USC University of Southern California

PTE-572 Geostatistics

Parametrization and generation of geological facies models with DCGAN based on a training image

Mahammad Valiyev, PhD Student

USC Viterbi School of Engineering

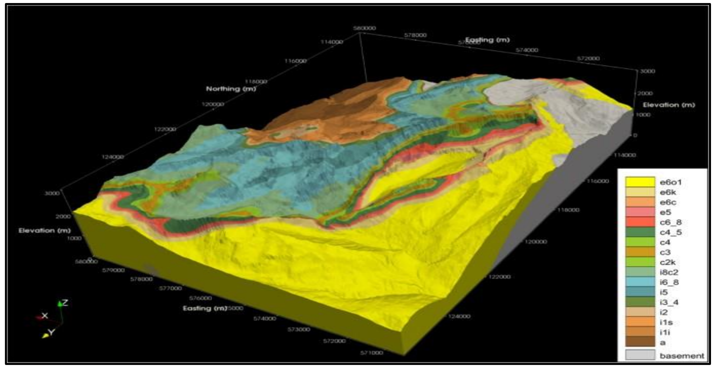
Contents

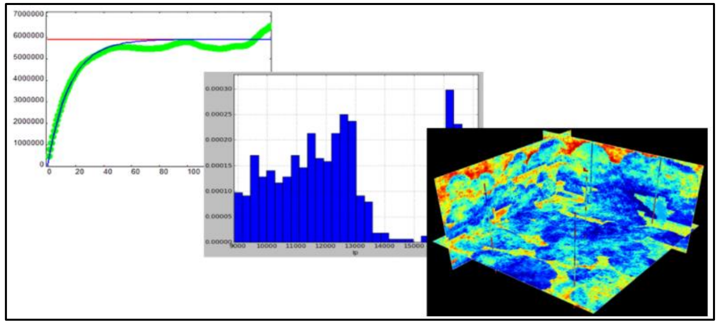
Problem:

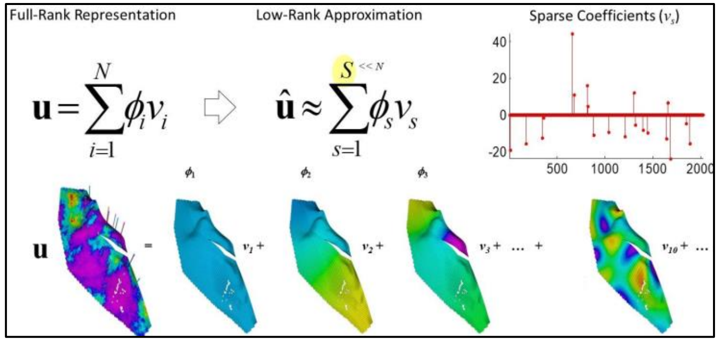
- Generation of **geologically realistic** and **reliable** predictive models, **honoring** both **prior conceptual geological** model and **field measurements** for field development and reservoir management applications

Related subproblems

1. Development of tools enabling generation of **geologically complex and realistic prior** models effectively **capturing uncertainty**
2. Development of **effective low-dimensional parametrization** workflows for computationally efficient **data conditioning**


Geological model (Credit: Nature.com)


Geostatistical workflow (Credit: Serengeo)



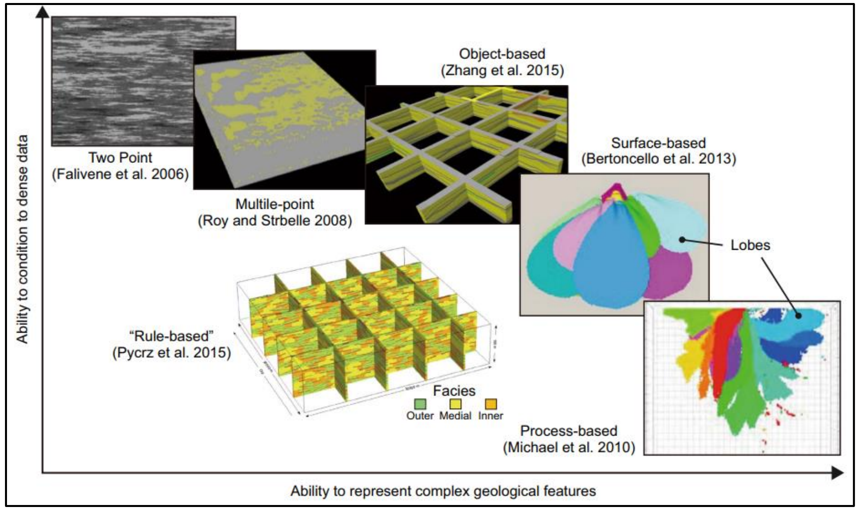
Parametrization (Credit: SEES lab)

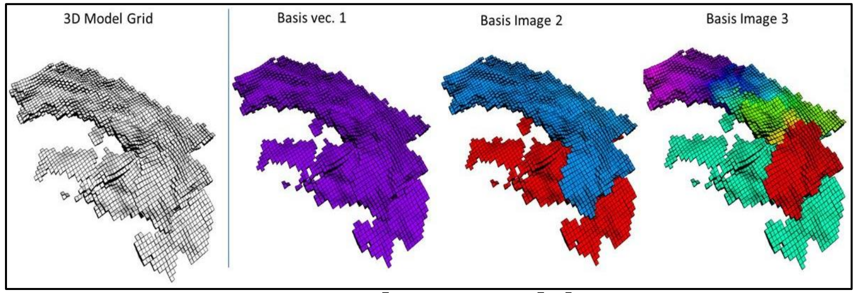
Existing tools and workflows for geological model development and parametrization

- Some tools for **generation of geological models** and their **drawbacks**
- Variogram-based: too simplistic, not geological
- OBM: limited to specific environments, not flexible for data conditioning
- MPS: construction of TI, capturing variability and uncertainty is challenge

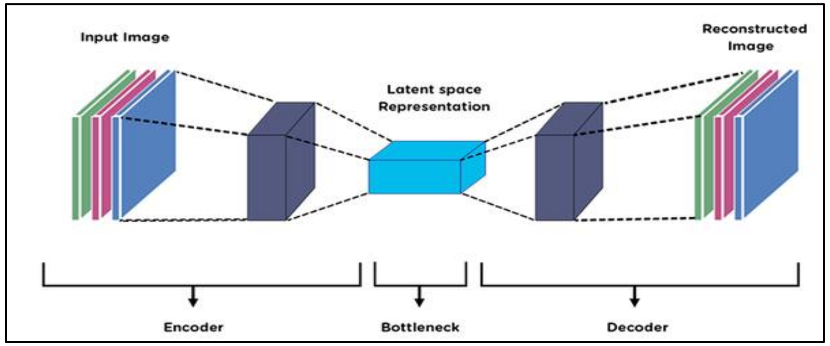
- Some tools for **parametrization** of geological models and their **drawbacks**
- Transform-domain parametrization methods: PCA/Kernel PCA, Fourier, Wavelet: **not very effective reconstruction of complex geology** or a need for inclusion of **relatively large number of parameters** (Datta-Gupta et al., 2016)
- Autoencoders, Variational Autoencoders: **not very crispy images** due to **user defined loss function** (AE, VAE) or **fixed parametrization** (VAE) (Rosca et al., 2017)

- Potential tool for parametrization and (possibly) generation of geological models (Chan et al., 2017)
- GAN





