PoPGAN: Points to Plant Translation with Generative Adversarial Network

Yuki Yamashita
Department of Design, Kyushu University, Fukuoka, Fukuoka, Japan
yamashita.yuki.299@s.kyushu-u.ac.jp

Kenta Akita
Department of Design, Kyushu University, Fukuoka, Fukuoka, Japan
akita.kenta.633@s.kyushu-u.ac.jp

Yuki Morimoto
Department of Design, Kyushu University, Fukuoka, Fukuoka, Japan
morimoto@design.kyushu-u.ac.jp

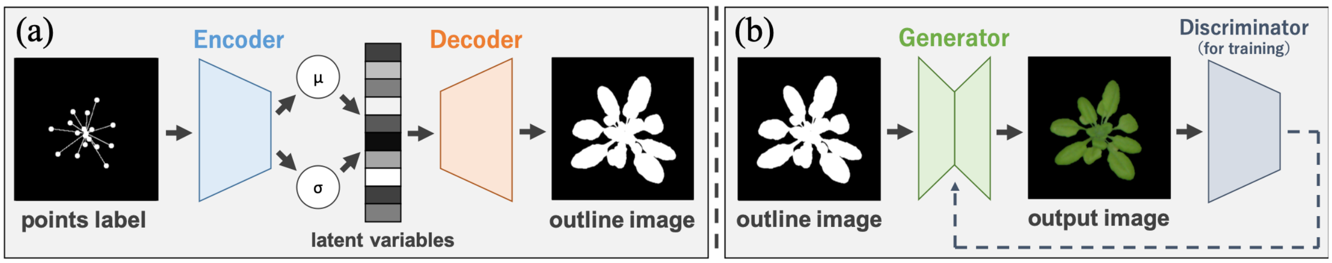

Figure 1: Overview of our proposed two-stage image generation model.

CCS CONCEPTS

• **Computing methodologies**; • **Image processing**;

KEYWORDS

Deep Learning, Image and animation generation, Plants

ACM Reference Format:
Yuki Yamashita, Kenta Akita, and Yuki Morimoto. 2019. PoPGAN: Points to Plant Translation with Generative Adversarial Network. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters (SIGGRAPH '20 Posters), August 17, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3388770.3407421

1 INTRODUCTION

We propose a method of 2D image generation for plant images and animations, based on deep learning. Our point-input interface allows users to specify the positions of leaves in a simple manner. Our neural network enables the generation of high-quality images and animations, using feature interpolation.

Existing image generation methods based on deep learning have problems with the input of image descriptions. Methods with input labels or parameters force users to prepare quite sensitive drawings or excessively rough specifications. Moreover, to generate animations, far more inputs are required.

The generative adversarial networks (GANs) [Goodfellow et al. 2014] method is one of the most effective methods for image generation. The conditional GANs (cGANs) [Mirza et al. 2014] method allows contents to be specified by inputting numerical values or labels. AriGAN [Giuffrida et al. 2017], which specializes cGANs for generating plant images, can specify the number of leaves only. With pix2pix [Isola et al. 2017], which applies cGANs with a color-coded label image as input, it is sometimes necessary to specify the target shape in detail.

In this work, we propose a two-stage deep learning model which generates high-resolution plant images and animations using relatively little training data, and point-input method, to facilitate simple input.

SIGGRAPH '20 Posters, August 17, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7973-1/20/08.
https://doi.org/10.1145/3388770.3407421

2 METHOD

The training data, the input points label images, and the output plant images are all 256×256 RGB images. We used Arabidopsis, which is one of the frequently used model plants, as the target of image generation. More details about this chapter such as the architecture of our networks, GUI editor for points labels and the training dataset are provided in the supplemental material.

2.1 Neural Networks

We propose a two-stage deep learning network, based on GANs and IntroVAE [Huang et al. 2018]. In the first stage (Figure 1a), the encoder extracts the latent variables representing image features from the points label, and the decoder translates the latent variables to an outline image, in which the inside of the plant contour shape is filled in white. In the second stage (Figure 1b), the generator creates an image, including detailed patterns and shadings on leaves, from the outline image. Finally, the user can obtain the specified plant image by the points label.

For training, we applied the loss functions of IntroVAE as a base to make encoders and decoders perform adversarial training, similarly to GANs, so that we could obtain higher-quality images from a smaller amount of training data.

2.2 Points Label

As an input to our image generation model, we apply the points label: this is an image with white points placed at arbitrary positions on a black background. The number and positions of points correspond to the number of leaves and the center position of the plant in the generated image, respectively. Users create the points label, using our original editing GUI, to specify the positions of leaves. Our points label method enables simple and intuitive input.

