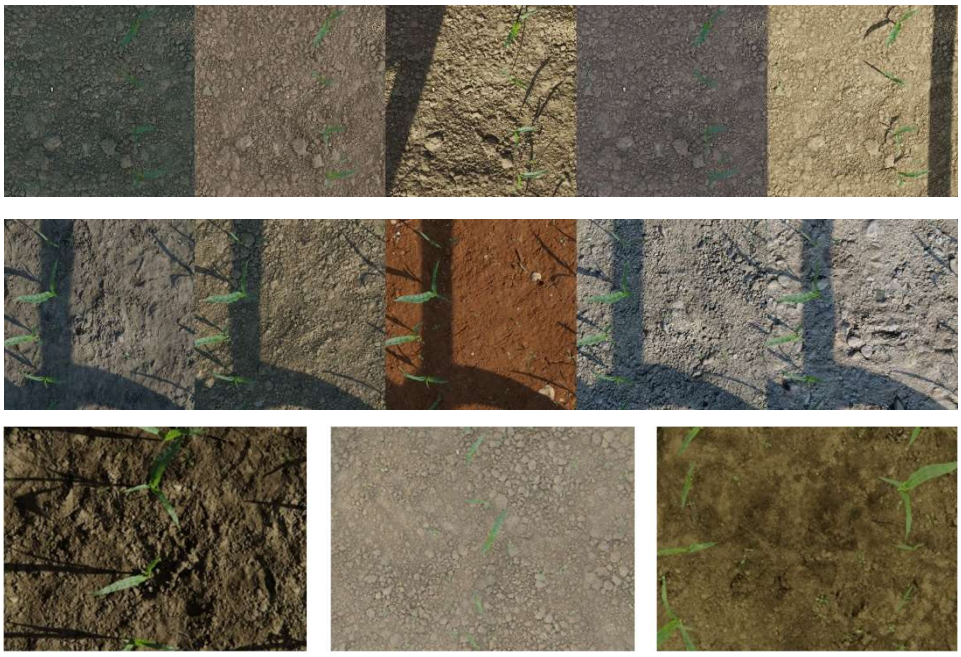

Fig. 2: Example images from the generated synthetic data set, illustrating different lighting conditions for identical perspective (top), different ground textures for identical lighting with simulated shadows from hardware (middle) and overall variability (bottom row)

2.2 Real data

Several data sets of maize for ML applications have been published, e.g. [La22], [Li22], and [MS22]. However, existing data sets exhibit limitations such as low variability of growth stages, controlled constant lighting, missing annotations for weeds, low weed levels due to conventional herbicide treatments or simply small data set size. We therefore made use of an experimental field established at Hof Fleming (Löningen, Germany) within the Agri-Gaia project to record our own training and reference data. The experimental field consists of 32 rows of maize with a length of 40 m. Two measuring campaigns were conducted between May and August 2022, capturing early growth stages of maize following two seeding applications. Herbicide treatment was varied within the rows to create variable weed pressure.

A manually driven wheeled sensor carrier was constructed to serve as a non-invasive phenotyping platform, consisting of an aluminum frame mounted on bike wheels (Fig. 3). A sensor box is mounted at the center, containing a JAI FS3200T multispectral camera in nadir view and a processing unit running ROS Noetic. The overall carrier is lightweight and maneuverable and can be operated by a single person, while the large tires make it possible to leave taller crops undisturbed. Additional equipment includes a tablet to control

the ROS setup, an RTK-GPS receiver for georeferencing and an LTE/WIFI router.

