

Deep Leaf Segmentation Using Synthetic Data

Daniel Ward
Daniel.Ward@data61.csiro.au
Peyman Moghadam
Peyman.Moghadam@data61.csiro.au
Nicolas Hudson
Nicolas.Hudson@data61.csiro.au

Robotics and Autonomous Systems
The Commonwealth Scientific and
Industrial Research Organisation
(CSIRO), Data61
Brisbane, Australia

Abstract

Automated segmentation of individual leaves of a plant in an image is a prerequisite to measure more complex phenotypic traits in high-throughput phenotyping. Applying state-of-the-art machine learning approaches to tackle leaf instance segmentation requires a large amount of manually annotated training data. Currently, the benchmark datasets for leaf segmentation contain only a few hundred labeled training images. In this paper, we propose a framework for leaf instance segmentation by augmenting real plant datasets with generated synthetic images of plants inspired by domain randomisation. We train a state-of-the-art deep learning segmentation architecture (Mask-RCNN) with a combination of real and synthetic images of *Arabidopsis* plants. Our proposed approach achieves 90% leaf segmentation score on the A1 test set outperforming the-state-of-the-art approaches for the CVPPP Leaf Segmentation Challenge (LSC). Our approach also achieves 81% mean performance over all five test datasets.

1 Introduction

To achieve sustainable agriculture, we need to expedite the breeding of new plant varieties which consume less water, land or fertilizer and have greater resistance to parasites and diseases while producing greater crop yields. Plant phenotyping is the process of relating plant varieties to genotype and environmental conditions and how these affect the observable plant traits.

In recent years, several computer vision and machine learning techniques [1, 2] have been proposed to increase the throughput of non-destructive phenotyping. Automatic, non-destructive extraction of plant parameters such as plant height, shape, leaf area index (LAI) or growth rate throughout the crop growth cycle are essential for rapid phenotype discovery and analysis. Phenotypic measurements are often done at two levels: plant-level measurements (projected LAI, height, plant architecture) or leaf-level measurements (individual leaf area, leaf count, leaf growth rate). This paper focuses on the former category.

Despite substantial progress, segmentation of individual leaves (leaf instance segmentation) remains extremely challenging owing to the variability in leaf shapes and appearance

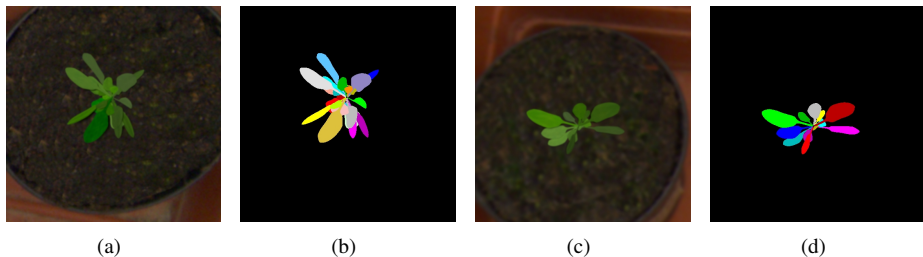


Figure 1: Example top down view of synthetic *Arabidopsis* plants (a, c) and their segmentation labels (b, d).

over the life-cycle of the deformable object (plant). Another major challenge is that leaves partially overlap and occlude each other as new leaves grow.

Currently, leading instance segmentation techniques [6] based on deep convolutional neural networks require huge amounts of annotated training data. Manual annotation of images for a task such as instance segmentation is often time consuming and expensive. For leaf instance segmentation there are only a few annotated datasets available and the size of these datasets is very small (a few hundred images) [14]. In this paper we use the CVPPP dataset from the leading Leaf Segmentation Challenge (LSC). It is a small dataset of top-down 2D visible light images of *Arabidopsis* and tobacco plants from an indoor high throughput plant phenotyping system. This dataset only contains 27 images of tobacco and 783 *Arabidopsis* images with pixel-level leaf segmentation labels.

