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, counting, leaf, plant phenotyping, segmentation**

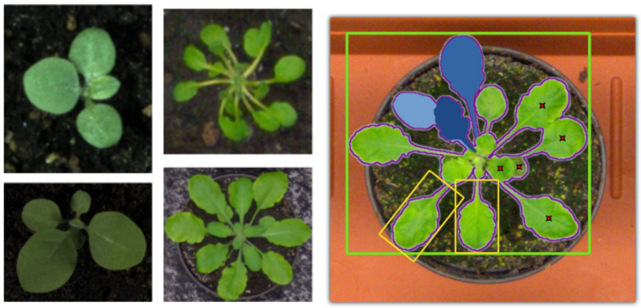
I. INTRODUCTION

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 in 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 the field 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 implementation 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 proposing methods to compare their solutions, either for working on different data sets or evaluating performance by different metrics.

Mobilise 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 realising the data set the authors not only gave the annotated segmentation masks of the plants and individual leaves but also metrics for benchmarking solutions while using their data. The tasks stipulated for model evaluation included 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]. Some examples of the images of the CVPPP data set are illustrated in Fig. 1.

With an objective benchmark and data set released, researchers started to take an interest in the task of leaf segmentation and counting. Some suggested novel approaches, 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]. Between recent works that have allegedly beaten



(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].

benchmarks when proposed, however, mainly all did it by implementing novel ways to augment the limited CVPPP data set. The implementation mostly comes from generating synthetic data that can be added to the training data in an attempt to increase the model's ability to generalise. For example, cutting instances of the leaves and pasting them into similar backgrounds like the ones present in the training data [12]. Others can be complex with intricate pipelines for plant 3D modelling with the subsequent rendering of 2D images of plants in the same view as the real data [13]. The latter, to be the best of the authors' knowledge, sets the current benchmark for the task of leaf segmentation in the CVPPP data set.

This paper has the goal of presenting and discussing these innovative and ingenious strategies for data augmentation while also proposing some evidence of their significance in plant phenotyping. The methods discussed here were all presented in the past 3-4 years. They were classified into three modelling classes: cut and paste, graphical modelling, and generative networks. The experiments, set to highlight the importance of data augmentation practices, include training and evaluating models on the CVPPP data set and their relative comparison when data augmentation is present or not. The relative performance can be used to discuss some data characteristics and exemplify overfitting and the importance of regularisation when the data are limited. Therefore, the contribution here is two-fold: to inform the reader on novel data augmentation practices proposed in recent years and to provide evidence of the potential importance of such practices in tasks of plant phenotyping, which usually contemplates limited data sets. The authors hope that such discussion will inform and inspire readers working on similar problems and highlight potential gaps deserving of further research.

II. AUGMENTATION TECHNIQUES

A. *Cut and paste methods*

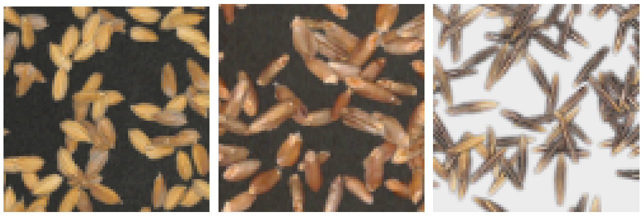
The technique presented by [14], called cut, paste and learn, is a simple but yet effective data augmentation method. As the name suggests, the application relies on automatically cutting instances of objects and pasting them into random backgrounds to synthesise data. These images are then used to augment the training data and improve performance. The main advantage is that it is a rapid and automatic way to collect new data for instance detection and segmentation tasks since the location or mask of the object are known by the generator. 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 and no synthetic data.

(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.

