Domain Targeted Synthetic Plant Style Transfer using Stable Diffusion, LoRA and ControlNet

Zane K. J. Hartley
University of Nottingham
Wollaton Rd
NG8 1BB
zane.hartley@nottingham.ac.uk

Rob J. Lind
Syngenta
Jealott's Hill International Research Centre
RG42 6EY
rob.lind@syngenta.com

Michael P. Pound
University of Nottingham
Wollaton Rd
NG8 1BB
michael.pound@nottingham.ac.uk

Andrew P. French
University of Nottingham
Wollaton Rd
NG8 1BB
andrew.p.french@nottingham.ac.uk

Abstract

Synthetic images can help alleviate much of the cost in the creation of training data for plant phenotyping-focused AI development. Synthetic-to-real style transfer is of particular interest to users of artificial data because of the domain shift problem created by training neural networks on images generated in a digital environment. In this paper we present a pipeline for synthetic plant creation and image-to-image style transfer, with a particular interest in synthetic to real domain adaptation targeting specific real datasets. Utilizing new advances in generative AI, we employ a combination of Stable diffusion, Low Ranked Adapters (LoRA) and ControlNets to produce an advanced system of style transfer. We focus our work on the core task of leaf instance segmentation, exploring both synthetic to real style transfer as well as inter-species style transfer and find that our pipeline makes numerous improvements over CycleGAN for style transfer, and the images we produce are comparable to real images when used as training data.

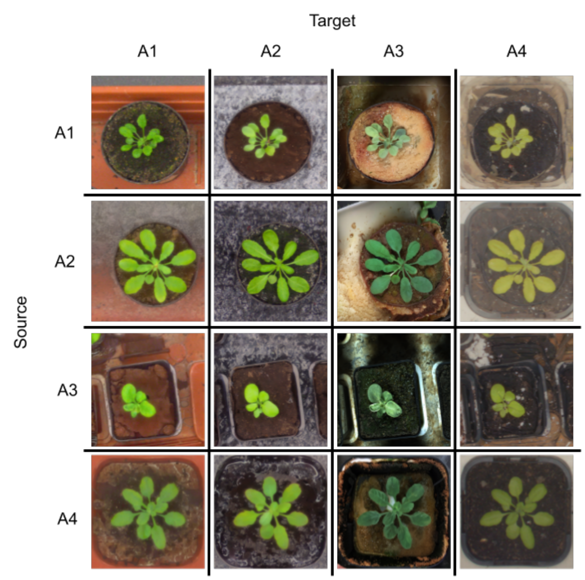


Figure 1. Our LoRA and ControlNet based style transfer pipeline demonstrates transferring between different rosette plant datasets, showing the ability to realistically transfer style whilst accurately maintaining both geometry and shape. Note the constrained geometry of the leaves despite the style differences in each row. Images across the diagonal show the original images from each dataset.

1. Introduction

The ability to apply today's most powerful deep learning models to a given problem is often constrained by the availability of high quality, domain specific, and annotated training data. In diverse areas such as digital plant phenotyping this is exacerbated by a wide range of different growth environments, plant species and growth stages. This leads to an extremely varied set of domains, which in turn can be difficult and expensive to capture in training data. Often, data augmentation can allow small datasets to perform as though a greater number of images was provided, however this is still inferior to increasing the number of unique samples. In this work we therefore focus on the creation and domain style transfer of images of handcrafted artificial plants, using 3D rendering to generate new images, and a diffusion based AI pipeline to improve the photo realism of the generated images using few shot LoRA models to target specific test sets.

Synthetic generation has been proposed as a way of creating low-cost training data for deep learning models for many years. More significantly, synthetic image generation pipelines such as presented here allow for the automated annotation of these images as well, further reducing time and effort cost, especially in the case of complex polygon-based labels required for instance segmentation. The challenge faced in using synthetic data is domain shift, where synthetic images are unable to train a model that generalises onto the real images they seek to emulate. In order to solve this we present a pipeline comprising multiple cutting-edge denoising models, enabling us to substantially increase the visual and feature similarity between our computer generated plants and real plants that we hope to digitally phenotype.

The ultimate goal is to produce training sets that lead to a model's accuracy which surpasses that of a model trained on real images alone. This could be achieved by making a blend of real and synthetic training images, or purely using synthetic images. The process aims to challenge a model during training to a wide diversity of images that ensure a robust and generalisable model which doesn't suffer from erratic behaviour and has consistent high accuracy.

We select leaf instance segmentation as an application on which to showcase our new pipeline. We chose this problem first because it is of interest to many researchers in digital plant phenotyping, with the segmentation of individual plant components being both an interesting problem and an important first step to many downstream analyses, such as plant health monitoring. This problem domain also benefits from high availability of datasets of rosette plants, including multiple of *Arabidopsis Thanalia*, which makes up the focus of our work.

Our motivation is the high cost of creating novel plant datasets for different phenotyping problems. Capturing real world images is problematic, as it is very difficult to capture the wide selection of morphological distributions and environmental factors needed to create a truly diverse dataset. By contrast, after its initial creation, a synthetic plant model can be easily reconfigured to create a much wider range of data, improving generalisation. Moreover, blending real and synthetic imagery into a training set has been shown to give a beneficial boost of accuracy compared to either component alone [6].

We are also motivated to take advantage of rapid improvements in generative AI technology. Models based on denoising diffusion have emerged in recent years to outperform previous GAN-based methods of AI image generation [13]. In this paper we therefore hope to leverage this new technology as part of a powerful, novel synthetic data pipeline.

In summary our contributions are as follows:

1. We describe our method for synthetic data generation using L-Systems within Blender to capture both images and instance segmentation masks.
2. We present our complete pipeline for domain style transfer using Stable Diffusion, ControlNet and LoRA models.
3. We compare our model both qualitatively and quantitatively against CycleGAN, a common GAN based style transfer model.
4. We conduct a series of experiments evaluating our data against real data for instance segmentation, demonstrating our methods comparable results to real data, particularly when limited real data is available.

Figure 2. Example images from each of the five real datasets we use as targets for our style transfer.

2. Related Work

2.1. Phenotyping with Synthetic Data

Researchers have shown an interest in using synthetic data to train digital phenotyping models for the past 10 years. Early attempts using 3D modelling include Ward et al [19] [18], Barth et al [3] and Napier et al [11] who focus on rosette plants, pepper plants, and wheat crops respectively. In these publications, 3D modelling tools such as Blender are used to create detailed 3D scenes and large datasets of synthetic images.

Other techniques for generating synthetic data have included GAN based images such as Giuffrida et al's [17] model ARIGAN which generates Arabidopsis from noise, or Toda et al [16] and Gao et al [5], who each create new images by compositing elements of existing real images together.

3D modelling is generally popular because it allows the greatest level of control over the synthetic plants, alongside automatic capture of annotations. However, images generated in this way often suffer most from the domain gap problem, where their artificial quality generalises poorly onto real images. Forms of domain adaptation such as feature space alignment have been used in to combat domain shift between synthetic and real images, such as Ayalew et al [2] in 2020, in which a gradient reversal layer enforces domain invariance in the CNN's feature extractor. Alternatively, style transfer approaches that attempt to copy the visual style of a target domain onto an image are a common method used to counter synthetic to real domain shift. Here, CycleGAN [22] has often been the first choice for unpaired style transfer since its publication. CycleGAN enables style transfer from a source to target domain and back again, ensuring image content stays consistent and has been used in a large number of publications [3] [7] [10]. In this paper we show that compared to modern approaches such as ours, CycleGAN is limited by its domain specificity, difficulty in training, and need for larger training datasets.

2.2. Diffusion Models

Diffusion models are a type of generative AI which are able to produce highly detailed images from noise using a iterative denoising process. Since their popularisation in 2020 by Ho et al [8], such models have become common for creating AI artwork, photo editing and most recently for generating synthetic data such as in Nguyen et al [12] in 2023 who generates images using classes from MSCOCO. The rapid advances in this technology have allowed high quality images to be produced, among them Latent Space Diffusion introduced by Rombach et al [13], incorporating a VAE and cross attention layers into the overall architecture.

2.3. Conditioning Diffusion Models

In order to be used for synthetic data generation, a major challenge to diffusion models is the requirement to produce accurate pixel label annotations that closely align with the generated images. Approaches that condition models with an input that also serves as an annotation have risen in popularity. ControlNet by Zhang et al [21] has quickly become a very popular way to condition diffusion models on a wide range of different annotation formats, and Anagnostopoulou et al [1] have demonstrated it as effective approach for conditioned style transfer of synthetic data of mushrooms.

