# USING SYNTHETIC DATA TO BOOST AUTOMATED IMAGE-BASED PLANT PHENOTYPING

Daniel Ward, Peyman Moghadam, Nicolas Hudson
Robotics and Autonomous Systems, CSIRO Data61

**www.data61.csiro.au**

We present a framework for leaf instance segmentation by augmenting real plant datasets with generated synthetic plant images of rosette-shaped appearance. We present a synthetic data generation pipeline for obtaining labelled data required for leaf segmentation algorithms. The data pipeline produces 3D plant models from which a 2D top down colour image is rendered. We train a deep learning segmentation architecture (Mask-RCNN) on a combination of real and synthetic images of Arabidopsis plants. Our proposed approach achieves 90% leaf instance segmentation score outperforming the state-of-the-art approaches for the Computer Vision Problems in Plant Phenotyping (CVPPP) Leaf Segmentation Challenge (LSC) [2].

## Problem

- Deep learning algorithms require large amounts of labelled data.
- Compared to common datasets, limited annotated leaf data is available.
- Different plants and species differ significantly in shape and appearance.
- Collecting sufficient training data for different plants is intractable.

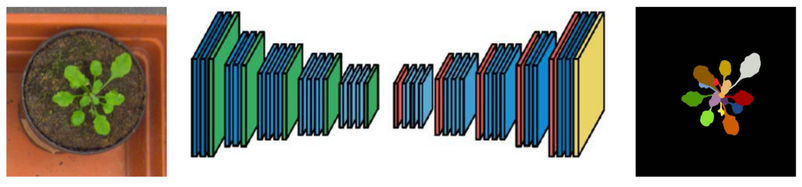


**Figure 1:** A visualisation of a deep learning segmentation approach. Given sufficient training data (left) and corresponding labels (right), a network (centre) learns a mapping from the input image to the leaf instance segmentation labels.

### The CVPPP Leaf Segmentation Challenge

- Contains five datasets of two plant species: Arabidopsis plants (A1, A2, A4) and Tobacco plants (A3). A5 contains a mix of Arabidopsis and Tobacco plants.
- A1, A2, A3, A4 contain 128, 31, 27, 624 training images and 33, 9, 65, 168 testing images respectively.

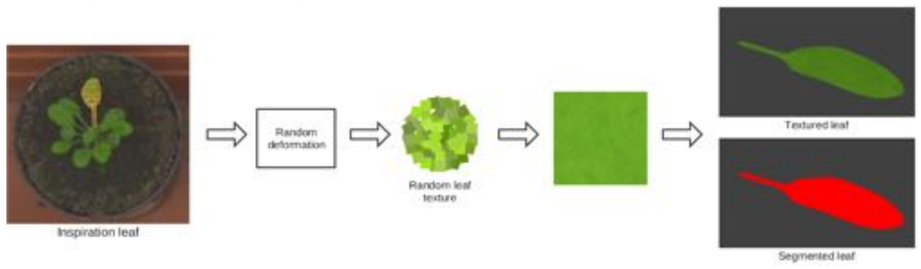
**Table 1:** The number of images in common computer vision datasets. Popular datasets contain significantly more images than the amount of data available for the CVPPP leaf segmentation challenge.

| Dataset | Application | Number of images |
|---|---|---|
| Pascal VOC | Object segmentation | 10 000 |
| MS COCO | Object segmentation | 330 000 |
| ImageNet | Object Recognition | 500 000 |
| CVPPP | Leaf segmentation | 1000 |

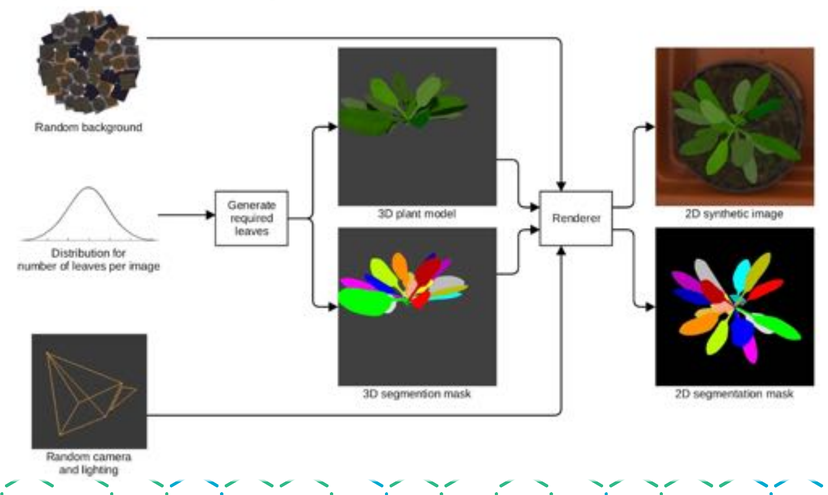
## Synthetic Data Generation

- We propose a framework to generate training data and segmentation labels.
- The synthetic data comes without the cost of data collection or annotation.

