Encoding large scale cosmological structure with Generative Adversarial Networks

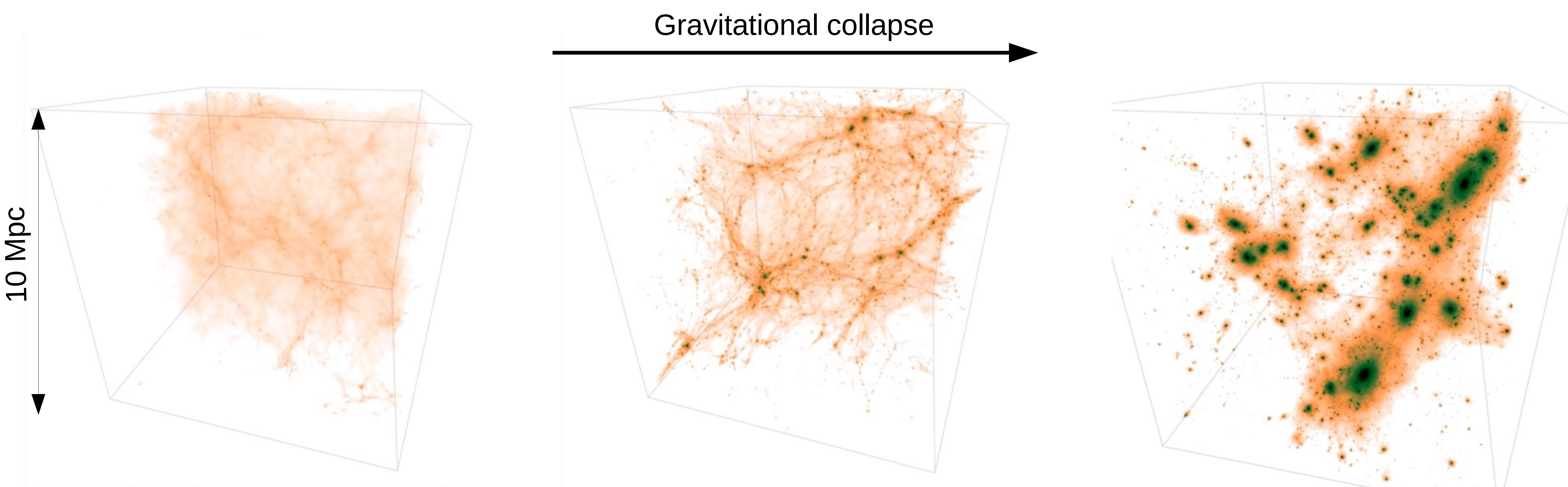
Marion Ullmo (IAS)
Nabila Aghanim (IAS) – Aurélien Decelle (LRI, UCM)

The Cosmic Web and Large Scale Structures

Initial Conditions :
quasi-homogeneous
matter distribution

Structures form through
gravitational collapse

Current state :
virialized overdense
halos have formed



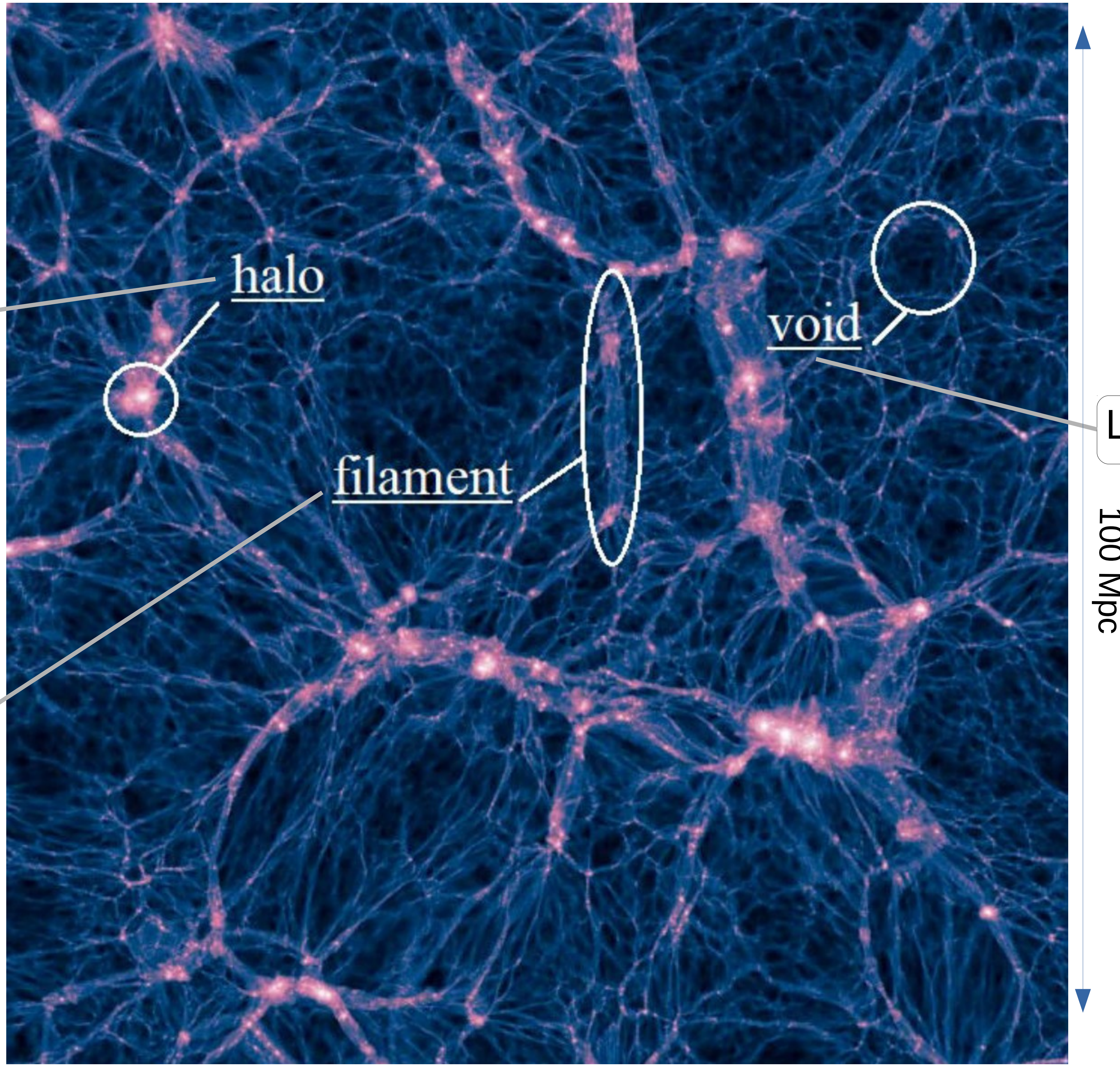
Redshift:	z =10	z=5	z=0
Time since Big Bang:	0.5 billion years	0.9 billion years	13.8 billion years

Illustris Simulations - (10 Mpc)^3 snapshot   Vogelsberger *et al*, 2014

The Cosmic Web and Large Scale Structures

Dense region: contains clusters, massive galaxies

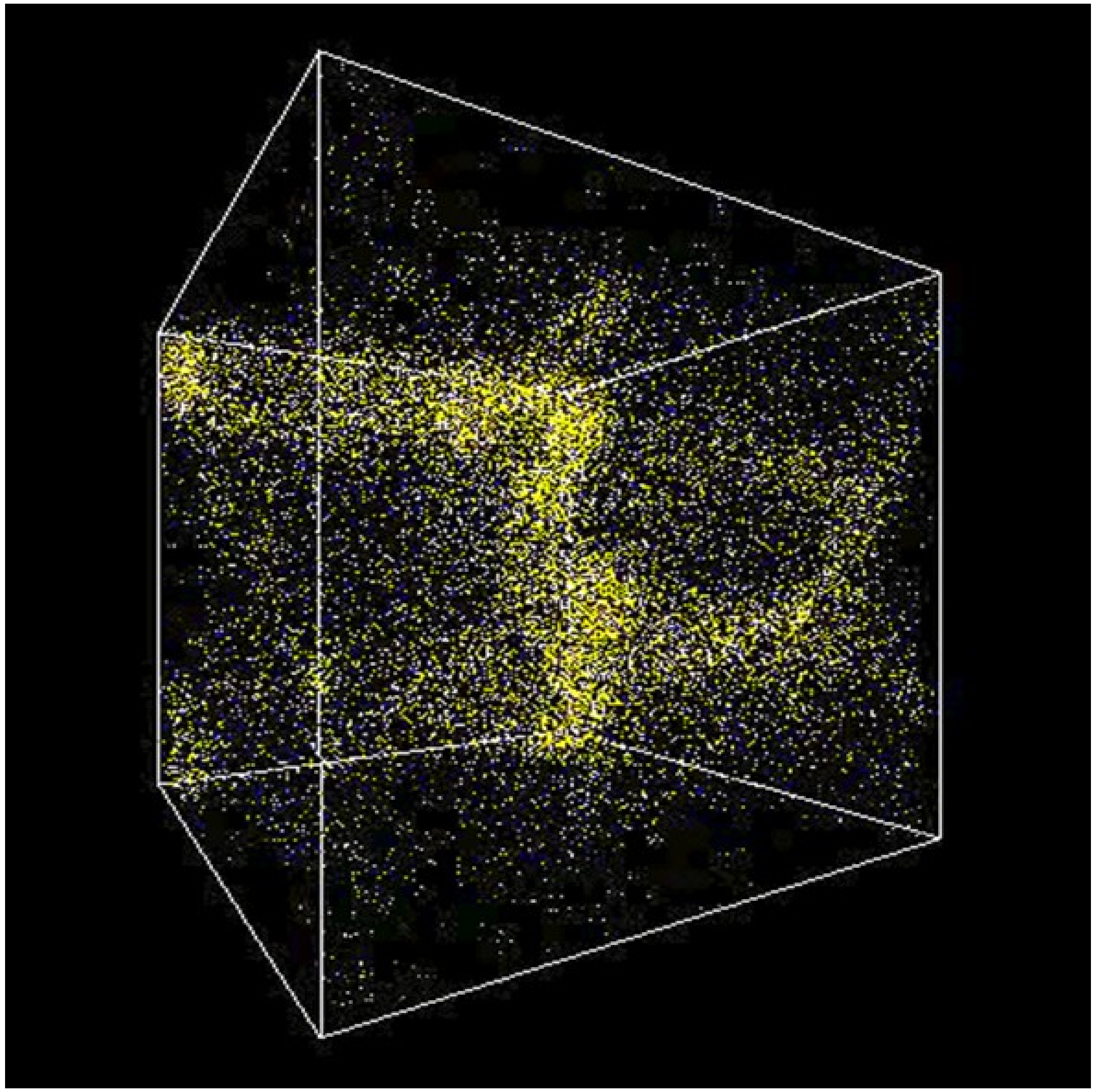
Elongated structure: contains galaxies & diffuse gas



Large, empty region

Illustris simulations – (100Mpc)² - Vogelsberger *et al*, 2014

Simulations - a costly necessity



- Simulations →an essential tool to compute the non-linear structuration of matter
- Typically N-body simulations with 10^6-10^{10} particles

  A few examples:

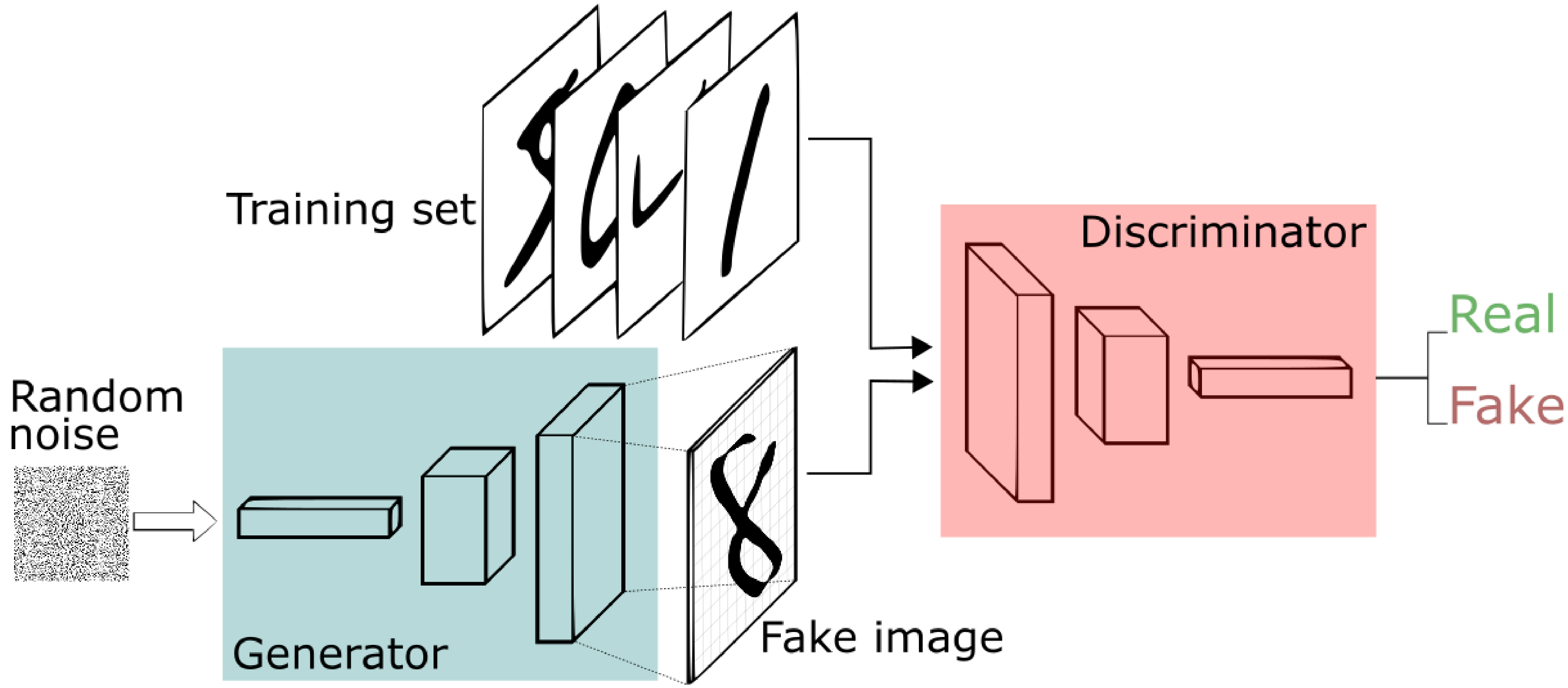
  - Gravitation only: Millenium, 250 000 CPU hours, 28 days runtime
  - Hydrodynamical: Illustris, 19 million CPU hours, 3 months runtime
- Tradeoff between large structures and fine detail

Goals

- We build and train a network for fast generation of simulation-like datasets

- We make use of this trained network to construct a simple autoencoder (AE) as a first step towards building a predictive model

GANs* in a nutshell

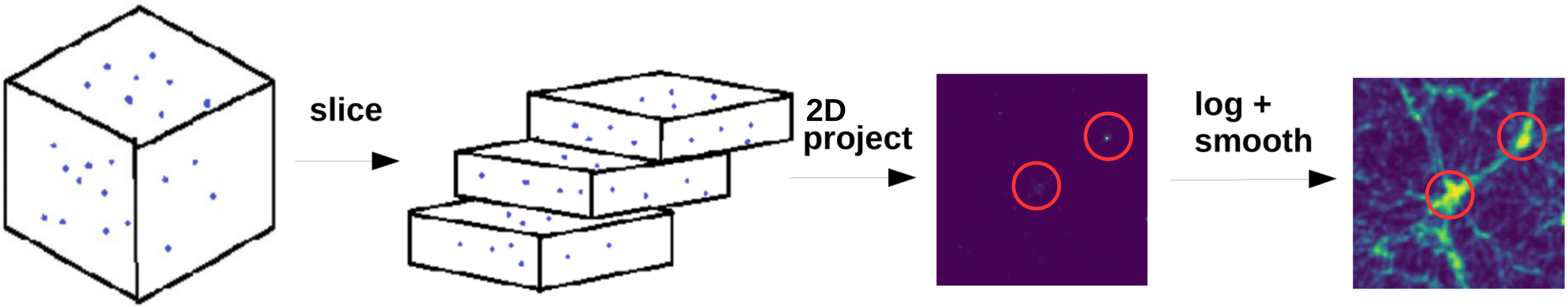


- GAN : Generative Adversarial Network
- Two competing networks :
  - the generator, generates new images
  - the discriminator, outputs the certainty (0 to 1) with which it believes an image is from the training set (rather than from the generator)
- A simple loss function! $L_D = -\frac{1}{2}\mathbb{E}(logD(I_R) + log(1 - D(I_G)))$

* *Goodfellow, 2014*

Datasets

2 types of datasets : 3D & 2D simulations



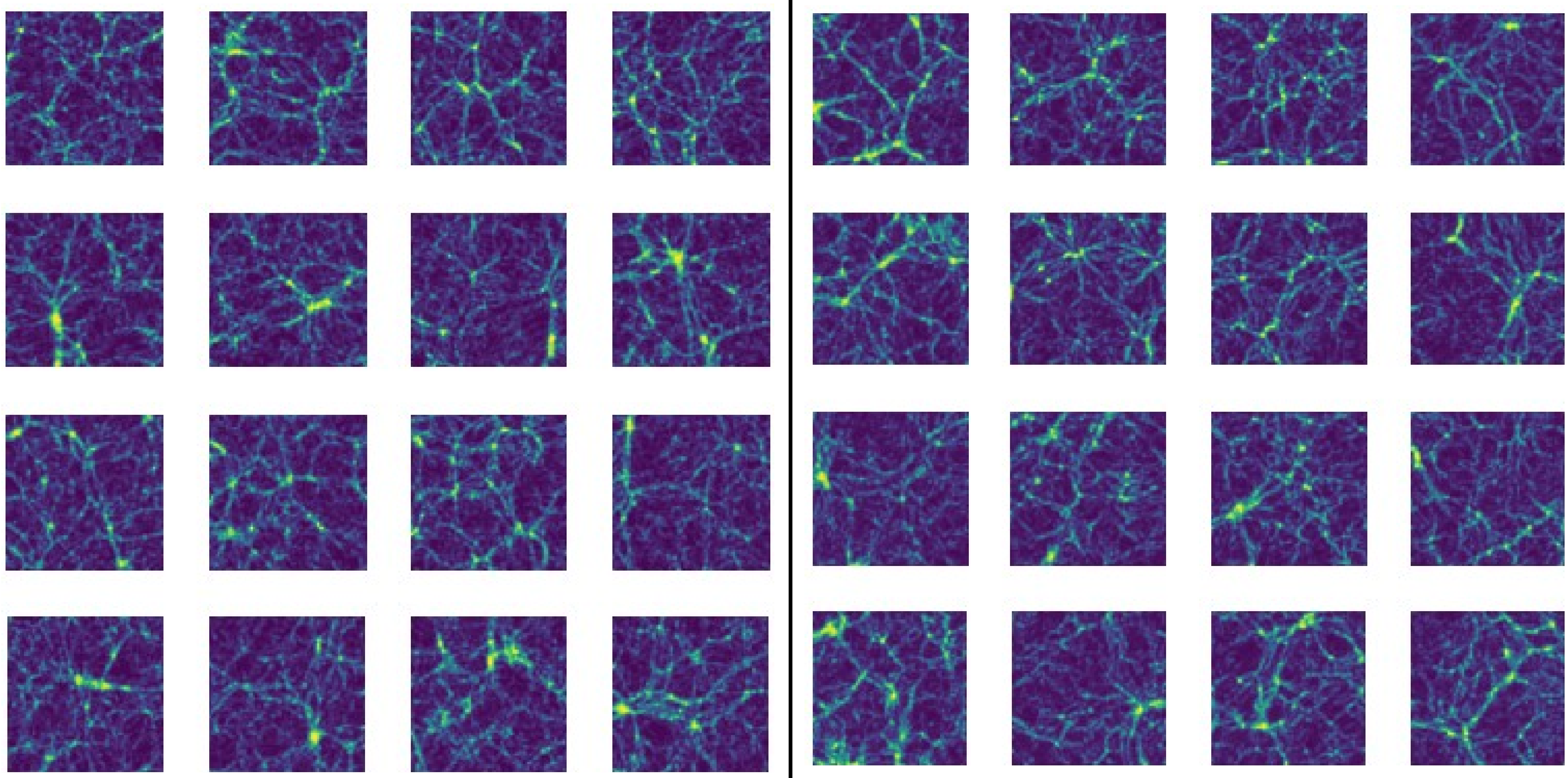
- 3D simulations :
  - Run with GADGET2
  - Box size (100 Mpc)^3, (512)^3 particles
  - Snapshot is divided into slices, from which we estimate a set of 2D log-density maps (76 000 images, size (50 Mpc)^2, 128^2 pixels) after augmentation

- 2D simulations :
  - Run with a 2D particle-mesh N-body code *
  - 1000 simulations, size (100 Mpc)^2 , 512^2 particles each
  - From them, estimate 2D log-density maps
    (76 000 images, size (50 Mpc)^2, 128^2 pixels) after augmentation

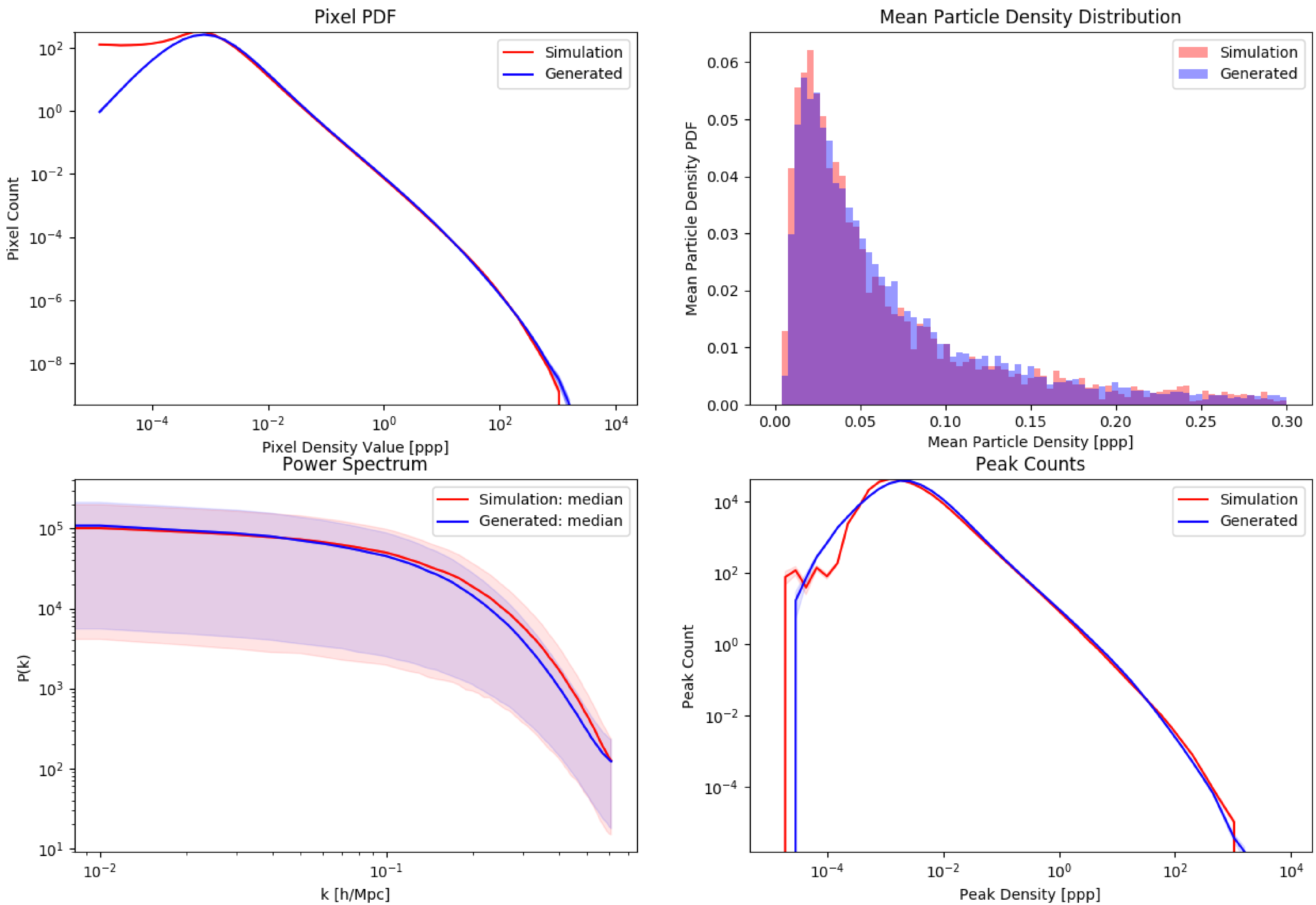
→ We show GAN results for 3D simulations

Results – log density images from 3D simulation slices

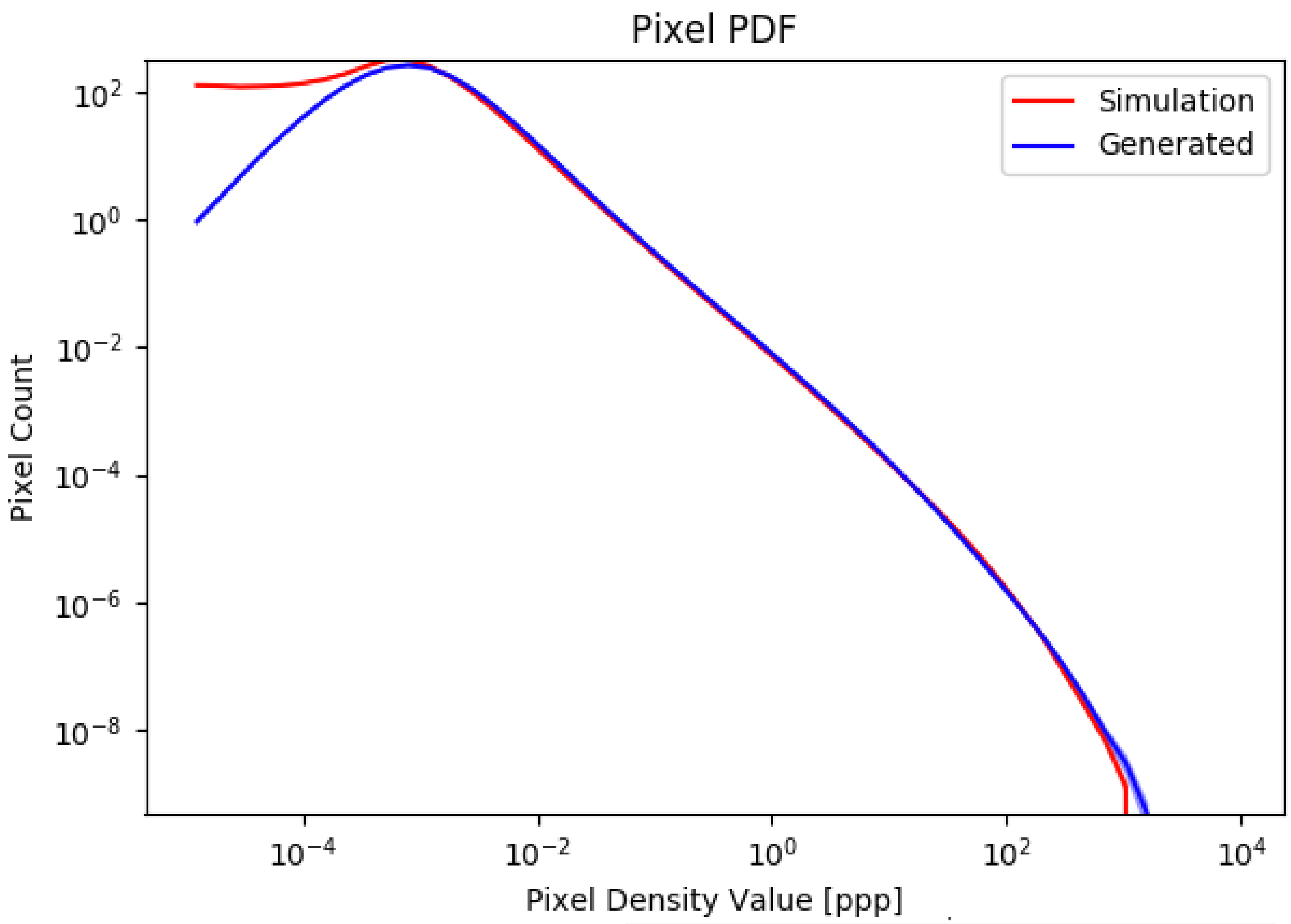
Simulation images ("true") | GAN images ("fake")

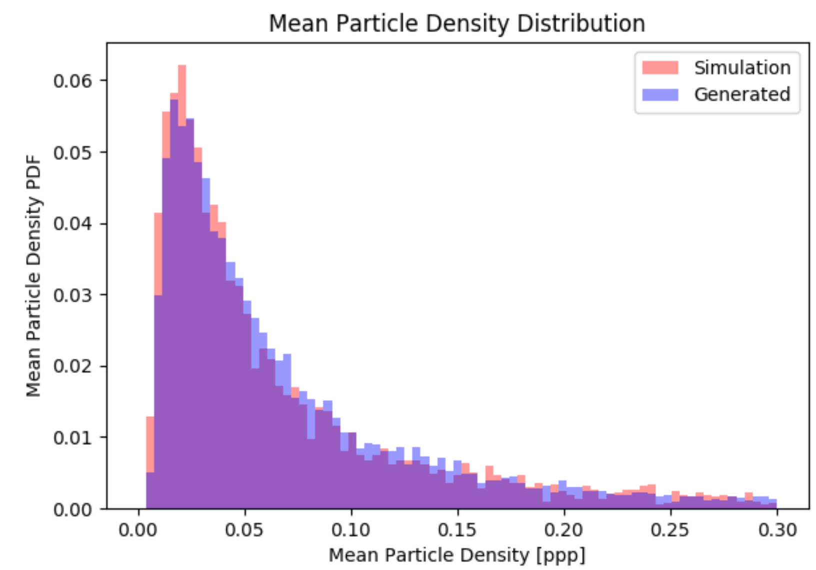


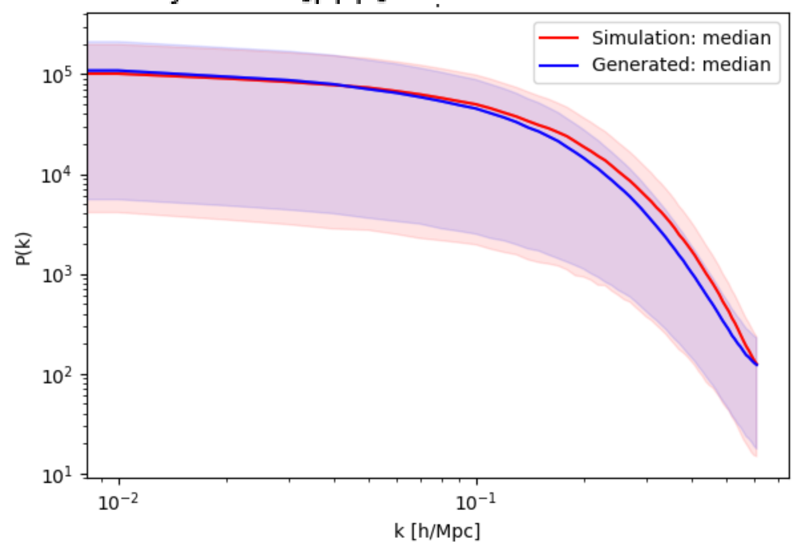
Comparing simulated images & generated images

