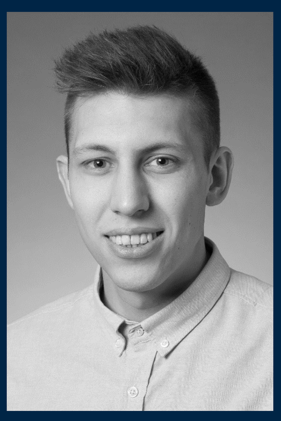
Using Semi-supervised GANs to Disentangle Information in Artificial Images of Plant Seedlings (Preliminary results)



Simon Leminen Madsen
PhD Student

slm@eng.au.dk
+45 50409949

Anders Krogh Mortensen
Postdoc

anmo@agro.au.dk

Henrik Karstoft
Supervisor 1
Professor (Docent)

hka@eng.au.dk

Rasmus Nyholm Jørgensen
Supervisor 2
Senior Researcher

rnj@eng.au.dk

INTRODUCTION

Machine learning and especially deep learning have become increasingly popular in recent years. However, these algorithms are dependent on the availability of annotated training data, as they are still primarily trained using supervised learning.

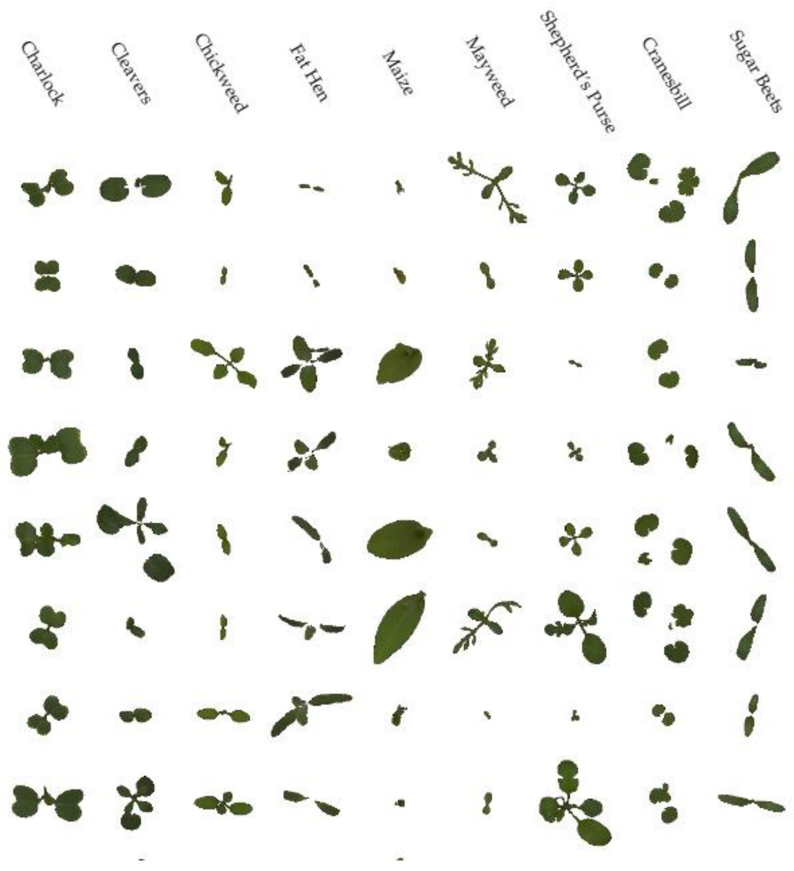
There exists many publicly available dataset online. However, for domain specific applications such as plant detection and recognition the availability of annotated training data is limited (Tsaftaris et al. 2016). Due to this, it is challenging to train deep learning algorithms for plant seedling classification.

It is relatively easy to collect new data in the domain of agriculture. However, the annotation process is often very time-consuming and require a high level of expertise, since plant seedlings are visually similar across multiple species (Minervini et al. 2017). Even with expert annotators, the process is prone to errors (Dyrmann et al. 2016).

Alternatively, a generative model could be used to produce artificial samples that mimic properties of real data samples.

In this work we explore the possibility of using a generative adversarial networks (GANs) model (Goodfellow et al. 2014), to produce artificial samples of nine different species of plant seedlings. The model is trained semi-supervised to produce samples that are distinguishable across the different species and to allow some user-control over the appearance of the artificial samples through a set of latent variables.