### Leaf Synthesis Pipeline




### Plant Synthesis Pipeline




## Leaf Instance Segmentation Results

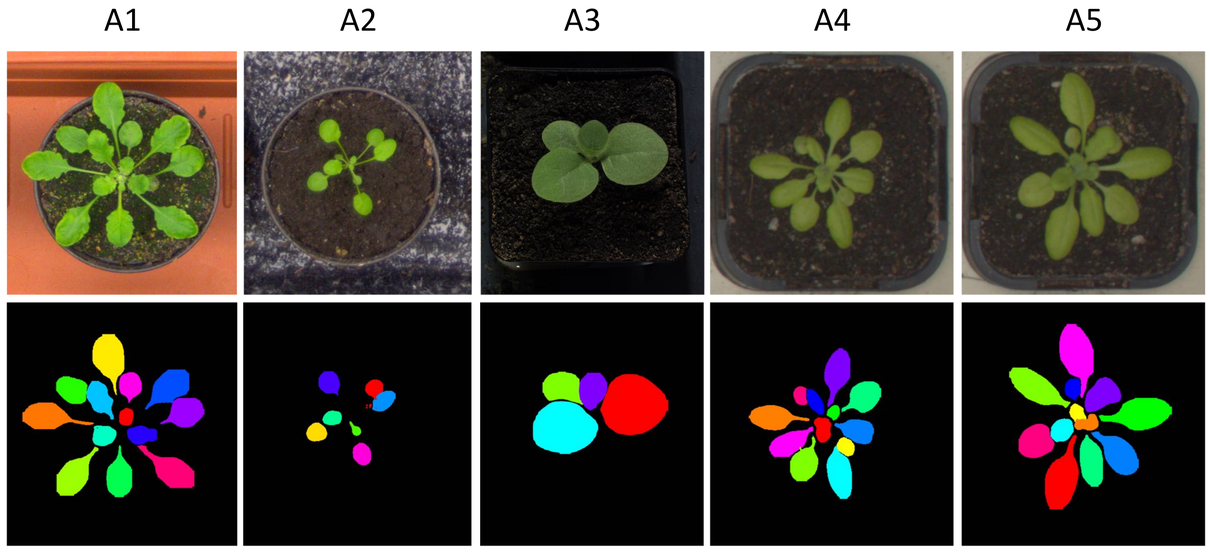



**Figure 2:** The first row shows images from the CVPPP LSC datasets. Our proposed leaf instance segmentation predictions are shown in the second row. We train Mask-RCNN on a combination of real and synthetic plant images.

**Table 2:** Leaf segmentation results for the CVPPP LSC test datasets compared to the state-of-the-art. Note that other approaches were individually trained for each test dataset (A1, A2, A3) while our method was only trained using A1 plants (Arabidopsis). Dataset A3 contains tobacco plants.

| APPROACH | CVPPP LSC TEST SET PERFORMANCE (SBD) | | |
|---|---|---|---|
| | A1 | A2 | A3 |
| RIS + CRF [17] | 66.6 | - | - |
| MSU [20] | 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 et al. [19] | 74.7 | - | - |
| Brandenbare et al. [3] | 84.2 | - | - |
| Ren et al. [16] | 84.9 | - | - |
| Ours [1] | **90.0** | **81.0** | 51.0 |

## Synthetic Arabidopsis Plant Dataset

- 10,000 synthetically generated Arabidopsis plant images.
- Leaf instance segmentation labels included.
- Publicly available at: https://research.csiro.au/robotics/databases/synthetic-arabidopsis-dataset/

**Figure 3:** A selection of synthetically generated Arabidopsis plants and their corresponding leaf instance segmentation labels. Note the variation in leaf shape and size and the variation in backgrounds.

**FOR FURTHER INFORMATION**

Dr. Peyman Moghadam
**e** peyman.moghadam@csiro.au

Mr. Daniel Ward
**e** daniel.ward@data61.csiro.au

**REFERENCES**

[1]: D. Ward, P. Moghadam, and N. Hudson. Deep leaf segmentation using synthetic data. In Proceedings of the British Machine Vision Conference (BMVC) Workshop on Computer Vision Problems in Plant Phenotyping (CVPPP), 2018.

[2]: CVPPP Leaf Segmentation Challenge, https://competitions.codalab.org/competitions/18405

**ACKNOWLEDGEMENTS**

This research was funded by the CSIRO AgTech Cluster for Robotics and Autonomous Systems. https://research.csiro.au/robotics/

Our Paper:

Our Dataset: