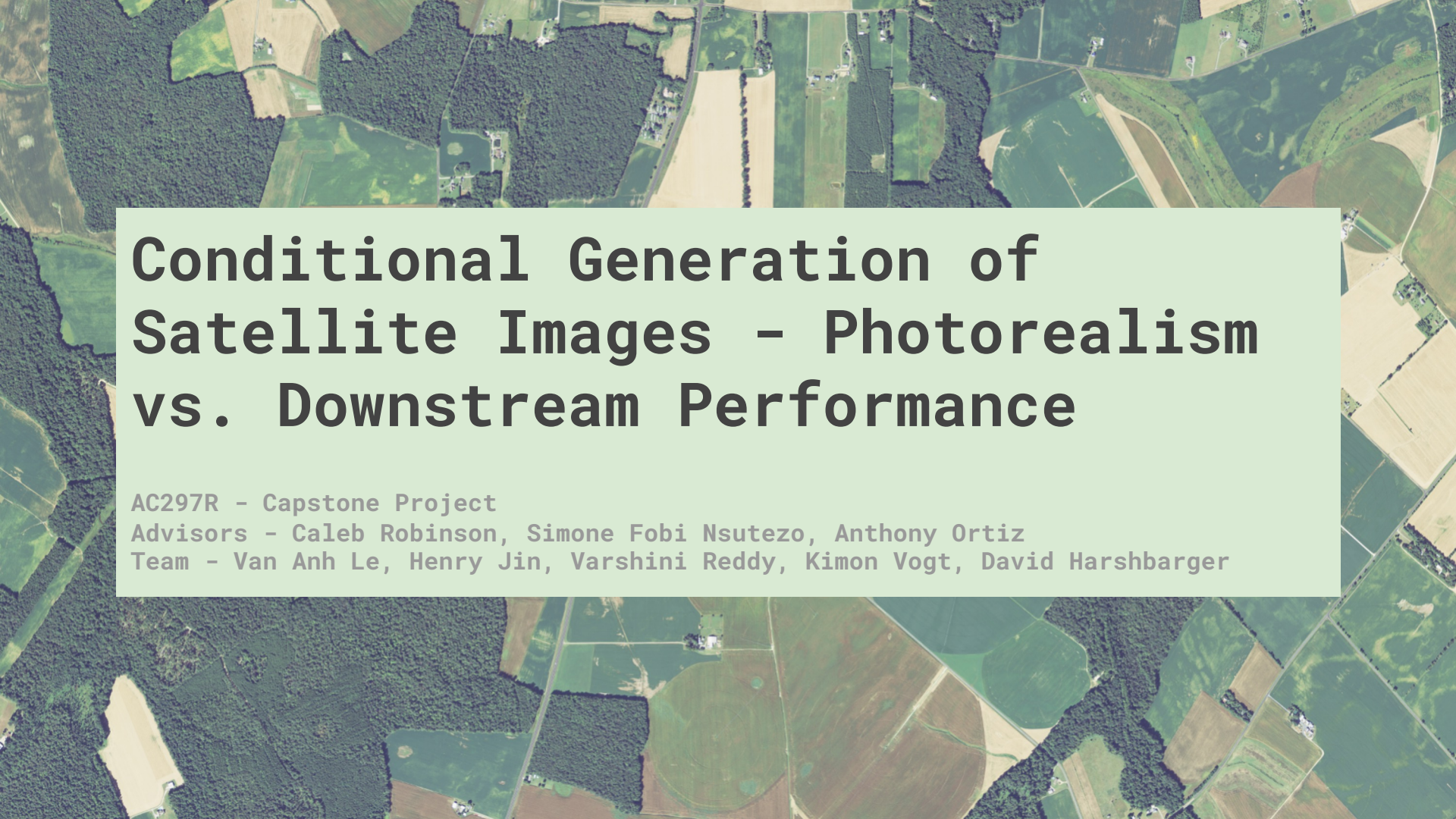


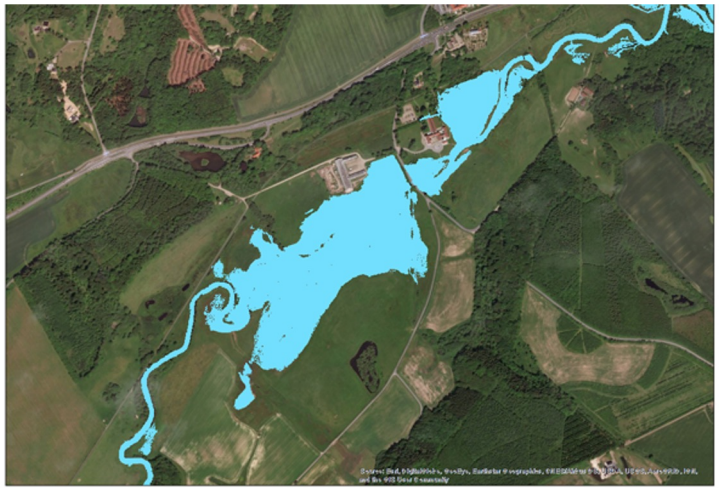
Conditional Generation of Satellite Images - Photorealism vs. Downstream Performance

AC297R - Capstone Project
Advisors - Caleb Robinson, Simone Fobi Nsutezo, Anthony Ortiz
Team - Van Anh Le, Henry Jin, Varshini Reddy, Kimon Vogt, David Harshbarger

Problem & Motivation

Motivation

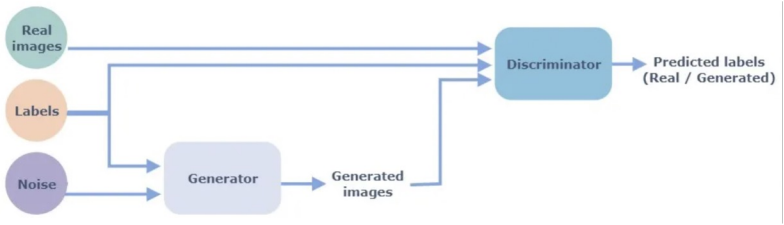
- Geospatial ML offers valuable insights to many fields of studies but licensing cost makes access to data difficult.
- **Improve access to geospatial ML -** Cannot release commercial datasets for social goods problems due to licensing concerns.
- **Improve model performance** - Synthetic data can augment existing real data to build better models.