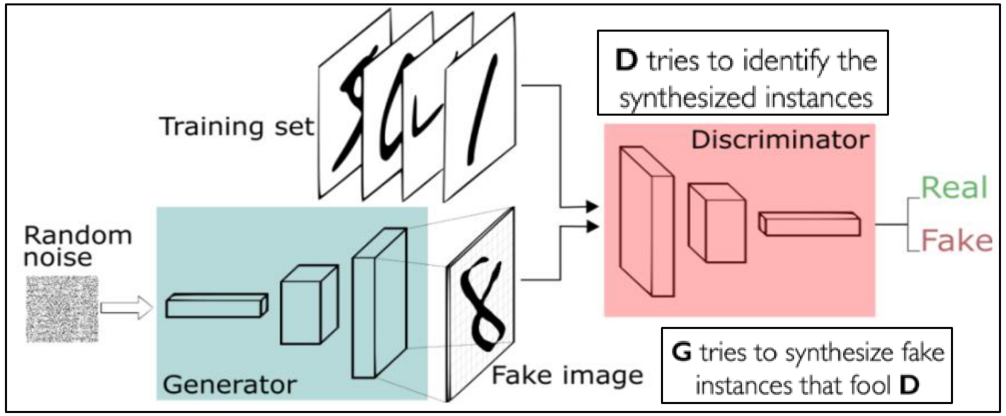
Credit: SEES lab



Credit: Edureka

GAN

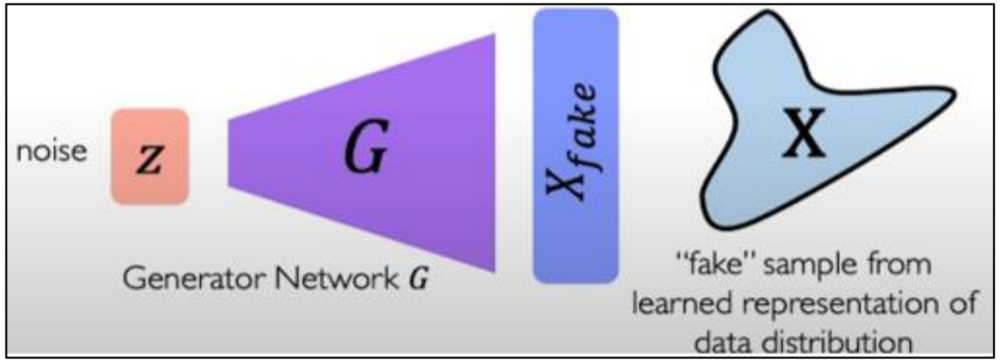
- **Generative Adversarial Networks (GANs)** are a way to make a **generative deep learning model** by having **2 neural networks** (generator: G and discriminator: D) to **compete** with each other
- Idea: if we **just want to sample** from data distribution / generate new data from **similar distribution** as of original one, maybe we **don't need to explicitly** (complex) **model** data PDF



- **Step 1**: **Train** coupled neural networks alternately in **adversarial manner** for some number of iterations

$$\arg\min_G \max_D \mathbb{E}_{\mathbf{z},\mathbf{x}}[\log D(G(\mathbf{z})) + \log(1 - D(\mathbf{x}))]$$

- **Step 2**: Sample from **latent vector z** and pass it to **generator G** and obtain new data samples **x=G(z)**



Credit: MIT 6.S191

Advantages and disadvantages of GANs

Advantages

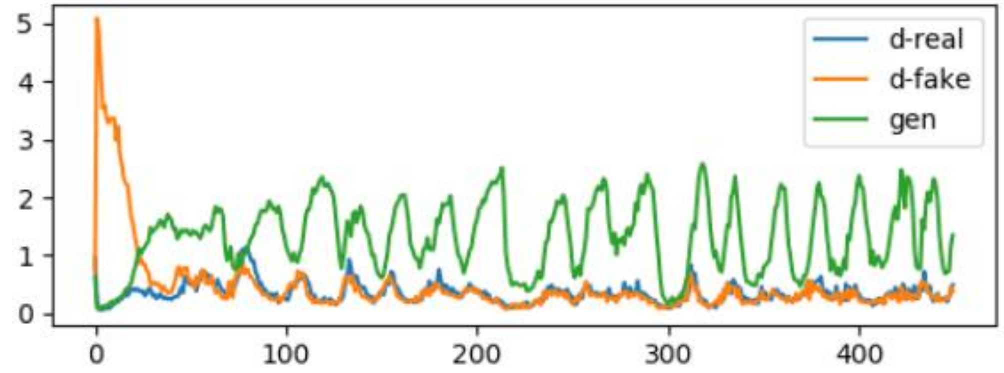
- **Highest resolution image generation** among available generative models
- **Fast inference** (fast generation of new data)

Disadvantages

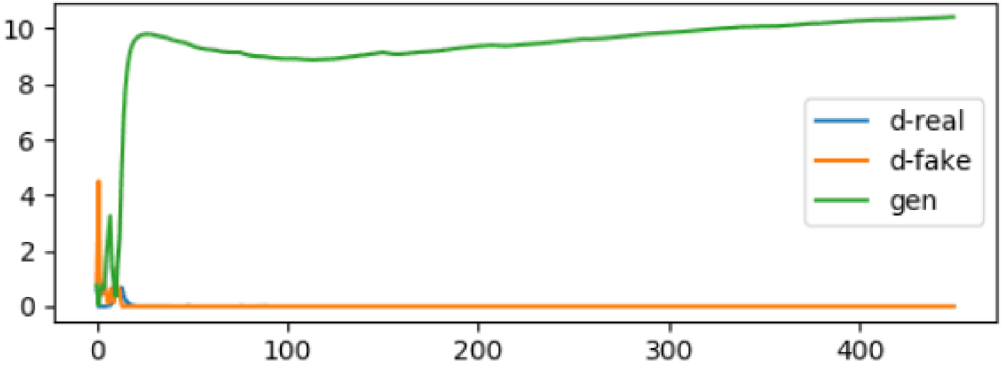
- **Unstable training** (open research problem): non-convergence, mode collapse
- **Lack of intrinsic evaluation metrics** (mainly by visual inspection): no explicit metric to optimize or compare generated and true images
- **No density estimation** (could be needed for anomaly detection like problems): formulation of problem is different compared to other generative models