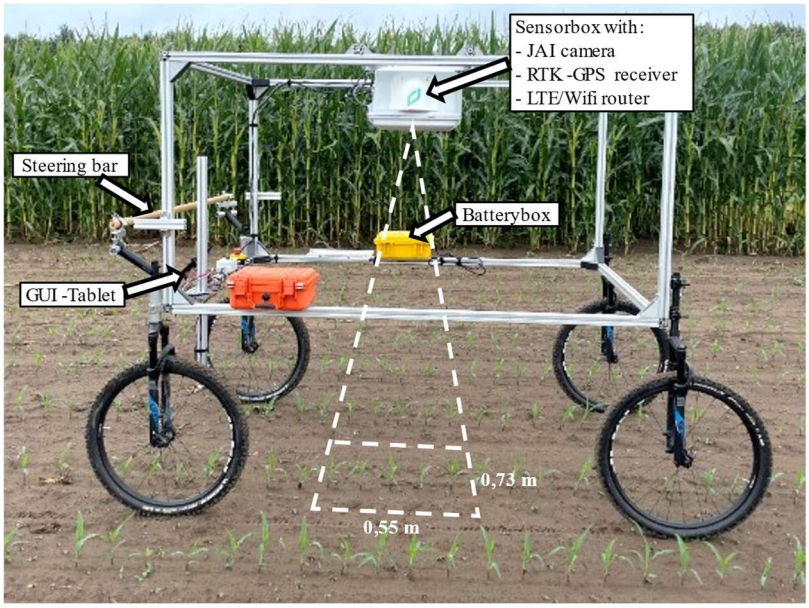

Fig. 3: The developed manually driven sensor carrier with various attachments

The image data was recorded with a resolution of 2047 x 1525 pixel, resulting in a spatial resolution at ground level of 0,36 mm/pixel. Data was captured on three time intervals after seeding (day 11, 15 and 17), covering different growth stages of the maize as well as environmental conditions such as weed coverage, soil texture or illumination (Fig. 4).

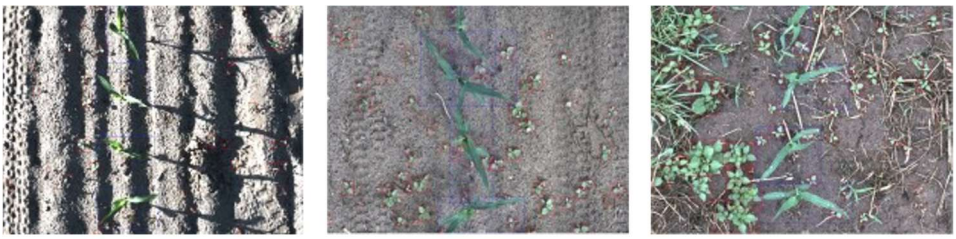

Fig. 4: Some example images of the recorded dataset showing the variability of the data regarding illumination conditions, weed intensity and growth stages

The open-source software CVAT [Cv22] was used for annotation. The data was labelled with bounding boxes of the same classes as the synthetic data, *Maize* and *Weed*. The labelled dataset contains 731 images with the two class IDs exported in both COCO annotation format and YOLO annotation format (Tab. 1). Most of the weed instances have areas of less than 30,000 pixels, while the maize instances have much larger areas, mostly around 100,000 pixels. It is intended to open-source this data set in the near future.

	Maize	Weed	Total
Day 11	1061	9743	10804
Day 15	696	14508	15204
Day 17	537	5435	5972
Total	2294	29686	31980

Tab. 1: Number of instances of Class IDs *Maize and Weed* across growth stages among Day 11, Day 15, and Day 17 after the initial seeding of the maize plants

3 Evaluation experiments

To conduct our deep learning experiments, we chose two state-of-the-art ML networks: Yolov5m (medium-sized) with ResNet50 backbone (see [Jo22], [RF18]) and FCOS with ResNet50 backbone (see [Ti20], [De22]). YOLOv5m uses anchors tuned to plant detection, while FCOS does not need any predefined anchors, thus comparing these models provides an insight into overall accuracy when there are no anchors involved. Both networks use a ResNet50 backbone which was pretrained on the ImageNet data set. We used the initialized weights from that pretraining as a starting point for all our experiments.

For each network, we ran eight experiments falling into four categories: training on synthetic data and testing on synthetic data (experiment Y1 and F1), training on real data and testing on real data (Y2/F2), training on synthetic data and testing on real data (Y3-7/F3-7), and training on a mix of synthetic and real data and testing on real data (Y8/F8).

