# Recent Data Augmentation Strategies for Deep Learning in Plant Phenotyping and Their Significance

1st Douglas Pinto Sampaio Gomes
*School of Computing and Mathematics*
*Charles Sturt University*
Melbourne, Australia
douglas.uf@gmail.com

2nd Lihong Zheng
*School of Computing and Mathematics*
*Charles Sturt University*
Wagga Wagga, Australia
lzheng@csu.edu.au

***Abstract*—Plant phenotyping concerns the study of plant traits resulted from their interaction with their environment. Computer vision (CV) techniques represent promising, non-invasive approaches for related tasks such as leaf counting, defining leaf area, and tracking plant growth. Between potential CV techniques, deep learning has been prevalent in the last couple of years. Such an increase in interest happened mainly due to the release of a data set containing rosette plants that defined objective metrics to benchmark solutions. This paper discusses an interesting aspect of the recent best-performing works in this field: the fact that their main contribution comes from novel data augmentation techniques, rather than model improvements. Moreover, experiments are set to highlight the significance of data augmentation practices for limited data sets with narrow distributions. This paper intends to review the ingenious techniques to generate synthetic data to augment training and display evidence of their potential importance.**

***Index Terms*—augmentation, leaf counting, leaf segmentation, synthetic data**

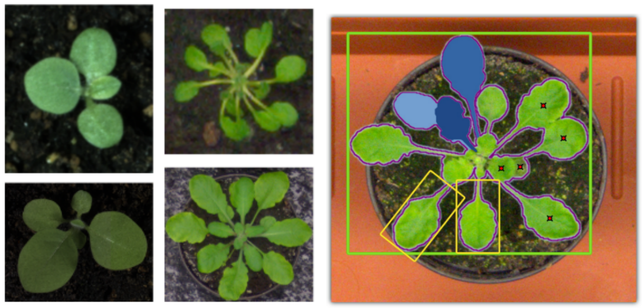
## I. Introduction

The field of plant phenotyping studies plants characteristics that resulted from their interaction with the environment [1]. The analysis of phenotypic traits can play a role in the advancement of plant science and the aspects of breeding and crop management. However, making the measurements necessary to perform a thorough analysis of the plant traits can be demanding and invasive. Such measurements were traditionally manually made, which results in low throughput and limits a comprehensive study of the plants' characteristics [2]. This inability is represented by term coined as phenotyping bottleneck [3] used to translate the factors that limit understanding and slow the field progress. Since image acquisition has become more accessible and processing power has experienced tremendous growth, a new bottleneck given by the lack of algorithms to analyse all the plant data effectively has been formed [4].

Since computer vision (CV) represents one of the most accessible and less invasive approaches for plant phenotyping, it has drastically increased in popularity in the field [5]. In the past five years, in particular, the application of deep learning in this field has become ubiquitous as in many others sub-field of computer vision. Tasks such as disease detection and plant-part segmentation, which was previously done by heuristics or hand-engineered feature extraction and symbolic machine learning, are now being mainly done in an end-to-end fashion with deep neural nets [6]. Nevertheless, although much of the current computer vision implementations show impressive results, the application of deep learning for plant phenotyping is still limited by the lack of large, public, labelled data sets and community agreed benchmarks. Such a scenario makes it difficult for comparing solutions from proposing methods, either for working with different data sets or evaluating performance with different metrics.

Mobilised by the deficiency of data and benchmarks, researchers decided to organise and distribute a well-annotated data set of Arabidopsis and tobacco plant images, due to their prevalent use in the field [7]. When releasing the data set, the authors not only gave the annotated segmentation masks of the plants and individual leaves but also benchmark problems for proposing works using their data. The set of problems includes plant detection and localisation, plant segmentation, leaf detection, segmentation, counting, tracking, and boundary estimation. The authors later organised the 'Leaf Segmentation Challenge' (LSC) at the Computer Vision Problems in Plant Phenotyping (CVPPP 2014) workshop, which resulted in many exciting solutions for the task of multi-instance segmentation of leaves [8]. The main benchmark metrics that the authors proposed, and now commonly used in works using such data, is the Symmetric Best Dice (SBD) and Difference in Count (DiC). The former is a mask-to-ground truth metric, which shares similarities with the intersection over union metric. The latter is an error measure in the leaf counting given by the difference of the number of predicted and real leaf instances. Some examples of the images of the CVPPP data set are illustrated in Fig. 1.

With an objective benchmark and data set released, there are growing research interest interest in the task of leaf segmentation and counting. Some suggested novel approaches,



(a) Examples of images of the CVPPP dataset.

(b) Examples of possible labels.