Problem Statement

Generate class conditional synthetic satellite imagery

- Conditional generative models **f(y) = x**, where **x** is satellite imagery and **y** is segmentation mask.



Research Question

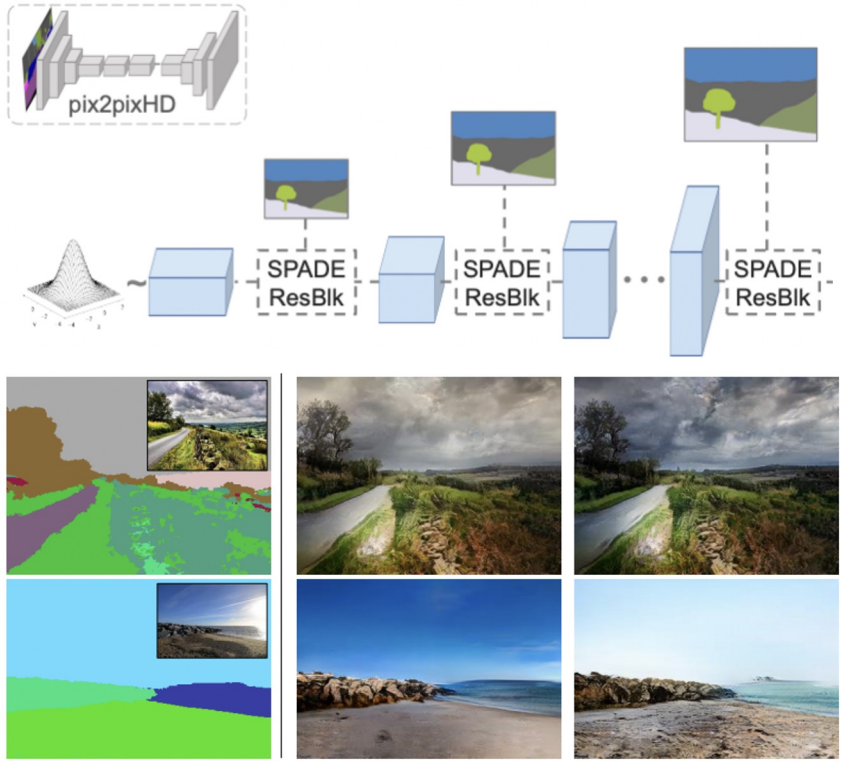
- Generate synthetic high-resolution satellite imagery from label prompts that are realistic and can improve downstream task performance
- **Examine trade-offs between diversity of synthetic data and downstream performance**

Focus of our group

Methods

Generative Model - SPADE

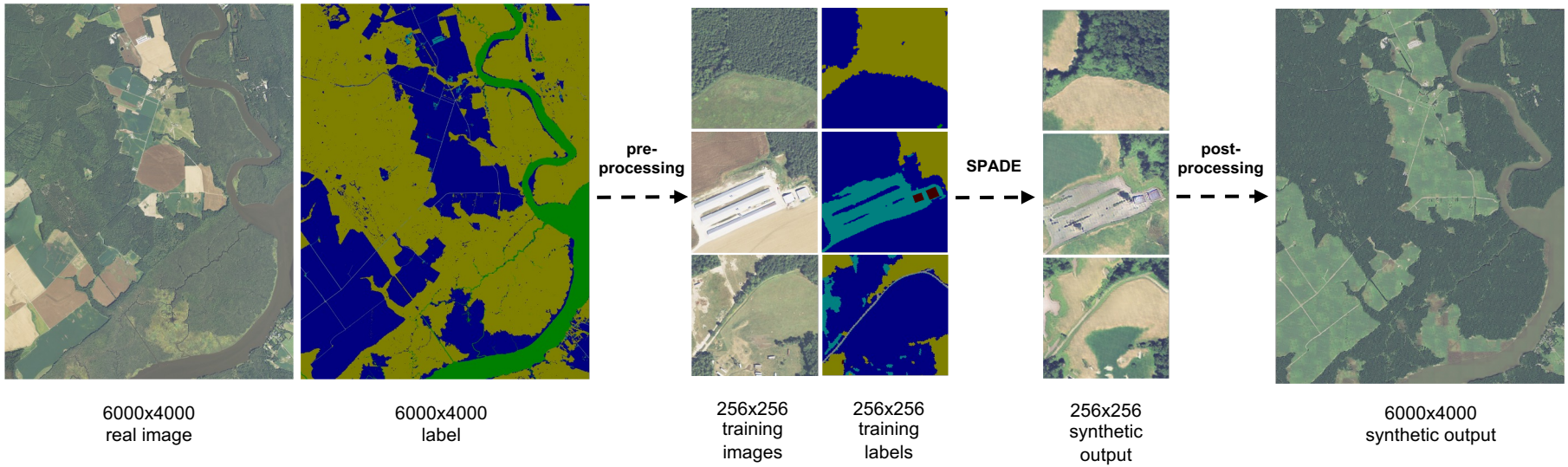
- Spatially-Adaptive Denormalization (Park et al. 2019)
- Based on pix2pixHD conditional GANs, with additional VAE structure.
- Label masks are denormalized
- Better performance due to retained information of “ground truth”



[1] Park, Taesung, et al. "Semantic image synthesis with spatially-adaptive normalization." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019

