Emulating cosmic web simulations with Generative Adversarial Networks

Marion Ullmo
Thesis Advisors: Nabila Aghanim (IAS) – Aurélien Decelle (LRI)

IAS Institut d'Astrophysique Spatiale Orsay

European Research Council
Established by the European Commission

The Cosmic Web and Large Scale Structures

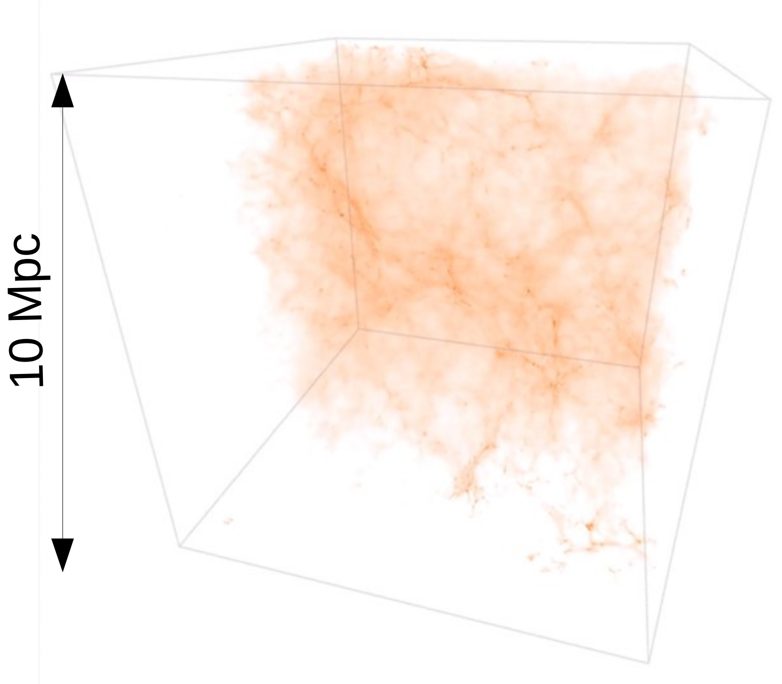
Early times : quasi-homogeneous matter distribution

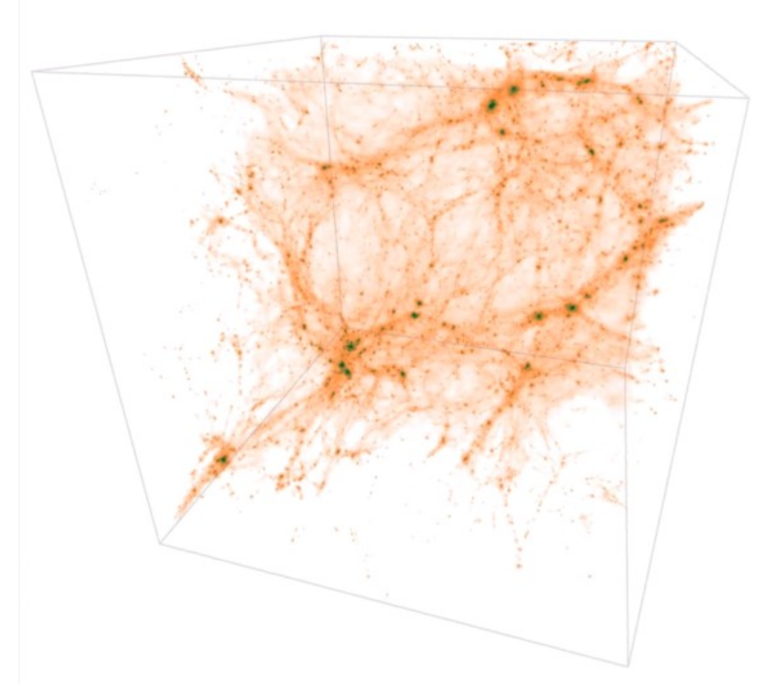
With time, a web-like structure forms

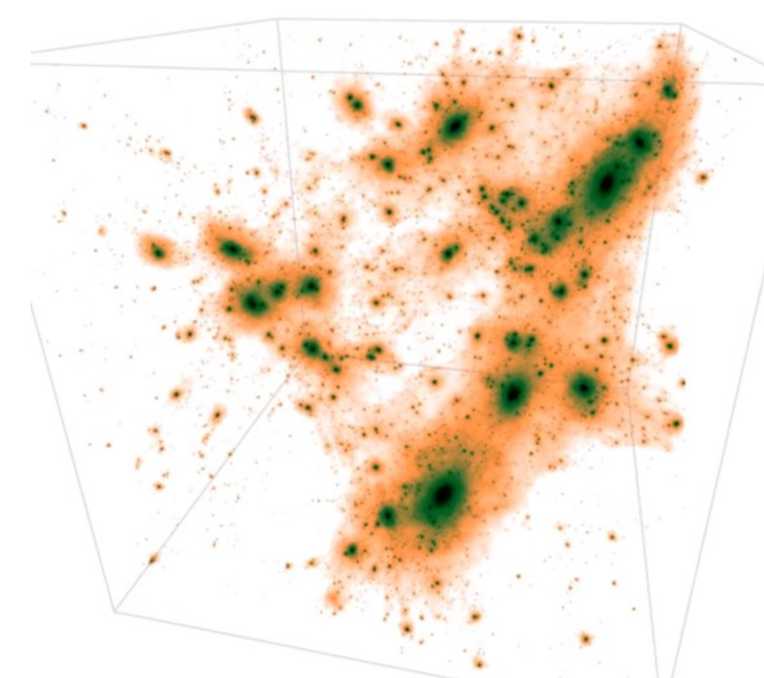
Current state : thermalized overdense halos have formed

