

Data-Efficient GAN for Synthetic Samples of Rare Classes

How well can GAN generate realistic samples for rare classes?



1. Background

- Camera traps are used for ecology research
- Classification models have bad accuracy without enough training data
- CCT-20 Dataset with an artificial rare class for deer.
- Using synthetic data can significantly improve the precision of rare classes [1]
- How well would synthetic data generated using a GAN improve precision of rare classes?

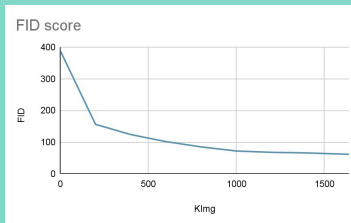
2. Method

- Train GAN with images from the training data-set
- Train classification model without added images as Base
- Train classification models with 1,000 and 10,000 added deer images generated by GAN
- Train classification models with 1,000 and 10,000 oversampled deer images

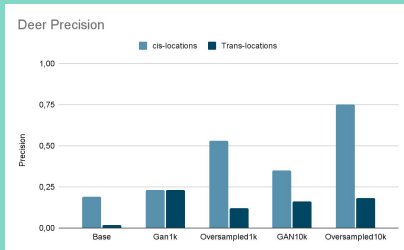
3. Results



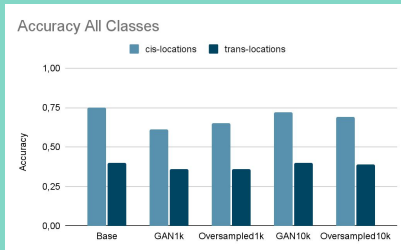
Images of deer generated after 1 day (left) and after 5 days of training (right)



Average FID score of 50 thousand generated images over training time.



Precision of the deer class per model



Accuracy of all classes per model

4. Comparison

- For locations seen during training: the models using GAN synthetic data do not outperform our oversampled models in deer accuracy.
- For locations not seen during training: the models using the GAN synthetic samples outperform the other models.
- For locations not seen during training: using synthetic samples generated using simulation software increases both the deer precision and the overall accuracy compared to using a GAN [1]

5. Conclusions

- Significant training time is required for a GAN to produce good results
- Synthetic data generated by a GAN does not outperform synthetic data generated using simulation software.
- Using a GAN to generate synthetic data is not worth it compared to other methods

References:
[1] S. Beery, Y. Liu, D. Morris, J. Plavis, A. Kapoor, M. Meister, N. Joshi, P. Perona. Synthetic Examples Improve Generalization for Rare Classes