2.4. Finetuning Diffusion Models

Since many popular diffusion models are trained across large and varied datasets, using them to produce images in

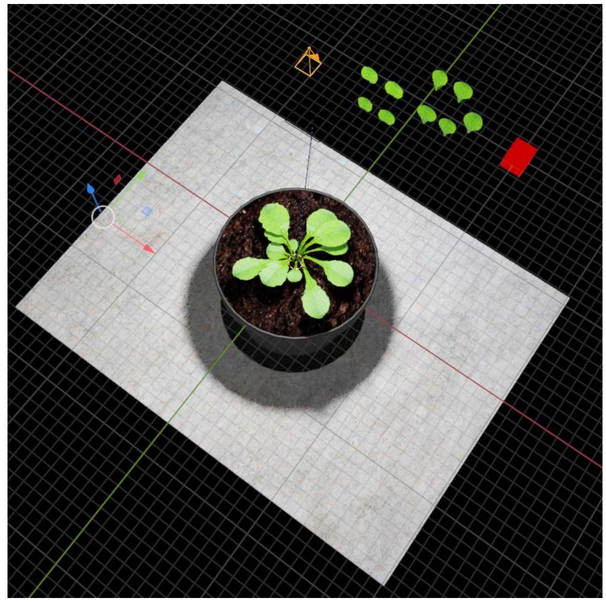


Figure 3. Synthetic plant being generated in Blender. To the top right you can see instances of different shaped leaves that are copied randomly into the new plant in addition to being rescaled.

the style of particular target domains is challenging. However, fine tuning large transformer models for specific domains and tasks can be costly and expensive, so instead there has been a focus on research that explores the best way to adapt or extend large base models at a lower cost. Dreambooth by Ruiz et al [15] allows the introduction of new subjects in text-to-image diffusion models using just a few examples by overriding a rare token in the models vocabulary with a new meaning. More recently low rank adapters [9] or *LoRA*s have been ported to diffusion models, after being first introduced by Hu et al in 2021 as a means of more efficiently fine tuning large language models. LoRA in particular is popular for allowing finetuning for specific tasks or domains without having to directly their hundreds of billions of parameters, making it computationally cheaper.

In this paper we propose combining LoRAs with ControlNets to create a domain transfer architecture that can be targeted at specific target domains. We combine this with 3D modelling to create a complete pipeline for creating highly realistic synthetic datasets using diffusion.

3. Materials and Methods

Here we describe our overall pipeline for generating our synthetic datasets. Starting with the creation of 3D modelled plants, we then apply a diffusion model to create realistic synthetic images. Our diffusion model is combined with multiple ControlNets, which enable pixel perfect alignment with our Synthetic annotations, and a LoRA that allows highly accurate style transfer to a learned target domain. Overall our pipeline allows us to create synthetic datasets that outperform previous methods, while having a much lower cost to create.

3.1. 3D Modelling Rosette Plants

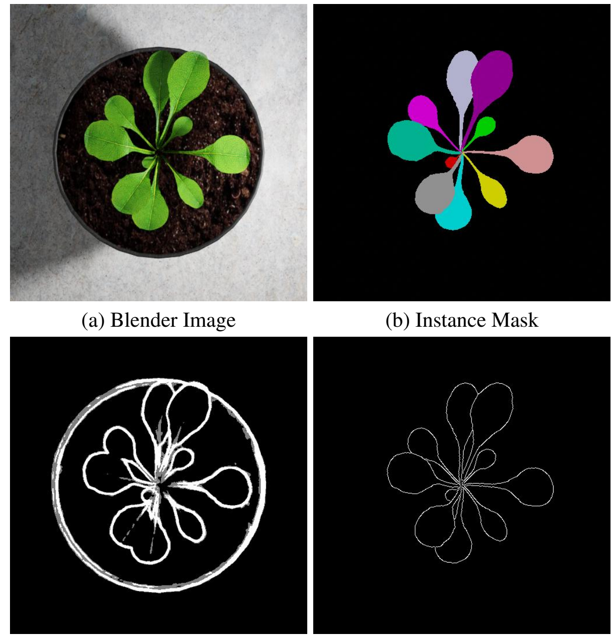
We create and render synthetic rosette plants using Blender [4], a popular 3D modelling tool for use in this work and shown in figure 3. We chose to design a generic rosette plants to maximise ability to generalise to different plants, and to fully leverage our style transfer pipeline. In order to create our digital plants we first created a number of 3D modelled *template* leaves using a simple mesh, and texture it with a photograph of an Arabidopsis leaf. We also model a basic pot, soil and background to make up the template for our scene.