Dataset / Modeling

- 100 train and 25 test *tiles* from Maryland, 7 land cover classes



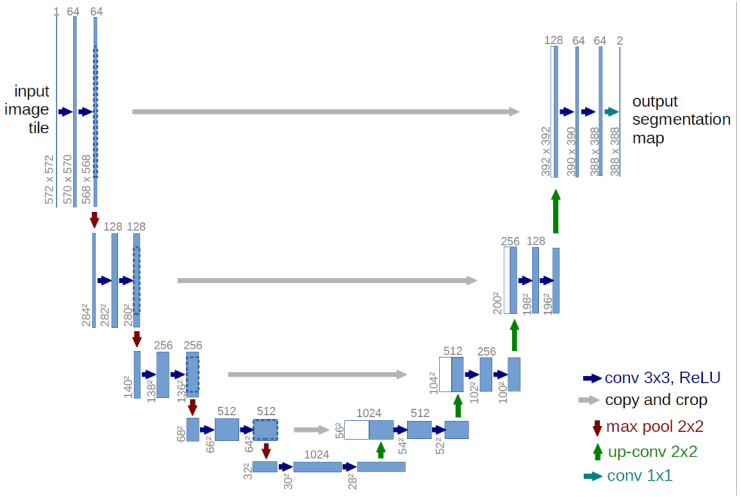
Evaluation: FID and mIoU

Photorealism:

- Frechet Inception Distance (FID) to measure distance between distributions of real vs synthetic data.

Downstream performance:

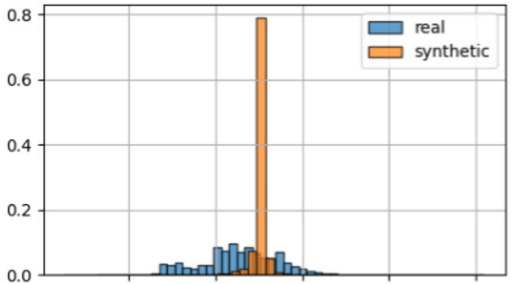
- Semantic segmentation using Unet architecture with ResNet encoder
- Train on synthetic training data (or mixture of real / synthetic training data)
- Test on real testing data
- Mean Intersection-over-Union and other metrics



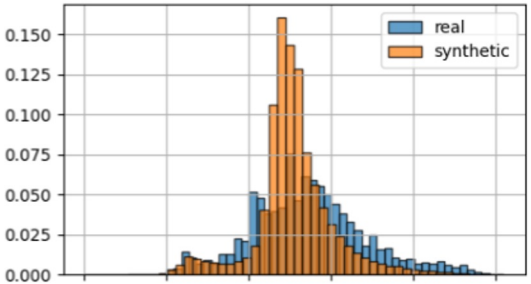
Evaluation

- **Pixel diversity**
  - Measure the distribution of values in the imagery for each class
  - Real imagery distribution
  - Synthetic imagery distribution