GAN training

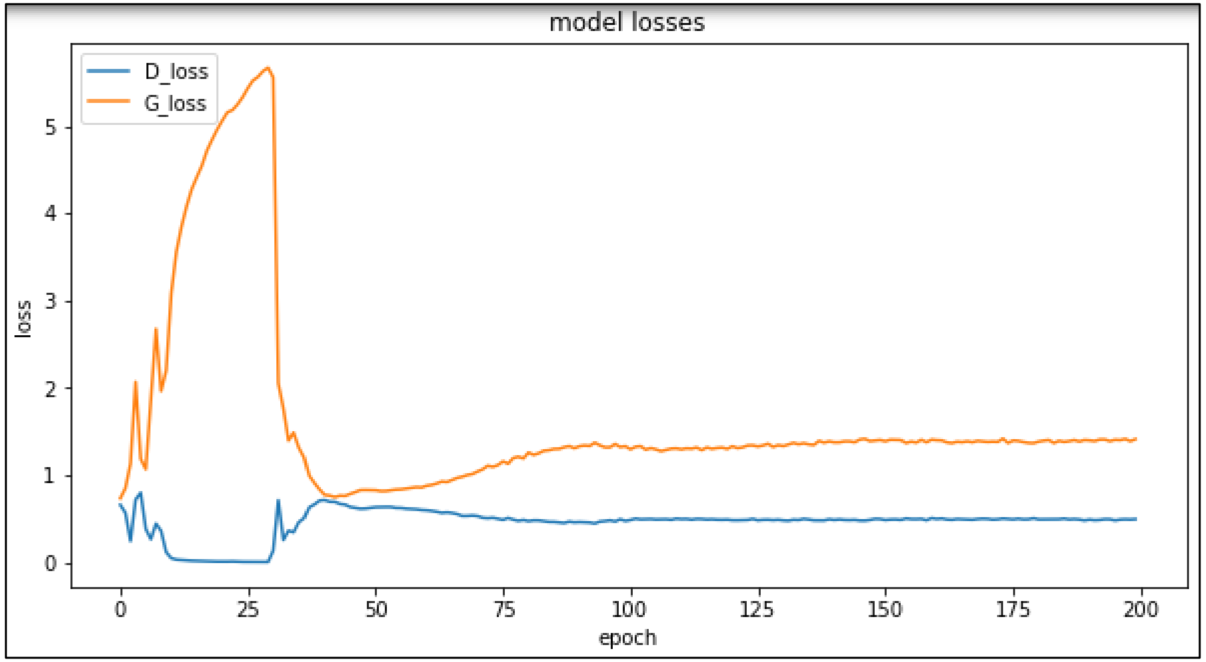
- **2 possible problems**: 1) mode collapse 2) convergence failure
- **Mode collapse**: a problem that occurs when the generator learns to map several different input z values to the same output point (NIPS 2016 Tutorial: Generative Adversarial Networks, 2016.
- **Convergence failure**: not finding an equilibrium between the discriminator and the generator.



Losses for a GAN with mode collapse (Credit: Machine Learning mastery)



Losses for a GAN with convergence failure (Credit: Machine Learning mastery)



Losses for a GAN used in this workflow

Workflow

Geological facies model definition:
470x470 binary channel TI

Training data generation:
Randomly crop 10 000 64x64 images from TI

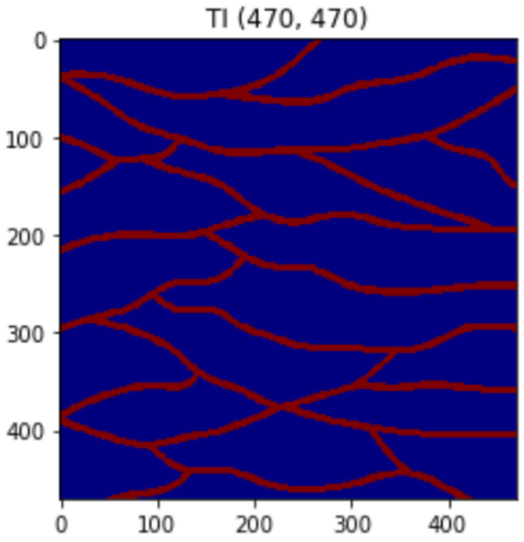
GAN training (base cases)

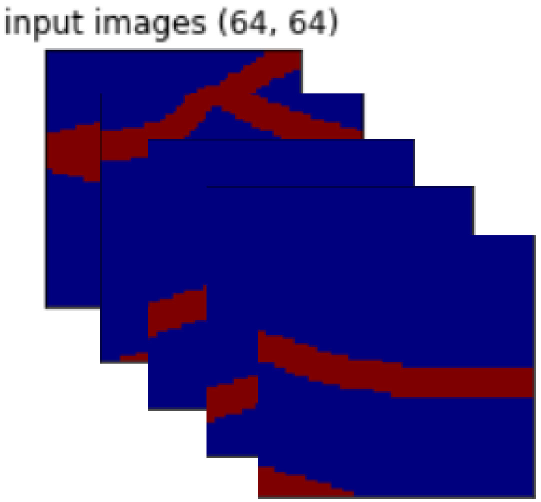
GAN training (alternative case)

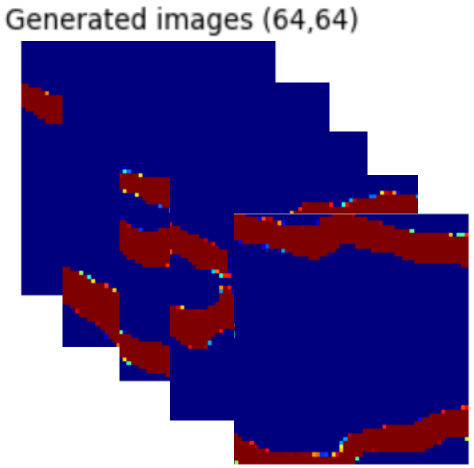
Generation of new images using GAN

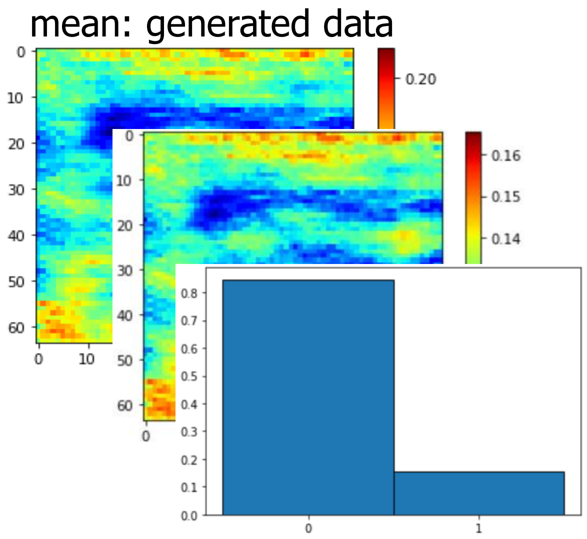
Evaluation of GAN-generated images and comparison with VAE and PCA-generated images

Workflow components



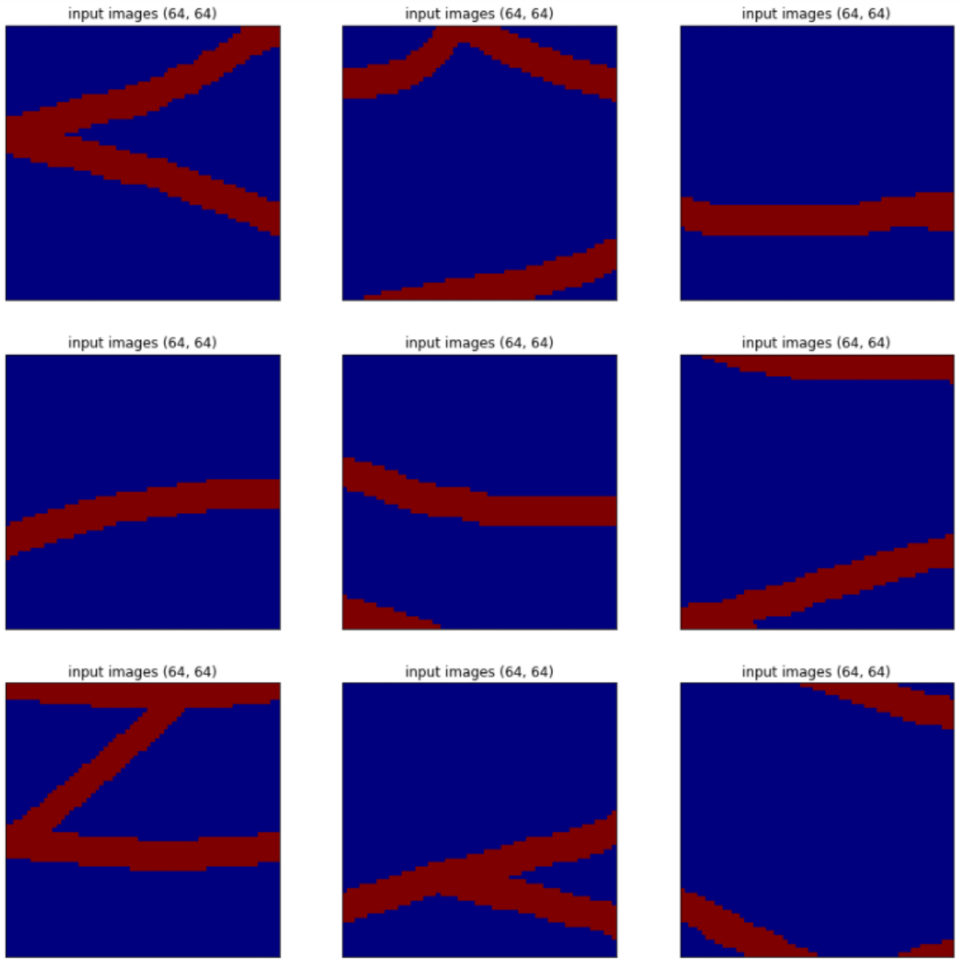




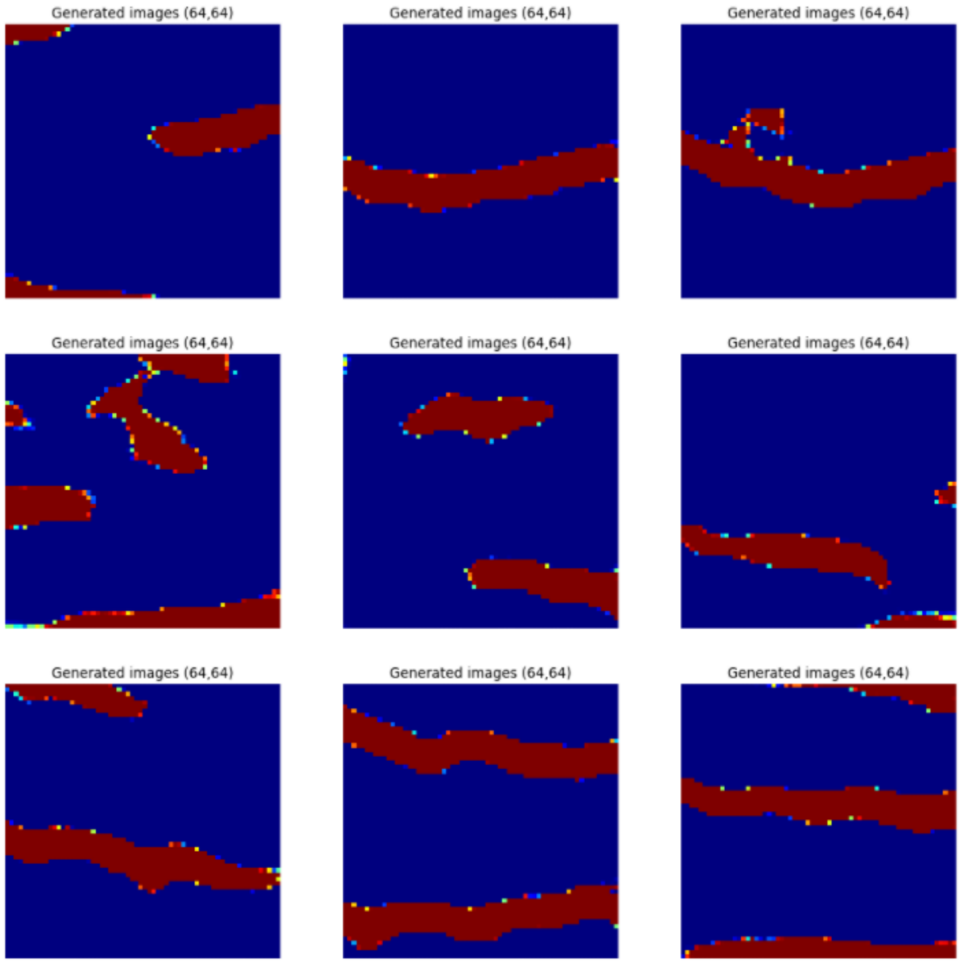


Experiment: base case results

Training data

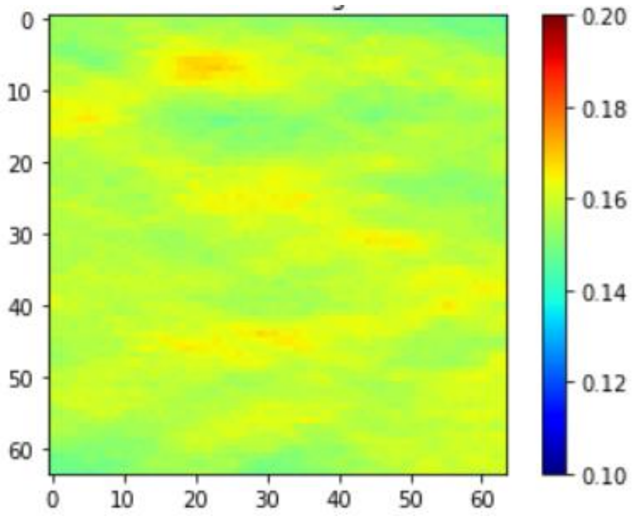


GAN-generated images

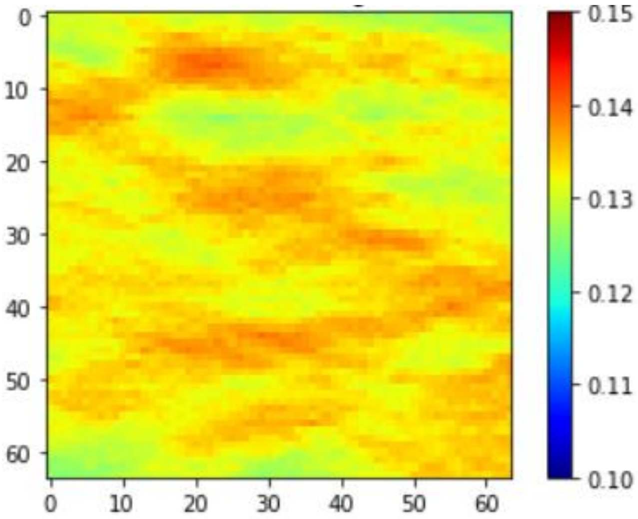


Evaluation of results / QC