In order to generate a large number of unique plants we use Blender's built in Python scripting tools to create a pipeline for automatic plant creation and image capture. This pipeline also utilizes L-systems [14], a method of modelling plants using iterative growth rules, which aid in making generated plants more natural. Creation of individual plants are achieved by duplicating our template leaves and rotating, scaling them and positioning them inside our pot based on the manner we would expect real plants to grow, for example leaves appearing in opposing pairs.

After each 3D plant is created we render an image from above using the CYCLES render engine, with random adjustments applied to scene lighting to increase variation. After rendering we then additionally capture an instance segmentation mask of the scene by re-texturing the leaves to unique RGB colours and disabling the rendering of background elements before capturing a second *label* render (shown in figure 4b. This process is repeated for the dataset, using random values for the number, size and rotation of individual leaves. As a result of this process we generate a dataset of 1400 synthetic rosette plant images alongside their associated label masks, creating 100 images each for leaf counts between 2 and 15. We then apply our Generative AI pipeline to manipulate these images to make them more realistic and appropriate to a particular phenotyping task.

3.2. Low Ranked Adapter

LoRA models allow the fine tuning of large diffusion models using only a small sample of images of a specific domain, subject or style. This is achieved by learning a set of weight changes, which, combined with the pretrained weights, make up the *finetuned model*. A low rank decomposition of the weight changes substantially reduces the number of trainable parameters, making LoRA models faster and computationally cheaper to train. Our pipeline uses a LoRA trained on unlabelled images from our target dataset, and can quickly learn to associate the style of the target domain with a trigger keyword.



(a) Blender Image (b) Instance Mask

(c) Soft Edge Mask of RGB Image (d) Canny Edge of Instance Mask

Figure 4. Example of the four synthetic images used in our style transfer pipeline.

For our experiments we select 10-30 images from each of our target datasets to use as training material for our style transfer LoRAs, examples of each are shown in figure 2. We use an implementation of LoRA training called *Kohya* to train a separate LoRA for each target dataset, using default settings initialised for Stable Diffusion v1.5, a popular implementation of Latent Diffusion Models. Each model is trained for 10 epochs using an NVIDIA Titan X GPU.

3.3. ControlNets

ControlNet models are extensions of a diffusion model that allows additional forms of conditioning to guide image synthesis. This is achieved by duplicating blocks from the encoder half of the models UNet that are conditioned by a new input and injecting their outputs into the models decoder. In order to ensure the generated images match the annotations from our synthetic data, we employ two ControlNet models to condition the diffusion model in different but complimentary ways.

ControlNet models support a wide range of different conditional inputs. We first use a Canny Edge detection ControlNet, that enforces the generated image to match edges given as input. Canny edge filters are a low cost technique that outputs hard edges detected in an image. As such we apply a Canny Edge filter to the segmentation mask of the source image (figure 4d), giving us a perfect outline of each individual leaf (excluding occluded edges). Using the

Target Dataset		$MaP_{50:95}$			
Name	Size	CycleGAN	LoRA-Synth	LoRA-All	Real
CVPPP A1	30	0.550	0.559	0.666	**0.701**
CVPPP A2	10	0.238	0.540	**0.656**	0.594
CVPPP A3	10	0.339	0.423	**0.669**	0.635
CVPPP A4	30	0.458	0.457	0.541	**0.633**
Komatsuna	30	0.081	0.594	0.635	**0.770**

Table 1. Quantitative Results of our experiments, highest scores for each dataset are shown in bold.

edges of the segmentation mask with a ControlNet enforces strict adherence to the geometry of the leaves and ensures that segmentation accuracy is preserved during style transfer. This process promotes a pixel perfect alignment between the new image and the synthetic labels.

We also utilize a SoftEdge Control Net to enforce a more relaxed geometry of the input image to apply during style transfer. Unlike Canny edge detection, SoftEdge is based on holistically-nested edge detection [20], a learned edge detector that captures edges while being less sensitive to noise and small shapes than Canny edge detectors. We apply a HED processor to each of our rgb synthetic images, capturing the outline of each plant as well as the pot and background elements 4c. By using the HED edges as input to the SoftEdge ControlNet we guide the diffusion model to generate high quality edges of the entire scene, even when only very small leaves are present.

3.4. Style Transfer Workflow

Our base model is Stable diffusion v1.5[13], selected as a high performing and general model that supports our target 512x512 resolution. This model is then adapted with our LoRA trained on the current dataset being targeted (see 4.2) and prompted with the trigger word for that dataset's style. For each input image we extract first a HED edge from the RGB image and a canny edge from the instance segmentation mask and apply these to a soft edge and canny Control Net each respectively. A specific advantage of using LoRA and Control Nets versus finetuning a specific checkpoint is the ability to apply them in this modular format enabling easy reuse at a lower cost of resources.