Gravitational collapse →

* 1pc ~ 3.3 light years





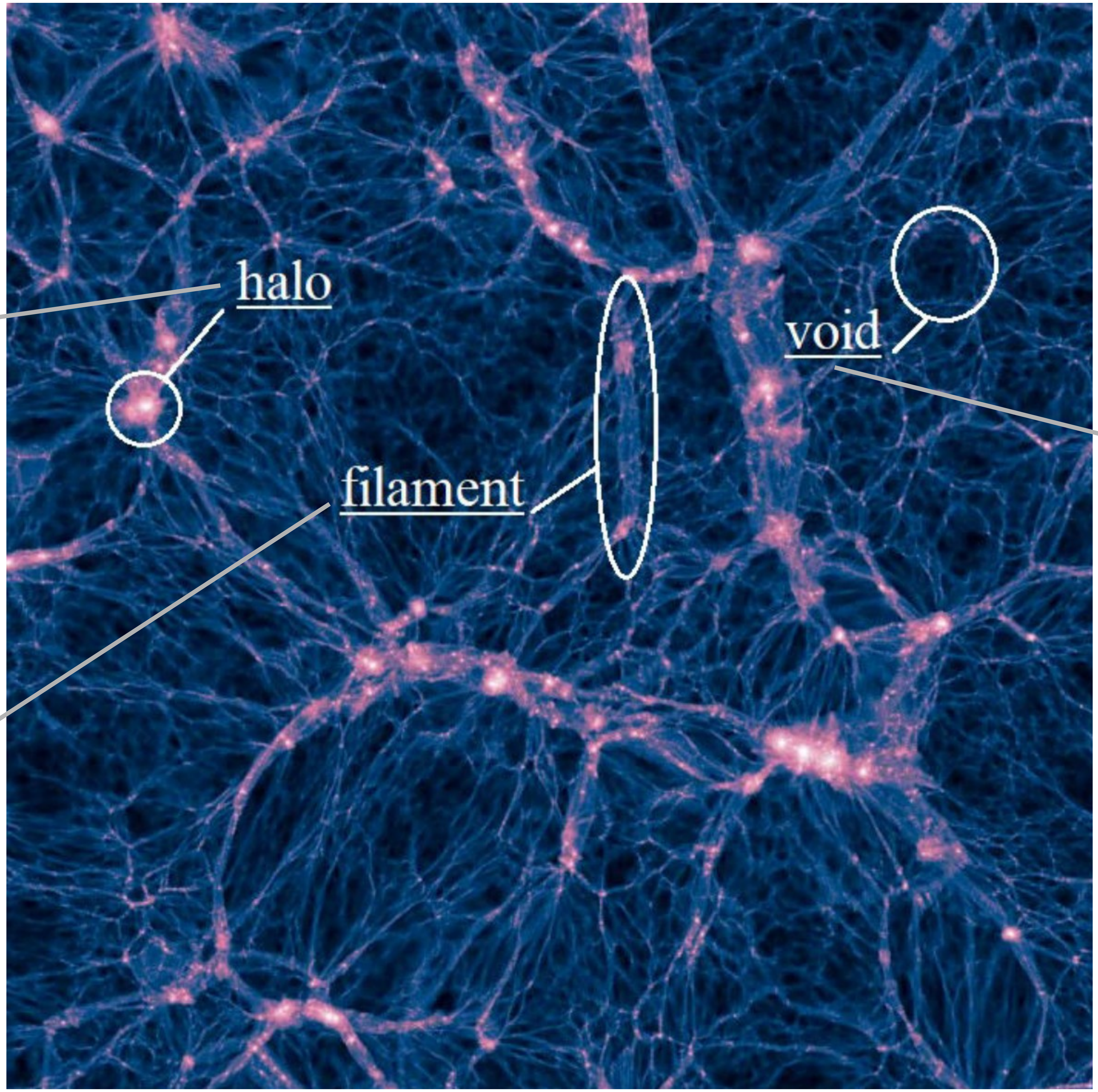


Redshift:	z =10	z=5	z=0
Time since Big Bang:	0.5 billion years	0.9 billion years	13.8 billion years

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

The Cosmic Web and Large Scale Structures

*1pc ~ 3.3 ly



Dense region: contains clusters, massive galaxies

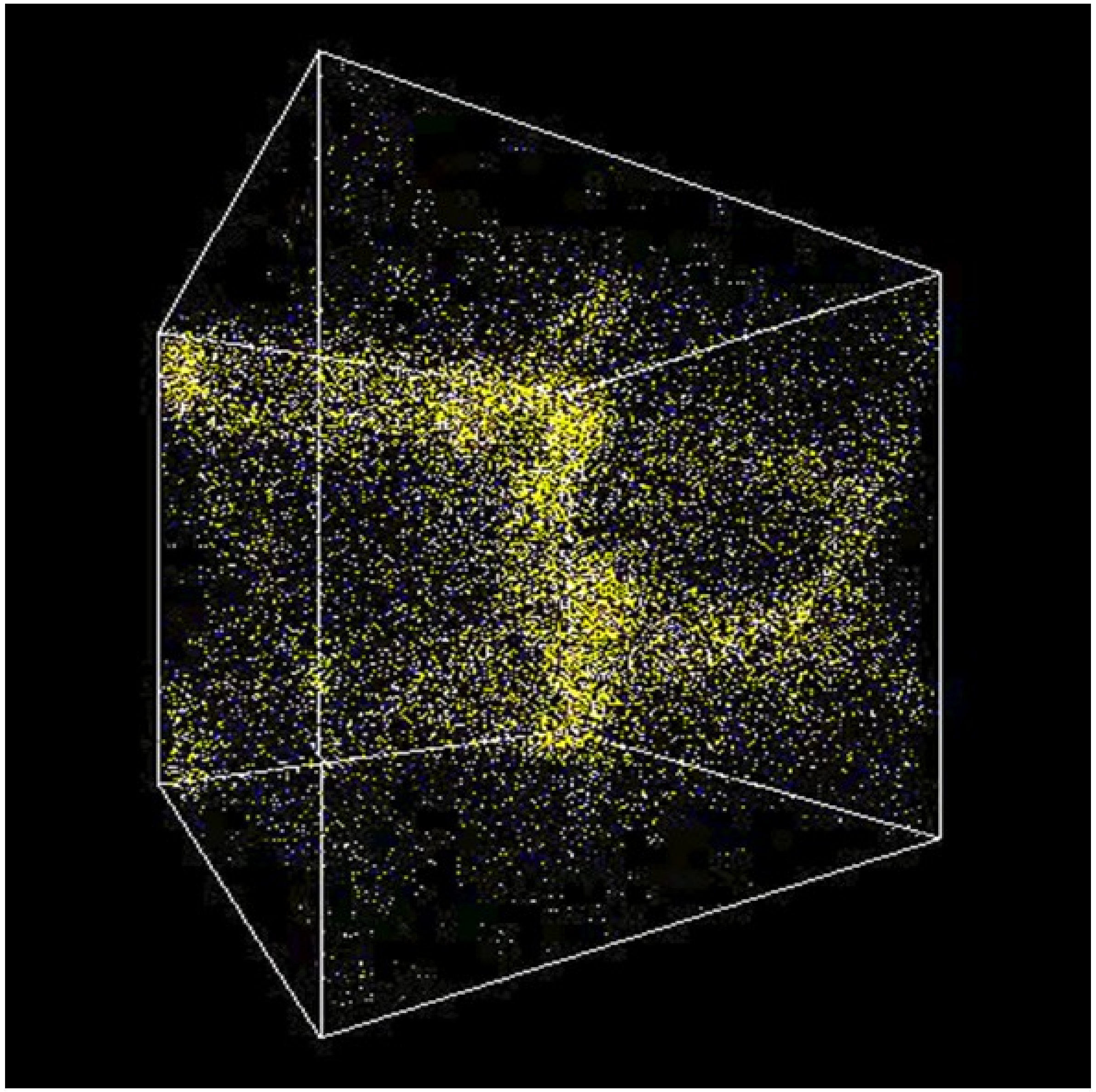
Elongated structure: contains galaxies & diffuse gas

Large, empty region

100 Mpc

Illustris simulations – $(100\text{Mpc})^2$ - Vogelsberger *et al*, 2014

Simulations - a costly necessity



We use simulations for a theoretical view of cosmic web structures

- 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, 3 million CPU hours,3 months runtime

- Very costly, with scales too small (100 Mpc) for good sampling of larger structures (>50 Mpc)

Goal, Motivations & Means

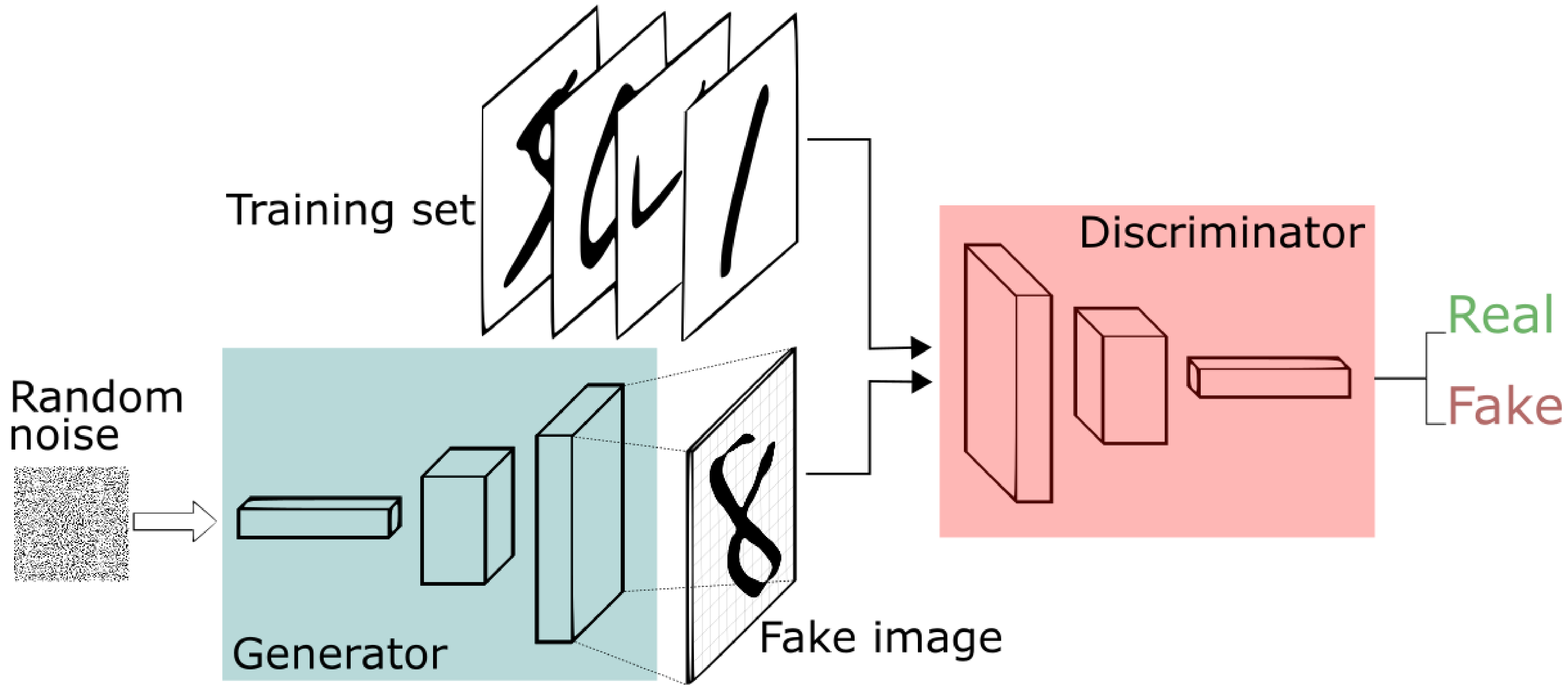
- Goal: emulate cosmic web simulations with the help of machine learning

- Motivations:
  We aim to characterize cosmic structures statistically
  → we need to generate large amounts of data
  → we must find faster alternatives to simulations

- Means: Deep Neural Networks

Images generated by StyleGAN -Karras *et al* & Nvidia 2018

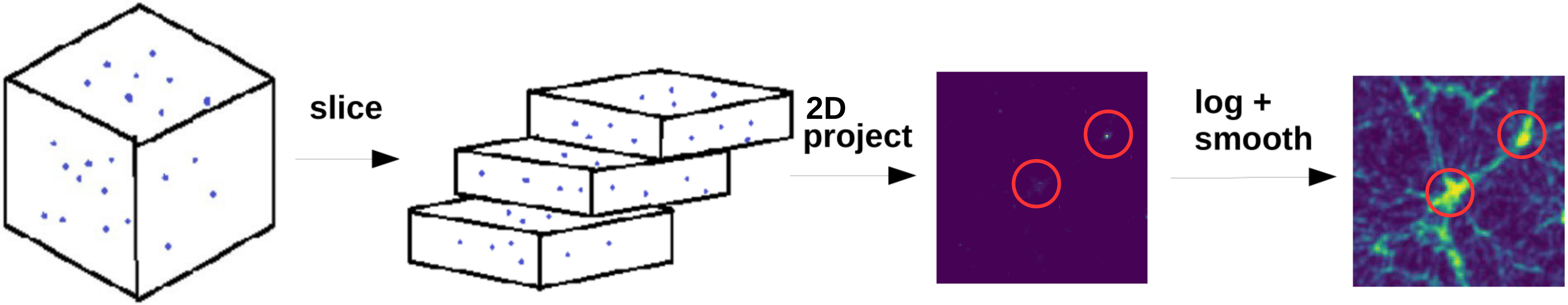
GANs* in a nutshell



- GAN : Generative Adversarial Network → Generative model
- Two competing networks :
  - the generator, generates new images
  - the discriminator, determines the probability for an image to come from the dataset or the generator
- An easy to compute loss but hard to find a working architecture