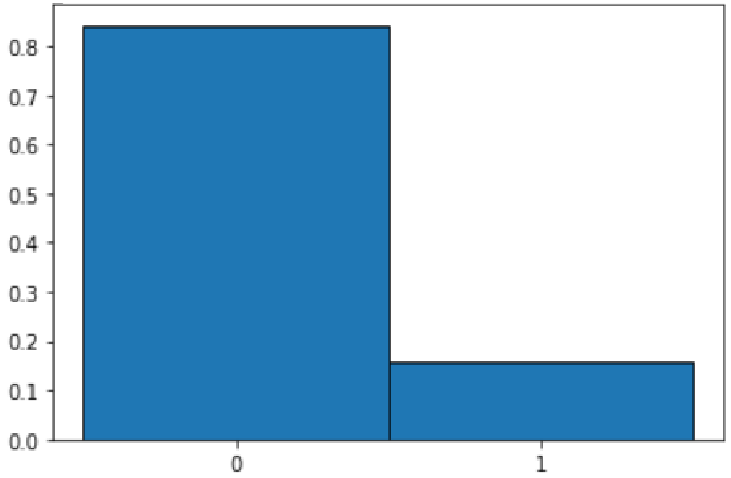
mean: training data



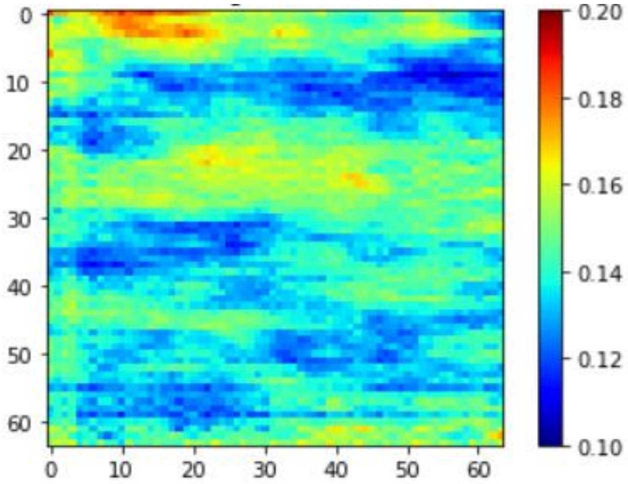
var: training data



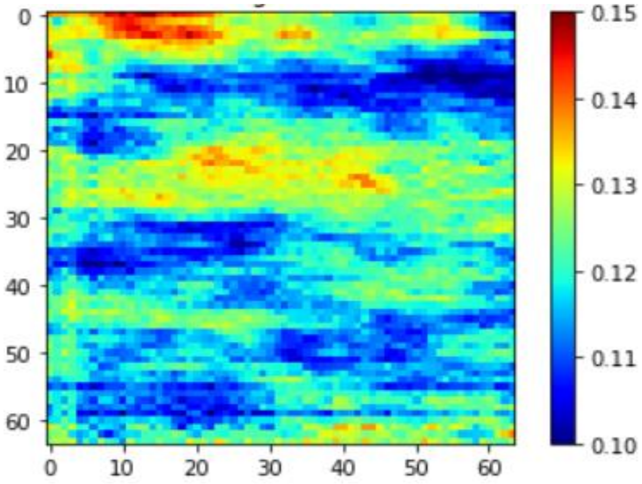
histogram: training data



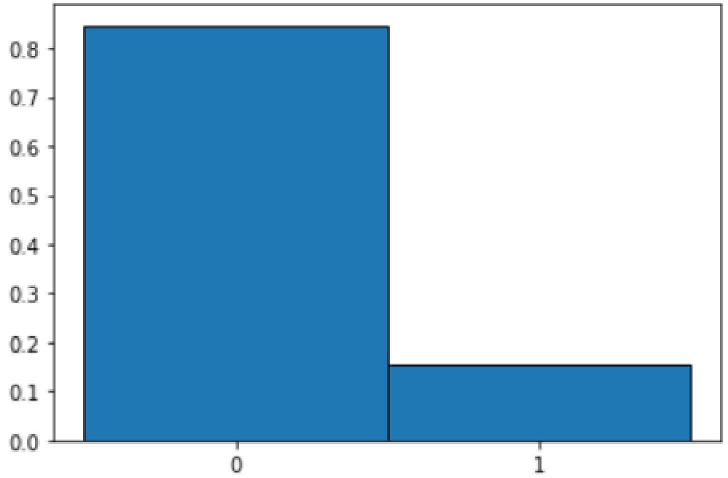
mean: generated data



var: generated data