We pass images from our source dataset along with their annotations into our workflow as inputs. Our model then outputs the style transferred image, which combined with out input annotations make up our LoRA style transferred datasets.

4. Experimental Results

4.1. Target Datasets

We employ a variety of different *real* target datasets listed below, covering both examples of different datasets of the same plant species, and alternative rosette style plant species. An example of each dataset is shown in figure 2.

- **CVPPP A1, 2 and 4**. Arabidopsis thaliana plants from the CVPPP dataset, these datasets contain 128, 31, and 625 images respectively each grown in different environments and captured at different stages of growth.
- **CVPPP A3**. Tobacco Plants from the CVPPP dataset containing 27 images.
- **Komatsuna**. Komatsuna spinach plant dataset of which we are using a random sample of 300 images.

Note that Tobacco (A3 dataset) and Komatsuna plants are both plants that produce rosettes of leaves during early growth stages similar to Arabidopsis, but differ in their leaf shape, colour, and texture. By designing our synthetic data to be a generic rosette plant, we hope to be able to use it to obtain strong results across multiple species, demonstrating the efficiency of our approach.

With each dataset we create a training split of images to use as our target for domain style transfer, as well as validation during training. The larger datasets (A1, A4, Komatsuna) we use 30 images as our target and for smaller datasets (A2, A3) we take only 10 images to ensure a larger number of images are available for testing. All remaining images from each dataset are then used during our experiments as an unseen test set.

4.2. Experiments

For each target dataset we conduct a series of experiments to evaluate the effectiveness of different approaches to style transfer. Each experiment produces a training dataset from which we trained a Mask-RCNN model as a means of evaluating that particular approach to style transfer. We train each Mask-RCNN model for 100 epochs using a default hyper parameter configuration for instance segmentation using an NVIDIA A6000 GPU. and evaluated using the $MaP_{50:95}$ metric, where precision is averaged across a range of IoU values from 50 to 95.

For all experiments images were resized to 512x512 resolution for consistency and to limit computational expense on the higher resolution datasets.

- **CycleGAN.** Each target dataset of *real* real images was used to train a CycleGAN model for 100 epochs. This model was used to convert our dataset of 1400 synthetic images to the *real* which was in turn used to train Mask-RCNN.

- **LoRA-Synth.** Each training split is used to train a LoRA for stable diffusion that adapts the base model to a new domain. Our synthetic dataset is then style transferred using the LoRA and ControlNet pipeline described in 3.4. This creates a dataset of 1400 images which are used to train Mask-RCNN.
- **LoRA-All.** Here we use the same LoRA as the experiment above, however in addition to applying our pipeline to our synthetic data we also convert all real images from the other domains, by inputting these images and their annotations into our style transfer pipeline. Images of real-to-real style transfer are shown in Figure 1 and this combined dataset was then used to train Mask-RCNN for instance segmentation.
- **Real.** We train Mask-RCNN using our real data split. Note that we expect this experiment to perform very well, even with a small dataset due to the relatively simple nature of this problem. Achieving comparable or better scores than this with our other experiments would be a major metric of success for our work.

4.3. Qualitative Results

In Figure 5 we can see that CycleGan performs poorly when performing domain style transfer compared to our LoRA based approach. In the examples shown we see that the CycleGAN model consistently adds green leaf texture outside the bounds of the instance segmentation mask. Additionally we see that our LoRA based model creates much more defined edges between leaves that occlude each other compared to CycleGan where edges are less clear.

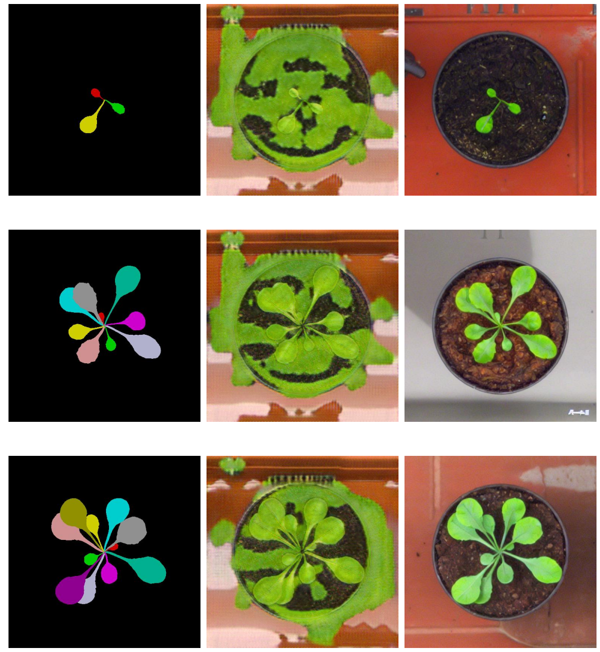
We also show results from our LoRA model when converting from Synthetic to each real domain in figure 6. Here we can see by comparison to a real example that the LoRA model learns to generate visually distinct styles for each target domain that closely resemble the real images (shown in the first column).