The intention of experiments Y1/F1 and Y2/F2 was to establish a baseline of performance that could be achieved with our data sets and chosen models (with the added element of sanity-checking our synthetic data). Most experiments, however, are focused on evaluating how well the models performed when mainly or exclusively trained on synthetic data in this use case, thus exploring the impact of the domain gap between real and synthetic data. Following the recommendation of Nowruzi et. al. [No19], we implemented the idea of bootstrapping the network, i.e., first training the network only on synthetic data and then retraining it on a very small subset of the target domain data.

The networks were trained for at maximum 150 epochs, with some training cycles being stopped early when no progress was made on the loss values. We used the default hyperparameter values during training which are provided in the model repositories (see [De22], [Jo22]). Due to the small area of the weeds in the image, we downscaled the images to a resolution of 1312 x 1312 pixel. Any further downscaling compromises the detection accuracy by a large margin. Therefore, all of the following experiments are performed at a resolution of 1312 x 1312 pixel for both training and testing.

We split the synthetic data set into 5000 images for training and 1500 images for testing. For some of the experiments only the first 500 out of the 5000 images for training were used. The images were not shuffled in order to ensure comparability.

The 731 real images from all days were split into fixed subsets of 400 images for training, 150 images for validation and 150 images for testing. This was done as opposed to randomly picking subsets to prevent data leakage. For each day, images from different rows of the field were selected to exclude the possibility of individual plants re-occurring between days. Within these rows, data was spatially split between training and validation/test images to avoid overlap between consecutive images affecting separation of the data subsets. To ensure comparability, we used the same subsets of real data for all experiments. We combined training, validation and testing images into one subset for those experiments where the models were purely trained on synthetic images to increase the amount of available real images.

4 Results and discussion

Table 2 shows the results of all experiments, while Figure 5 shows an example output of a trained YOLO model. For training and validation on synthetic data, both YOLO and FCOS achieve high accuracy for both classes (see rows Y1 and F1), with higher values overall in the case of synthetic data. This establishes an upper bound on model accuracy in the absence of a domain gap and validates our synthetic data generation setup. Similarly, when the models are trained on real data, it yields high accuracies on the real validation data (see Y2/F2). The values of e.g., 95.6 and 90.5 on maize for YOLO and FCOS, respectively, are a high baseline to be reached for mixed training setups.

Overall, experiments with the models trained on synthetic data show a lower performance on the real test data compared to the experiments using only real data, indicating a substantial domain gap between the two data sets. This may indicate an insufficient level of realism in the synthetic data in some respects, e.g., the variability of the rendered plant models. However, it is also to be noted that the synthetic data contains a larger variability in backgrounds and illumination, given that the real data set was recorded on only a few days and on the same soil. This may also be an explanation for the observation that training on larger synthetic data sets does not lead to an increase in performance (see row Y3, Y4 and F3, F4), again also indicating the limited variability and thus number of features in the plant models.

Model	Exp	Training data set	Eval data set	mAP50 All	mAP50 Weeds	mAP50 Maize
YOLO v5m	**Y1**	5000 syn	1500 syn	97	98.1	**95.9**
	Y2	400 real	150 real	85.6	75.3	**95.9**
	Y3	5000 syn	700 real	52.4	29.6	75.1
	Y4	500 syn	700 real	61.8	44.6	**78.6**
	Y5	500 syn	304 real d11	63.1	37.2	89.0
	Y6	500 syn	240 real d15	62.3	51.2	72.3
	Y7	500 syn	187 real d17	53.9	42.8	65.0
	Y8	500 syn + 31 real	700 real	75.3	66.1	**84.5**
FCOS	**F1**	5000 syn	1500 syn	93.09	94.5	**92.05**
	F2	400 real	150 real	79.15	67.8	**90.5**
	F3	5000 syn	700 real	38.89	20.70	57.60
	F4	500 syn	700 real	47.33	24.00	**70.66**
	F5	500 syn	304 real d11	53.60	25.40	81.80
	F6	500 syn	240 real d15	49.76	26.23	73.28
	F7	500 syn	187 real d17	23.49	12.95	34.04
	F8	500 syn + 31 real	700 real	55.99	48.86	**63.12**