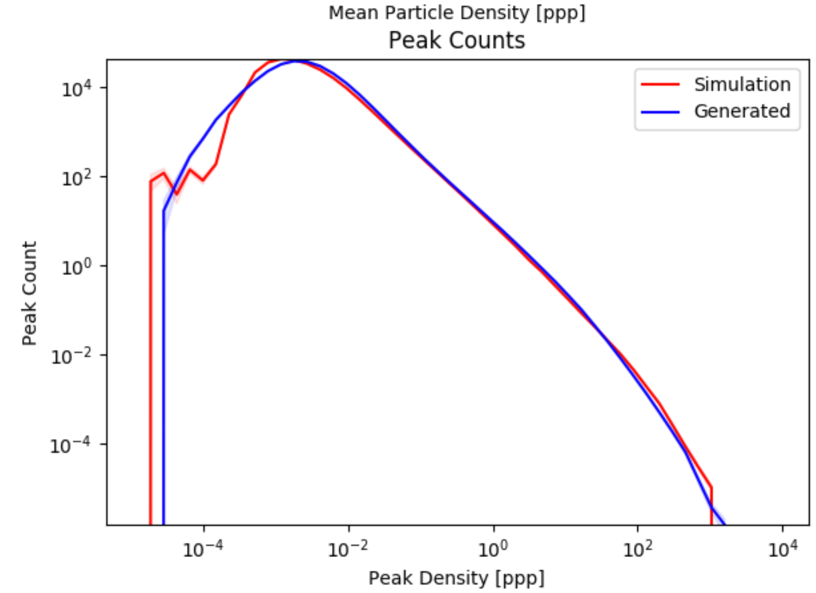


Comparing simulated images & generated images

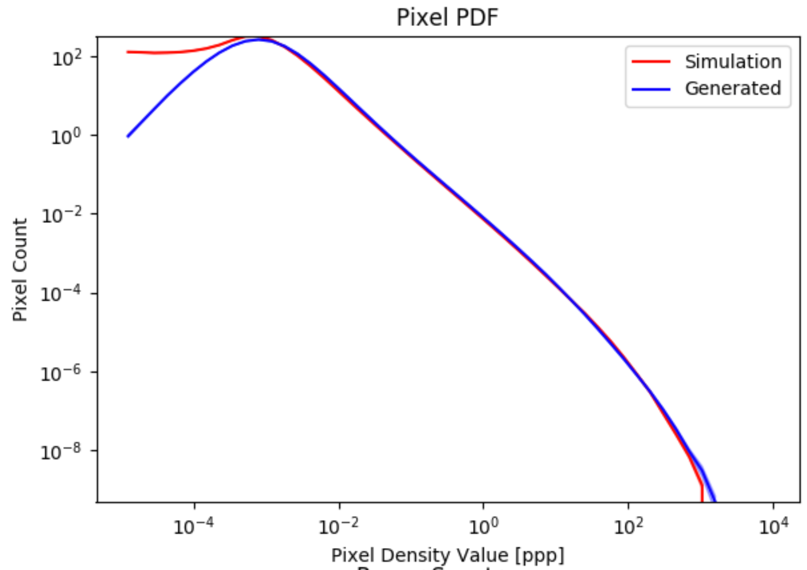


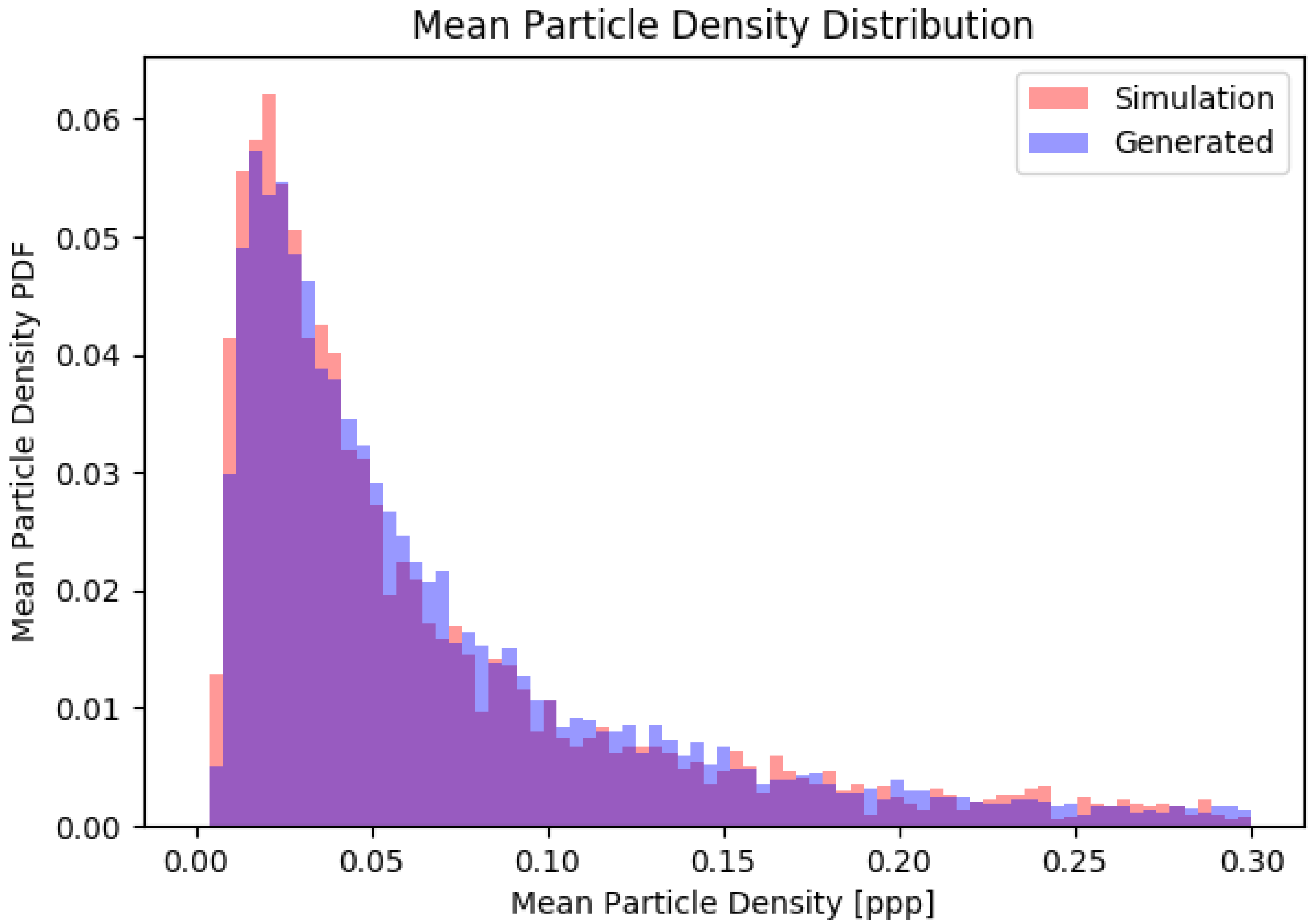


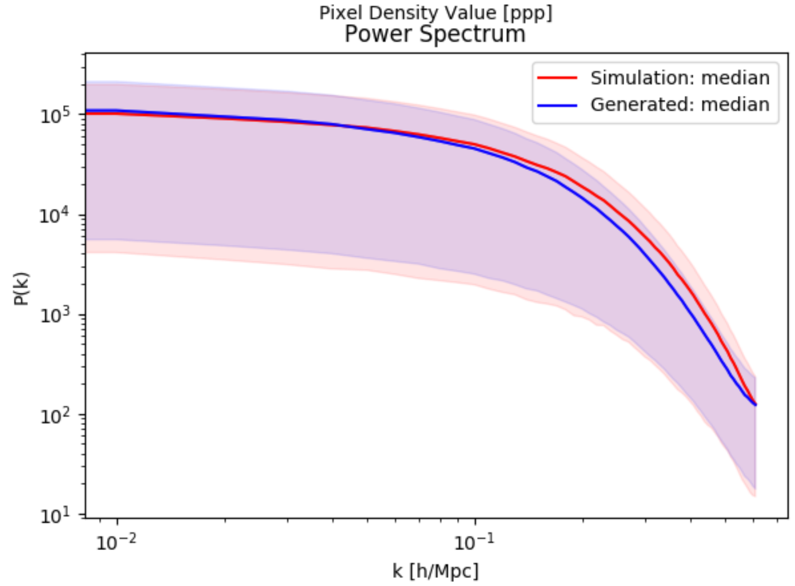


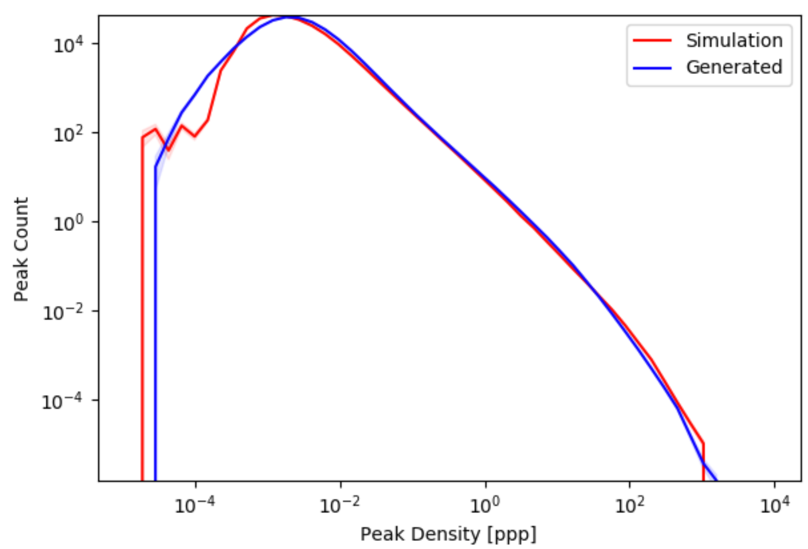