Fig. 1: Examples of plant images in the CVPPPP dataset. Adapted from [7].

like the use of fully convolutional networks for plant segmentation and recurrent networks for leaf counting [9]. More intricate pipelines in a similar approach soon followed through different authors [10]. Others used CNNs as features extractors while performing counting by regression in later fine-tuned layers [11]. However, mostly all the works reporting to recently surpass benchmarks on the LSC and LCC (leaf counting challenge) did it through novel ways to augment the CVPPP data. The implementation mostly comes from generating synthetic data that can be added to training, as an attempt to increase the model's ability to generalise. There are, nevertheless, many ways to generate such data, which can significantly vary in complexity. An example of a more straightforward approach is cutting instances of the leaves and pasting them into similar backgrounds like the ones in the training data [12]. More complicated methods contemplate intricate pipelines for plant 3D modelling with the subsequent rendering of 2D images of plants in the same view as the training data [13]. Although higher complexity does not equate to higher performance, the latter example [13] tops the current leaf segmentation benchmark in the CVPPP data set regarding works that propose a novel data augmentation as their primary contribution.

This paper has the goal of presenting and discussing these innovative and ingenious strategies for data augmentation while also showing some evidence of their significance in plant phenotyping. The methods discussed here were all presented in the past 3-4 years. Their main merits were thoroughly discussed in three modelling classifications: *cut and paste*, *graphical modelling*, and *generative networks*. To highlight the importance of data augmentation practices, the experiments were set to include training and evaluating pre-trained, fine-tuned models on specific CVPPP data set splits with different augmentation strategies. The relative performance can be used to discuss some data characteristics and exemplify overfitting and the importance of regularisation when the data are limited. Although it is certainly desirable to compare all the strategies discussed here, very few works follow the practice of making their data available. The one synthetic data set from previous authors employed in the experiments is, nevertheless, a strong candidate to translate such concepts, as it resulted in a top-ranking performance model. Therefore, the contribution of this paper is two-fold: to inform the reader on novel data augmentation practices proposed in recent years and to provide evidence of their potential importance in plant phenotyping tasks of limited ground truth data. The authors hope that such discussion will inform readers working on similar problems and highlight potential gaps deserving of further research.

## II. Augmentation techniques

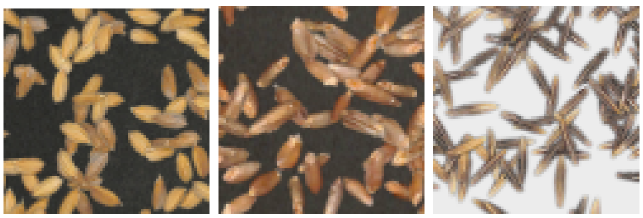
### A. Cut and paste methods

The technique presented by [14], called *cut, paste and learn*, is a simple but yet effective data augmentation method. As the name suggests, its application relies on automatically cutting instances of objects and pasting them into random backgrounds as a method to synthesise data. These images are then used to augment the training data and improve performance. The main advantage is that it is a rapid and automatic way to generate data for the tasks of instance detection and segmentation as the generator knows the position and mask of the created object. The authors in [14] showed that a model trained on a combination of real and synthetic data from the GMU Kitchen Dataset [15] resulted in a performance gain of approximately 3% in mAP values. Perhaps even more impressive, the technique showed significant results in a domain adaptation approach where the GMU Kitchen Dataset was used for training, and the Active Vision Dataset [16] was used for testing. By combining the synthesised dataset with just 10% of the real data, the model performed better than when using all the real but no synthetic data.

The idea of 'cut and paste' was replicated and did generalise for leaf segmentation tasks in plant phenotyping. A recent work [12] applied it for increasing performance on the task of segmenting and counting rosette plants. The application of the technique consists in the segmentation of non-occluded leaves to create synthetic data from two different datasets: (i) mature avocado and banana plantlets (80 images), and (ii) the CVPPP dataset. Examples of data generated from both of these data sets are illustrated in Fig. 2. The authors of the paper established two main methods for generating synthetic data: *naive and structured collage*. The naive approach was used in the avocado data set, and it comprised random collages of 10 to 40 segmented leaves in backgrounds similar to their real environment. To surpass benchmarks on the LSC, however, a more intricate approach was needed. Such a method comprised heuristics to mimic the leaves positions on the CVPPP data set, which was needed due to the images' particular characteristics: shot from the top, similar backgrounds (plant pots), and leaves emerging from the centre. The heuristics are composed of parameters that generate plant images with a different number of leaves, rotation angle, and size. The authors allegedly surpassed the benchmark at the LSC by using a pre-trained version of Mask RCNN [17] fine-tuned to their augmented data set of synthetic images.

(a) Naive collage. (b) Structured collage.