The idea of 'cut and paste' was replicated and did generalise for leaf segmentation tasks in plant phenotyping. The paper by [12] uses the method for segmentation and counting of 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 leaf segmentation challenge dataset. Both of these generation techniques are illustrated in Fig. 2. A naive approach is used in the avocado data set, comprising random collages of 10 to 40 segmented leaves in backgrounds similar to their real environment. To beat the benchmarks on the LSC, however, a more intricate approach named Structured collage was used. It is made of heuristics to mimic the leaves positions on the CVPPP data set and needed because it had particular characteristics: shot from the top, similar background (plant pot), and leaves emerging from the centre. The heuristics are composed of parameters that generate images with plants formed by leaves with different number, rotation angle, and size. The authors allegedly beat the benchmark at the LSC by using a pre-trained version of Mask RCNN [17] while fine-tuning on their augmented data set of synthetic images.

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.



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

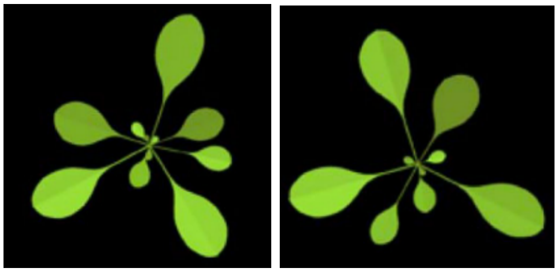
Fig. 3: Examples of the synthetic images of seeds from different species generated by the method presented in [18].

The 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 the technique applied to these three different crops. Although the performance gain between training on real and synthetic was not investigated, the authors showed that the 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 are not new, but their use to augment data for computer vision tasks has been recently explored due to the recent developments in deep learning and processing power. Original mathematical models, known as Lindenmayer systems (L-systems), date back to the 60s [20], where the 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]. It is arguable that one of the main advantages of such an augmentation approach is that many phenotypes can be modelled, increasing the chance of generalization if the models indeed represent the distribution of real plants.

In a recent paper [22], the authors went as far as saying that the real and synthetic data could be interchangeably used to train deep learning models in the task of leaf counting. The authors model rosettes of Arabidopsis using an L-systems-based plant simulator software fitted with probabilistic curves for different phenotypic traits. As possible to see in the examples illustrated in Fig. 4, this implementation did not generated images with a defined background. The proof of generalization came from leveraging the fact that the dataset had two splits of Arabidopsis (CVPPP) from different years, thus representing two distinct distributions. The loss in absolute count difference when one of them is used for training, while the other is used for testing, is reduced when using the synthetic data to augment training. The authors also showed that training in synthetic images-only did generalize to a reasonable level when testing on real images.

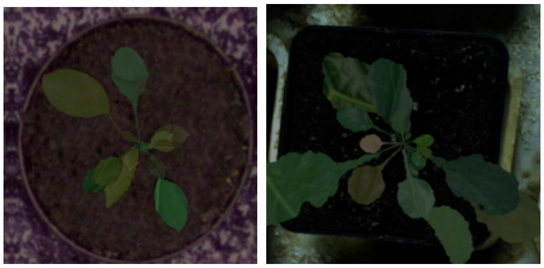


(a) (b)

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

The authors in [23] also showed that the idea of graphically modelling the plant could be effective at augmenting the data in the task of leaf segmentation. They allegedly beat the benchmark, reaching 90% in segmentation scores on the A1 CVPPP data set. The authors decided to model the plant on a leaf-by-leaf approach. As the data set is formed by rosette plant images taken from the top, the modelled leaves were arranged circularly, arising from the centre of the pot, with 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. At their best result, 10 000 synthetic images were used in the augmentation. When used to fine-tune a MarkRCNN model pre-trained on the COCO dataset, performance improvements of up to 20% were seen in one of the dataset splits. Performance increases were not seen on one of the five data set splits, which 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 with graphical modelling for data augmentation with a bold aim: bridging the species gap in plant phenotyping. As plant species greatly vary in phenotypic characteristics, and models from trained on one species do not necessarily generalise for others, closing such a gap would indeed be much valuable. By proposing a novel and elaborate plant generation pipeline, the authors attempted to close this generalisation gap for the task of leaf segmentation. On a leaf-by-leaf basis, the pipeline involved steps of leaf and texture generation and processing, background processing from images of the CVPPP dataset, and overall plant assembly. One can see the result of the pipeline implementation by the examples showed in Fig. 5. The comparison between training with synthetic data or not showed interesting improvements in performance. The authors claim to have beaten the state-of-the-art methods, which was a title held by previously mentioned works here [12], [23] whose also proposed novel data augmentation techniques. The 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 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