Finally, Figure 1 shows the effectiveness of our approach at translating different real images between their respective styles and species. Again we see that our approach is successful in transferring the aesthetic of different datasets to one another, including where the plant is of a different species, whilst maintaining the geometric structure of the leaves.

4.4. Quantitative Results

In table 1, we report MaP results for each of our experiments and datasets.

For our CycleGAN experiments, we achieved a wide spread of results with a worst MaP score of 0.081 for Komatsuna and a best score of 0.550 for CVPPP A1. We also see that among the CVPPP datasets, the results for A2 and A3, which importantly contain only 10 training images, were notably lower amongst the CycleGAN results.



(a) Instance Mask (b) CycleGAN Output (c) LoRA Output

Figure 5. Comparison of CycleGan and LoRA style transferred images from the same annotation, showing the higher quality domain adaptation achieved with LoRA and the better adherence to the annotation geometry.

LoRA-Synth produces a narrower range of results (0.423-0.559). The majority of LoRA-Synth results outperformed equivalent CycleGAN experiments, with set A4 being the only exception to this, which we note was still within 0.001 of the CycleGAN result. In all other experiments the LoRA dataset was substantially more effective, especially for the Komatsuna dataset.

LoRA-All showed the strongest performance of all of the non-real tests, scoring between 0.541 and 0.669; it outperforms CycleGAN and LoRA-Synth in every instance. For experiments A2 and A3, the LoRA-All model outperformed even the real data experiments, showing its robustness to even very limited data. For all other experiments LoRA-All performed well while trailing the real results by only a small amount each time.

5. Discussion

5.1. Analysis of results

Overall the results show that our LoRA model performs better style transfer than CycleGAN, with a downstream instance segmentation model scoring higher in most cases.

We also see that LoRA is more robust when used with small training datasets of real images, as compared to CycleGAN we do not see a performance drop for sets A2 and A3, when only 10 images are supplied. We observe that LoRA-all is even able to outperform *Real* for this dataset, suggesting that in some circumstances LoRA models could even be a suitable replacement for new, real-image datasets.

CycleGAN performs poorly on the Komatsuna dataset, while LoRA is still able to perform relatively well, scoring 0.594 (LoRA-Synth) compared to 0.081 (CycleGAN). We hypothesise this is because, of the datasets tested, Komatsuna is the most distinct in appearance from the Arabidopsis and Tobacco plants in the other datasets, causing CycleGAN to struggle to perform style transfer between such different objects. By contrast, our combination of LoRA and two ControlNet models appears more robust.

5.2. Advantages over CycleGAN

In addition to our LoRA model outperforming the CycleGAN model overall in our experimental results, we also favour this newer method due to its greater flexibility, robustness and better visual performance.

As shown in figure 5, CycleGAN-generated images are much more likely to hallucinate plausible leaves instead of those in the source image, and is much less consistent overall than the ControlNet conditioning. In addition to this, the experiments with the A2 and A3 datasets show that CycleGAN's ability to generate accurate style transfer almost completely collapses when provided with only 10 training images, whereas images generated from the LoRA approach appear visually acceptable even with so few training images.

A key advantage of our pipeline not being restricted to a single source domain is that it is much easier to incorporate new data into a training pipeline. CycleGAN is source-domain specific, and as such is more computationally expensive when combining data from multiple sources, as additional models will need to be trained for each source dataset. In contrast LoRA can be applied to images from any domain without retraining making it easy to apply to new datasets as demonstrated in our LoRA-All experiments.

5.3. Future Work

Having demonstrated the effectiveness of this pipeline on leaf segmentation, we hope to go further and explore more challenging problems in digital phenotyping, for example *Panoptic Segmentation of Crops and Weeds*, where we believe that synthetic data is even more important due to the increased cost of annotation, and we also theorise that without a similar pipeline, a larger amount of real images would be needed for equal performance due again to complexity.

In addition to further applications within plant phenotyping, we also believe that the use of diffusion model text prompts could also be used to curate an even more bespoke dataset. Examples of how this could be used could involve prompting the model to visualise plant stresses or other environmental elements. This control of impressive but general purpose diffusion models will become increasingly important if we hope to generate more scientifically accurate and plausible training data in the future.

6. Summary

In this paper we have presented a new pipeline for image to image style transfer, and applied it to digital plant phenotyping training datasets. We have demonstrated its effectiveness at both *synthetic-to-real*, *inter-species* and more conventional domain adaptation between datasets. Overall, our new pipeline has shown improvements in performance over CycleGAN style transfer, demonstrated through our experimental results. Furthermore we have reported improvements to robustness, shown by its consistency across varied datasets, and flexibility, shown by its ability to use a single model for style transfer from both synthetic and multiple real datasets, requiring only small amounts of training data. Overall we have shown that a style transfer approach such as ours would be preferable to a CycleGAN based approach for a wide range of more challenging phenotyping tasks, both where synthetic data is used, and where other real data can be leveraged in place of creating new annotations.

References

[1] Dafni Anagnostopoulou, George Retsinas, Niki Efthymiou, Panagiotis Filntisis, and Petros Maragos. A realistic synthetic mushroom scenes dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6281–6288, 2023. 3

[2] Tewodros W Ayalew, Jordan R Ubbens, and Ian Stavness. Unsupervised domain adaptation for plant organ counting. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*, pages 330–346. Springer, 2020. 3

[3] Ruud Barth, Jochen Hemming, and Eldert J Van Henten. Optimising realism of synthetic images using cycle generative adversarial networks for improved part segmentation. *Computers and Electronics in Agriculture*, 173:105378, 2020. 2, 3

[4] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. 4

[5] Junfeng Gao, Andrew P French, Michael P Pound, Yong He, Tony P Pridmore, and Jan G Pieters. Deep convolutional neural networks for image-based convolvulus sepium detection in sugar beet fields. *Plant methods*, 16:1–12, 2020. 3

[6] Zane KJ Hartley and Andrew P French. Domain adaptation of synthetic images for wheat head detection. *Plants*, 10(12): 2633, 2021. 2

[7] Zane KJ Hartley, Aaron S Jackson, Michael Pound, and Andrew P French. Ganana: Unsupervised domain adaptation for volumetric regression of fruit. *Plant Phenomics*, 2021. 3