Figure 1: Examples of plant seedlings from nine different species. The examples are taken from Giselsson et al. 2017



REFERENCES

1. Tsaftaris, S. A., Minervini, M., & Scharr, H. (2016). Machine learning for plant phenotyping needs image processing. Trends in plant science, 21(12), 989-991.
2. Minervini, M., Giuffrida, M. V., Perata, P., & Tsaftaris, S. A. (2017). Phenotiki: An open software and hardware platform for affordable and easy image-based phenotyping of rosette-shaped plants. The Plant Journal, 90(1), 204-216
3. Dyrmann, M., Midtiby, H. S., & Jørgensen, R. N. (2016). Evaluation of intra variability between annotators of weed species in color images. In International Conference on Agricultural Engineering 2016.
4. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).
5. Giselsson, T. M., Jørgensen, R. N., Jensen, P. K., Dyrmann, M., & Midtiby, H. S. (2017). A public image database for benchmark of plant seedling classification algorithms. arXiv preprint arXiv:1711.05458

METHODOLOGY

The applied GANs model, WacGAN-info is inspired by three previous GANs designs:

- WGAN-GP (Gulrajani et al. 2017), ACGAN (Odena et al. 2017) and infoGAN (Chen et al. 2016).

Network design:

- The network designs are inspired by Odena et al.:
  - The discriminator, $D(\boldsymbol{x})$, is implemented as a convolutional neural network with three parallel outputs: D_S, D_C and D_I
  - The generator, $G(z, y, u)$, is implements as a deconvolutional neural network.

Objective function:

The WacGAN-info objective is divided into three parts: a source loss, L_S, a class loss, L_C, and an information loss, L_I:

- The source loss, L_S, is used to distinguish between the real and artificial data distribution. L_S is implemented as a Wasserstein loss with gradient penalty as described by Gulrajani et al. :

$$L_S = \mathbb{E}_{\boldsymbol{x}\sim p_r}[D_S(\boldsymbol{x})] - \mathbb{E}_{\widetilde{\boldsymbol{x}}\sim p_g}[D_S(\widetilde{\boldsymbol{x}})] + \lambda\, \mathbb{E}_{\hat{\boldsymbol{x}}\sim p_{\hat{\boldsymbol{x}}}}[(\|\nabla_{\hat{\boldsymbol{x}}} D_S(\hat{\boldsymbol{x}})\|_2 - 1)^2]$$

- The class loss, L_C, is used to ensure the generator produces distinguishable samples for each class. L_C is implemented as the cross-entropy loss between the expected output, y, and the output of the discriminator's classification branch, $D_C(\boldsymbol{x})$:

$$L_C = \mathbb{E}_{\boldsymbol{x}\sim p_r}\left[-\sum_{i\in N_C} y_i\, D_{C_i}(\boldsymbol{x})\right] - \mathbb{E}_{\widetilde{\boldsymbol{x}}\sim p_g}\left[-\sum_{i\in N_C} \tilde{y}_i\, D_{C_i}(\widetilde{\boldsymbol{x}})\right]$$

- The information loss, L_I, is used to ensure information preservation through the GANs setup, thus enabling control over the sample appearance. L_I is implemented as the mean square loss between the intended info variables, u, and the predicted info variables of the discriminator, $D_I(\boldsymbol{x})$:

$$L_I = \mathbb{E}_{\widetilde{\boldsymbol{x}}\sim p_g}\left[-\sum_{j\in N_I} \tilde{u}_j\, D_{I_j}(\widetilde{\boldsymbol{x}})\right]$$

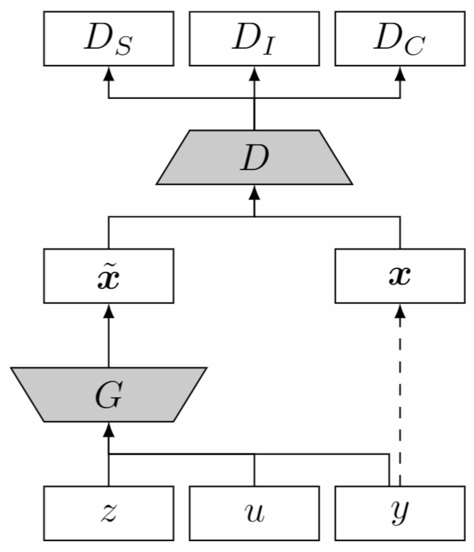


Figure 2: WacGAN-info model configuration. z, y and u are the model inputs corresponding to noise vector, class vector and latent info vector respectively. The signals D_S, D_C and D_I are parallel model outputs corresponding to the distribution estimation (true or artificial samples), the class estimation and the info variable estimation.