Water

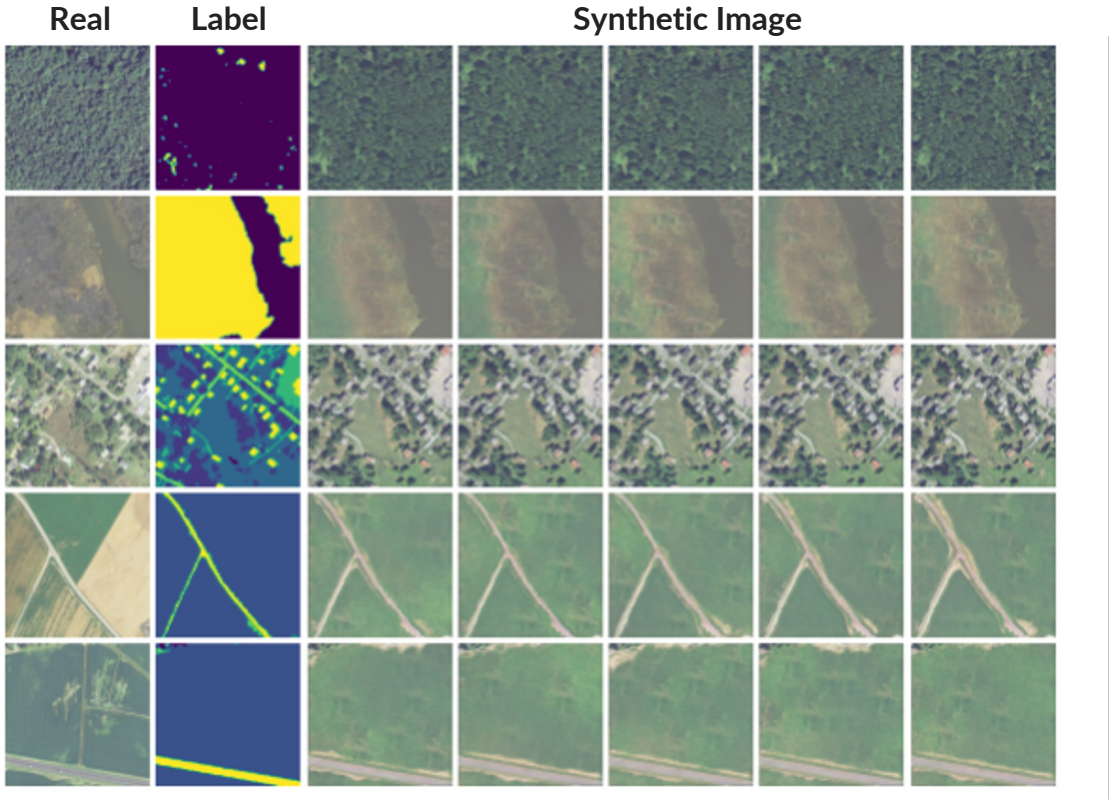


Low Vegetation



Baseline Performance

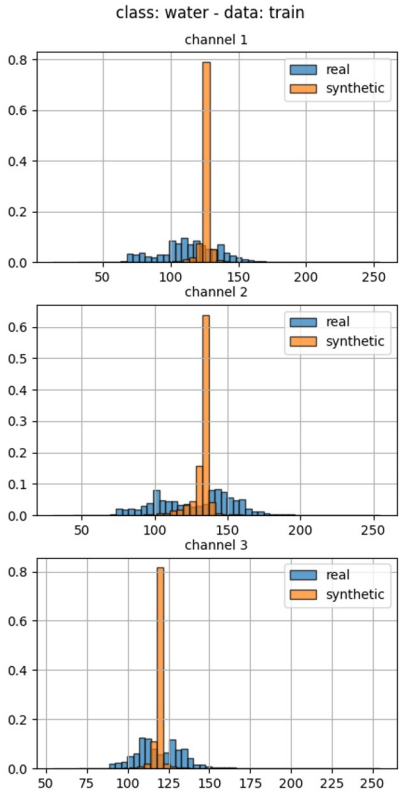
Baseline

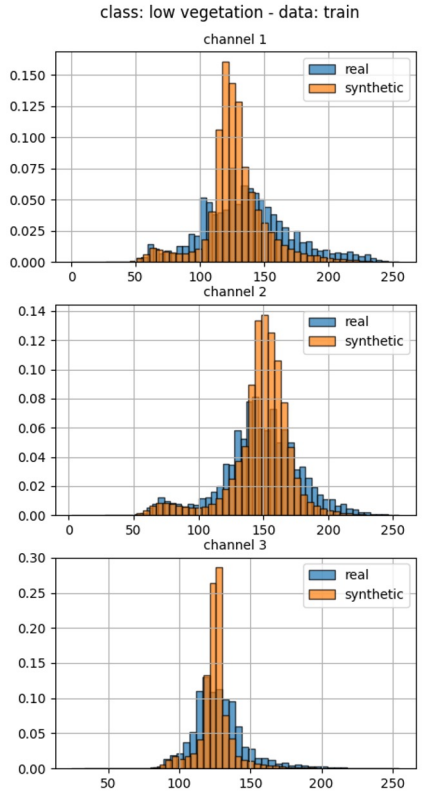


Baseline

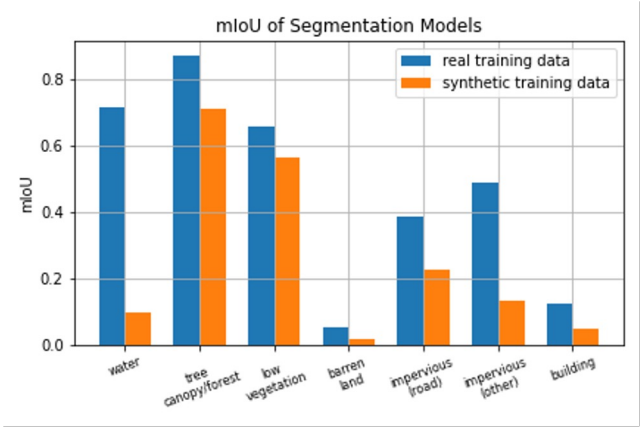
Results on test set after training for 4 epochs

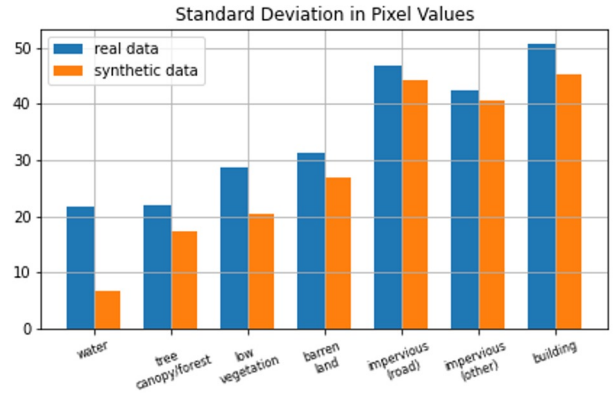
- Test FID = 68.53
- Test mIoU = 0.579
- Lack of diversity in output

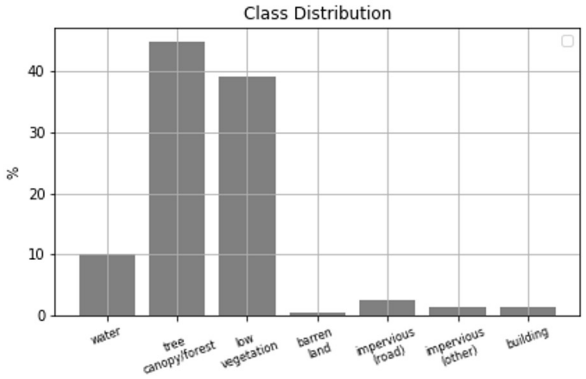




Baseline







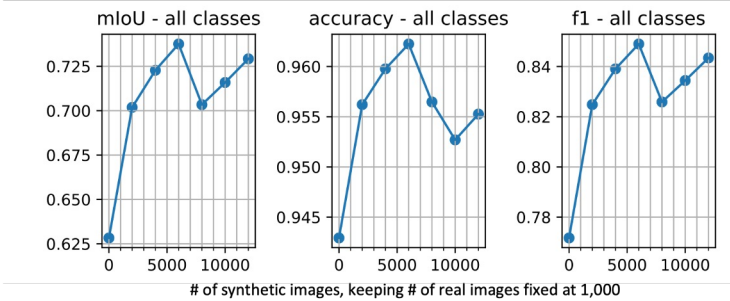
Segmentation results by class:

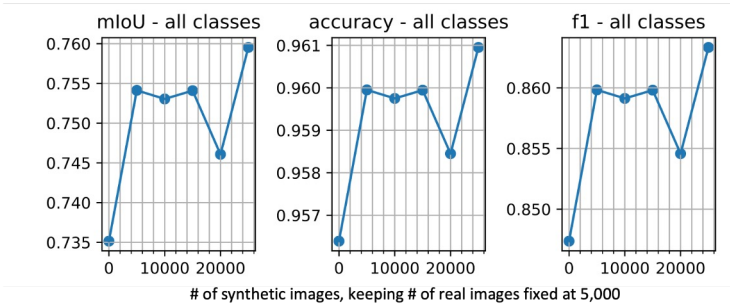
- Poor performance in water class - issues with lack of diversity
- Other poor performing classes - issues with imbalance representation.