2.3 Creation of the Dataset

We obtained the training data of our image generation model from a public dataset [Scharr et al. 2014] comprising 783 images of Arabidopsis. This dataset contains raw plant images, their leaf positions, and white-masked images of the plant shapes without background. Our self-constructed dataset consists of 783 sets of three images: the ground-truth plant images, their corresponding points labels, and the outline images. In our dataset, points labels are created by placing white points at leaves and root locations. The leaf points are connected to the root points with white lines. To create the outline images, we applied mask images in the public dataset to separate the foreground from the background. The ground-truth plant images were created by removing the background from the plant images in the public dataset.

Preprocessing—comprising random rotation, scaling, and parallel translation—were applied to our dataset images, to increase the robustness of the neural network.

3 RESULTS AND CONCLUSION

We compare the results of our method and pix2pix in Figure 2. The same original dataset and points labels were provided as input to both of the methods for training and generating plant images. The results show that both methods reproduced the shape of the ground-truth plant images. However, our method represented the details, such as patterns on leaves, more clearly.

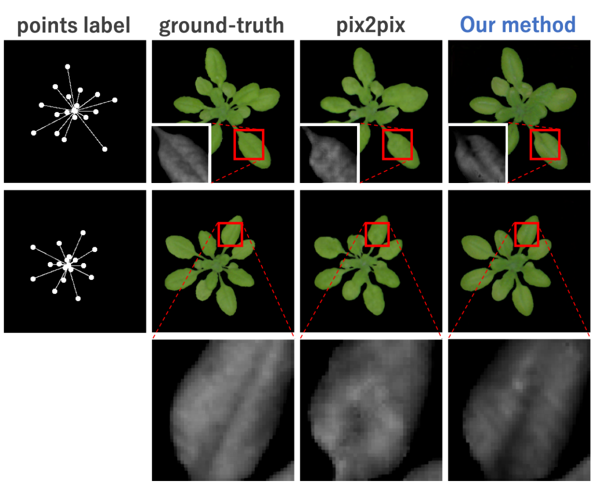

Figure 2: Comparison of the results with points label as input.

Table 1: Results of the evaluation of reproducibility.

Performance indicator	pix2pix	Our method
SSIM (the higher, the better)	0.855	0.907
FID (the lower, the better)	162.565	142.595

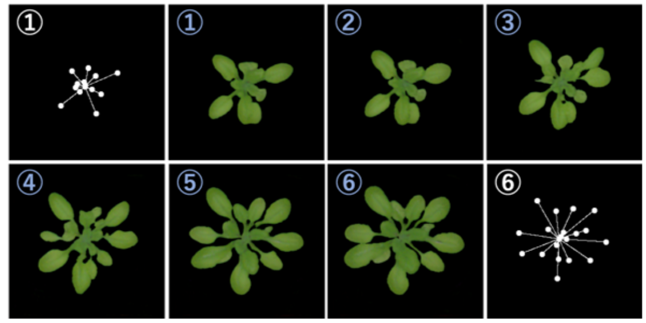

Figure 3: Animation generation of plants using our method.

To evaluate the reproducibility of image generation by our method, we performed quantitative evaluation using structural similarity (SSIM) and Fréchet Inception Distance (FID) The evaluation results (Table 1) show that our method produced higher-quality images (more similar to the target plant images) than pix2pix, for both of the performance indicators.

Our neural network can generate animations by inputting two points labels. The network linearly interpolates between the two latent variables extracted from the points labels. The resulting animations change smoothly between the two inputs (Figure 3). This reproduces the appearance of the growth process of the plants. Thus, our network can generate meaningful animations, in contrast to simple morphing. More results are provided in the supplemental material.

Our future work is to increase the variety of training data. In addition, we aim to improve the flexibility of our neural network and interface, to generate parts of plant other than leaves, and plants of other species. Formatting math statements and extracts.

REFERENCES

Ian J. Goodfellow, *et al.* 2014. Generative Adversarial Nets. NIPS (2014).
Mehdi Mirza, *et al.* 2014. Conditional Generative Adversarial Nets.
Mario Valerio Giuffrida, *et al.* 2017. ARIGAN:Synthetic Arabidopsis Plants using Generative Adversarial Network. ICCV CVPPP Workshop (2017).
Phillip Isola, *et al.* 2017. pix2pix: Image-to-Image Translation with Conditional Adversarial Networks.CVPR (2017).
Huaibo Huang, *et al.* 2018. IntroVAE: Introspective Variational Autoencoders for Photographic Image Synthesis. NeurIPS (2018)
Hanno Scharr, *et al.* 2014. Annotated Image Datasets of Rosette Plants Technical Report No.FZJ-2014-03837 (2014).