Model updates:

Discriminator is optimized to minimize: $-L_S + \omega_C L_C + \omega_I L_I$

Generator is optimized to minimize: $L_S + \omega_C L_C + \omega_I L_I$

REFERENCES

1. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of wasserstein gans. In *Advances in neural information processing systems* (pp. 5767-5777).
2. Odena, A., Olah, C., & Shlens, J. (2017, August). Conditional image synthesis with auxiliary classifier gans. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70* (pp. 2642-2651). JMLR. org.
3. Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in neural information processing systems* (pp. 2172-2180).

PRELIMINARY RESULTS

Generating artificial images of plant seedlings

- Producing discriminable samples from multiple classes, while maintaining relative high intra-class variance.

Figure 3: Artificial samples of plant seedlings generated using the WacGAN-info model. Each column represents a different species and each row is generated using the same unique noise vector.

Class discriminability

- The class discriminability is assessed using an external ResNet101 classifier trained for the plant seedling classification.
  - The accuracy measure indicated how often a WacGAN-info sample is recognized as the intended species.
- Recognition accuracy on all WacGAN-info samples:

Species	Test Set			
	Real Data		Artificial Data	
	N Samples	Accuracy	N Samples	Accuracy
Charlock	21–35	$91.6 \pm 6.6\%$	10 000	$80.1 \pm 9.6\%$
Cleavers	16–34	$93.4 \pm 6.4\%$	10 000	$73.3 \pm 8.3\%$
Common Chickweed	52–69	$92.4 \pm 5.4\%$	10 000	$87.9 \pm 8.8\%$
Fat Hen	34–50	$93.1 \pm 4.0\%$	10 000	$71.7 \pm 17.0\%$
Maize	10–20	$91.0 \pm 10.9\%$	10 000	$64.9 \pm 10.5\%$
Scentless Mayweed	41–60	$88.3 \pm 11.1\%$	10 000	$74.7 \pm 17.8\%$
Shepherd's Purse	19–30	$85.2 \pm 8.5\%$	10 000	$53.8 \pm 10.2\%$
Small-flowered Cranesbill	35–55	$95.9 \pm 3.7\%$	10 000	$84.4 \pm 6.2\%$
Sugar Beets	12–29	$93.0 \pm 7.0\%$	10 000	$78.0 \pm 6.6\%$
Total	310	$91.8 \pm 1.9\%$	90 000	$66.9 \pm 4.0\%$

- Exploring the relationship between the discriminator's confidence in the artificial samples and the recognition accuracy reported by the external classifier:

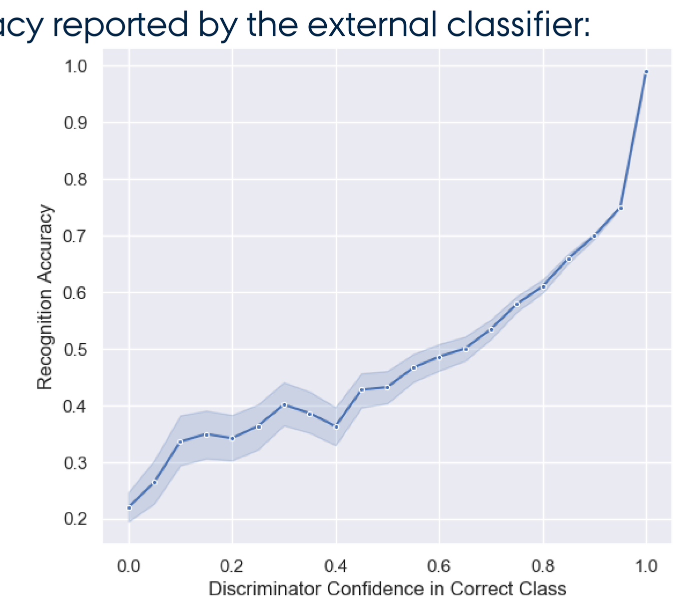


Figure 4: Average accuracy of recognizing the intended species from the WacGAN-info samples, when evaluated using an external classifier, as a function of the discriminator's confidence in the sample belonging to the correct class ($D_C(\boldsymbol{x})$).