Data Augmentation

- **Substitutability**: Performance drop when substituting real training data for synthetic (first plot)

- **Utility for augmentation:** augmenting limited real data with synthetic improves performance (second and third plot)





Increasing Diversity

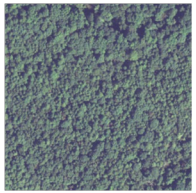
Proposed Improvement

- **Increase diversity of output** using additional regularization term in training, proposed by Yang et al. (2019)
- Additional loss term that measures L1 distance between 2 forward passes by the generator.
- Experiment with varying weights on diversity loss

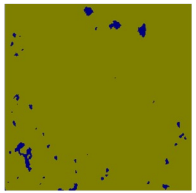
$$\max_G \mathcal{L}_z(G) = \mathbb{E}_{z_1, z_2} \left[\min \left(\frac{\|G(x, z_1) - G(x, z_2)\|}{\|z_1 - z_2\|}, \tau \right) \right]$$

[1] Yang, Dingdong, Seunghoon Hong, Yunseok Jang, Tianchen Zhao, and Honglak Lee. "Diversity-sensitive conditional generative adversarial networks." *arXiv preprint arXiv:1901.09024* (2019).

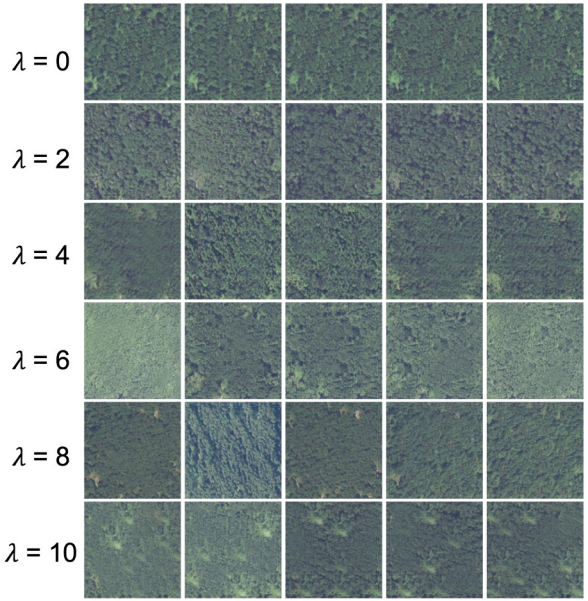
Diversity Regularization



Real Image

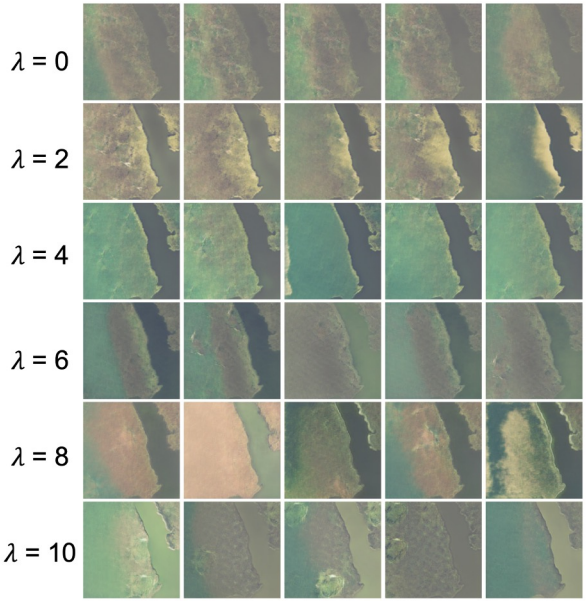


Label

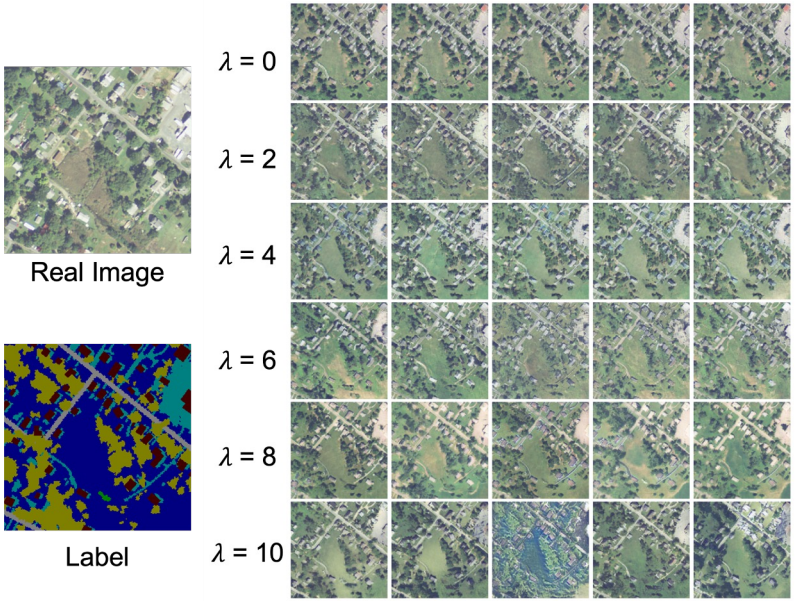


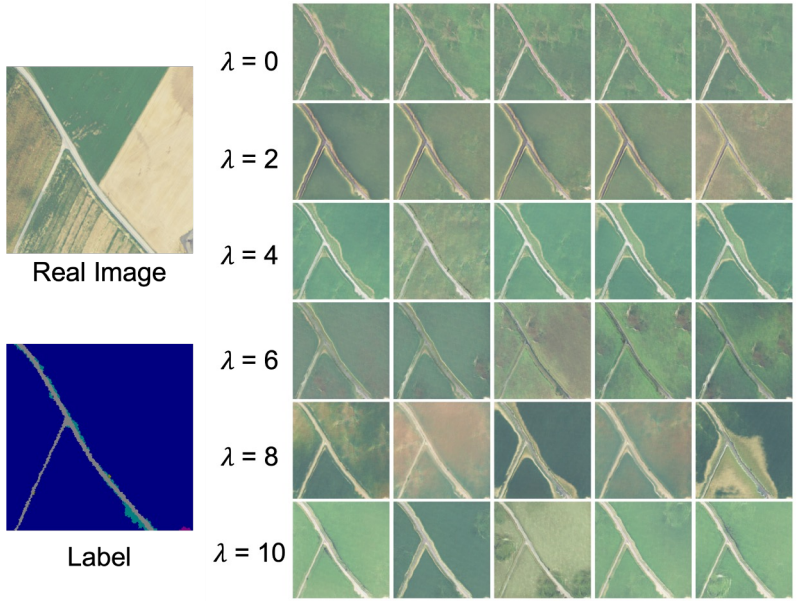
Real Image

Label



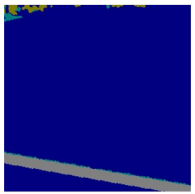
Diversity Regularization



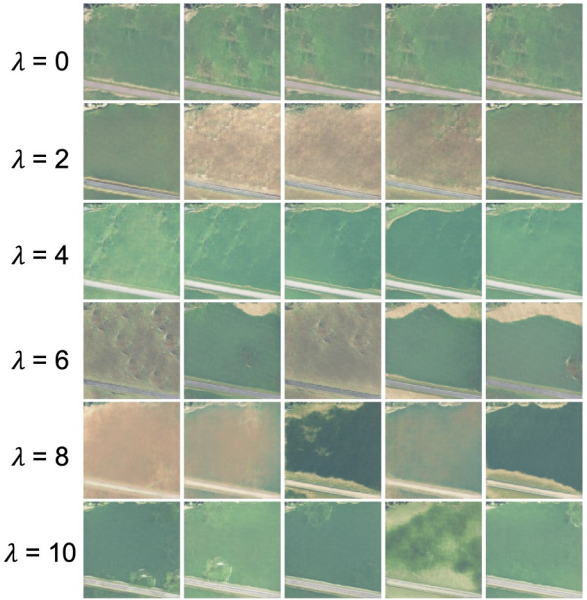


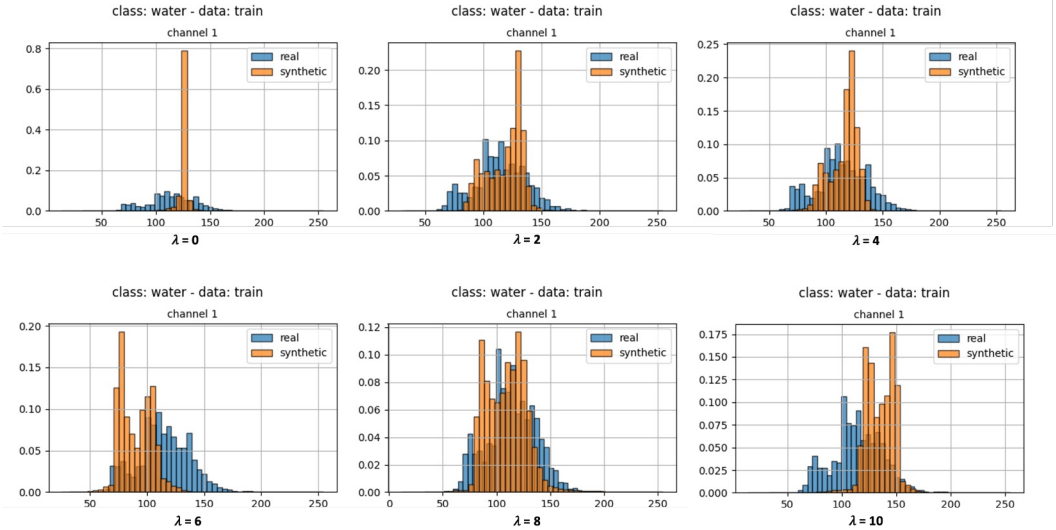
Diversity Regularization

Real Image

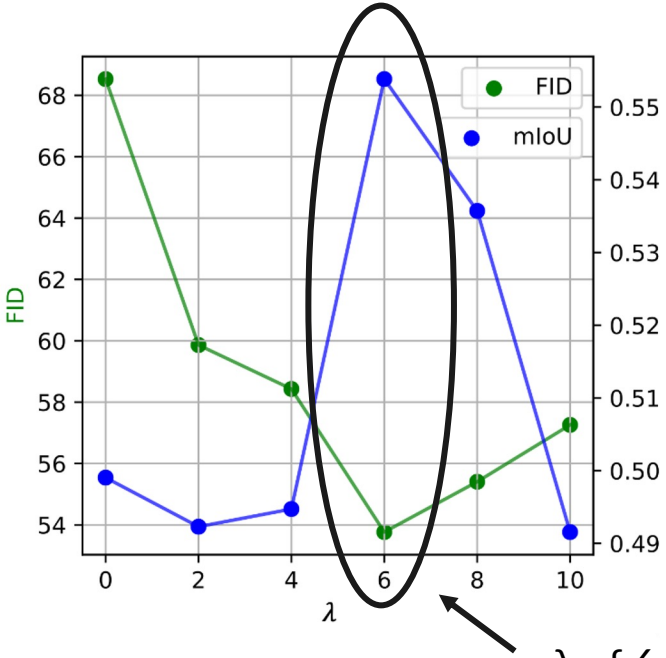