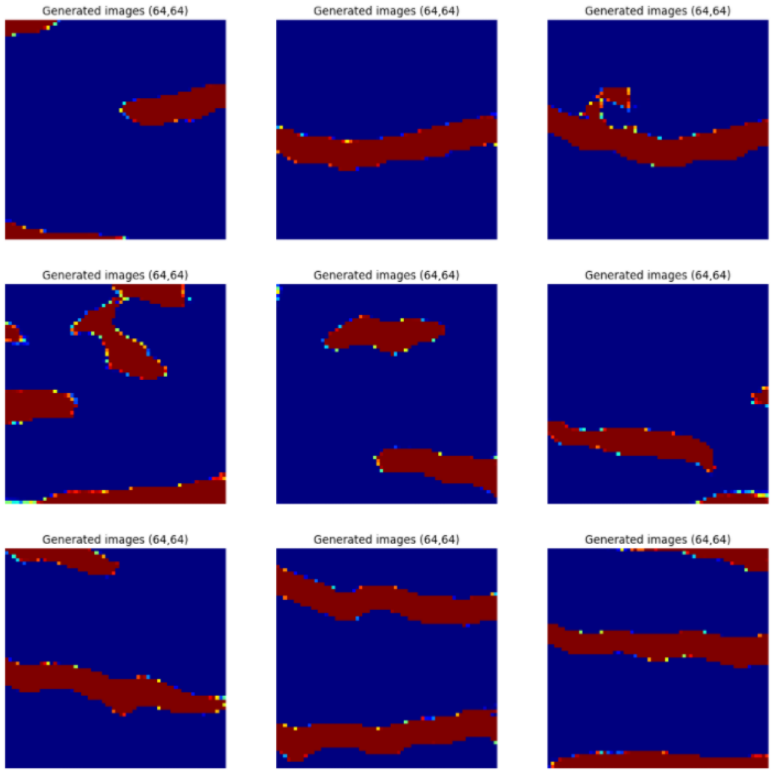


histogram: generated data

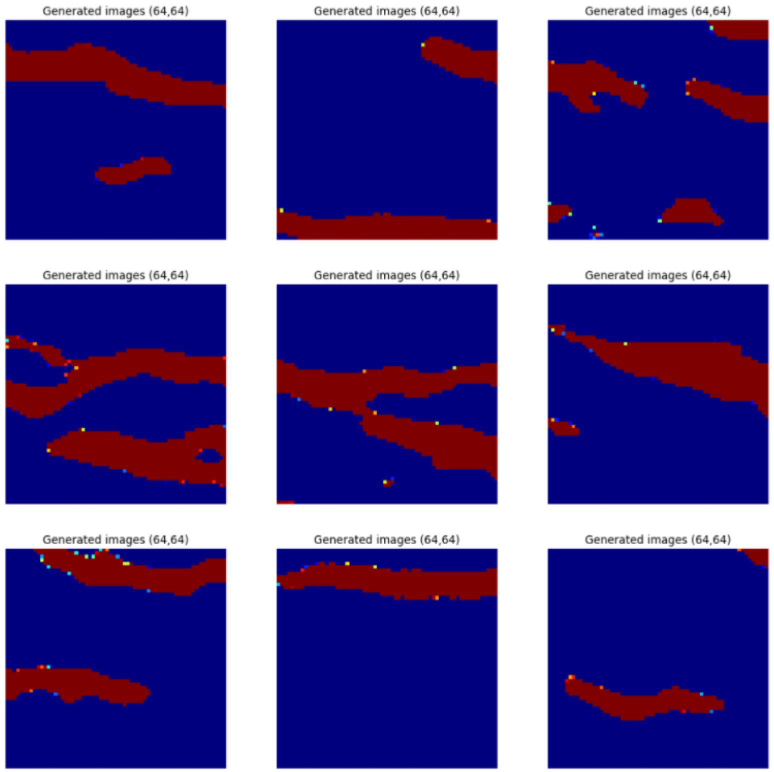


Sensitivity to gaussian noise vector dimension

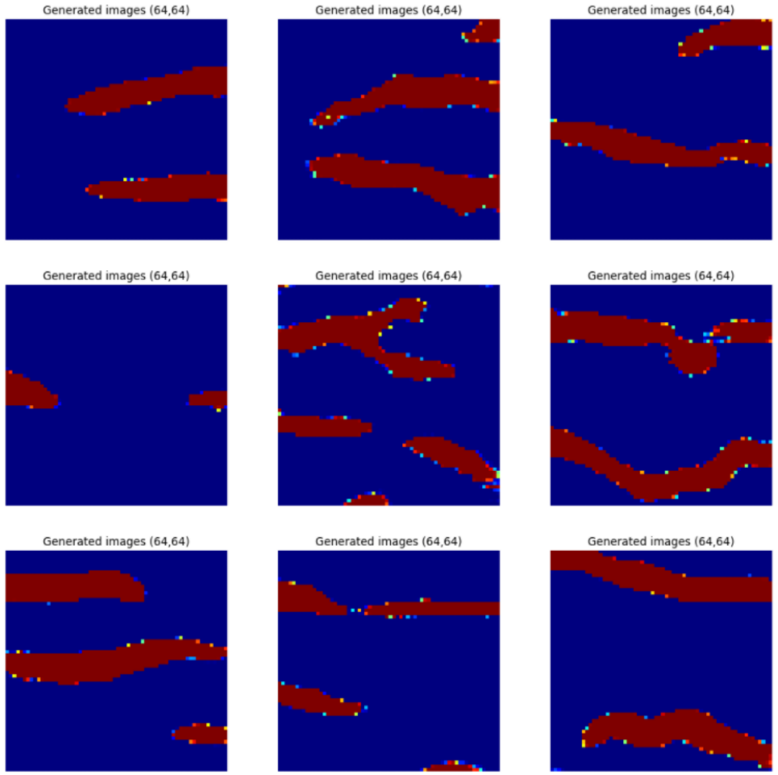
z=500



z=100

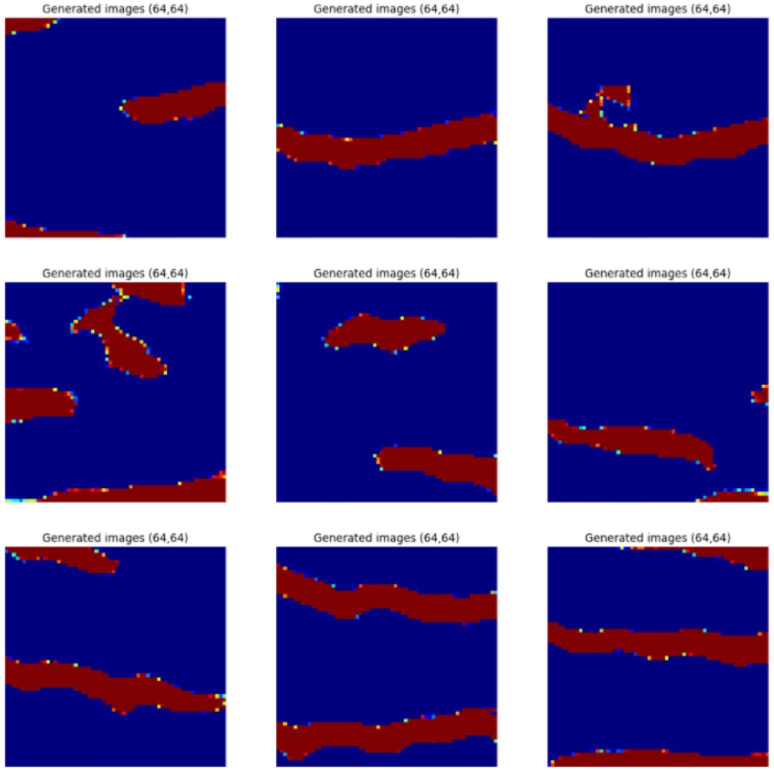


z=1000

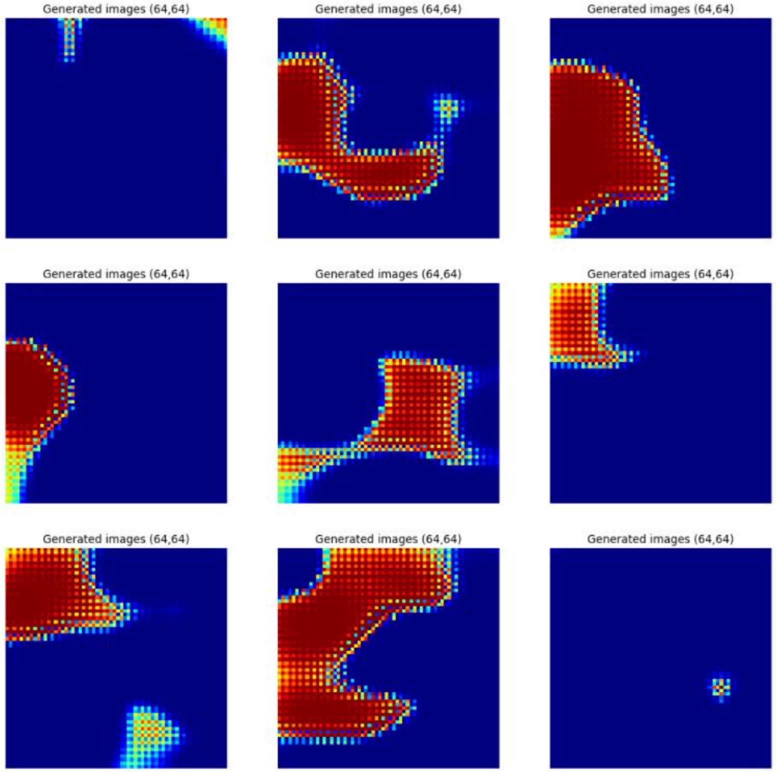