To address these problems, we propose a framework to augment limited annotated training data of real plants with synthetic plant images inspired by domain randomisation [21]. Synthetic plant images come without the cost of data collection and annotation (examples shown in Figure 1). Our main contributions can be summarised as follows:

- We outperform the state-of-the-art leaf segmentation approach by augmenting real data with synthetic data.
- This is the first time synthetic data has been used for the CVPPP Leaf Segmentation Challenge using a novel plant modeling system.
- We investigate the application of domain randomisation concepts on deformable objects (plants) and found that sampling realistic foreground and background textures improved results.

2 Related Work

There has been significant progress in the vision community on object instance segmentation algorithms, which requires the detection of each object (or leaf) in the image, and then the precise segmentation of each instance.

In this paper, we consider the Mask-RCNN [6] framework, as it achieved state of the art single-model results on the COCO instance segmentation task [14] in 2017, and feature pyramid derivatives of it remain the competition leaders. With the addition of the branch for predicting segmentation masks, Mask-RCNN is a low overhead extension of Faster-RCNN, which recently also demonstrated the best detection performance among a family of object detectors [8].

Training detection or segmentation algorithms takes vast amounts of labeled data: COCO instance segmentation training has 123,287 images; ImageNet [18] detection has over one million images with annotated object bounding boxes, although some classes in it, such as *strawberry*, will appear in only hundreds of those images. One of the few large scale data sets where objects are broken down further into segmented components is ADE20K [25], with 20,210 images with all objects, and components within segmented.

In comparison, the CVPPP LSC uses 810 images for training, each with a single plant with multiple leaf instances. Increasing the size of this training data set would lead to increased test set performance. Recent work in classifying birds [23], while varying the training data set from 10 to 10,000 images, suggested a rule where the test set error drops by a factor of 2x each for every 10x increase in the number of training images.

Given a limited training dataset, where it is hard or expensive to annotate more real world data, recent work in computer vision looks at using simulated data to train networks. It is known that the *reality gap*, or the inability of renders to create the statistics of the real world, lead to results where naive training and evaluations on synthetic images do not transfer to the real world. The literature has several approaches to incorporating synthetic imagery including: Mixing generated objects over real world scenes [11, 9]; domain randomisation [21], where parameters of the simulator creating the images are varied in non-realistic ways to make the networks invariant to them; creating photo-realistic renders [10], where substantial effort is made to reduce the reality gap directly; and General Adversarial Networks (GANs) [2] to automatically learn and compensate for important differences between simulated and real world data. Freezing a pre-trained feature extractor layer in the networks and using simulation to only train output layers [2] has shown amazing results for a limited set of rigid objects, when in fact the pre-trained features are expressive enough. Hinterstoisser *et al.* [2] found that this resulted in features closer to those extracted from real images being extracted from the synthetic images.

To the best of our knowledge, synthetic data has been used twice before in conjunction with the CVPPP datasets. In the absence of generating synthetic segmentation labels, both approaches focused on estimating the count of leaves in each image. Furthermore, each method’s synthetic data was rendered without a background. ARIGAN [6] presents a data driven approach where a GAN learns the distribution of the CVPPP A1 subset dataset and generates new samples to an expected leaf count. GANs are renowned for more accurately modeling texture than geometry and this is evident in the ARIGAN synthetic *Arabidopsis* dataset. In the second synthetic data approach, Ubbens *et al.* [22] leverages a model based on the L-system formalism. L-systems are commonly used to describe the structure and development of plants. The design of their *Arabidopsis* L-system model involved defining plant attributes including leaf shape and inclination angle. These attributes were defined using functions drawn by manually placing control points within an L-systems library [15].

3 Method

This work augments and compares performance of real world training to using synthetic leaf images. Leveraging ideas in the domain randomization literature, leaf textures are randomised and simulated data is overlaid onto random real world scenes. In addition, we evaluate freezing feature layers in models like in [2]. The work here extends these ideas to deformable plant models where each object is made up of many leaf components, and does not adhere to the rigid body models used in most synthetic data approaches.