(a) (b)

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

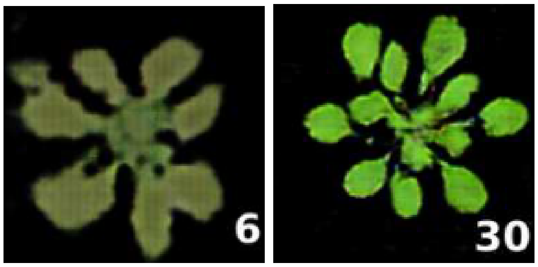
the data set. For further validation, the authors also tested the method trained on the CVPPP and other data sets. The latter was made of capsicum and Komatsuna [24] images. Such results showed an average of 51% performance increase of best dice when the synthetic images were used to augment the training data.

C. Generative networks

Another approach to generating data for augmentation is by using Generative Adversarial Networks (GANs) to synthesize plant images. Proposed in [25], such networks are able to learn a latent space and use it to generate new data representative of a 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 is trained to maximize the probability of the latter making a mistake. The GANs framework resulted in realistic natural scenes and faces images, and variations of it are still considered state of the art in image generation [26].

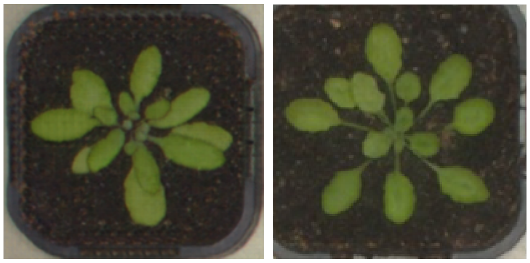
Adopting the popularity of the CVPPP data set, the authors in [27] applied GANs to generate Arabidopsis images for data augmentation. They were inspired by [28], which proposed a method for unsupervised representation learning with GANs. The authors coined their Arabidopsis image generation method as ARIGAN [27]. 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 certain controls when generating images. In the case of generating the images of plants, the authors used such condition to stipulate the number of leaves in the plant. Despite resulting in an alleged increase in performance, this approach appears to not capture the high-frequency features and texture details of the plants in the data set, as shown in examples in Fig. 6. The authors decided to test their augmentation technique in the task of leaf counting. For that, 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 [29] aimed at taking the generative approach of [27] further by generating more realistic plant images in 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 the size of masks 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, however, improvements in the best dice metric did not achieve 1%. It appears that the authors decided not to release their synthetic data set to the public.



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

Fig. 6: Examples of Arabidopsis images generated by the ARIGAN method proposed in [27] with the correspondent number of leaves.



(a) (b)

Fig. 7: Examples of Arabidopsis images generated by the GAN-based approach presented in [29].

III. EXPERIMENTS

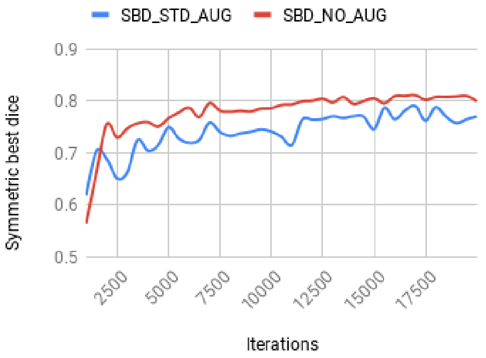
To contribute to the evidence of the effectiveness of data augmentation techniques, the authors would like to present some revealing experiments made on the CVPPP data set. The idea of such experiments was to attest the importance of augmentation in the regularisation and potential generalisation of models trained on this data.