Fig. 2: Examples of the images in the two data sets generated by cut and paste methods in [12] and their respective methods.



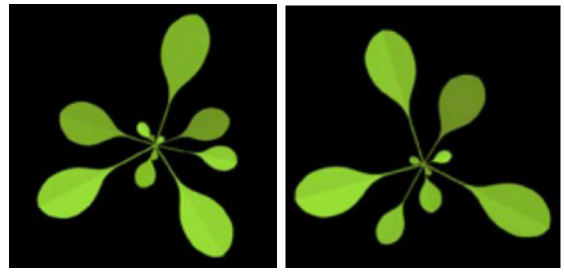
(a) Rice. (b) Wheat. (c) Oat.

Fig. 3: Examples of the synthetic images of seeds from different species generated by the method presented in [18].

The replication of this technique in agricultural phenotyping has also been attested in crop seed segmentation. Presented in [18], the cut and paste technique showed to be useful and to generalise to many types of seeds in segmentation tasks. The synthetic dataset was created by randomly rotating and pasting seed instances into background extracted from the real images. The methodology, initially set up for barley seeds, also performed well when applied to rice, wheat, oat, and lettuce seeds. Fig. 3 shows examples of the technique when applied to these three different crops. Although the performance comparison between training on real and synthetic was not investigated, the authors showed that models trained only on synthetic data resulted in AP50 values of 0.95 when averaged over many data sets. The high-throughput automatic analysis of seeds is crucial since it has been shown that their shape and sizes are essential predictors of quality and yield of crops [19].

### *B. Graphical modelling*

The idea of graphically modelling plants is not recent, but using it to augment data in computer vision tasks has been recently explored due to the current developments in deep learning and computing power. Original mathematical models, known as Lindenmayer systems (L-systems), date back to the 60s [20] where the main focus was to represent plant topology. Graphical rendering of such models came much later [21], as well as their much recent application to augment data [22]. Arguably, one of the main advantages of such an augmentation approach is that many different phenotypes can be modelled and simulated. That ability could increase its potential for generalising, especially if the models are able to represent the distribution of real plants more precisely.



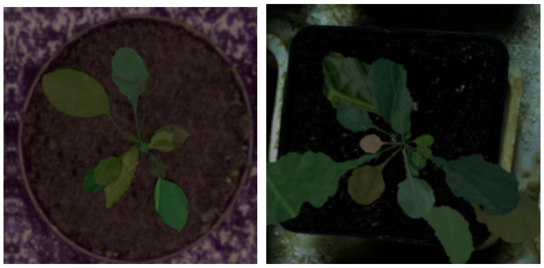
(a) (b)

Fig. 4: Example of generated images by the L-systems-based technique presented in [22].

The authors of a recent paper [22] went as far as stating that the real and synthetic data could be interchangeably used to train deep learning models in the task of leaf counting. Their method modelled rosettes of Arabidopsis using an L-systems-based plant simulator software fitted with probabilistic curves from different phenotypic traits. As seen in the examples illustrated in Fig. 4, this implementation did not generate images with a defined background. The presented evidence of generalisation came from leveraging the fact that the data set had two splits of Arabidopsis (CVPPP) from different years, representing two distinct distributions. The absolute leaf count difference when one of them is used for training, while the other is used for testing, is reduced if the synthetic data are considered in training. The authors also showed that a model trained with synthetic images-only did generalise to a reasonable level when testing on real images.

The authors in [23] also showed that the idea of graphically modelling plants could be effective at augmenting the training of models for the task of leaf segmentation. They allegedly surpassed the benchmark, reaching an SBD score of 90% on the A1 CVPPP data set. In their work, the method used to generate plant images follows a leaf-by-leaf approach. The modelled leaves were arranged circularly, arising from the centre of the pot, and had dimensions scaled independently along each axis. Each leaf came from applying random deformations and textures to an inspiration 3D leaf model. The position and rotation parameters were sampled from a uniform distribution for each leaf instance. The images were rendered from the top angle, as in the CVPPP images. In achieving their best result, 10,000 synthetic images were used to augment training, which was performed by fine-tuning a Mask RCNN model pre-trained on the COCO data set [24]. The authors reported performance improvements of up to 20% in one of the dataset splits, but there was no improvement in one of the rest five data set splits. Such split is composed of young tobacco plants, differing from the Arabidopsis plants that the method tried to model.