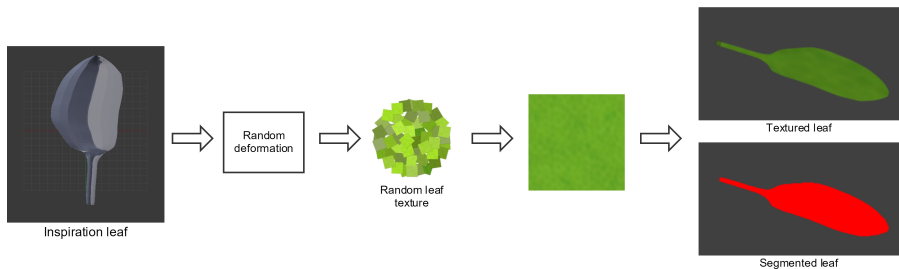


Figure 2: The synthetic leaf generation pipeline. Each leaf in a synthesised plant is randomly deformed, textured and positioned.

3.1 Synthetic Data Generation for Arabidopsis

Our synthetic *Arabidopsis* generation pipeline (Figures 2 and 3) begins with manual specification of geometric properties. This is done by defining an *inspiration leaf* rather than multiple functions like in L-systems [22]. First, an inspiration leaf was designed by tracing a randomly chosen *Arabidopsis* leaf image in Blender to produce a 3D mesh by adding and adjusting manipulation points. This was done only once for all synthetic data used in this study. Figure 2, displays the steps involved in generating a single leaf from the original inspiration.

In order to model leaves of different shape and size, every leaf is randomly scaled along each axis independently. This process differs from other applications such as synthetic cars, where the resulting significant changes to object aspect ratio would be undesirable. To provide variation in texture between leaves, each one is rendered with a different texture. Thirty leaf textures were extracted from the CVPPP training datasets and augmented using the data augmentation as described in Section 3.2 with the addition of random adjustments to image exposure.

Being rosettes, the leaves of an *Arabidopsis* plant are arranged circularly. Furthermore, all leaves are at a similar height and remain close together [23]. Following this, our model can be intuitively thought of as an arrangement of leaves stemming out from a sphere. The location of leaf stems is most likely at the equator and equally likely at any point around the equator in the xy plane. With the leaf initialised at the origin in the xy plane, the final position of the leaf is defined by rotating the mesh around the x , y and z axes independently. To produce the rosette, the rotation around the z axis is sampled from a uniform distribution, $U(0, 2\pi)$. Similarly, the leaf pitch and roll are sampled from $U(-\frac{\pi}{4}, \frac{\pi}{4})$ and $U(0, \frac{\pi}{4})$ respectively to subtly randomise leaf pose. Uniform distributions were used to provide a wide variety of leaf positions. This process is shown in Figure 3 and repeated multiple times according to a normal distribution of leaves per plant centered around the mean leaf count in the CVPPP A1 subset dataset but with wider variance (Figure 6). Figure 3 defines the plant generation pipeline. The number of leaves in a generated plant is first sampled from a provided distribution and the leaf generation pipeline (Figure 2) is called. A top down view of the 3D plant model is then rendered to simulate the CVPPP data. Like [21], the camera and lighting positions were varied however only a single light source was used. The camera position randomisations were restricted to ensure a top down view was still obtained. Using a sample from each CVPPP dataset, A1-A4, we extracted the background images to obtain 4 plant pot background samples. These were then processed using the data augmentation described in Section 3.2 and used for the render backgrounds. Following [21], we applied Gaussian blurring to the rendered plant to better integrate it into the real world scene (back-

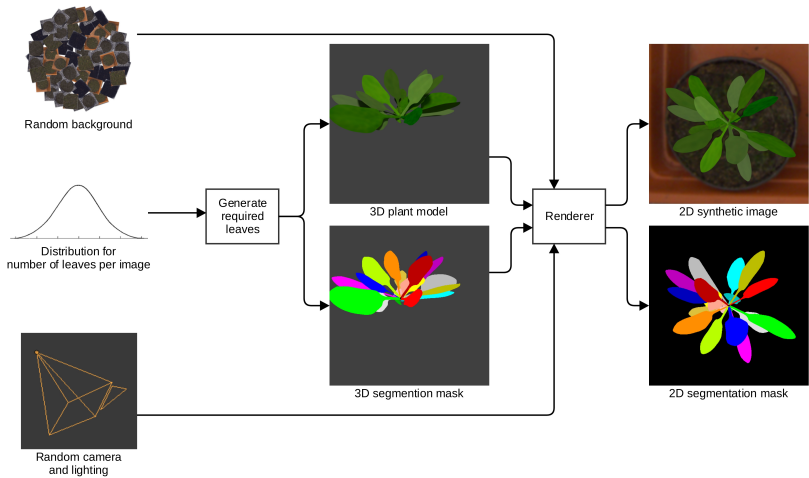


Figure 3: The synthetic data pipeline. A 3D plant model of synthetic leaves (Figure 2) is assembled and rendered over random plant pot background.

ground).