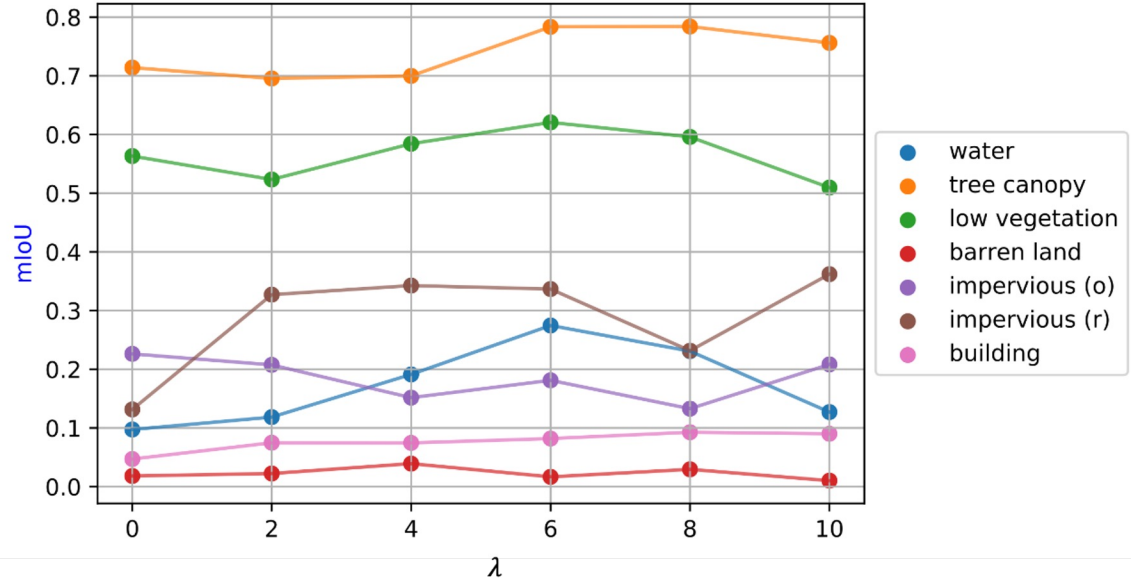
Label





Diversity improves FID and mIoU

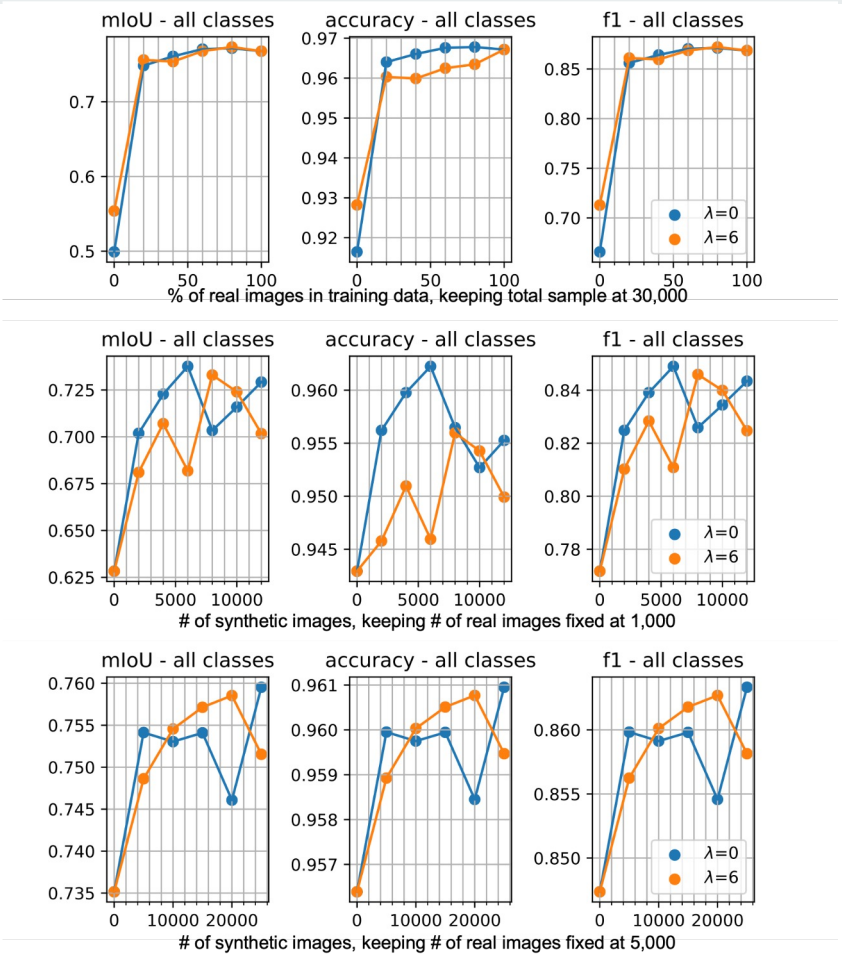




λ of 6 maximizes performance

Data Augmentation

- Despite improvement in FID and mIoU, more diversity does not necessarily improve utility in augmentation.

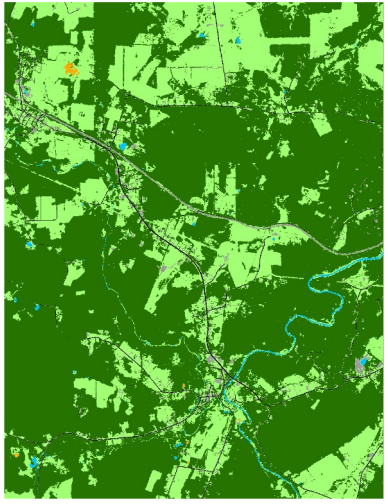


Real Image

$\lambda = 0$

$\lambda = 2$

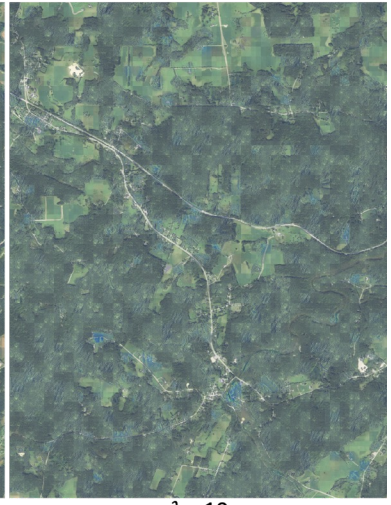
$\lambda = 4$



Label

$\lambda = 6$

$\lambda = 8$



$\lambda = 10$

Conclusion

1. **Synthetic images are useful as an augmentation strategy** in low-data regime, **but cannot fully replace real ones** in downstream task.
2. **Increasing diversity of synthetic output can improves both FID and mIoU**.