* *Goodfellow, 2014*

Modus Operandi



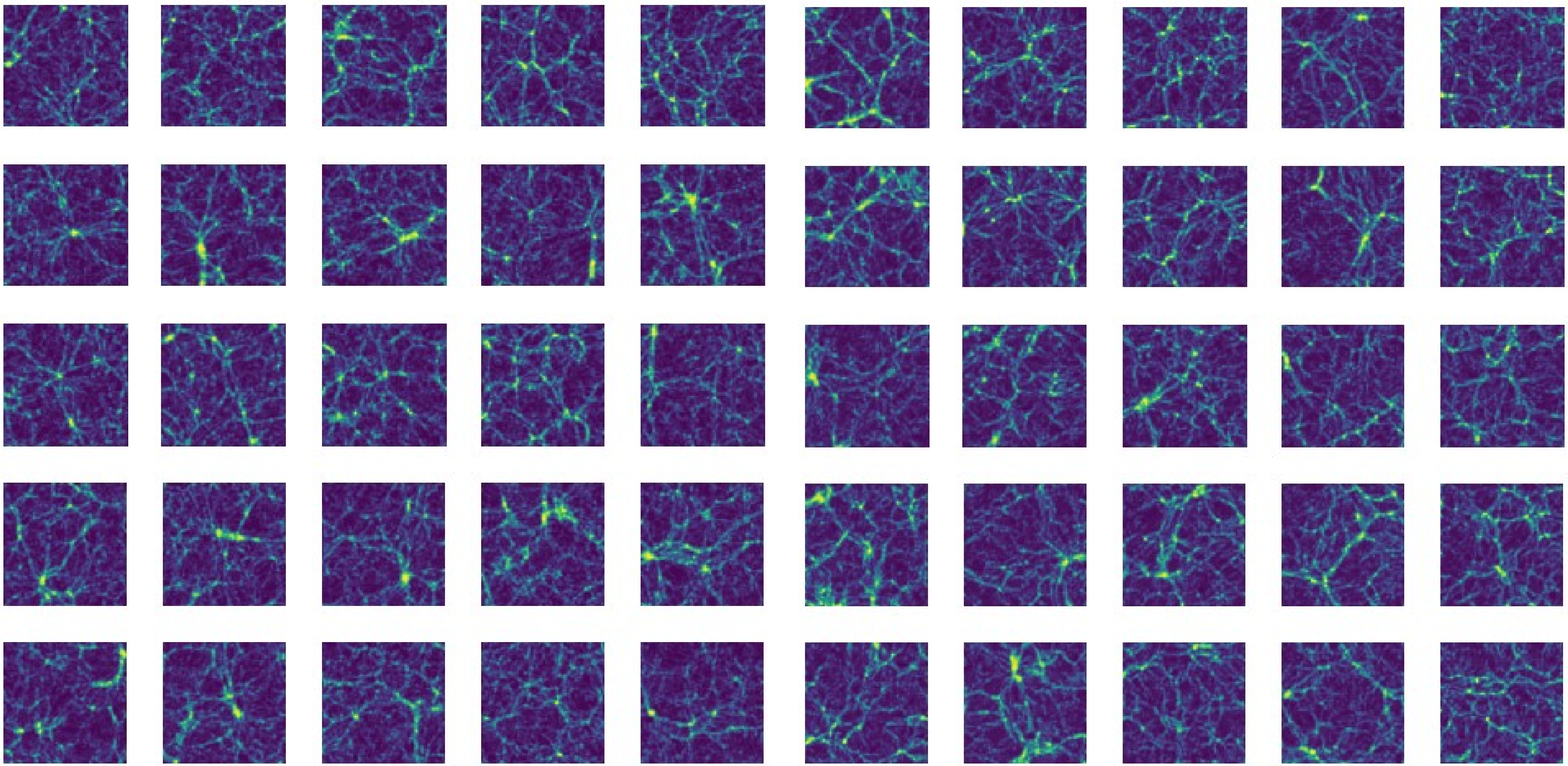
- <u>Step 1:</u> Run simulation:
  - -box size is (100 Mpc)^3,
  - -(512)^3 particles
  - -runtime is 4 days

  → a snapshot (3D box containing particle positions) is obtained

- <u>Step 2:</u> Extract slices from the snapshot to create a set (96000) of 2D images (128x128 pixels)