Comparing simulated images & generated images

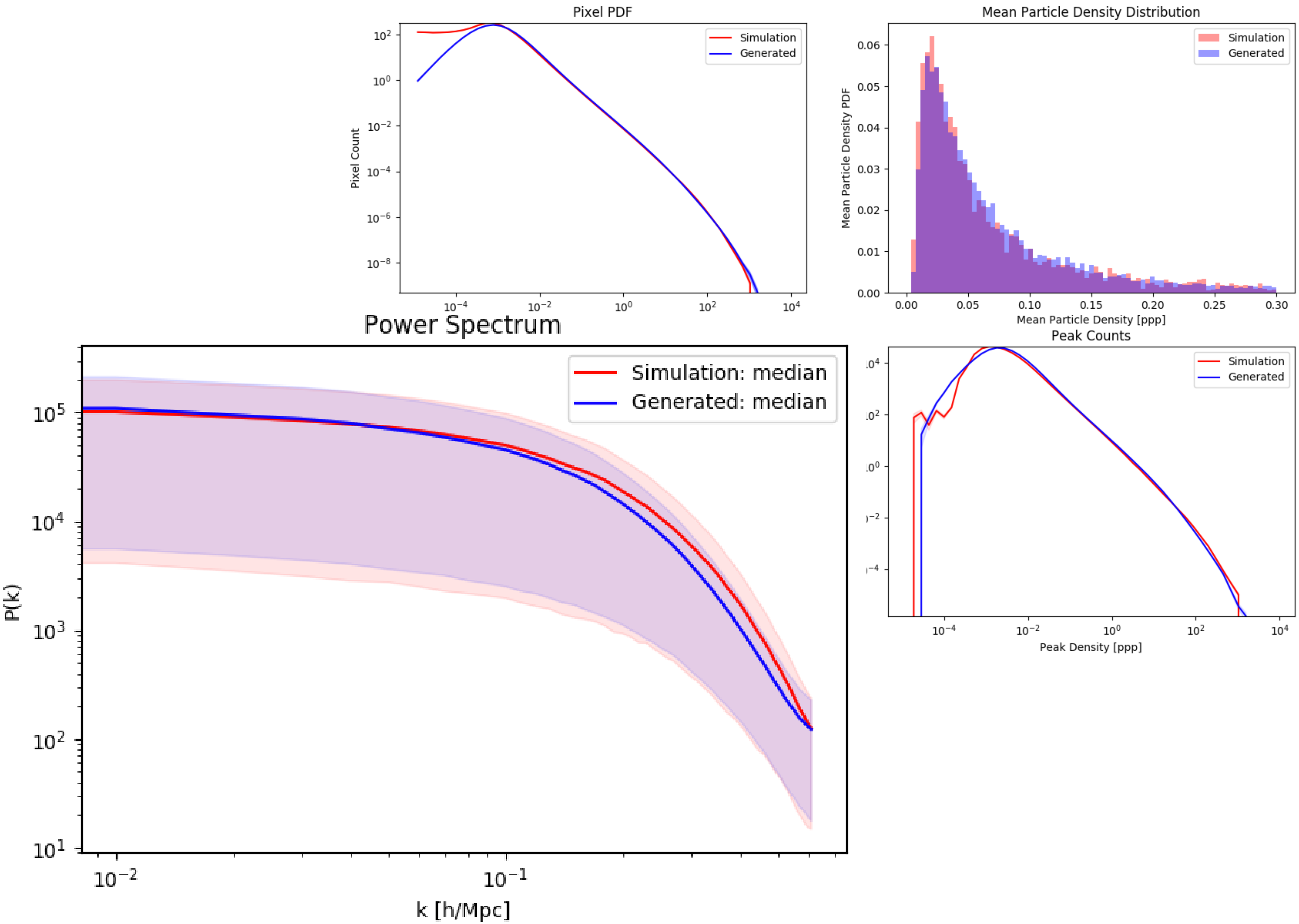




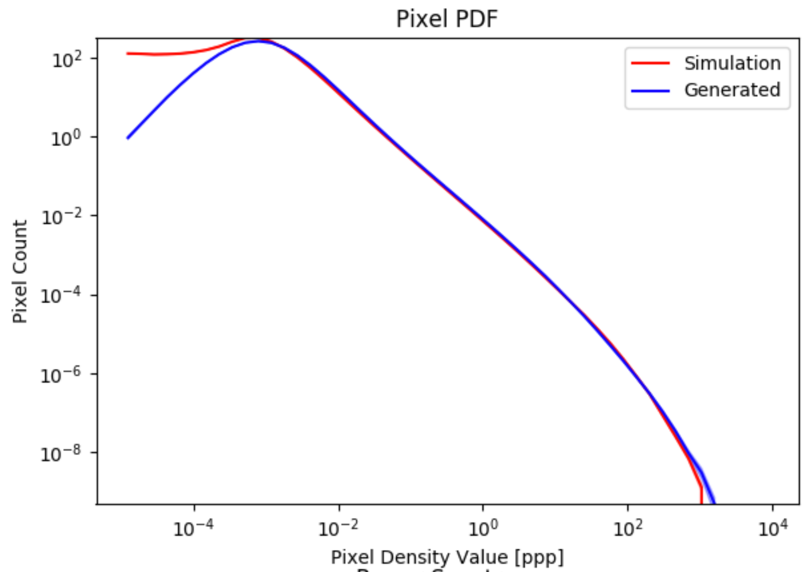


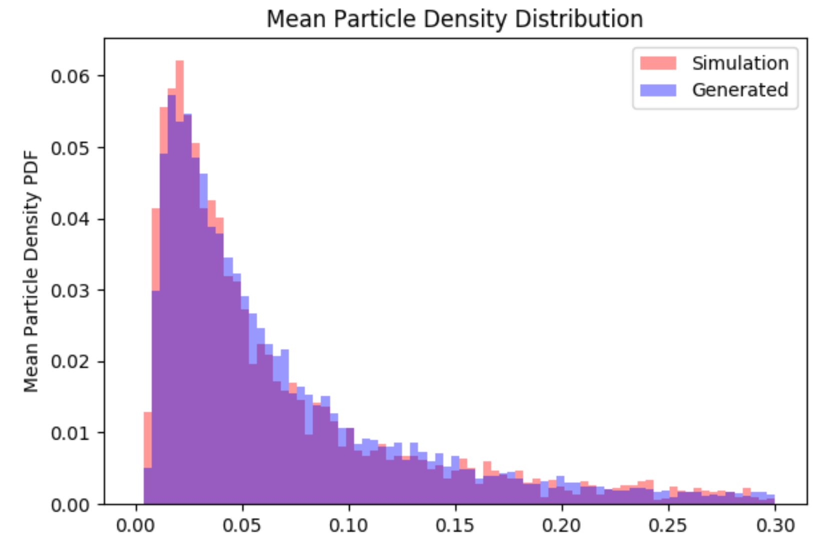


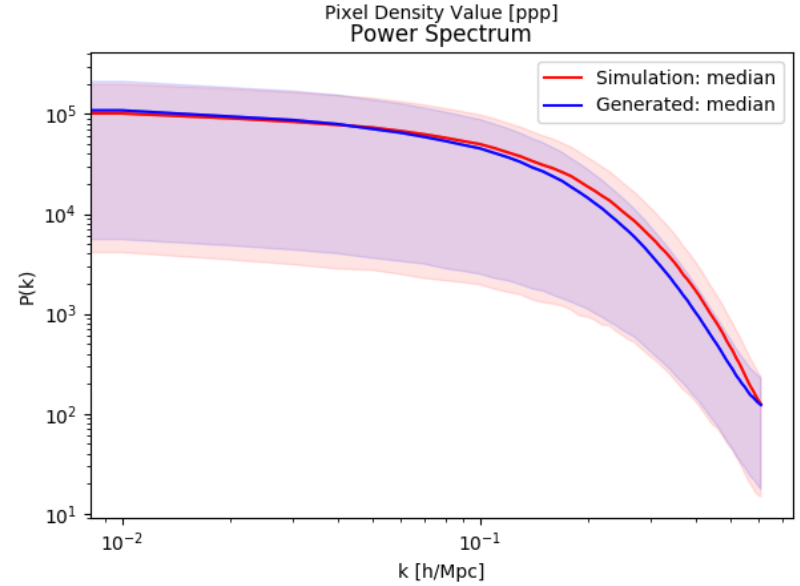
Comparing simulated images & generated images