Controlling the visual content using latent variables

- The WacGAN-info model is trained to preserve information, stored in the latent variables, u.
- The latent variables seem to capture dominating modalities in the data, such as rotation/orientation or size of the plant seedlings.
- The latent variable thus enables control over these modalities:

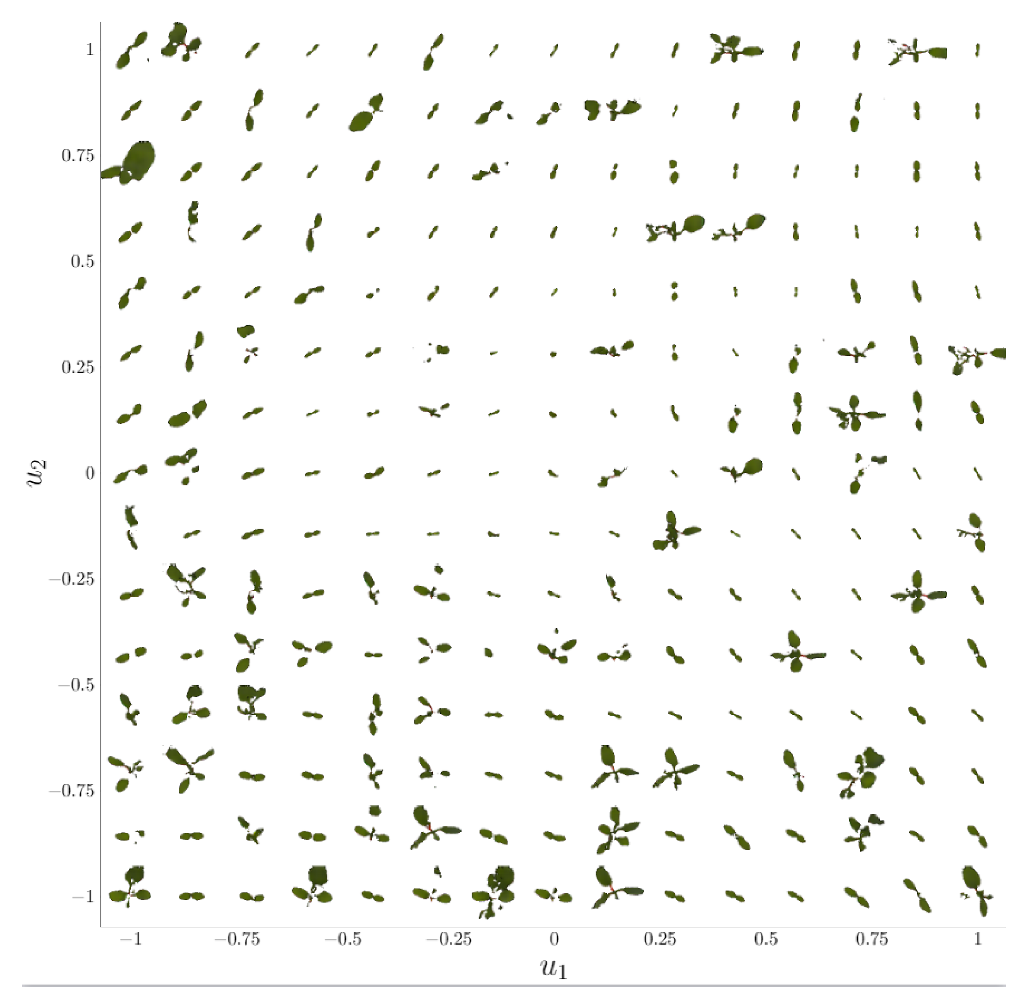


Figure 5: Visualization of the information space using two uniformly distributed continuous variables. The visualization show that specific combinations can be used to control the rotation and size of the generated plant seedlings. All samples presented in the plot is supposed to represent Common Chickweed.

AARHUS UNIVERSITY
DEPARTMENT OF ENGINEERING

ELECTRICAL AND COMPUTER ENGINEERING
SIGNAL PROCESSING