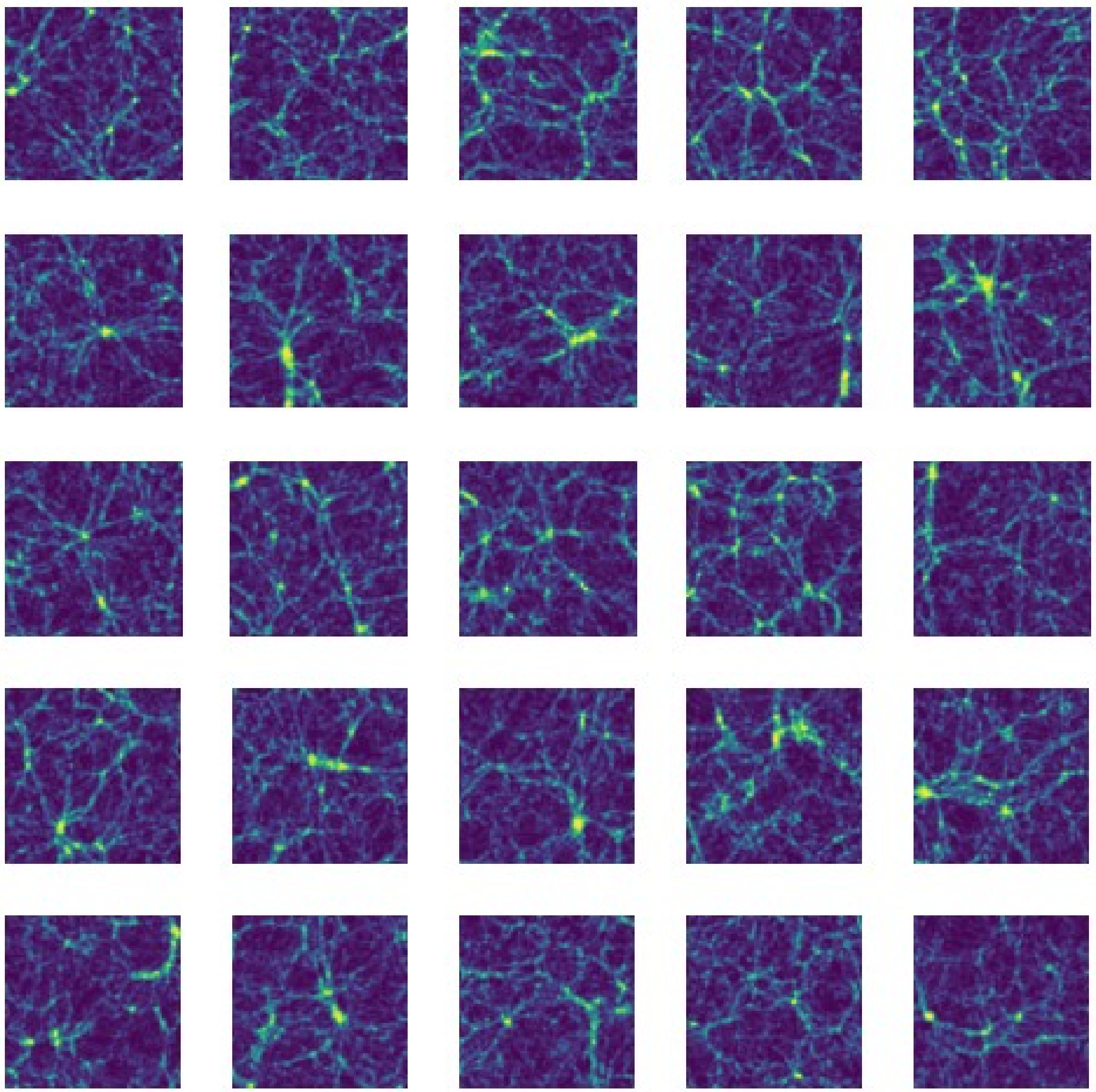
- <u>Step 3:</u> Train the GAN on the dataset

Results – 128x128 pixel images of log-density

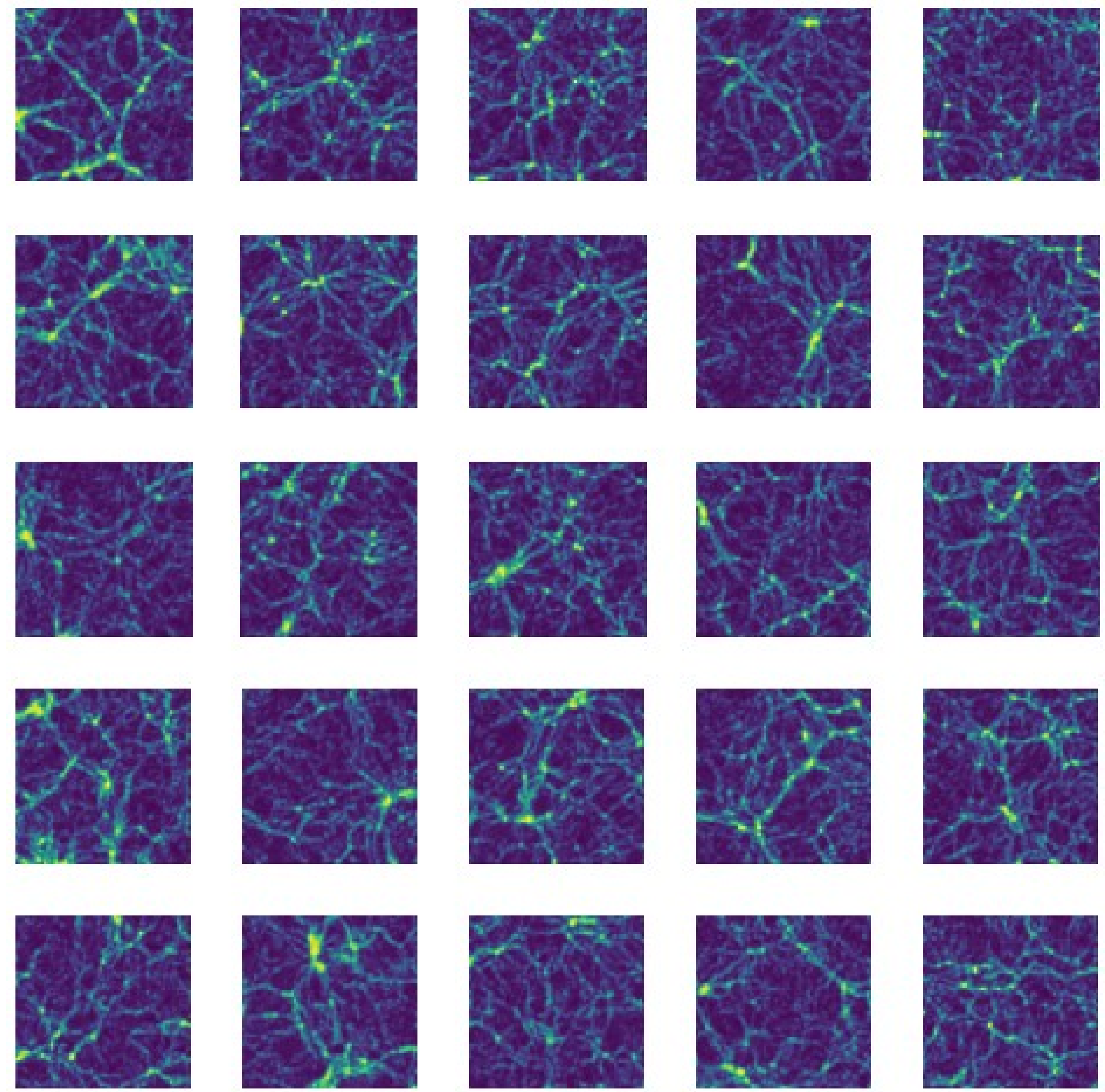


Results – 128x128 pixel images of log-density

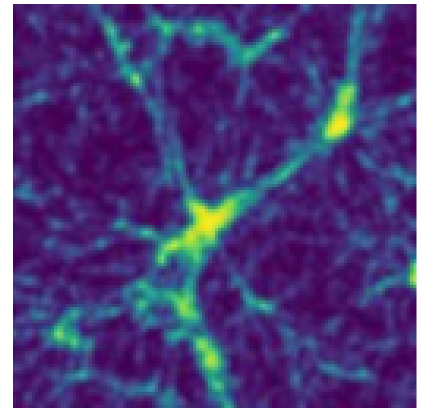
Images from the simulation ("real")