On a more recent work [13], the same authors raised the bar for graphical modelling with a bold aim: bridging the species gap in plant phenotyping. As plant species greatly



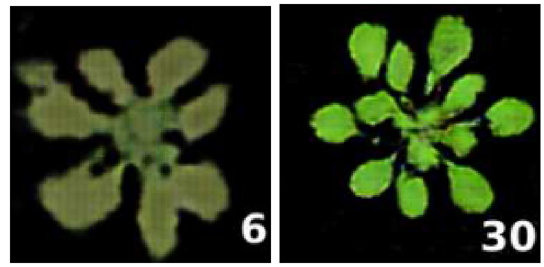
(a) (b)

Fig. 5: Examples of the images generated by the pipeline for 3D leaf modelling presented in [13].

vary in phenotypic characteristics, and models from trained on one species do not necessarily generalise for others, closing such a gap would represent a significant contribution. By proposing a novel and detailed plant generation pipeline, the authors attempted to close this generalisation gap in the task of leaf segmentation. Again on a leaf-by-leaf basis, the pipeline involved steps of leaf and texture generation, background processing (extracted from CVPPP images), and overall plant assembly. One can see the result of the pipeline implementation by the examples showed in Fig. 5. By comparing the results between training with synthetic and real data, interesting improvements in performance were shown. The authors claim to have surpassed state-of-the-art methods, which were claimed by previously mentioned works here [12], [23], whose also proposed novel data augmentation techniques. The symmetric best dice improvement of 31% compared to their previous method [23] on the A3 data set shows that some improvements across species were indeed achieved. This improvement is noticeable because the A3 data set is under-represented in the CVPPP data, with just a few tens of images. It is made of Tobacco rosettes rather than Arabidopsis, which represent the great majority of images of the data set. For further validation, the authors also tested their method trained on the CVPPP on another data set. Such an external data set is composed of capsicum and Komatsuna [25] images. The results showed an average of 51% performance increase of symmetric best dice when the synthetic images were used to augment training data. It is worth mentioning that the authors did make their synthetic data set of images public and easily accessible; their data are used in the experiments presented in the next section.

### *C. Generative networks*

Another approach to generating data for augmenting training is to use Generative Adversarial Networks (GANs) to synthesise plant images. Proposed in [26], such networks are able to learn a latent space and use it to generate new data representative of a given data set distribution. The framework is composed of two models: a generator and a discriminator. The latter is trained to distinguish real from synthetic data, while the former to maximise the probability of the latter making a mistake. When first proposed, the GANs framework resulted in



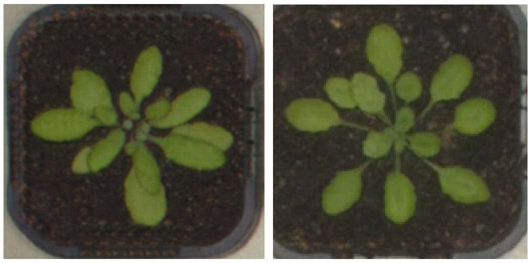

(a) A 6-leaf Arabidopsis. (b) A 30-leaf Arabidopsis.

Fig. 6: Examples of Arabidopsis images generated by the ARIGAN method proposed in [28] with the correspondent number of leaves.

realistic natural scenes and faces images, its recent variations are still considered state of the art in image generation [27].

Adopting the popularity of the CVPPP data set, the authors in [28] applied GANs to generate Arabidopsis images for data augmentation. They were inspired by [29], which proposed a method for unsupervised representation learning with GANs. The authors coined their Arabidopsis image generation method as ARIGAN [28]. An interesting aspect of this implementation is the use of a GAN variation called Conditional GAN (cGAN). Differently from conventional GANs, which generates images with a random noise seed, cGANs introduce a conditional vector that allows for a level of control when generating images. In the case of generating the images of plants, the authors used such condition to stipulate its number of leaves. Despite resulting in an alleged increase in performance, this approach appears not to capture the high-frequency features and texture details of the leaves in the data set, as shown in examples in Fig. 6. The improvements in performance were presented for the task of leaf counting, and leaf segmentation was not assessed. The presented results showed a decrease in absolute difference counting error of 5.4% and 14.4% reduction in mean squared error.

The authors in [30] aimed at pushing the previous generative approach [28] further by generating more realistic plant images with higher resolutions. The method described in the paper also uses cGANs to control an aspect of the generative process; in this case, leaf segmentation masks. The process of generating the leaf segmentation masks starts with extracting instances of mask leaves from the A4 CVPPP data set and sorting them by size. Then, an algorithm with heuristic rules assembles the masks to create a range of image generation seeds. The masks instances beginnings are centred on the image, and their size was chosen by the number of leaves (input). Each new mask is set up with a rotation of 140-200 degrees and then fed to the generator. As compared to the previous method, the images generated by this approach seems more realistic with apparent greater capabilities of capturing leaf textures, as illustrated in Fig. 7. The images were used to augment the training of a pre-trained MaskRCNN model. The authors reported an average leaf counting error reduction of 16.67% when augmentation was used. For the segmentation,



(a) (b)

Fig. 7: Examples of Arabidopsis images generated by the GAN-based approach presented in [30].

however, improvements in the best dice metric did not achieve 1%.