Sensitivity to size of training data

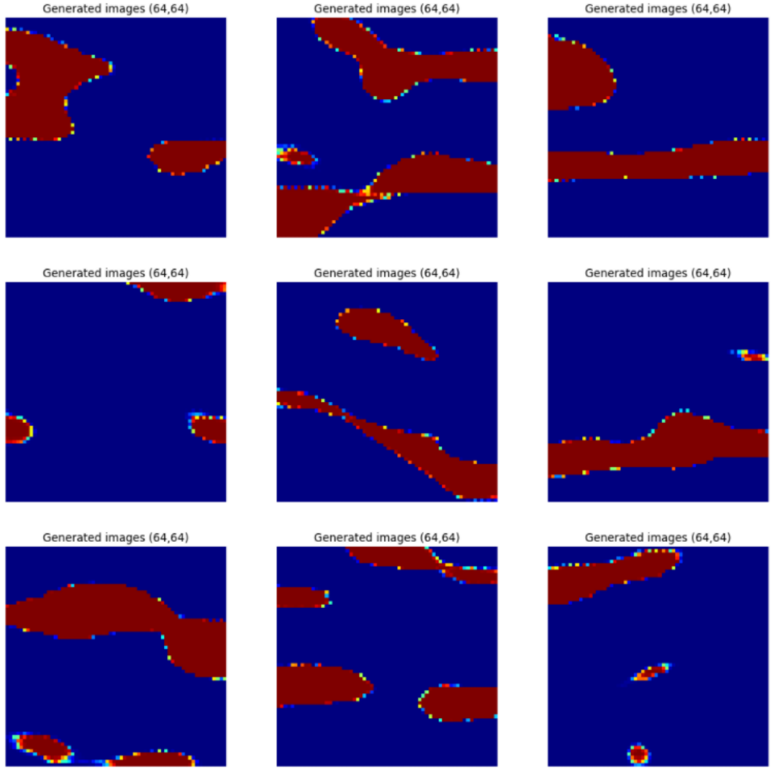
N=10 000



N=1000

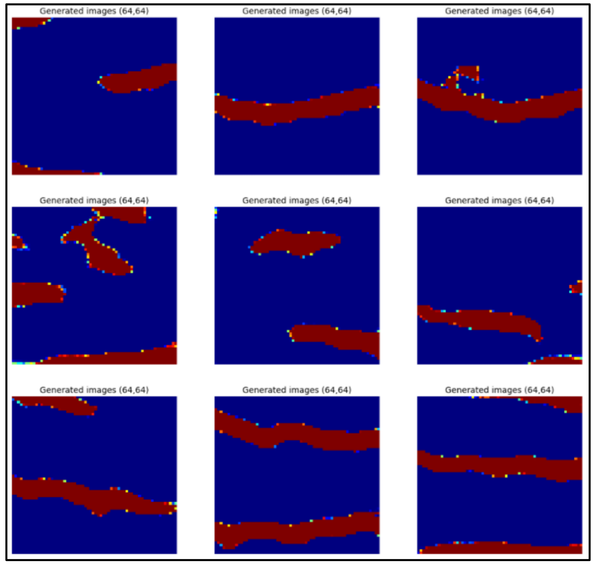


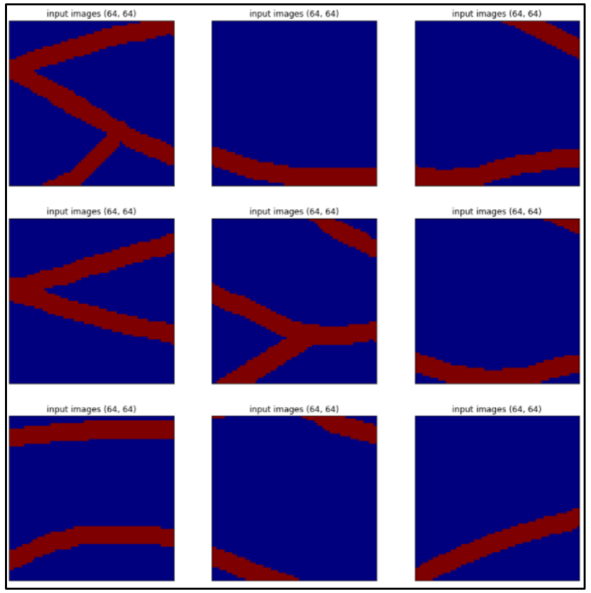
N=5000



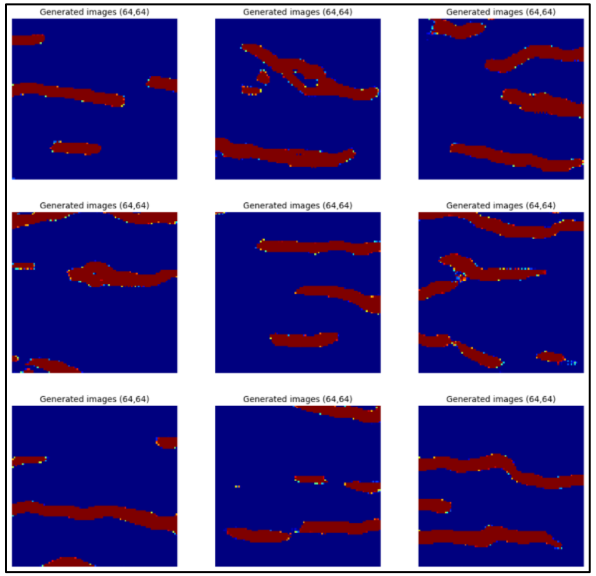
Sensitivity to resolution of training data

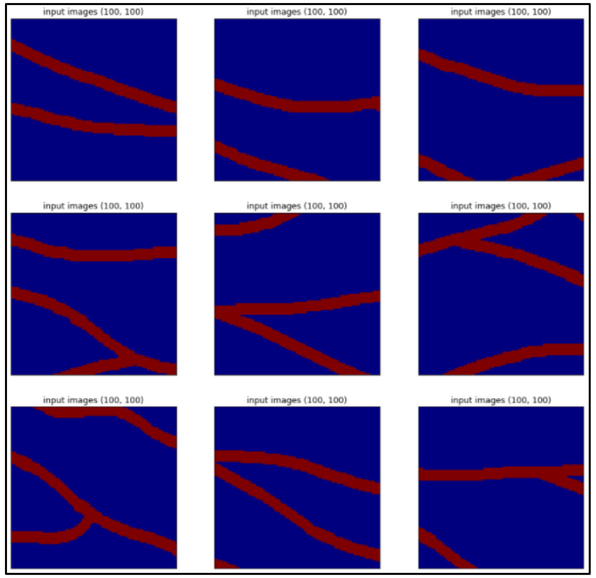
64x64



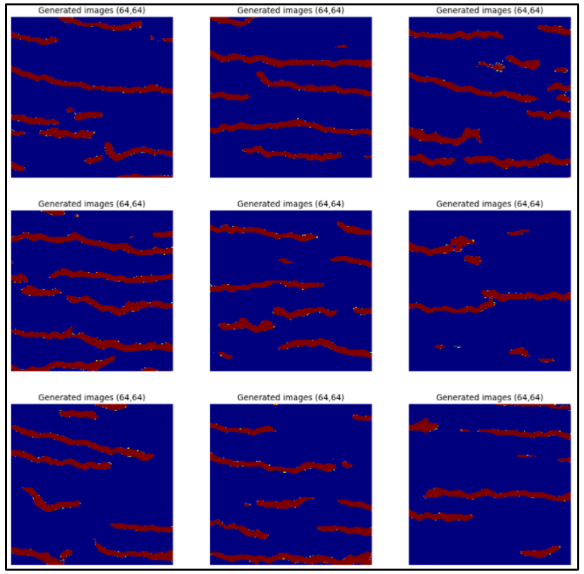


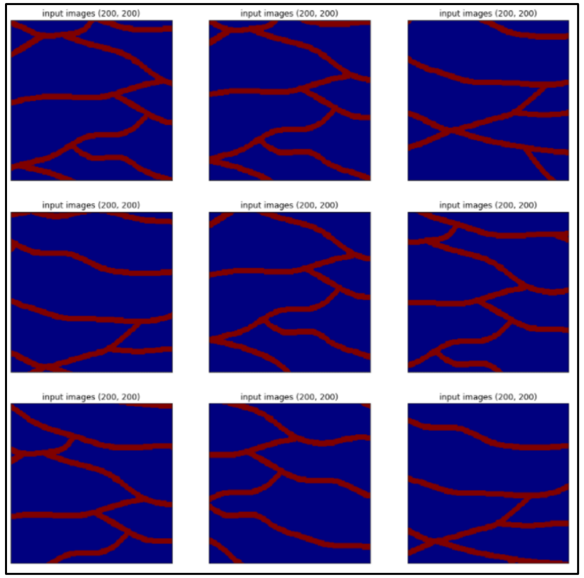
100x100





200x200

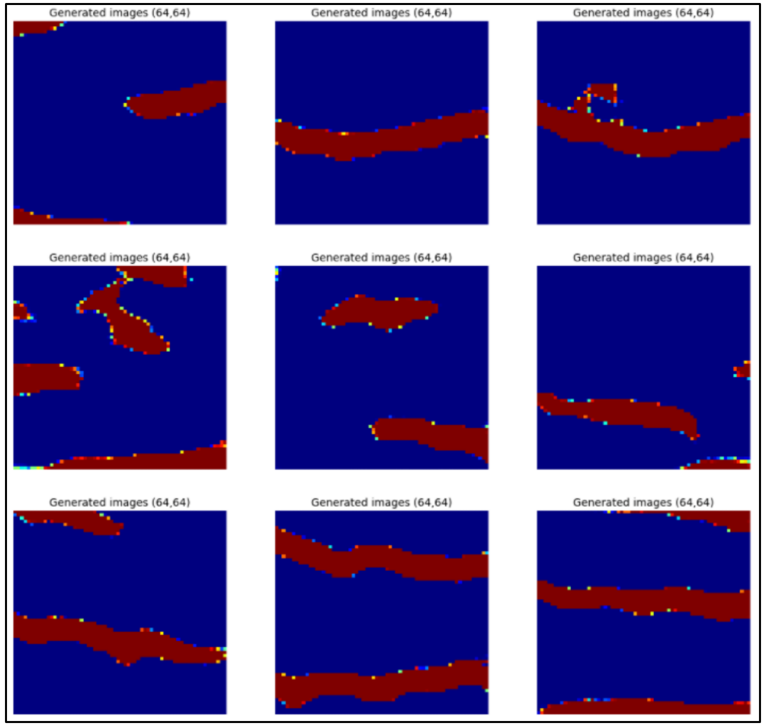




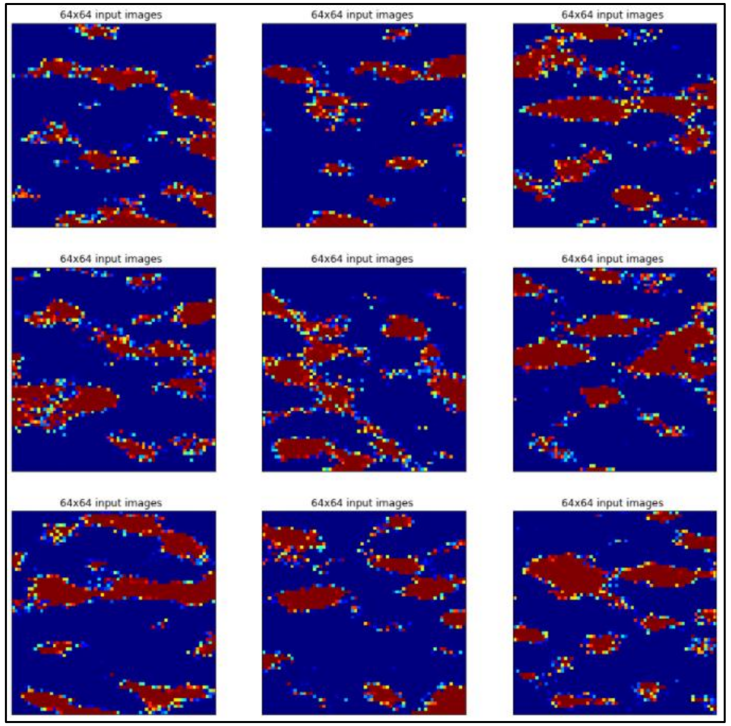
GAN vs VAE vs PCA

- Setup: 10 000 64x64 training images
- GAN: z=500; VAE: z=500 PCA: 75% variance retained (37 components)

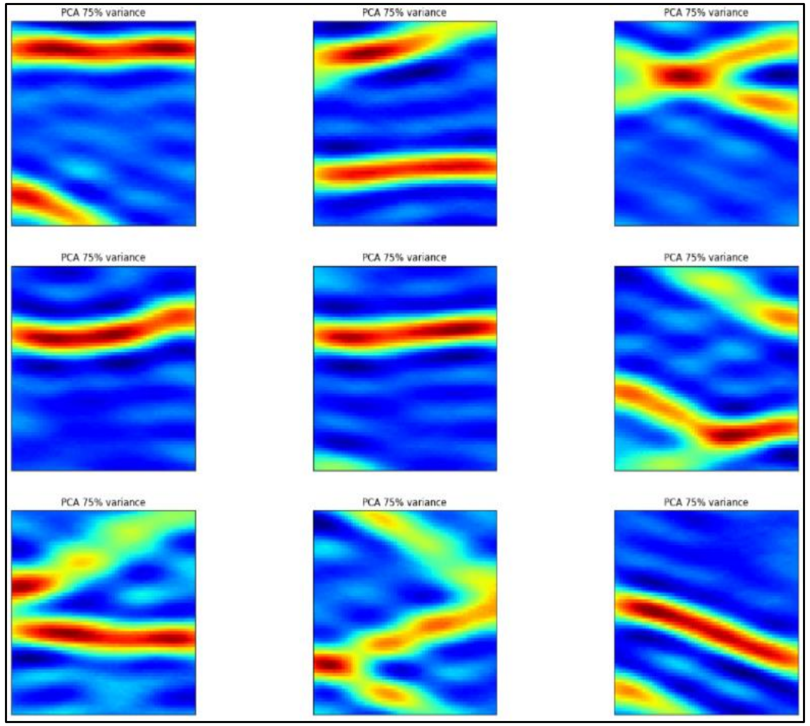
GAN



VAE



PCA



Conclusions and further work (Tuesday)

Conclusions

- GAN-based workflow for unconditional geological model generation and parametrization has been introduced
- GAN has been able to reproduce the data histogram and channel patterns from TI, but there are some artifacts introduced, and spatial distribution of channels in generated samples does not match the ones in training data
- Overall, GAN makes it possible to parametrize the geological model with low number of parameters and generate very fast moderately realistic geological models
- Results from workflow are superior to the ones generated based using VAEs and PCA
- Main challenges/issues in the workflow are GAN-training complexity and effective data generation/sampling from TI

Future work

- Test different data generation workflows: 1) feed OBM based models directly into GAN 2) sample differently from TI (sampling method, window size, multigrid) 3) use multiple TIs to capture more uncertainty
- More systematic and thorough training of GAN: different loss function formulations, different architecture configurations
- Conditional GAN models
- GAN-based inversion workflow

References

- Chan, S. and Elsheikh, A.H., 2017. Parametrization and generation of geological models with generative adversarial networks. arXiv preprint arXiv:1708.01810.
- Han, Jichao , Chen, Rongqiang , and Akhil Datta-Gupta. "Multiscale Method for History Matching Channelized Reservoirs Using Level Sets." Paper presented at the SPE Annual Technical Conference and Exhibition, Dubai, UAE, September 2016
- Rosca, M., Lakshminarayanan, B., Warde-Farley, D. and Mohamed, S., 2017. Variational approaches for auto-encoding generative adversarial networks. arXiv preprint arXiv:1706.04987.
- Zhang, L.F., Pan, M. and Li, Z.L., 2020. 3D modeling of deepwater turbidite lobes: a review of the research status and progress. Petroleum Science, pp.1-17.

Questions?

USC Viterbi School of Engineering