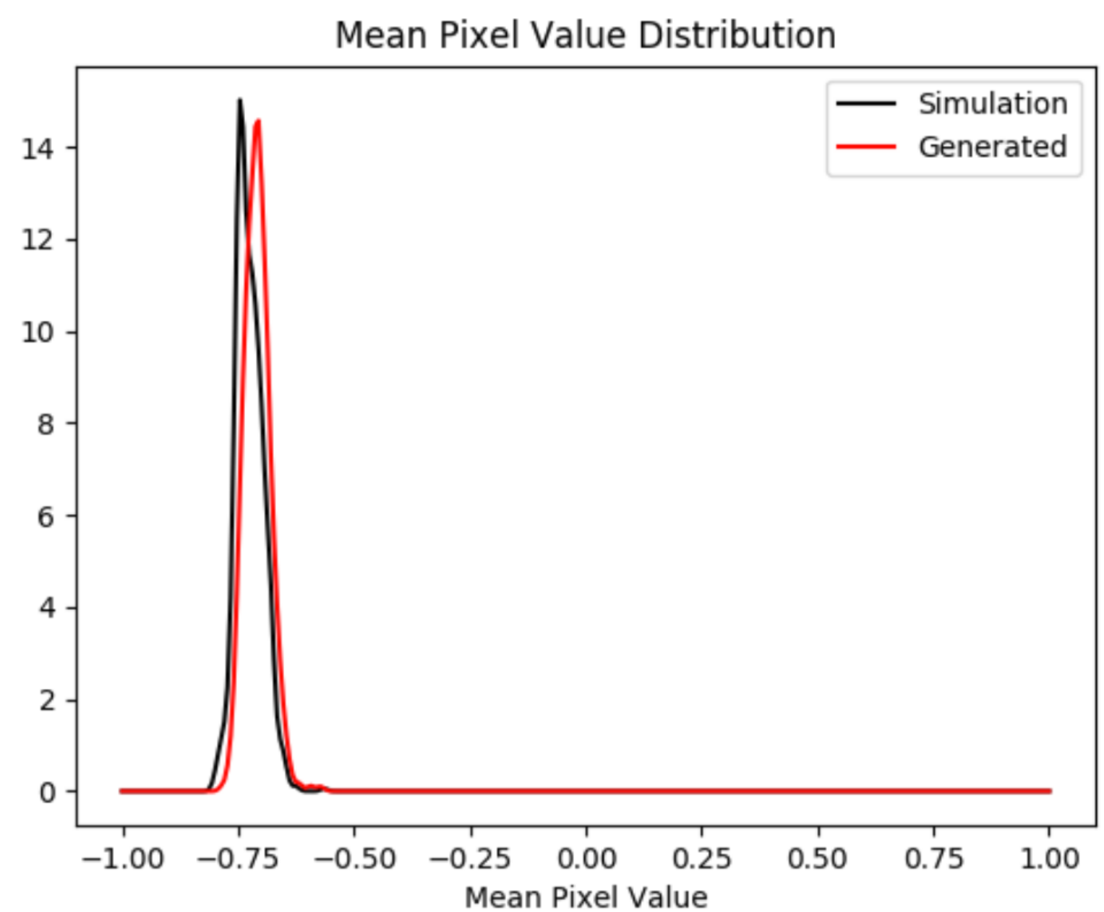
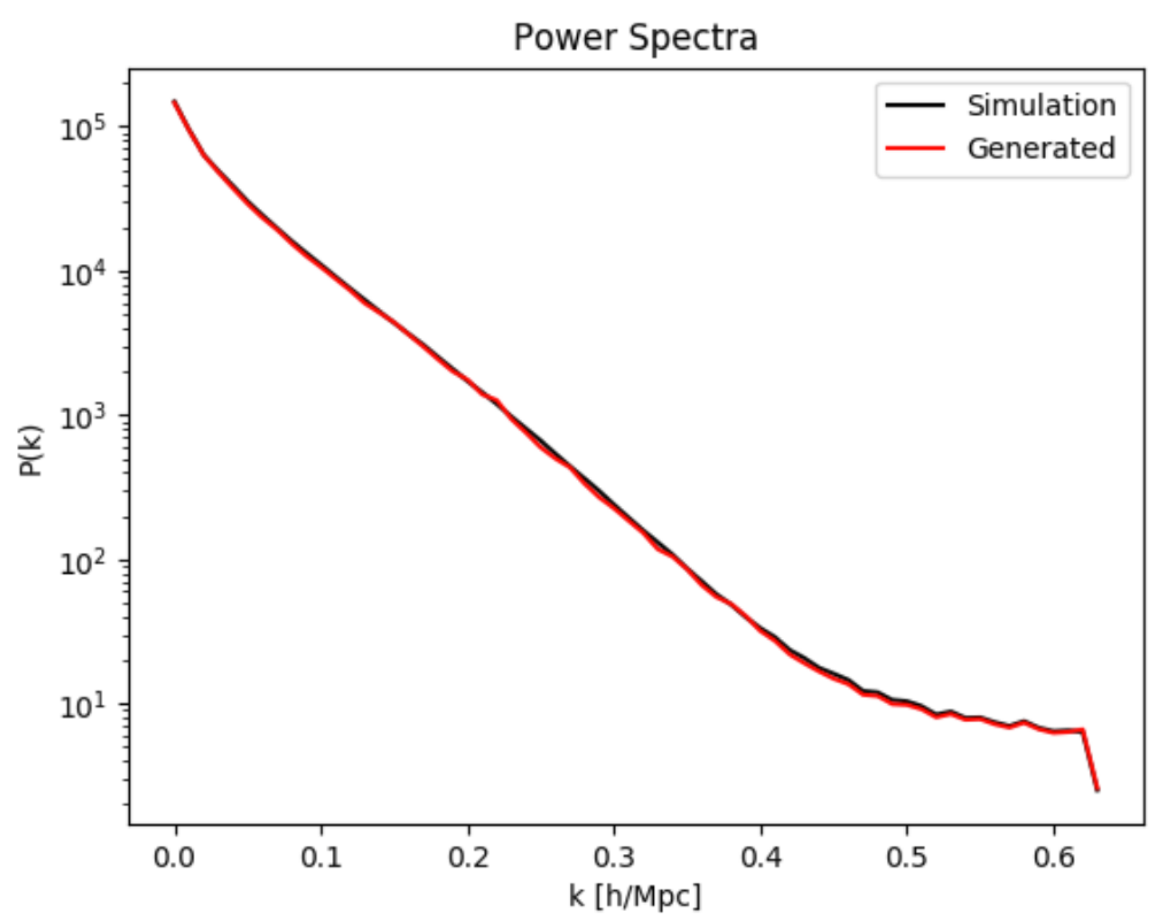


Images from the GAN ("fake")

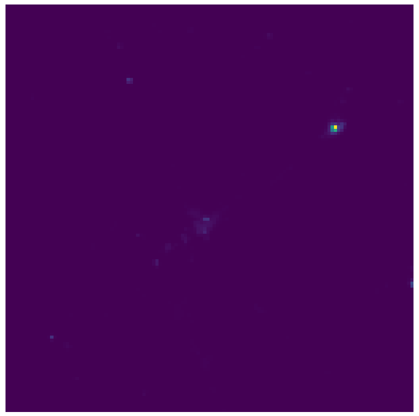


Comparing dataset and generated images (statistics on 1000 images) – log density images

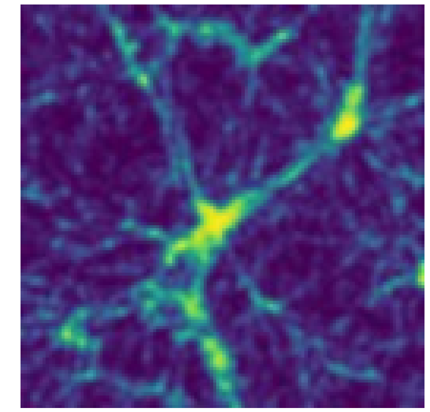


Mean Pixel Value Distribution	Averaged Power Spectra
	

Comparing dataset and generated images (statistics on 1000 images) – density images