## III. Experiments

To contribute evidence to the effectiveness of data augmentation techniques, the authors would like to present some revealing experiments made on the CVPPP data set. The goal of such experiments is to replicate and attest the importance of data augmentation and synthetic data in training, and their potential to aid generalisation and regularisation. The experiments were composed of training the same model on three different scenarios: no augmentation, standard augmentation, and augmentation with the synthetic data from [13]. Regarding the data augmentation techniques used in the second scenario, only two simple operations were applied: rotation and flipping. Random rotation was applied online (while training) with a 50% chance, in a range of -45 to 45 degrees. Flipping was also applied with a rate of 50%; no test-time augmentation was used in evaluation.

The comparison of the three scenarios was made contrasting the difference in leaf segmentation and counting metrics from such models when evaluated in two data sets: the development set (same distribution as training) and the test set (slightly different distribution). For both scenarios, the training data is an 80% split of the A4 data set from the CVPPP. The development set in the first scenario is the remaining 20% of the images in the A4 split. The choice of using the A4 as the training and development data sets is due to the fact that it is the split with the highest number of images (624), followed by the A1, with 128 images. The test set is the A1 split of the CVPPP, which has a slightly different distribution and is thus used as a relative measure of generalisation.

It worth noting that the A1 data set is similar to the A4 in many ways: they both contemplate images of Arabidopsis plants taken from the top view on a controlled environment. Therefore, changes in performance from the development (A4 split) to the test set (A1), if drastic, can indicate that the model's ability to generalise could be severely compromised. For the fact that they are similar in many ways, one should expect that a model that evaluates well on the development set (A4), would also present decent performance on the other test set (A1).

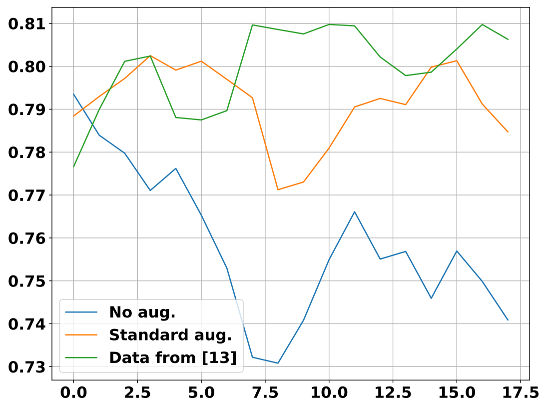
For the model architecture, a MaskRCNN approach with a ResNet backbone was used. The model was loaded with pre-trained weights training on the COCO data set. The weights were made accessible by the object detection framework by *Facebook AI Research* (FAIR) group named *Detectron II*. Their framework was also used to train and evaluate the models in the scenarios considered. The models in all three scenarios were trained for 100,000 iterations, on a learning rate of 1e-4, and evaluated every 5,000 iterations. Only a single image was used at each batch to avoid running out of memory. A decision for reporting the models' performance with the metrics adopted by the leaf segmentation and counting challenges (SBD and DiC) were made since they are the benchmarks used in this field. When presenting the evaluation results, the curves were slightly smoothed with a running average of three samples to make the learning trend clearer.

## IV. Results

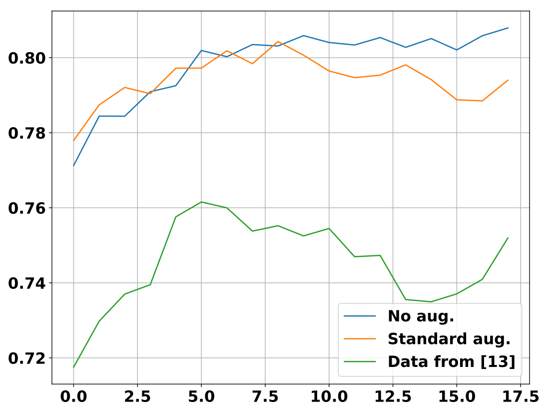
The SBD and leaf count metrics, resulted from training the models on the three scenarios are illustrated in Fig. 8 and 9, respectively. In Fig. 8, it is noticeable that, for the evaluations made on the development set (same distribution as training), augmentations scenarios under-perform. Such results show how augmentations have a regularising effect on models, effectively making it harder to fit the training distribution. Nevertheless, when the models are evaluated on the test set, it is clear that the scenarios contemplating augmentations perform better. Given that the test set (A1) is only slightly different from the development set, such an effect is clear evidence of how data augmentation can prevent overfitting. Small changes in the test data distribution greatly affect the ability of the method to generalise. The model that did not have data augmentation fits the training data easily and fast, with the first evaluation point being its maximum score. From this point onward, its performance decreases, showing that it is overfitting the training data while becoming worse at generalising to out-of-sample observations. The same contrasting effect can be observed in Fig. 9, which illustrates the leaf count evaluations; in this metric, the smallest the value the better.