Extra stuff

Sampling diversity in default model

Over a run of 500 samples from the same X the two samples that were farthest from each other in latent space look nearly identical


Learned shadows

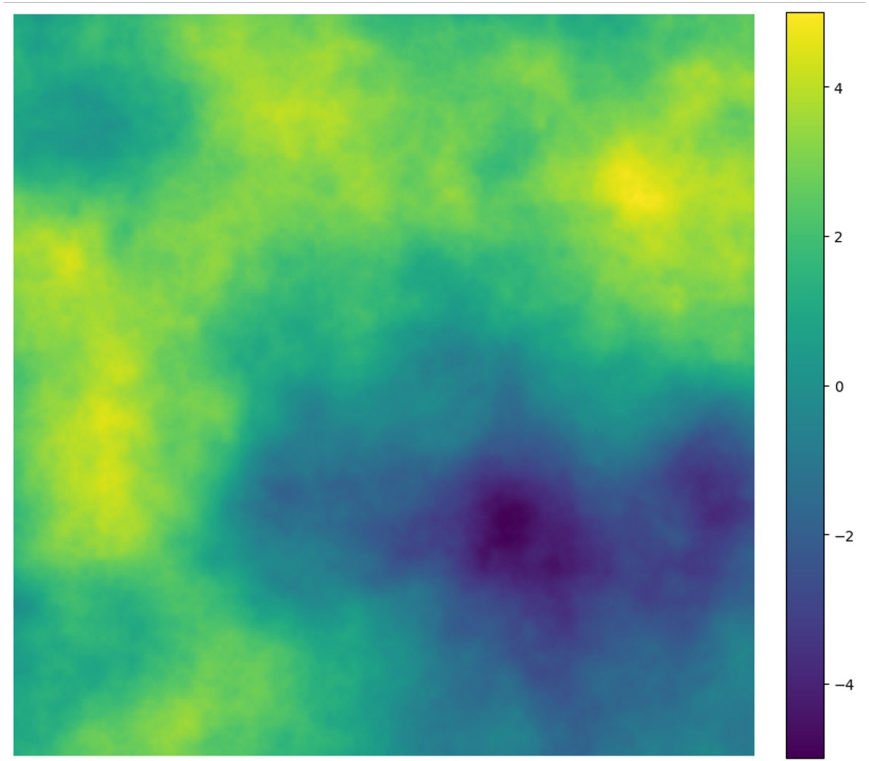
Reconstructing large scenes from generated patches

- Naively sampling z's from the encoder will generate checkerboarding artifacts when reconstructing large inputs (especially as the encoder learns diversity)
- Methods
  - Smoothing in output (pixel) space -- generate overlapping patches and average the pixel values
  - Smoothing in latent (input) space -- set *z'* = average of *z* in a window with patch radius *r*
  - Make *z* a function of structured noise

Structured noise

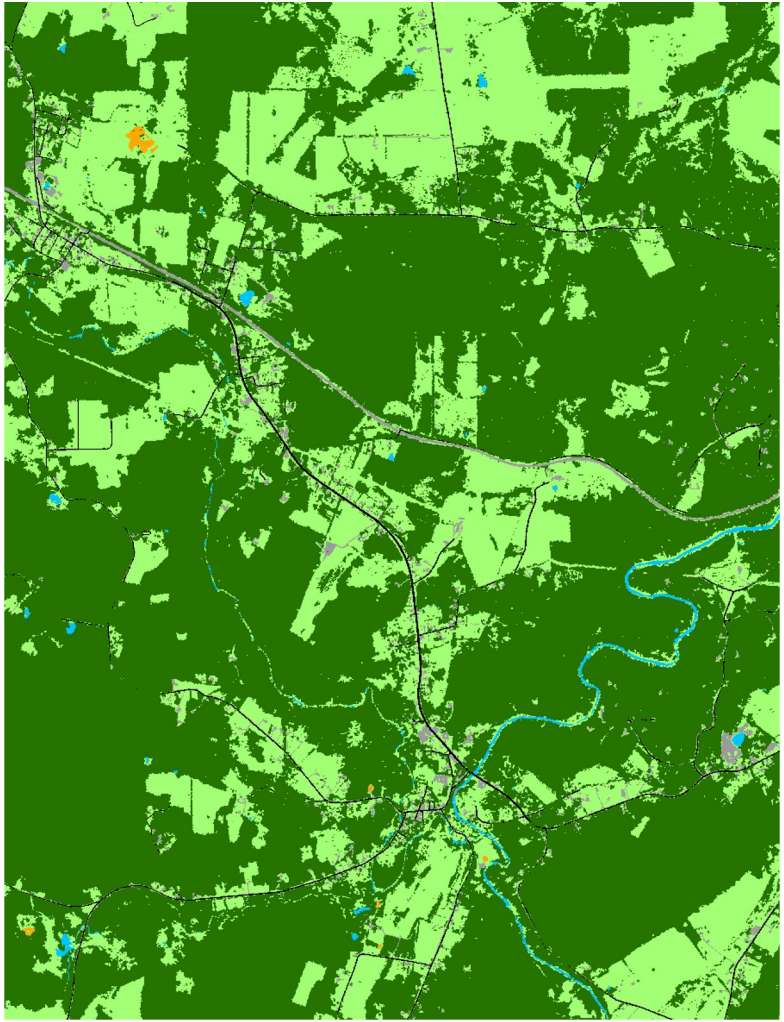
- E.g. perlin or simplex noise -- 2D random maps with local smoothness
- Notation
  - Let f(X) = z be the encoder
  - Let g(z, Y) = P be the decoder
  - Let *R* be a noise map
  - Let X_ij be the input patch at location i,j
- Method
  - z_ij = f(X_ij) * R_ij
  - P_ij = g(z, Y_ij)

Noise map, *R*



Real NAIP

Land cover



Legend

- Water
- Tree Canopy
- Low Vegetation
- Barren
- Impervious Surfaces
- Impervious Roads

Synthetic NAIP

Synthetic NAIP
z + N(0,1)*3

Synthetic NAIP
z * R