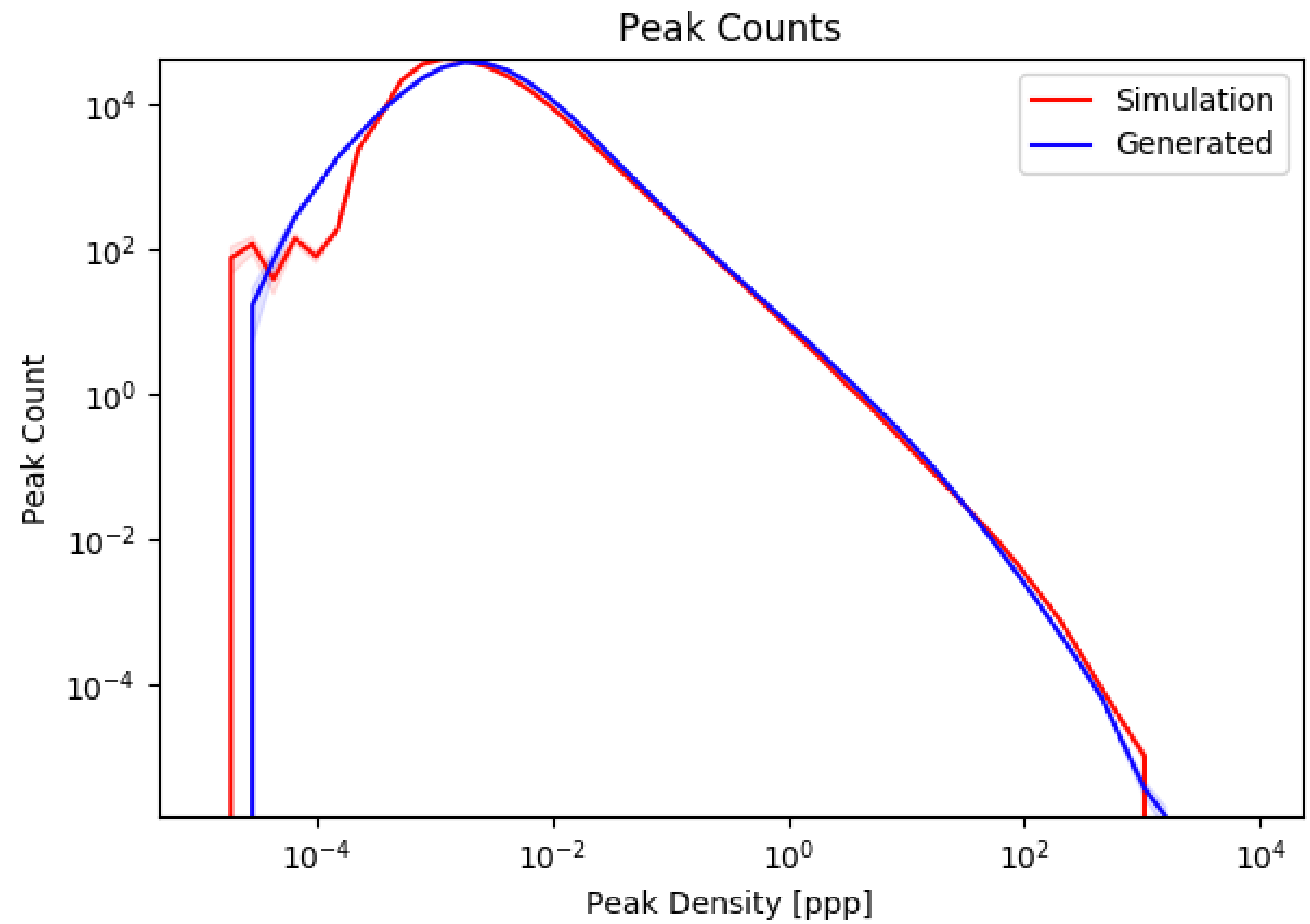


Comparing simulated images & generated images

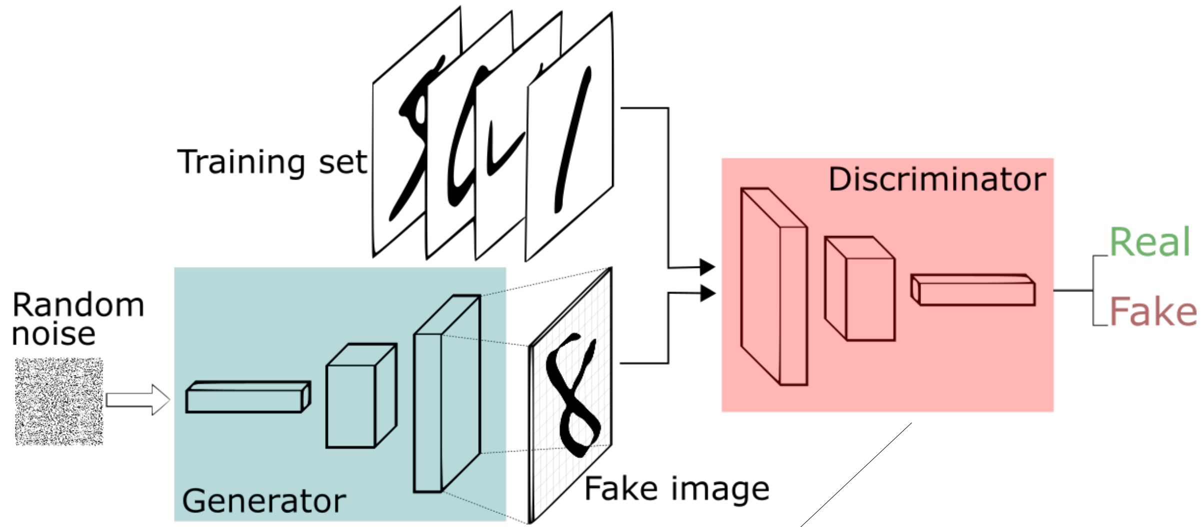






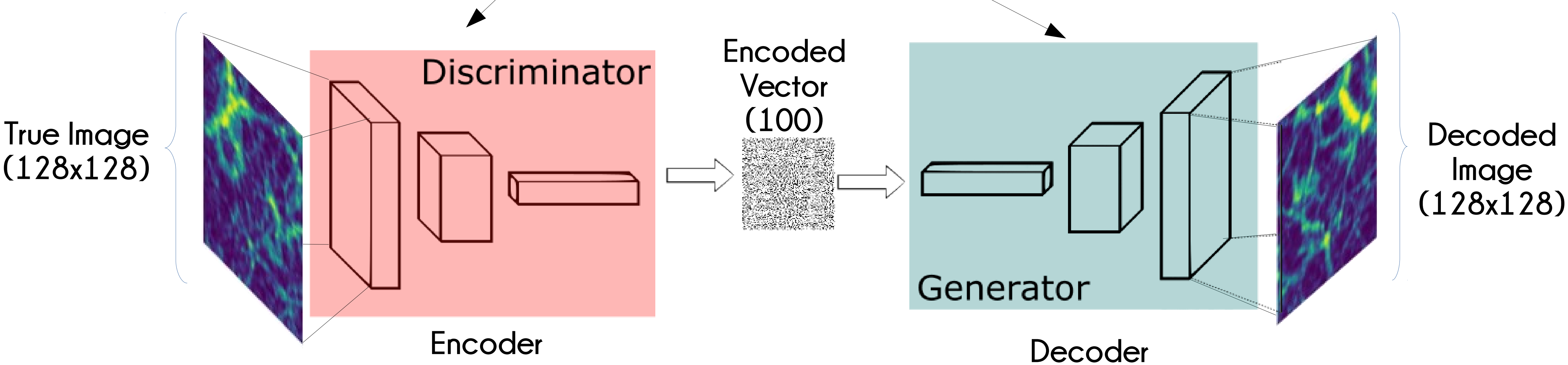


From GAN to Autoencoder



The discriminator breaks down an image to extract its essential information: we keep its structure but change its weights

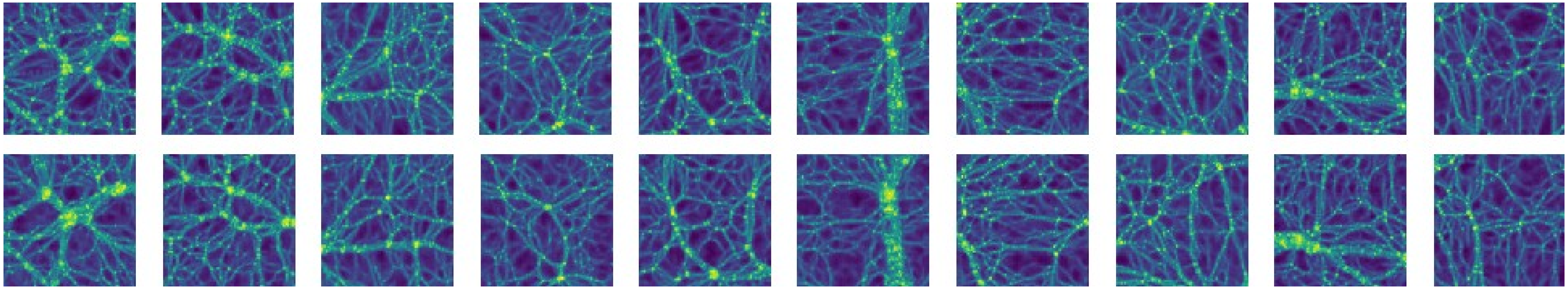
The generator intakes a vector and outputs a simulation-like image: we keep it unchanged, fixing its weights, to use as a decoder