For all the discussed scenarios, better results came from the model where training was augmented with synthetic data from the previous mention work [13]. It over-performs only using flipping and rotation, but not by much. With a single percentage gain in performance when compared to the scenario with standard augmentation, the results presented here show that these techniques are close to being interchangeable. Nevertheless, it is possible that a higher difference in performance would appear in case the models were evaluated on images from other plant species.

Table I summarises the best results from the three scenarios considered regarding the metrics of leaf segmentation and counting. It is interesting that despite the parameters or architecture of the model, one can infer interesting insights from the comparison of scenarios with data augmentation present or not. The model without augmentation converges much faster and outperforms by overfitting the A4 data set.




(a) Test set evaluations.



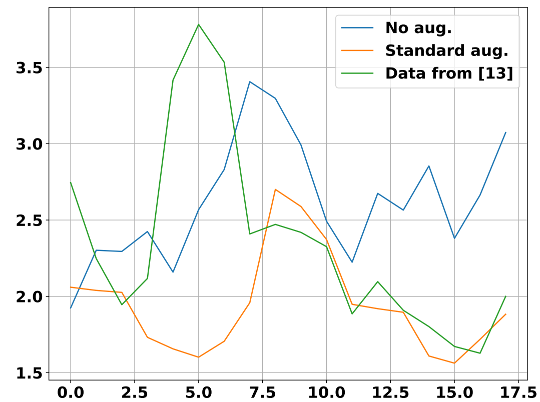

(b) Development set (same distribution) evaluations.

Fig. 8: Symmetric best dice (SBD) metric from models trained on the three scenarios with a) test set (A1) and b) A4 (20% split).

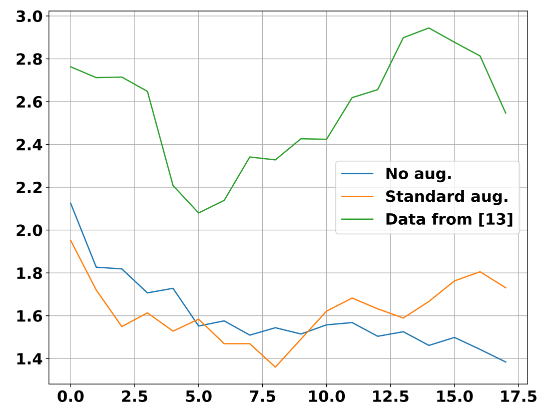
Meanwhile, the model with augmentation is regularised and performs drastically better on the test set, which has a slightly different distribution. Although one cannot make qualitative statements about the model or data augmentation practices with such experiments, they do reflect some characteristics of the data. The fact that the model quickly overfitted the A4 data, despite being the largest split, shows that it has a narrow distribution and probably should not be used to derive statements regarding generalisation, unless proper regularisation techniques are adopted.

## V. Discussions

The methods previously reviewed here refer to a class of data augmentation techniques called domain adaptation. Methods under such class augment their training data with examples that were not necessarily extracted from the same distribution as the test set, but that resembles it somehow.




(a) Test set evaluations.




(b) Development set (same distribution) evaluations.

Fig. 9: Absolute difference in leaf count from models trained on the three scenarios with a) test set (A1) and b) A4 (20% split).

TABLE I: Performance comparison between the best-performing models on the test set (A1) for each scenario.

| | SBD | $\lvert$ DiC $\rvert$ |
|---|---|---|
| No Aug. | 0.7982 | 1.7344 |
| Aug. | 0.8109 | 1.4687 |
| Data from [13] | 0.8201 | 1.3125 |

The closer the resemblance, the higher are the chances of the method generalising to the distribution of the test set, which is attested when evaluating the method. To that end, it is noteworthy that even simple techniques such as cut and paste can be effective at increasing a method's performance. It is arguable that these ideas for data generation are a way of transferring human knowledge (the idea for domain

TABLE II: Performance comparison between the works that proposed novel data augmentation techniques and performed evaluation on the CVPPP data set by any of the metrics.