Tab. 2: Comparison of mAP50 for two object detectors trained on real and synthetic data sets and evaluated on different evaluation splits

Furthermore, the networks perform with varying degrees of accuracy on different growth stages of maize plants (Y5-7/F5-7), hinting at different degrees of accuracy to which the growth stages in the real data were reproduced with our set of 3D models. That performance is overall worse on weeds than on maize (best mAP 66.1 for YOLOv5m and 48.86 for FCOS) is an indicator for the higher variability in the weeds. The presence of grassy weeds (monocotyledons) in the real data which were not labelled and are part of the background may contribute to this, as they are both misclassified as weeds or maize, in both cases with low confidence values.

When the models are trained on synthetic data and then trained on a small subset of the real data set (see row Y8 and F8), the performance of training the model on large sets of real data (Y2/F2) is not reached. YOLO reaches higher accuracies compared to purely training on synthetic data and testing on the whole real data set (Y8: 84.5 vs. Y4: 78.6). Similar behavior is not observed for FCOS (F8: 63.12 vs. F4: 70.66). One possible explanation for this is the anchor-less architecture of the FCOS. Since FCOS has no prior knowledge of the target domain in terms of anchors, it needs more data to perform well on that data. Compared to FCOS, YOLO learns faster with a small number of data points, which could be explained by its use of anchors.

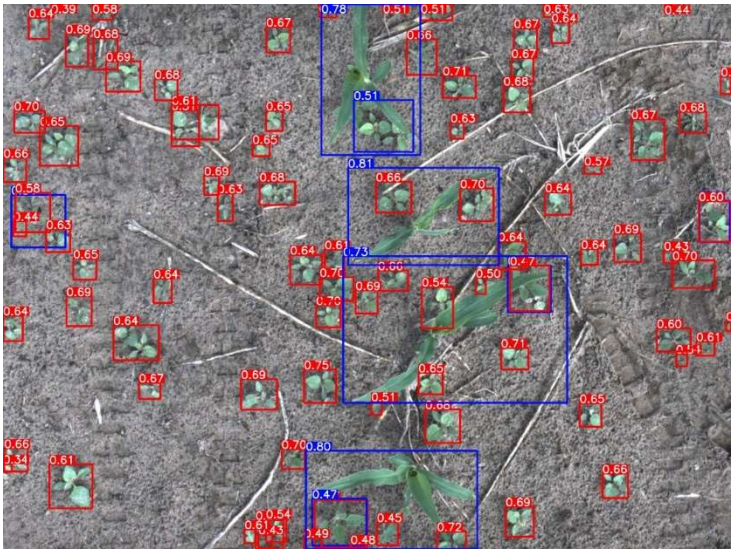

Fig. 5: Example image from real data set showing predictions for both Maize (blue) and Weeds (red) by YOLOv5m trained on 500 synthetic images, with the respective confidence values

5 Conclusion

We presented two data sets of a selective weeding use case, one obtained with established methods on a real field, the other generated using a synthetic data pipeline based on state-of-the-art 3D rendering tools. Our results show lower performance of current object detection models trained in a naïve way on synthetic data compared to models trained on real data, indicating a substantial reality gap. However, when using combinations of synthetic data with small sets of real data, accuracy can be improved, at least if the underlying model is able to adapt quickly to new data. It can be expected that reducing the domain gap apparent in the data from other experiments can further bolster performance, but more research is required to quantify the effects of the domain gap or the ratio of

synthetic and real data in mixed data sets on the achievable results.

Future work in this area should focus on identifying adequate metrics to quantify the domain gap between synthetic and real data sets, as well as on developing methods to further eliminate it, e.g. by improving the utilized 3D models – work that is under way in the Agri-Gaia project. This may open the door to significantly reducing the effort to collect and label real data if the performance of mixed training setups can consistently reach the performance of models trained on large amounts of real data. Even if this cannot be reached, the level of accuracy demonstrated here already makes this approach useful for the purpose of assisting in labelling real data, thus reducing overall labelling effort.

