Generating high-fidelity HI maps using score-based diffusion models

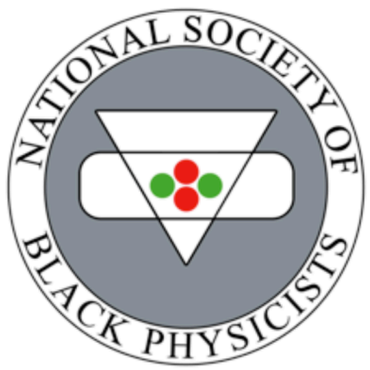
Sultan Hassan (he/him)

NHFP Hubble Fellow, New York University

Guest researcher, CCA/UWC

Galaxy Formation and Evolution in the Data Science Era

GALEVO23/KITP/March, 2023.



UNIVERSITY *of the*
WESTERN CAPE

Some of my greatest hits in generative models

Sultan Hassan (he/him)

NHFP Hubble fellow, New York University

Guest researcher, CCA/UWC

Galaxy Formation and Evolution in the Data Science Era

GALEVO23/KITP/March, 2023.

UNIVERSITY *of the*
WESTERN CAPE

Some of my greatest hits in generative models

- Emulating radiation transport on large scales using denoting U-Nets (Masipa, Hassan, Santos, Contardo, Cho 2023, accepted to ICLR 2023).

- Towards a non-Gaussian generative model of large scale reionization maps (Lin, Hassan, Blancard, Eickenberg, Modi 2022, accepted to NeurIPS 2022).

- Invertible Mapping between fields in CAMELS (Andrianomena, Hassan, Paco 2023, accepted to ICLR 2023).

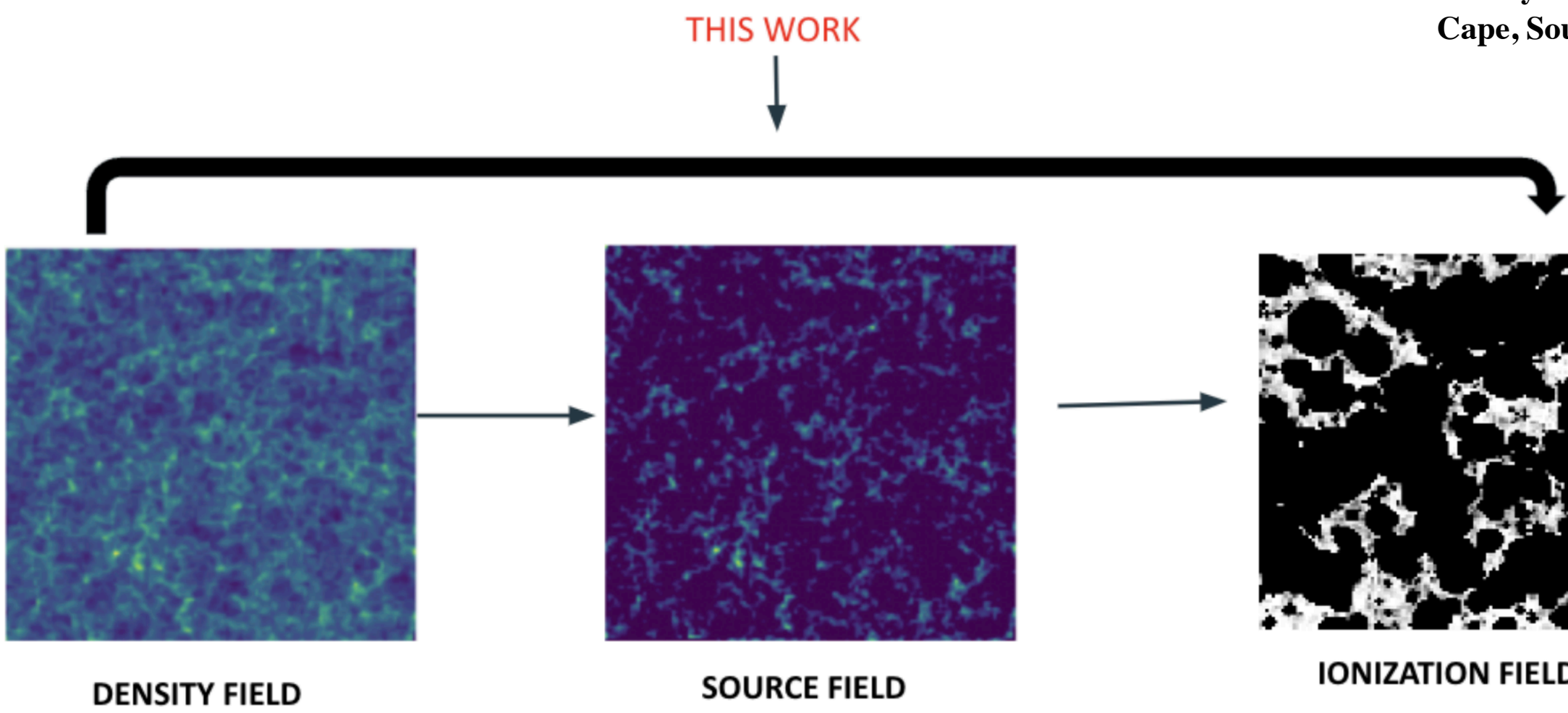
- Generating high-fidelity HI maps using score-based diffusion models (Started here, Hassan, Wu, Lovell, Cooray +, join us in slack @ galevo23-p11, to be submitted to ICML 2023 in Hawaii!)