| | | **SBD** | | | | | **DiC** | **\| DiC \|** |
|---|---|---|---|---|---|---|---|---|
| Ref. | Method | A1 | A2 | A3 | A4 | A5 | | |
| [12] | CAP | 0.88 | 0.84 | **0.80** | 0.87 | **0.85** | | |
| [23] | GM | **0.87-0.9** | 0.71-0.81 | 0.59-0.51 | **0.73-0.88** | 0.70-0.82 | | |
| [13] | GM | 0.81-0.89 | **0.81-0.88** | 0.84-0.86 | **0.86-0.88** | | | |
| [28] | GN | | | | | | 0.15-0.19 | 0.94-0.89 |
| [30] | GN | | | | **0.87-0.88** | | **(-0.22)-0.12** | **0.87-0.72** |

adaptation) to the deep learning model, reducing empirical risk without collecting more data. Table II is an ensemble of performance results reported by the methods discussed here that evaluated their model on the data sets A1-A5 with the Symmetric Best Dice (SBD) or Difference in Count (DiC) metrics. The table is sparse due to the fact that works usually focus on only one of either leaf segmentation or counting tasks. For the works that focus on segmentation, not all evaluate their methods on all the data sets splits. Where two numbers are presented in Table II, it depicts the changes in performance from only using real data to using synthetic and real data combined; bold letters highlight the maximum performance in each metric.

It is worth noting that the discussed methods are evaluated with a testing set with a narrow distribution, which can have significant consequences for generalisation. For the images in the CVPPP data set, for example, they are all taken from the top, with mostly the same plant species, having very similar backgrounds. There is probably much to be argued about the capability of generalisation of models using similar data, which future works will have to address. This assertion is not the take away from the efforts of the works cited here but to highlight possible future paths in the field of plant phenotyping. The results of the experiments performed on the CVPPP data set on its A1 and A4 splits is presented as evidence of such claims. The metrics on leaf segmentation and counting showed that it is not hard to overfit on data sets with a narrow distribution. Such an outcome could result in potential failure if methods that were trained without the proper regularisation care were used for inference in out-of-sample data. The results highlight the value of data augmentation in tasks of plant phenotyping, which often suffer from limited data sets with a narrow distribution.

With the number of works evaluating their method on the CVPPP data set in the past 3-4 years, it is arguable that it has become a standard for the tasks of multi-instance segmentation and counting of leaves. While using it, methods can objectively compare their performance on these tasks with has agreed upon metrics in a fixed domain. The leaf segmentation challenge (LSC) can now be found in CodaLab [31] where anyone can evaluate their performance against other benchmarks. Nevertheless, it should be reiterated that only two of the works presented here made their data available and public [13], [28], which is detrimental to the progress of the field. It is arguable that, for works having novel data augmentation techniques as their primary contribution, making the resulting data available it is as important for the field as having established benchmarks data sets and metrics that make methods comparison possible.

## VI. Conclusions

Novel data augmentation strategies proposed in recent years for the use of deep learning algorithms in plant phenotyping have been reviewed. The methods were divided into three classes: cut and paste methods, graphical modelling, and generative methods. A cut-and-paste approach is the simplest and comprises the extraction of instances of objects in the training data, followed by their ensemble in canvas with a similar background. Graphical modelling is probably the most complex, requiring intricate pipelines with possibly many rules, but it appears to result in the best performances in leaf segmentation. Methods based on generative networks leverage the intrinsic optimisation given by the training process of adversarial models to generate synthetic data, showing relevant improvements in the task of leaf counting. The many papers discussed represent pilling evidence that such ideas for data augmentation are effective to improve the models' performance in the test set. Nevertheless, more than increasing performance, experiments showed that data augmentation is significant to regularise the models trained on limited data sets. The methodology of the experiments, with development and test sets, showed that it is easy to overfit on plant image data sets with a narrow distribution, and that augmentation techniques are likely to help to generalise to out-of-sample images. Differently from traditional augmentation techniques, the works reviewed here are not basic copies of the training images with spatial or colour transformations applied, they are rather the result of the application of domain adaption by generating synthetic data. The increase in performance shows that these methods are a specialised way of circumventing limited amounts of training data present in problems of plant phenotyping. Such ideas could translate to other domains that suffer from the same problem and help increase the generalisation of future models.

## References


[1] R. Pieruschka, U. Schurr *et al.*, "Plant phenotyping: past, present, and future," *Plant Phenomics*, vol. 2019, p. 7507131, 2019.

[2] J. R. Ubbens and I. Stavness, "Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks," *Frontiers in plant science*, vol. 8, p. 1190, 2017.

[3] R. T. Furbank and M. Tester, "Phenomics–technologies to relieve the phenotyping bottleneck," *Trends in plant science*, vol. 16, no. 12, pp. 635–644, 2011.