The experiments were composed of two main comparisons of models trained with and without data augmentation. The comparison is on the difference in leaf segmentation and counting performance when evaluated on a development data set, and their difference on a test set with a slightly different distribution. For both models, the training data is an 80% split of the A4 data set from the CVPPP. The development set of the first experiment 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. For the second experiment, both methods (with and without augmentation) are evaluated on the A1 data set (different distribution) as a relative measure of generalisation. It must be noted here that, the A1 data set is similar to the A4 in plenty of ways: they both contemplate images of Arabidopsis plants taken from the top on a controlled environment. Therefore, any changes in this relative measure of evaluating on the A4 and A1 data sets, if drastic, can indicate that the model's ability to generalise could be 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).

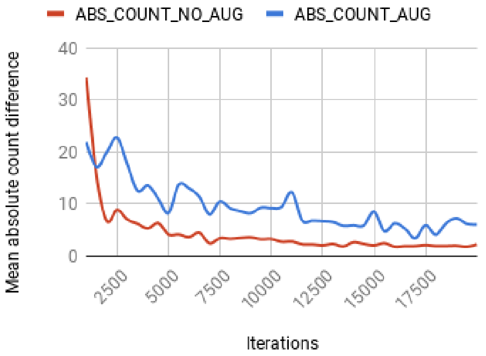
For the data augmentation techniques used on the model that contemplated it, only standard strategies were used. By standard one means the set of well-discussed image transformations present in many open-source frameworks: randomly flipping, rotating, cropping, changing contrast, brightness, and saturation of the images. The random rotation was applied in a range of -30 to 30 degrees, flipping had a rate of 50%, and contrast, brightness, and saturation were scaled in factors from 0.5 to 1.5. No test-time augmentation was used in evaluation; it only considered non-processed images. For the model architecture, a MaskRCNN approach with a ResNet backbone was used. The model was pretained on the COCO data set and made acessible by the object detection framework by *Facebook AI Research* (FAIR) group, which was used to train and evaluate the models compared here.

The results of the first experiment — training and testing on the 80-20 split of the A4 data set — are illustrated in Fig. 8. It is noticeable that the model with data augmentation suffered from a hard regularising effect, taking longer to converge. One could infer from the charts that data augmentation could be actually being detrimental to the model, reducing its performance in both tasks when compared to the model with no data augmentation.

However, such a picture drastically changes when the metrics of the second experiment, which involves evaluation on the different test set, are taking into consideration. Table I summarises the metrics of leaf segmentation and counting, showing the substantial difference in performance when a different data set is used in the evaluation. 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. Meanwhile, the model with augmentation is regularised and performs drastically better on the test set with a slightly different distribution. Although one cannot make qualitative statements about the model or practices of data augmentation 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 without any care for regularisation.



(a) Performance comparison of leaf segmentation with SBD metric.



(b) Performance comparison of leaf counting with the average of the absolute difference of predicted leafs as metric.

Fig. 8: Performance of models trained with (blue) and without (red) data augmentation on leaf segmentation and leaf counting metrics.

TABLE I: Performance comparison between the models with and without data augmentation on a different test set (A1).

	SBD	**| DiC |**
Aug.	0.74 (0.10)	2.39 (2.86)
No Aug.	0.55 (0.12)	11.50 (6.7)

IV. 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. 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 the way for transferring human knowledge, as the idea for the domain adaptation, to the deep learning model without the need to collect 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. Where two numbers are presented, 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.

Nevertheless, it is also 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 and narrow distribution resulted from controlled environments.

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 [30] where anyone can evaluate their performance against other benchmarks.

V. 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. A training and evaluation split on the same limited data set gave the impression that not having augmentation could increase performance, but only to be discredited by a different, yet similar, test set. 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.

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.	A1	A2	A3	A4	A5		
[12]	0.88	0.84	**0.80**	0.87	**0.85**		
[23]	**0.87-0.9**	0.71-0.81	0.59-0.51	**0.73-0.88**	0.70-0.82		
[13]	0.81-0.89	**0.81-0.88**	0.84-0.86	**0.86-0.88**			
[27]						0.15-0.19	0.94-0.89
[29]				**0.87-0.88**		**(-0.22)-0.12**	**0.87-0.72**

[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] 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.

[25] 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.

[26] 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.

[27] 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.

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

[29] 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.

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