Our synthetic data pipeline was designed in Blender (v2.79b). We use the Cycles renderer to produce the synthetic images with Lambertian and Oren-Nayar diffuse reflection shading. The segmentation labels were rendered using the native Blender renderer with all shading and anti-aliasing disabled to ensure their integrity.

3.2 Training & Evaluation

We used the Matterport¹ Mask-RCNN implementation for our experiments [6]. Mask-RCNN is a two-stage model, with a feature extractor feeding into a Region Proposal Network (RPN) and then into three heads producing box classification, box regression, and an object mask. We use a ResNet101 backbone with a Feature Pyramid Network for Mask-RCNN. Unlike the original implementation, we use 256 (opposed to 512) regions of interest per image during training.

In each experiment, the model was initialised with weights trained on the COCO dataset. It was then retrained on a combination of real world and synthetic data. Data augmentation is a common practice to improve model robustness and prevent over fitting. In all experiments random: left-right flipping; top-bottom flipping; rotation; zooming and cropping were applied to the training images. Data augmentation was applied to real and synthetic datasets. Our training procedure consisted of splitting the data into 80%, 20% training and cross validation sets respectively; batch normalisation across a batch size of 6 and leveraging early stopping. In experiments where both real and synthetic data were used for training, each batch was balanced such that 50% of images came from each dataset respectively.

To evaluate our methods, we compete in the CVPPP Leaf Segmentation Challenge (LSC)². It consists of three *Arabidopsis* plant (A1, A2, A4) datasets and a dataset of young tobacco plant images, A3. Training sets for A1, A2, A3, A4 contain 128, 31, 27, 624 images while the test sets have 33, 9, 65, 168 images respectively. Additionally, test set, A5, is also pro-

¹https://github.com/matterport/Mask_RCNN

²<https://competitions.codalab.org/competitions/18405>

vided which combines images from all testing datasets in order to evaluate the generalisation of proposed machine learning techniques. In order to compare our leaf instance segmentation performance with the current state-of-the-art approaches from the leaf segmentation Challenge (LSC), we only use the A1 dataset for training.

Using the method described in Section 3.1, three synthetic Arabidopsis datasets were generated. Each contained 10,000 labeled images and are defined as follows. In *Synthetic-plant*; all leaves in all plants contain the same green texture. In *Synthetic-leaf*; Each leaf in a plant is rendered with a random different green texture as shown in Figure 2. Finally, *Synthetic-COCO* is generated following the same process however the leaf textures are sampled from the COCO dataset. *CVPPP-A1* was the real dataset used during the experiments which consists of the 128 images from the CVPPP A1 dataset. The training procedure matching real and synthetic images per batch described above was named *Synthetic+real*.

4 Results & Discussion

In this section we present our performance on the CVPPP LSC, compare our results to the state-of-the-art and investigate the limitations of the synthetic datasets. Table 1 displays the results for each experiment and dataset in Section 3.2. The results shown are performances on the 5 test sets of the CVPPP LSC. Segmentation performances are quoted in the competition metric, symmetric best dice (SBD) score. Note, the column titled *Mean* presents the mean per test image performance across all test sets as provided by the competition and not the mean across the table row.

Training Regime	CVPPP test set (SBD)					
	A1	A2	A3	A4	A5	Mean
CVPPP-A1 (real)	87	71	59	73	70	71
Synthetic-plant	81	69	41	84	73	74
Synthetic-leaf	82	74	32	86	75	74
Synthetic-COCO	57	48	25	48	43	44
Synthetic+real	90	81	51	88	82	81

Table 1: CVPPP LSC results. Mask-RCNN trained on real data (CVPPP-A1), synthetic data (-plant, -leaf and -COCO) and both real and synthetic (Synthetic+real) data. Segmentation scores are quoted in symmetric best dice (SBD).

We achieve the best results on A1, A2, A4 and A5 when training on both synthetic and real data. Best performance on the A3 subset, however, is obtained by training on real data only. The generalisation of *Synthetic+real* is shown by a significantly higher mean result. Furthermore, the effect of varied texture can be seen in *Synthetic-plant*, *Synthetic-leaf* and *Synthetic-COCO*, where training on randomly textured leaves (*Synthetic-COCO*) achieved the lowest performance. A performance increase is seen when training on plants with different green textured leaves (*Synthetic-leaf*) on all test sets except A3 where *Synthetic-plant* excels.

Figure 4 compares the CVPPP A1 test set prediction for plant 117 from the A1 test set. The artifact in the background of the image, has been incorrectly segmented by both models trained on real only and synthetic data only. The model trained on both real and synthetic data, however, detects more of leaves and does not incorrectly segment the background arti-

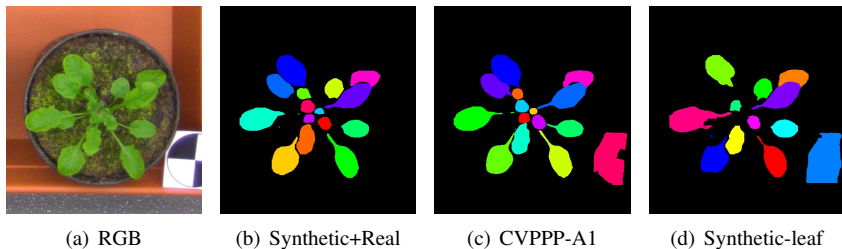


Figure 4: Segmentations of plant 117 (a) from the A1 test set. Predictions are presented from models trained on both synthetic and real data (b) and only real data (c) or synthetic data (d). Only the model trained on both synthetic and real data (b) avoided segmenting the background artifact (fiducial marker). Note that no fiducial markers appear in the real CVPPP A1 or any of the synthetic training datasets.

fact. No fiducial markers, the artifact, appear in any training data, real or synthetic. The false positive segmentation is likely explained by the curved geometry and edges of the object. Further, a possible reason for the model correctly not segmenting the artifact is the greater variance in geometry in the *Synthetic+real* dataset.

In addition to the results in Table 1, we experimented with freezing the feature extractor layers of Mask-RCNN when training on synthetic data as in [24] (see Section 2). This led to degraded results, similar to [24]. We are not confident if this is because of using flexible and deformable plant models as opposed to rigid CAD models that match the test set, or if there is some deficiency in our methodology, as we were constrained in producing synthetic sets at the scales used in [24].

Method	CVPPP test set (SBD)		
	A1	A2	A3
RIS + CRF [17]	66.6	-	-
MSU [24]	66.7	66.6	59.2
Nottingham [20]	68.3	71.3	51.6
Wageningen [24]	71.1	75.7	57.6
IPK [14]	74.4	76.9	53.3
Salvador <i>et al.</i> [19]	74.7	-	-
Brabandere <i>et al.</i> [8]	84.2	-	-
Ren <i>et al.</i> [16]	84.9	-	-
Ours	90.0	81.0	51.0

Table 2: CVPPP LSC results. We compare the performance of our best performing model (*Synthetic+real*(A1)) to existing approaches and outperform the A1 and A2 state-of-the-art. We achieve comparable results for A3 (*Tobacco*). Note the real data used to train our model came solely from A1 (*Arabidopsis*). To the best of our knowledge results for A4 and A5 have not yet been published and are hence omitted. Segmentation scores are quoted in symmetric best dice (SBD).

Our proposed approach (trained on real and synthetic data) achieves 90% SBD score outperforming the state-of-the-art leaf instance segmentation algorithms (See Table 2) for the A1 test set. The real data used to train our model came solely from the A1 data set.