[4] M. Minervini, H. Scharr, and S. A. Tsaftaris, "Image analysis: the new bottleneck in plant phenotyping [applications corner]," *IEEE signal processing magazine*, vol. 32, no. 4, pp. 126–131, 2015.

[5] A. L. Chandra, S. V. Desai, W. Guo, and V. N. Balasubramanian, "Computer vision with deep learning for plant phenotyping in agriculture: A survey," *arXiv preprint arXiv:2006.11391*, 2020.

[6] S. A. Tsaftaris and H. Scharr, "Sharing the right data right: A symbiosis with machine learning," *Trends in plant science*, vol. 24, no. 2, pp. 99–102, 2019.

[7] M. Minervini, A. Fischbach, H. Scharr, and S. A. Tsaftaris, "Finely-grained annotated datasets for image-based plant phenotyping," *Pattern recognition letters*, vol. 81, pp. 80–89, 2016.

[8] H. Scharr, M. Minervini, A. P. French, C. Klukas, D. M. Kramer, X. Liu, I. Luengo, J.-M. Pape, G. Polder, D. Vukadinovic *et al.*, "Leaf segmentation in plant phenotyping: a collation study," *Machine vision and applications*, vol. 27, no. 4, pp. 585–606, 2016.

[9] B. Romera-Paredes and P. H. S. Torr, "Recurrent instance segmentation," in *European conference on computer vision*. Springer, 2016, pp. 312–329.

[10] M. Ren and R. S. Zemel, "End-to-end instance segmentation with recurrent attention," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6656–6664.

[11] S. Aich and I. Stavness, "Leaf counting with deep convolutional and deconvolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 2080–2089.

[12] D. Kuznichov, A. Zvirin, Y. Honen, and R. Kimmel, "Data augmentation for leaf segmentation and counting tasks in rosette plants," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[13] D. Ward and P. Moghadam, "Scalable learning for bridging the species gap in image-based plant phenotyping," *Computer Vision and Image Understanding*, p. 103009, 2020.

[14] D. Dwibedi, I. Misra, and M. Hebert, "Cut, paste and learn: Surprisingly easy synthesis for instance detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1301–1310.

[15] G. Georgakis, M. A. Reza, A. Mousavian, P.-H. Le, and J. Košecká, "Multiview rgb-d dataset for object instance detection," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 426–434.

[16] P. Ammirato, P. Poirson, E. Park, J. Košecká, and A. C. Berg, "A dataset for developing and benchmarking active vision," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 1378–1385.

[17] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[18] Y. Toda, F. Okura, J. Ito, S. Okada, T. Kinoshita, H. Tsuji, and D. Saisho, "Learning from synthetic dataset for crop seed instance segmentation," *BioRxiv*, p. 866921, 2019.

[19] A. Shomura, T. Izawa, K. Ebana, T. Ebitani, H. Kanegae, S. Konishi, and M. Yano, "Deletion in a gene associated with grain size increased yields during rice domestication," *Nature genetics*, vol. 40, no. 8, pp. 1023–1028, 2008.

[20] A. Lindenmayer, "Mathematical models for cellular interactions in development ii. simple and branching filaments with two-sided inputs," *Journal of theoretical biology*, vol. 18, no. 3, pp. 300–315, 1968.

[21] P. Prusinkiewicz and A. Lindenmayer, *The algorithmic beauty of plants*. Springer Science & Business Media, 2012.

[22] J. Ubbens, M. Cieslak, P. Prusinkiewicz, and I. Stavness, "The use of plant models in deep learning: an application to leaf counting in rosette plants," *Plant methods*, vol. 14, no. 1, p. 6, 2018.

[23] D. Ward, P. Moghadam, and N. Hudson, "Deep leaf segmentation using synthetic data," *arXiv preprint arXiv:1807.10931*, 2018.

[24] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft coco: Common objects in context," 2014.

[25] H. Uchiyama, S. Sakurai, M. Mishima, D. Arita, T. Okayasu, A. Shimada, and R.-i. Taniguchi, "An easy-to-setup 3d phenotyping platform for komatsuna dataset," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 2038–2045.

[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[27] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 4401–4410.

[28] M. Valerio Giuffrida, H. Scharr, and S. A. Tsaftaris, "Arigan: Synthetic arabidopsis plants using generative adversarial network," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 2064–2071.

[29] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[30] Y. Zhu, M. Aoun, M. Krijn, J. Vanschoren, and H. T. Campus, "Data augmentation using conditional generative adversarial networks for leaf counting in arabidopsis plants." in *BMVC*, 2018, p. 324.

[31] CodaLab. Leaf segmentation challenge (lsc). [Online]. Available: https://competitions.codalab.org/competitions/18405