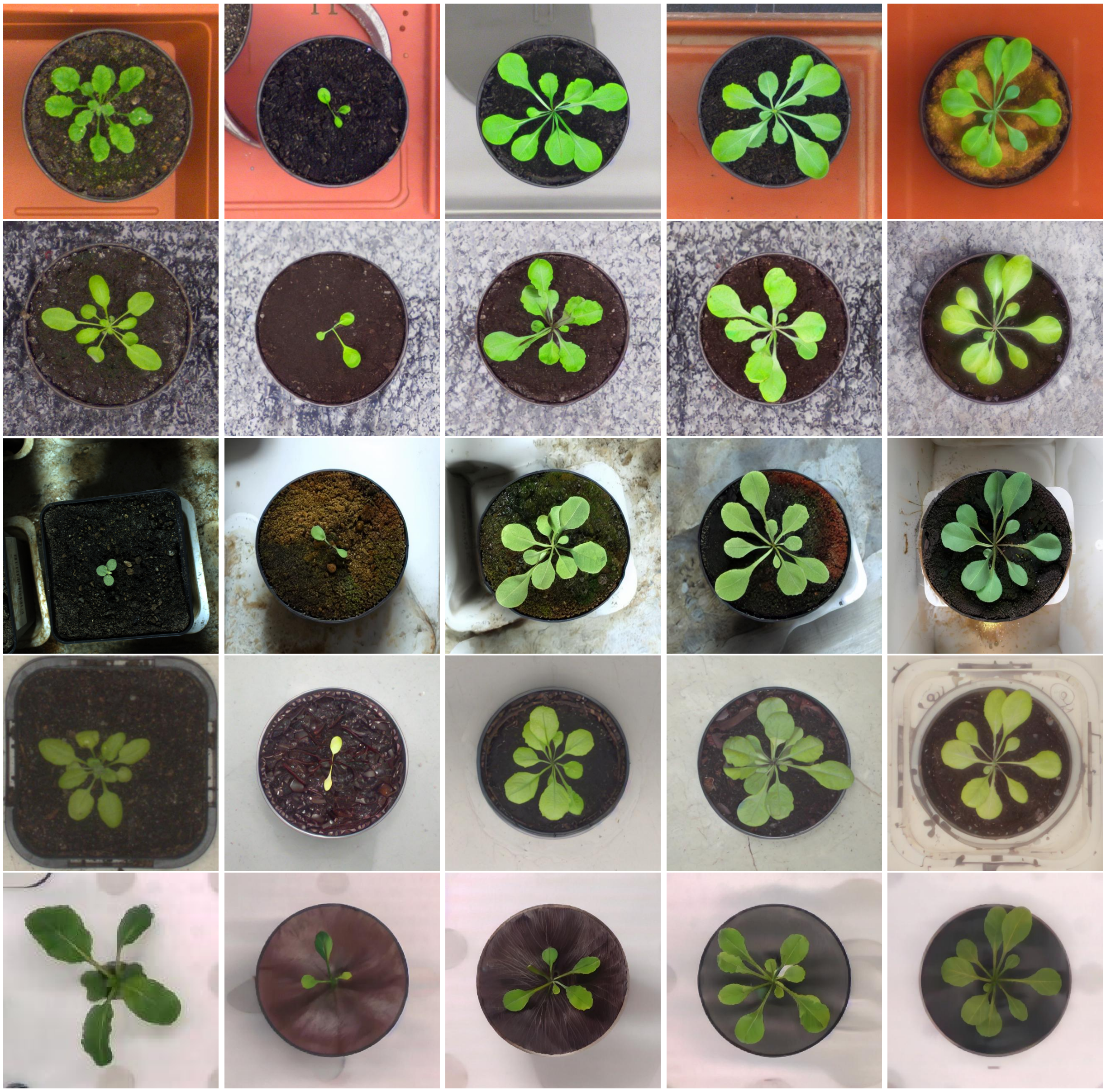


Figure 6. Figure showing an example real image from each dataset (column 1) following by 4 example images generated by style transferring real images using our diffusion pipeline (columns 2 - 5).

[8] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 3

[9] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021. 3

[10] Yinglun Li, Xiaohai Zhan, Shouyang Liu, Hao Lu, Ruibo Jiang, Wei Guo, Scott Chapman, Yufeng Ge, Benoit Solan, Yanfeng Ding, et al. Self-supervised plant phenotyping by combining domain adaptation with 3d plant model simulations: application to wheat leaf counting at seedling stage. *Plant Phenomics*, 5:0041, 2023. 3

[11] Chris C Napier, David M Cook, Leisa Armstrong, and Dean Diepeveen. A synthetic wheat l-system to accurately detect and visualise wheat head anomalies. In *3rd International Conference on Smart and Innovative Agriculture (ICoSIA 2022)*, pages 379–391. Atlantis Press, 2023. 2

[12] Quang Nguyen, Truong Vu, Anh Tran, and Khoi Nguyen. Dataset diffusion: Diffusion-based synthetic data generation for pixel-level semantic segmentation. In *Advances in Neural Information Processing Systems*, pages 76872–76892. Curran Associates, Inc., 2023. 3

[13] Robin Rombach, Andreas Blattmann, Dominik Lorenz,

Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2, 3, 5

[14] Grzegorz Rozenberg and Arto Salomaa. *The mathematical theory of L systems*. Academic press, 1980. 4

[15] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22500–22510, 2023. 3

[16] Yosuke Toda, Fumio Okura, Jun Ito, Satoshi Okada, Toshinori Kinoshita, Hiroyuki Tsuji, and Daisuke Saisho. Training instance segmentation neural network with synthetic datasets for crop seed phenotyping. *Communications biology*, 3(1): 173, 2020. 3

[17] Mario Valerio Giuffrida, Hanno Scharr, and Sotirios A Tsaftaris. Arigan: Synthetic arabidopsis plants using generative adversarial network. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 2064–2071, 2017. 2

[18] Daniel Ward and Peyman Moghadam. Scalable learning for bridging the species gap in image-based plant phenotyping. *Computer Vision and Image Understanding*, 197: 103009, 2020. 2

[19] Daniel Ward, Peyman Moghadam, and Nicolas Hudson. Deep leaf segmentation using synthetic data. *arXiv preprint arXiv:1807.10931*, 2018. 2

[20] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1395–1403, 2015. 5

[21] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023. 3

[22] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 3