Our model also achieves state-of-the-art and comparable performance on A2 and A3 respectively. The A3 dataset contains images of tobacco plants rather than *Arabidopsis* and unlike the compared methods, our model is only trained on images of *Arabidopsis* plants. The reduced performance is explored below. The networks trained only on synthetic data, on average outperformed the models using only the limited A1 training data set (128 images). This is likely due, in a large part, to the increase in variance across the distribution size and position of leaves in the synthetic set, and the much larger size of the data set itself (10,000 images vs 128). The positional variation in the synthetic dataset is also likely why the synthetically trained models perform particularly well on the A4 test set, which has larger range of plant age ranges than A1. These reasons also explain the incorrect segmentation phenomena shown in Figure 4.

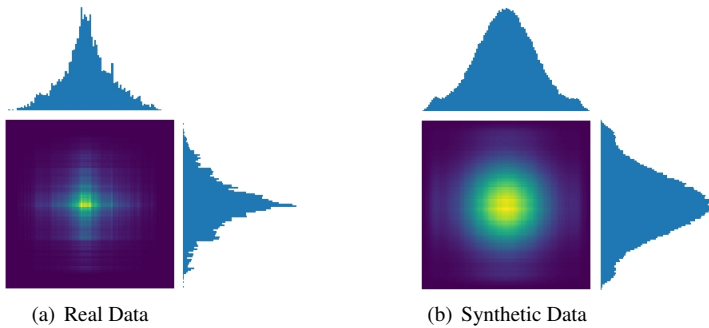


Figure 5: The distribution of leaf centroid locations in the training images. A wider distribution of leaf locations is presented in the synthetic data (b) while preserving the center bias. Note a uniform distribution across the image could be achieved by tuning the synthetic data pipeline.

Synthetic data combined with domain randomization provides the ability to increase the variability across a dataset compared to data augmentation (where the real data is cropped, stretched, blurred, etc). Figure 5 shows the distribution of leaf centroid locations between the CVPPP A1 and synthetic data sets. Clearly, in the real dataset (CVPPP A1), the majority of leaves appear around the center of the image. The synthetic data, while biased towards the image center, contains a wider distribution of leaf positions. A wider distribution of data to train on, improves a model’s ability to generalize. Figure 6 shows the distribution of number of leaves per image for CVPPP A1 and synthetic images. While the mean has been matched in the synthetic data, a larger variance and distribution tails demonstrates the wider distribution to learn from.

Within the models trained on only synthetic data, changing the texture per leaf (*Synthetic-leaf*), performed the best, except in the case of the tobacco plant (A3 in Table 1). The improvement of a per leaf texture over a uniform texture choice per plant led us to consider the basic premise of domain randomization: the model was likely over-fitting to the simplicity and unrealistic distribution of our simulated textures, so increasing the variance in the training data should make the model more invariant to the distribution difference in the test set. However, attempting to add extremely high variance textures (*Synthetic-COCO*), lead to extremely poor results, indicating that plant texture is an important aspect in segmenting leaves. This relevance of texture can also be seen in the comparatively poor performance for the A3 tobacco plant test set. The larger leaves and image resolution for A3 can lead

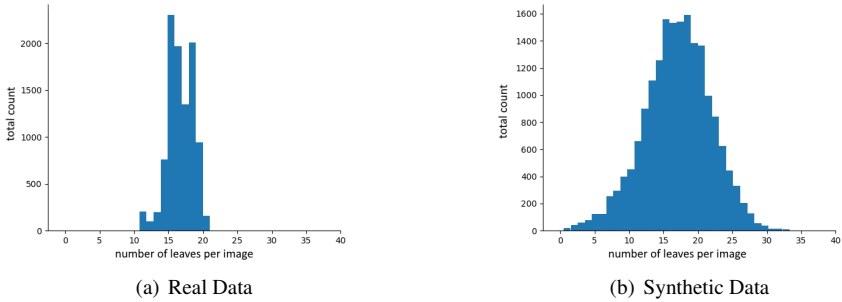


Figure 6: The distribution of the number of leaves per image in the training images. Greater variance is presented in the synthetic data (b) while preserving the mean.

to more visually prominent leaf veins. The over-segmentation of the leaf along a visually prominent vein can be seen in Figure 7. Blurring these images before segmentation saw an improvement in performance for these cases, but resulted in the under-segmentation of leaves on the smaller plants. Resilience to the distinct shadow on the largest leaf (Figure 7) can also be seen. The lack of over-segmentation is likely due to the synthetic data containing hard shadowing effects from the 3D geometric plant models and the rendering pipeline.



Figure 7: Plant 70 of the A3 test set. Prominent textural edges from the leaf veins lead to over-segmentation of the leaf. Further, invariance to shadowing can be seen as it has no effect on the segmentation.

The higher performance of the *CVPPP-AI* model trained on the tobacco (A3) test set, indicates the real data contains important textural information not present in the simulated data, which leads the model to not over-segment the tobacco leaves. The best results across all *CVPPP* LSC test sets were from training with mixed batches containing both real and synthetic data. We conclude this is due to the textural information from the real data, and the increased geometric diversity from the simulated plant data.

In addition we note that training on real and synthetic data also helps reject false positive detection on the background (e.g. Figure 4). In this case, the synthetic data did not contain any backgrounds with any black and white fiducial markers. The extra background variance across the combined set may have enabled the model to not detect this artifact.

These observations indicate that further performance benefits could be derived by increasing the variation of both the background images and the foreground textures. A simple

method to increase background variation would be to physically photograph the nursery before the plants start growing. In addition distractor objects (such as the white fiducial marker) could be added to the foreground [9]. The foreground leaf textures used in our experiments were sampled from a small, manually created, set of 30 leaf crops. Increased performance could then be expected by leveraging more sophisticated textures, both increasing the variation and perhaps by making the image more photo-realistic. There are several leaf databases that would be suitable for extracting this texture data [9].

Furthermore, to increase the variance in textures, further investigations will consider more sophisticated approaches synthesizing leaf and plant geometry leveraging L-systems. The methods used in this paper were all based off a single inspiration leaf of an *Arabidopsis* plant, with relatively compact leaf shapes. Tobacco plants (A3), however, have broader leaves where are not as well separated from each other. By creating training data sets with larger leaf and plant geometry variance, or at least matching the geometry to new target domains, we expect further increases in performance. This may also include incorporating leaf vein structures in the synthetic data to address the phenomenon presented in Figure 7.

Our generated synthetic dataset is publicly available at ³. The synthetic dataset contains 10,000 top down images of synthetic *Arabidopsis* plants and their corresponding 2D segmentation labels.

5 Conclusion

We have presented and evaluated a framework for generating synthetic *Arabidopsis* images and accompanying instance segmentation labels. By inspiring geometry and sampling real world textures for background and foreground, we have outperformed state-of-the-art results achieving 90% symmetric best dice score on the CVPPP A1 test set. Furthermore, this framework has the potential to be rapidly adapted to a new plant and automate leaf segmentation within a plant phenotyping setting. We achieved best results across all test sets, except CVPPP A3, when training on both the real data and synthetic data. Presumably, this is due to the extra textural information from the real data and the increased geometric diversity from the simulated plant data.

Future work will be directed at improving the variation in geometry and texture of the synthetic data. We would expect a performance and generalisation increase by sampling leaf textures from a wider, domain specific distributions. For geometric variation, supporting a structure description framework such as L-systems would alleviate the ability to simulate different plant species.

Acknowledgements: This research was funded by the CSIRO AgTech Cluster for Robotics and Autonomous Systems. The authors thank the Robotics and Autonomous Systems team for their valuable feedback and discussions, notably Mark Cox, Kamil Grycz, Terry Kung and Micheal Thoreau. Further, we thank Robert Lee for his advice and support.

³<https://research.csiro.au/robotics/databases>

References

- [1] Hassan Abu Alhaija, Siva Karthik Mustikovela, Lars Mescheder, Andreas Geiger, and Carsten Rother. Augmented reality meets deep learning for car instance segmentation in urban scenes. In *Proceedings of BMVC 2017 and Workshops*. British Machine Vision Association, 2017.
- [2] Konstantinos Bousmalis, Nathan Silberman, David Dohan, Dumitru Erhan, and Dilip Krishnan. Unsupervised pixel-level domain adaptation with generative adversarial networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, page 7, 2017.
- [3] Bert De Brabandere, Davy Neven, and Luc Van Gool. Semantic instance segmentation with a discriminative loss function. *arXiv preprint arXiv:1708.02551*, 2017.
- [4] Debidatta Dwivedi, Ishan Misra, and Martial Hebert. Cut, paste and learn: Surprisingly easy synthesis for instance detection. In *The IEEE international conference on computer vision (ICCV)*, 2017.
- [5] Mario Valerio Giuffrida, Hanno Scharr, and Sotirios A Tsaftaris. Arigan: Synthetic arabidopsis plants using generative adversarial network. In *Proceedings of the 2017 IEEE International Conference on Computer Vision Workshop (ICCVW), Venice, Italy*, pages 22–29, 2017.
- [6] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2980–2988. IEEE, 2017.
- [7] Stefan Hinterstoisser, Vincent Lepetit, Paul Wohlhart, and Kurt Konolige. On pre-trained image features and synthetic images for deep learning. *arXiv preprint arXiv:1710.10710*, 2017.
- [8] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, et al. Speed/accuracy trade-offs for modern convolutional object detectors. In *IEEE CVPR*, volume 4, 2017.
- [9] Neeraj Kumar, Peter N. Belhumeur, Arijit Biswas, David W. Jacobs, W. John Kress, Ida Lopez, and JoÃO V. B. Soares. Leafsnap: A computer vision system for automatic plant species identification. In *The 12th European Conference on Computer Vision (ECCV)*, October 2012.
- [10] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [11] Massimo Minervini, Andreas Fischbach, Hanno Scharr, and Sotirios A Tsaftaris. Finely-grained annotated datasets for image-based plant phenotyping. *Pattern recognition letters*, 81:80–89, 2016.

- [12] Peyman Moghadam, Daniel Ward, Ethan Goan, Srimal Jayawardena, Pavan Sikka, and Emili Hernandez. Plant disease detection using hyperspectral imaging. In *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–8, 2017.
- [13] Lars Mündermann, Yvette Erasmus, Brendan Lane, Enrico Coen, and Przemyslaw Prusinkiewicz. Quantitative modeling of arabidopsis development. *Plant physiology*, 139(2):960–968, 2005.
- [14] Jean-Michel Pape and Christian Klukas. 3-d histogram-based segmentation and leaf detection for rosette plants. In *European Conference on Computer Vision*, pages 61–74. Springer, 2014.
- [15] Przemyslaw Prusinkiewicz. Art and science of life: Designing and growing virtual plants with l-systems. In *XXVI International Horticultural Congress: Nursery Crops; Development, Evaluation, Production and Use 630*, pages 15–28, 2002.
- [16] Mengye Ren and Richard S Zemel. End-to-end instance segmentation with recurrent attention. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA*, pages 21–26, 2017.
- [17] Bernardino Romera-Paredes and Philip Hilaire Sean Torr. Recurrent instance segmentation. In *European Conference on Computer Vision*, pages 312–329. Springer, 2016.
- [18] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y.
- [19] Amaia Salvador, Míriam Bellver, Manel Baradad, Víctor Campos, F. Marqués, Jordi Torres, and X. Giró-i Nieto. Recurrent neural networks for semantic instance segmentation. In *CVPR 2018 DeepVision Workshop, 06/2018 In Press*.
- [20] Hanno Schar, Massimo Minervini, Andrew P French, Christian Klukas, David M Kramer, Xiaoming Liu, Imanol Luengo, Jean-Michel Pape, Gerrit Polder, Danijela Vukadinovic, et al. Leaf segmentation in plant phenotyping: a collation study. *Machine vision and applications*, 27(4):585–606, 2016.
- [21] Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaad Boochoon, and Stan Birchfield. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. *arXiv preprint arXiv:1804.06516*, 2018.
- [22] Jordan Ubbens, Mikolaj Cieslak, Przemyslaw Prusinkiewicz, and Ian Stavness. The use of plant models in deep learning: an application to leaf counting in rosette plants. *Plant methods*, 14(1):6, 2018.
- [23] Grant Van Horn and Pietro Perona. The devil is in the tails: Fine-grained classification in the wild. *arXiv preprint arXiv:1709.01450*, 2017.

- [24] Xi Yin, Xiaoming Liu, Jin Chen, and David M Kramer. Multi-leaf tracking from fluorescence plant videos. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 408–412. IEEE, 2014.
- [25] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.