Acknowledgements: We would like to thank Hof Fleming for their support during the field campaign and the data recording in the field. The assembly of the sensor carrier and the integration of the sensors into the ROS environment was supported by our Chief Technician Maik Ludwig and our Software Engineer Gurunatraj Parthasarathy. For their image annotation work we would like to thank our students Qalab Abbas, Vera Klütz, Reshma Khan, Lara Lüking, Dibyashree Nahak and Simon Zielinski.

This work was supported by the German Federal Ministry for Economic Affairs and Climate Action within the Agri-Gaia project (grant number: 01MK21004A). The DFKI Niedersachsen (DFKI NI) is sponsored by the Ministry of Science and Culture of Lower Saxony and the VolkswagenStiftung.

Bibliography

[Ba18] Barth, R., Ijsselmuiden, J., Hemming, J., Henten, E.J.V.: Data synthesis methods for semantic segmentation in agriculture: A Capsicum annuum dataset. Computers and Electronics in Agriculture. 144, 284-296 (2018). https://doi.org/10.1016/j.compag.2017.12.001

[Bo21] Botyanszka, L.: A Review of Imaging and Sensing Technologies for Field Phenotyping. Acta Horticulturae et Regiotecturae. 24, 58–69 (2021). https://doi.org/10.2478/ahr-2021-0011

[CPN20] Carbone, C., Potena, C., Nardi, D.: Simulation of near Infrared Sensor in Unity for Plant-weed Segmentation Classification. Presented at the January 1 (2020). https://doi.org/10.5220/0009827900810090

[Ci17] Di Cicco, M., Potena, C., Grisetti, G., Pretto, A.: Automatic Model Based Dataset Generation for Fast and Accurate Crop and Weeds Detection, 2016.

[Cv22] CVAT, https://www.cvat.ai/, last accessed 2022/10/13

[De22] facebookresearch/detectron2, https://github.com/facebookresearch/detectron2, last accessed 2022/10/13

[Ha22] Harling, J.: Graswald – State of the art 3D nature, https://www.graswald3d.com/, last

accessed 2022/10/13

[Jo22] Jocher, G.: YOLOv5 by Ultralytics, https://github.com/ultralytics/yolov5, (2020). https://doi.org/10.5281/zenodo.3908559

[La22] Lac, L., Keresztes, B., Louargant, M., Donias, M., Da Costa, J.-P.: An annotated image dataset of vegetable crops at an early stage of growth for proximal sensing applications. Data in Brief. 42, 108035 (2022). https://doi.org/10.1016/j.dib.2022.108035

[Ma22] Maxtree: Maxtree – 3D Plant Models | CG Assets, https://maxtree.org/, last accessed 2022/10/13.

[MS22] Milioto, A., Stachniss, C.: Bonnet: An open-source training and deployment framework for semantic segmentation in robotics using cnns. Presented at the 2019 International Conference on Robotics and Automation (ICRA), 2019.

[Ni21] Nikolenko, S.I.: Synthetic Data for Deep Learning, http://arxiv.org/abs/1909.11512, (2019). https://doi.org/10.48550/arXiv.1909.11512

[No19] Nowruzi, F.E., Kapoor, P., Kolhatkar, D., Hassanat, F.A., Laganière, R., Rebut, J.: How much real data do we actually need: Analyzing object detection performance using synthetic and real data. CoRR. abs/1907.07061, 2019.

[RF18] Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767; 2018.

[Ti20] Tian, Z., Shen, C., Chen, H., He, T.: FCOS: A simple and strong anchor-free object detector, http://arxiv.org/abs/2006.09214, 2020.

[Ub18] Ubbens, J.: The use of plant models in deep learning: an application to leaf counting in rosette plants. 10, 2018.

[Xu16] Xu, N., Price, B., Cohen, S., Yang, J., Huang, T.: Deep Interactive Object Selection, http://arxiv.org/abs/1603.04042, (2016). https://doi.org/10.48550/arXiv.1603.04042

[Za22] Zaal, G., Tuytel, R., Cilliers, R., Cock, J.R., Barresi, D., Mischok, A., Majboroda, S., Savva, D., Guest, J.: Poly Haven • Poly Haven, https://polyhaven.com/, last accessed 2022/10/13.