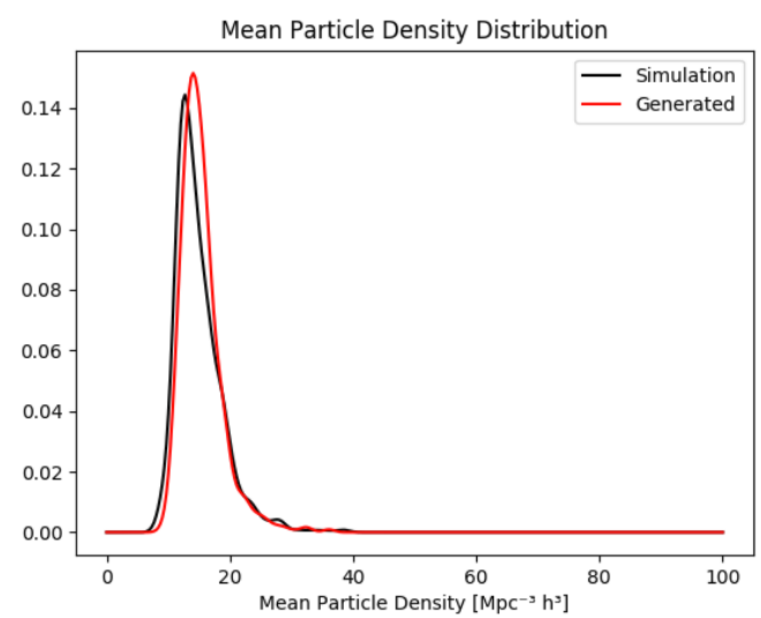
$= \log^{-1} ($



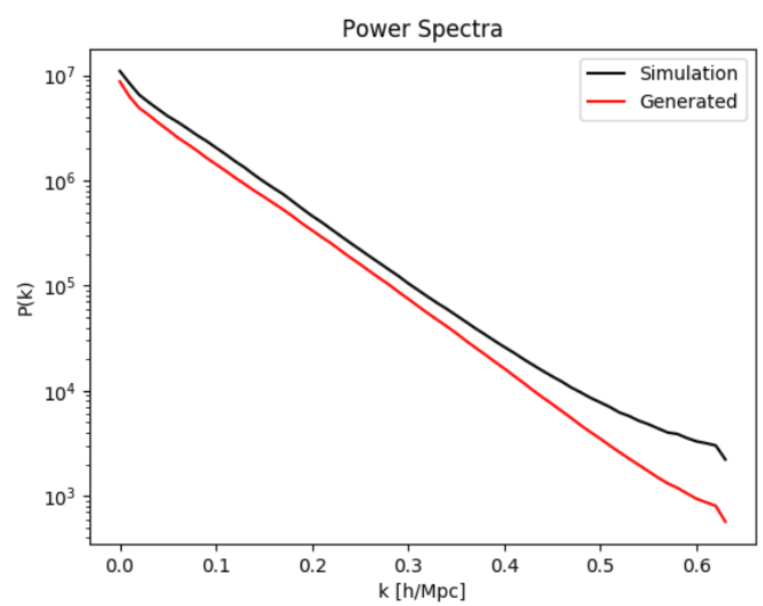
$)$

→ Discrepancy due to saturation effect

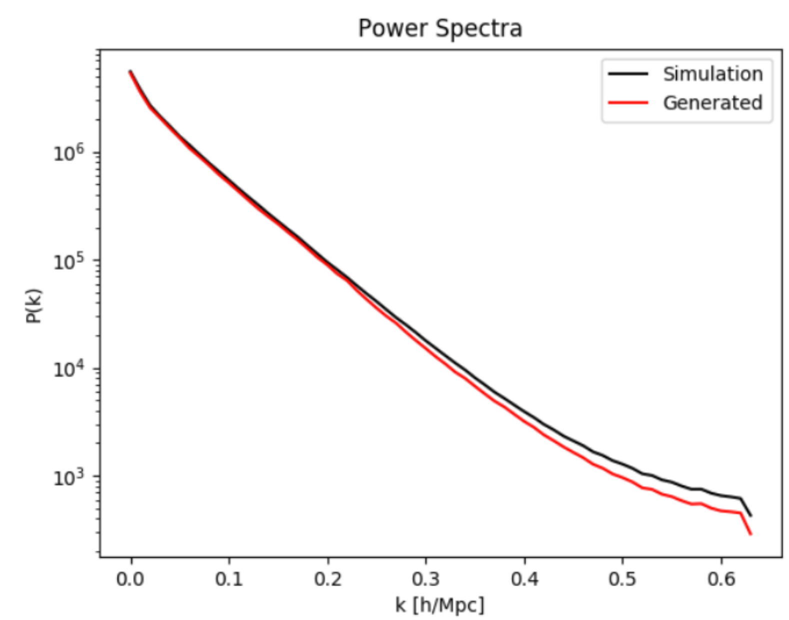
Mean Particle Density Distribution



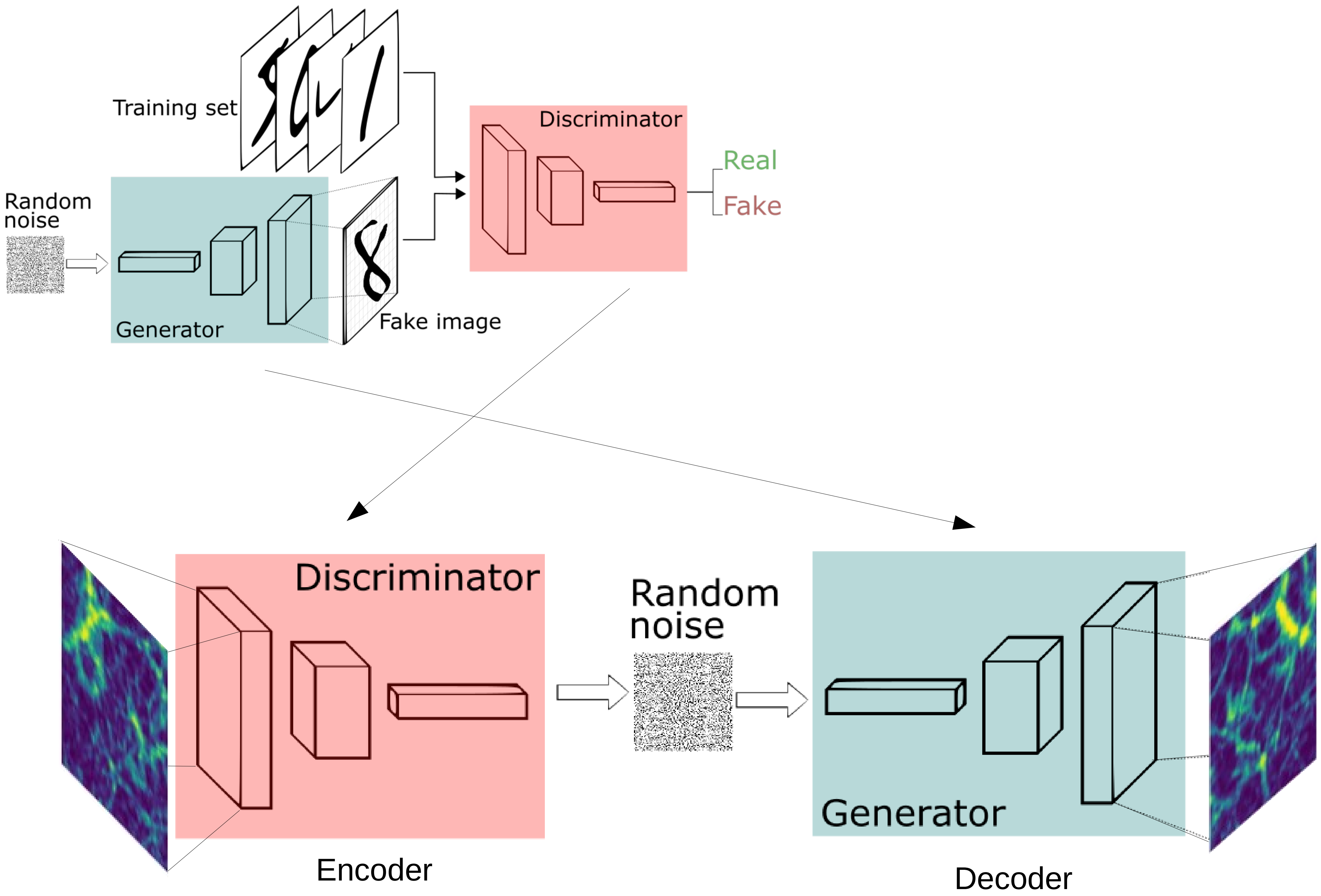
Averaged Power Spectra



Thresholded Averaged Power Spectra (top 1‰ removed)

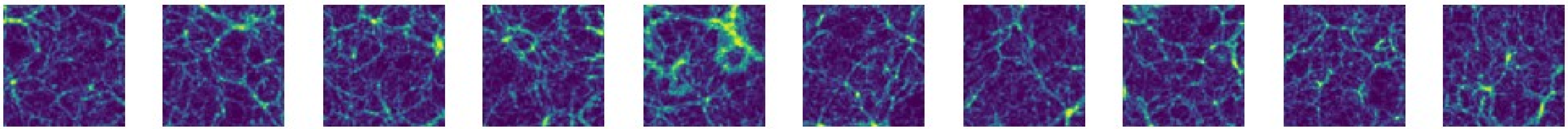


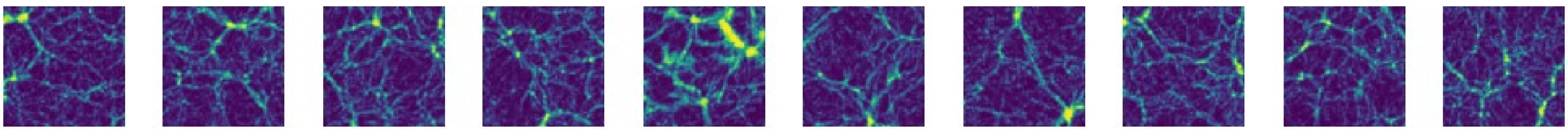
From GAN to Autoencoder



Checking Generator quality with the Autoencoder

10 “real” images from the simulations (taken at random)





And their 10 equivalents generated by the GAN

→ Recall: power spectrum is recovered with saturation effects on top 1‰ pixels

→ Shapes, texture and diversity are recovered
→ The phase distribution of the power spectra is also reproduced!

Conclusion

- The GAN can efficiently extract the inherent distribution of a given dataset
- It can then generate images from this distribution instantly
- GANs appear as a promising alternative to costly simulations
- We can make use of a GAN's learned features for new tasks