Autoencoders : visual inspection

→ We show AE results for 2D simulations

10 “true” images randomly chosen from the simulations



And their 10 counterparts as inferred by the AE

Image to image comparison

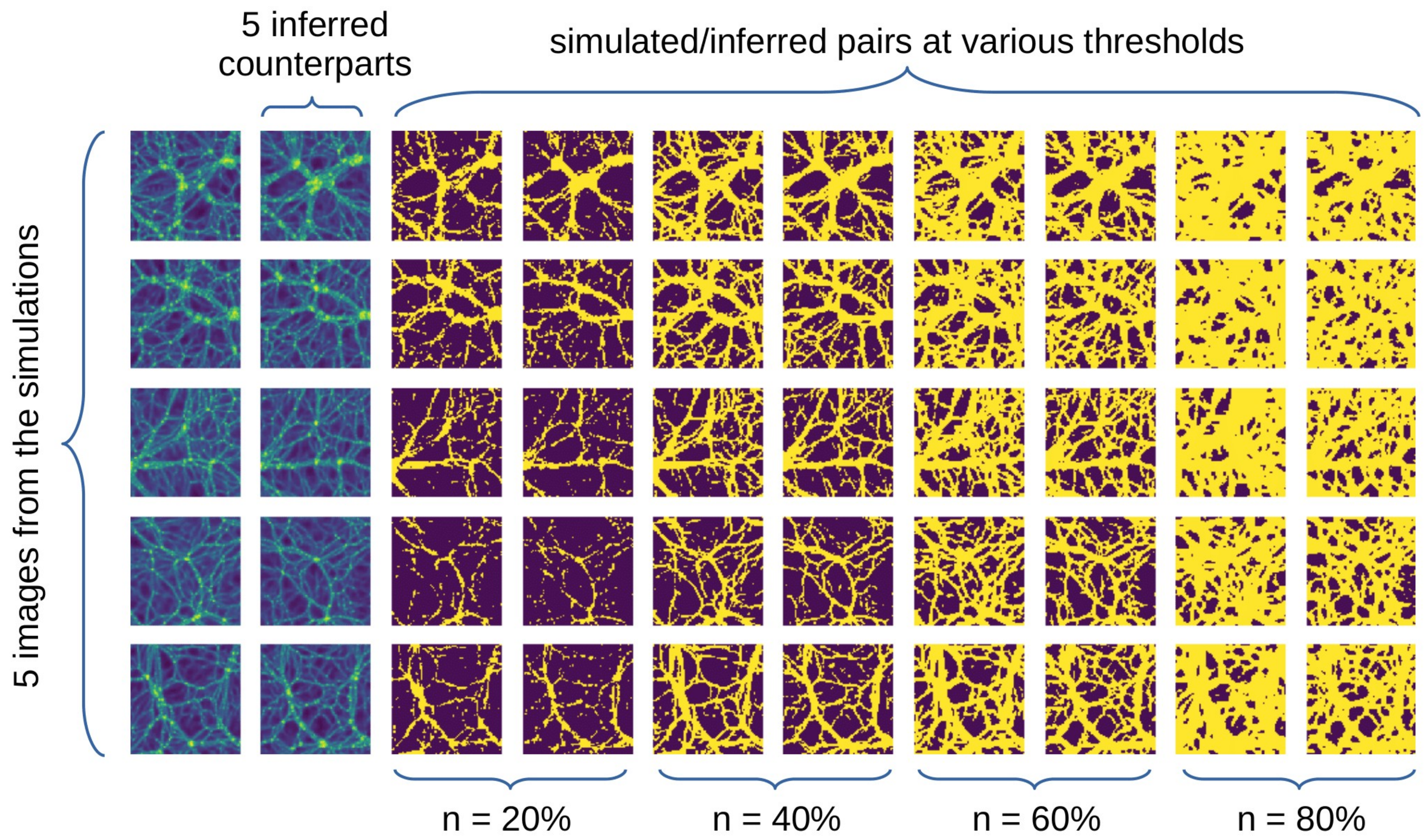
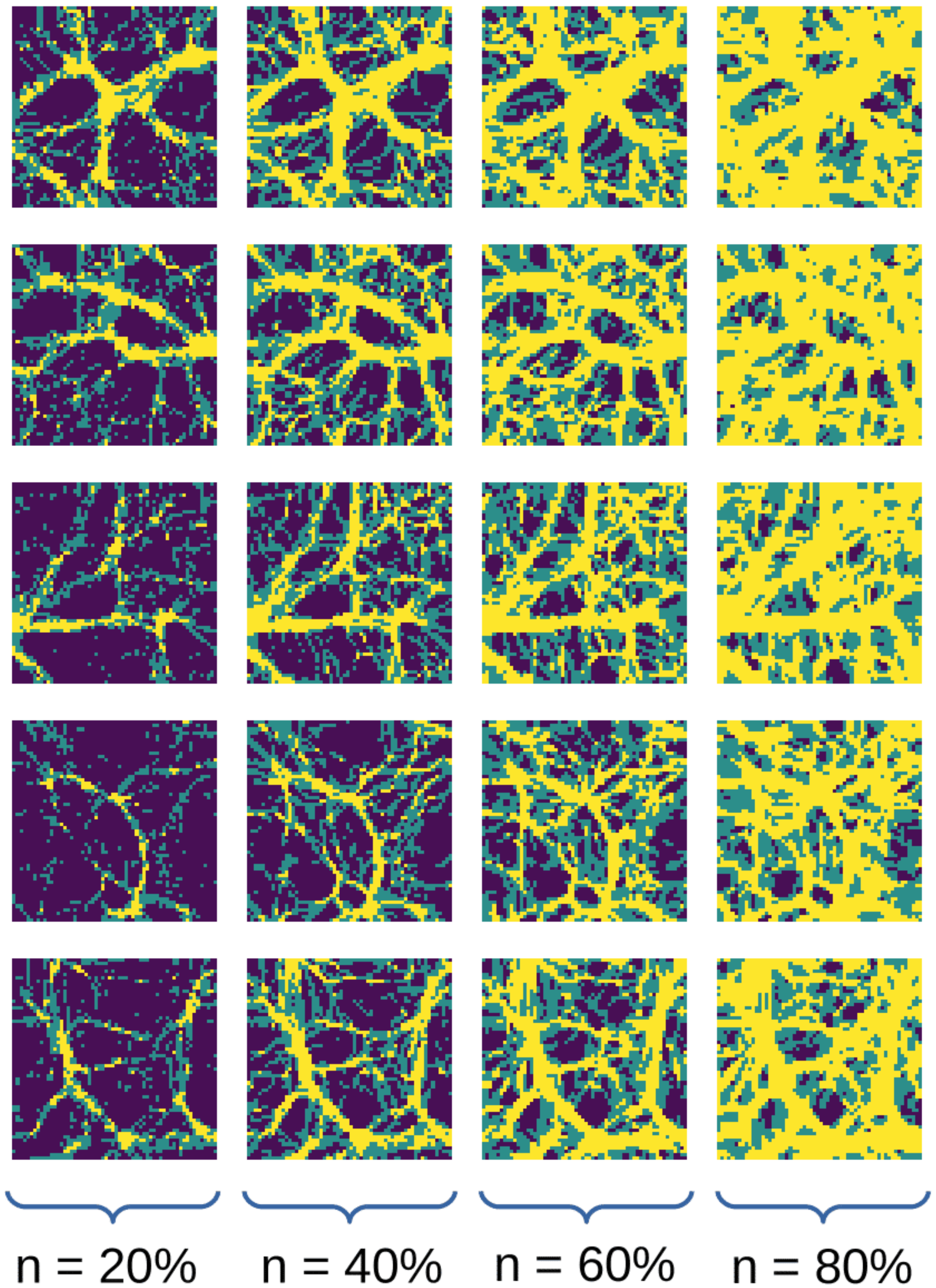
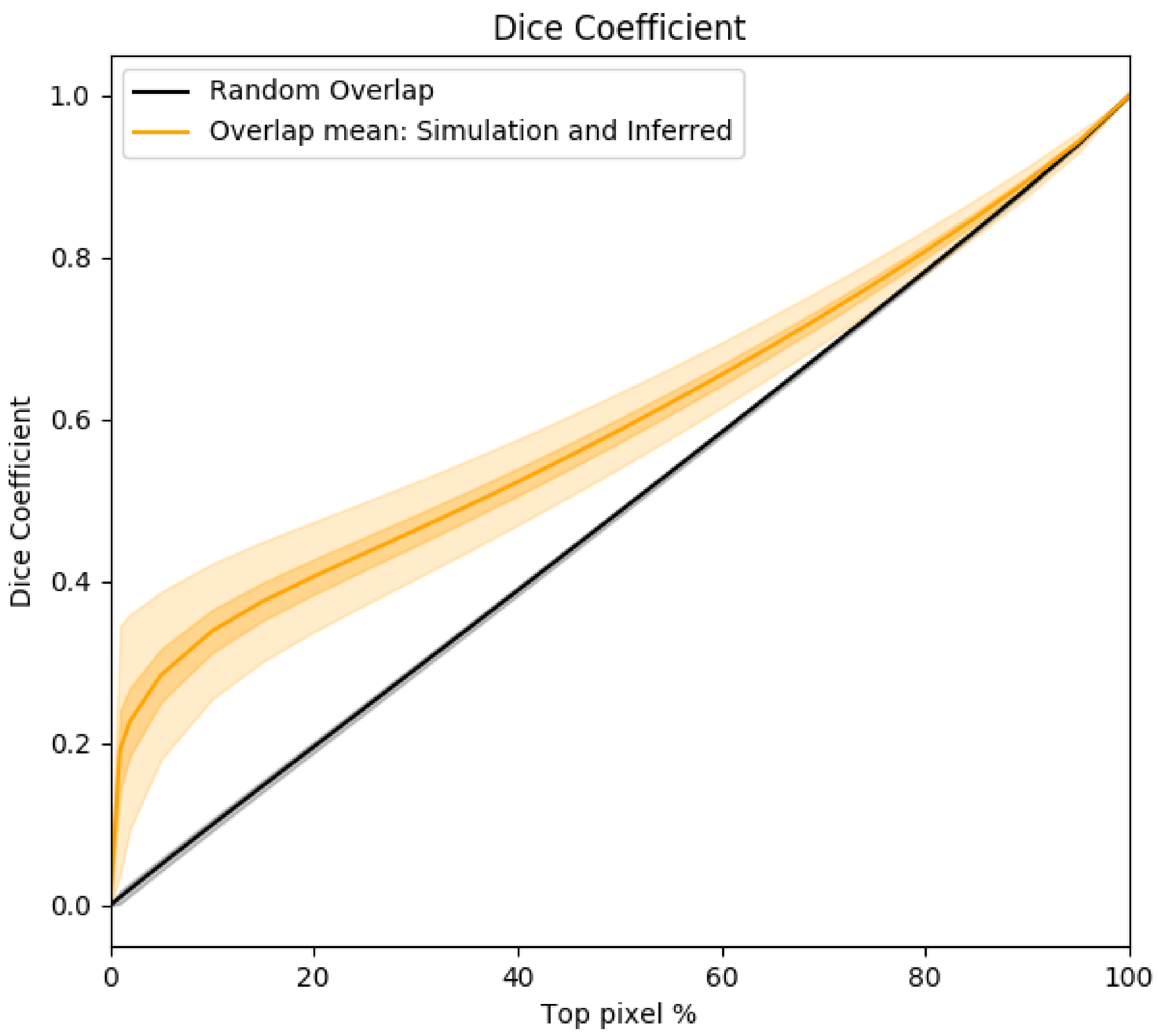


Image to image comparison

Overlap of the thresholded structures



Yellow: structures overlap   Green : structures differ



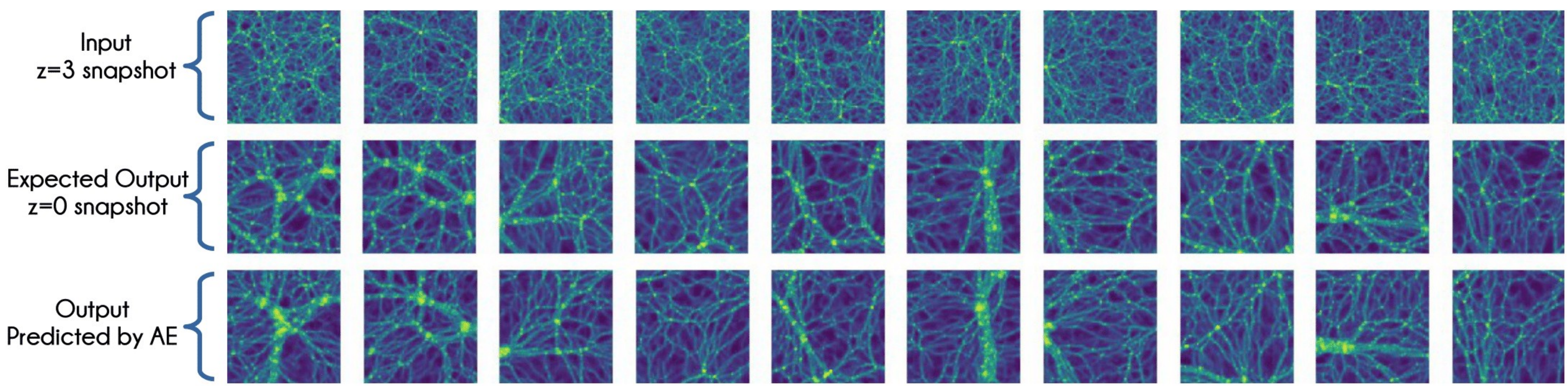
Conclusion

- GANs are a good alternative for fast generation of simulation-like datasets
- We have a working AE that is easy to build from a GAN and provides imperfect but satisfactory results
- We have a series of statistical estimators to quantify agreement between input and output beyond visual inspection → gives a basis of comparison for future work

Conclusion : Future Work

- Given GANs' great performance in 2D, we can expect this to translate well to 3D
- Using our working AE structure, we can now move on to use it for predictive purposes → encouraging results in 2D simulations already !

Sneak Peek : predicting z=0 images from z=3 inputs (2D simulations, test set results)



Thank You