Emulating radiation transport on large scales using denoting U-Nets (Masipa, Hassan, Santos, Contardo, Cho 2023, accepted to ICLR 2023)

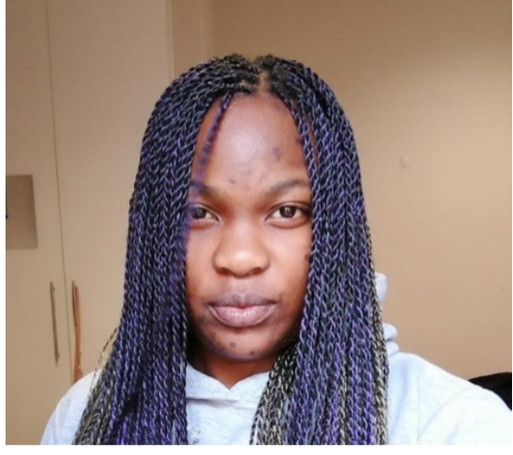
Emulating Radiation Transport on Cosmological Scales using a Denoising U-Net

Mosima Masipa, MSc student

University of the Western Cape, South Africa

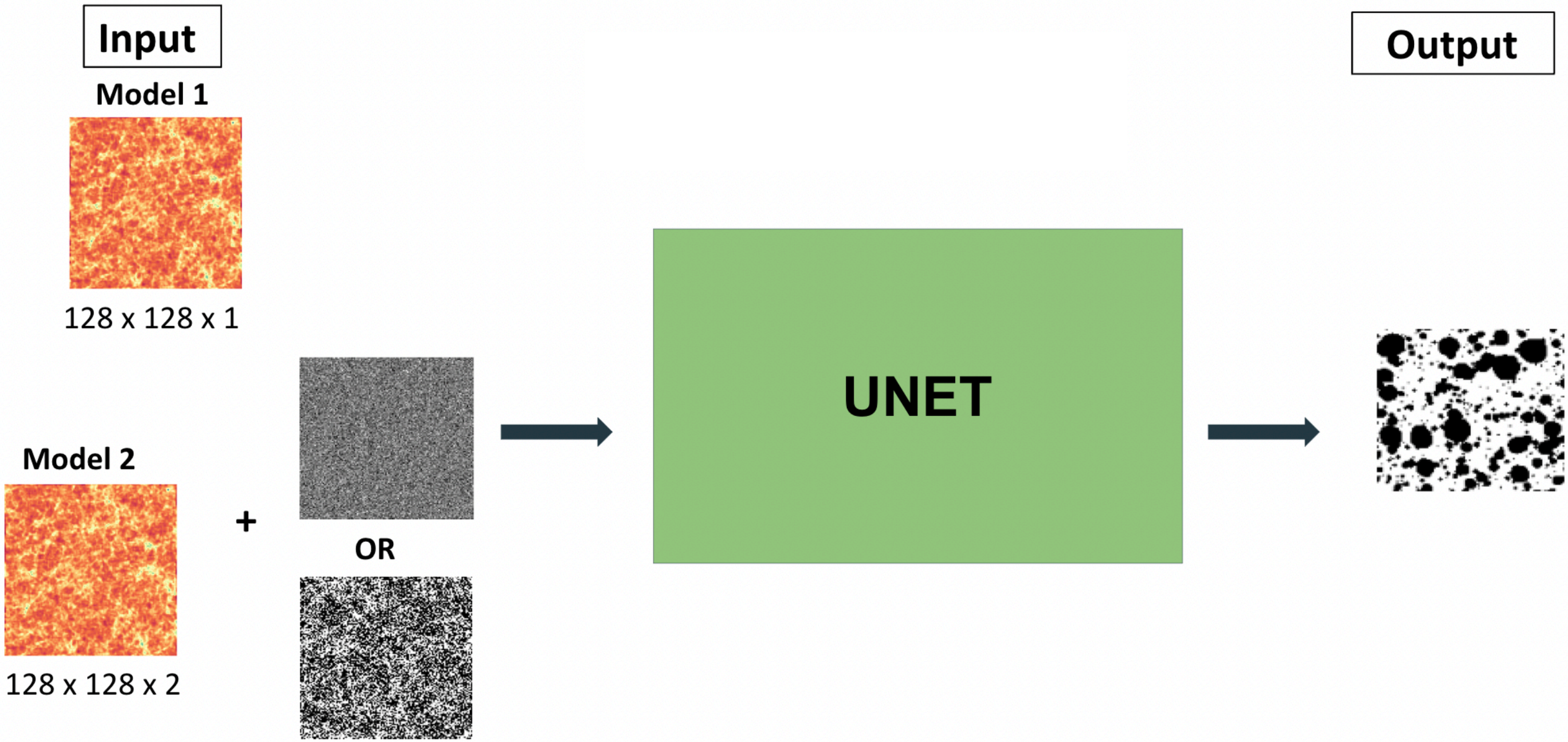


Emulating Radiation Transport on Cosmological Scales using a Denoising U-Net



Mosima Masipa, MSc student

University of the Western Cape, South Africa

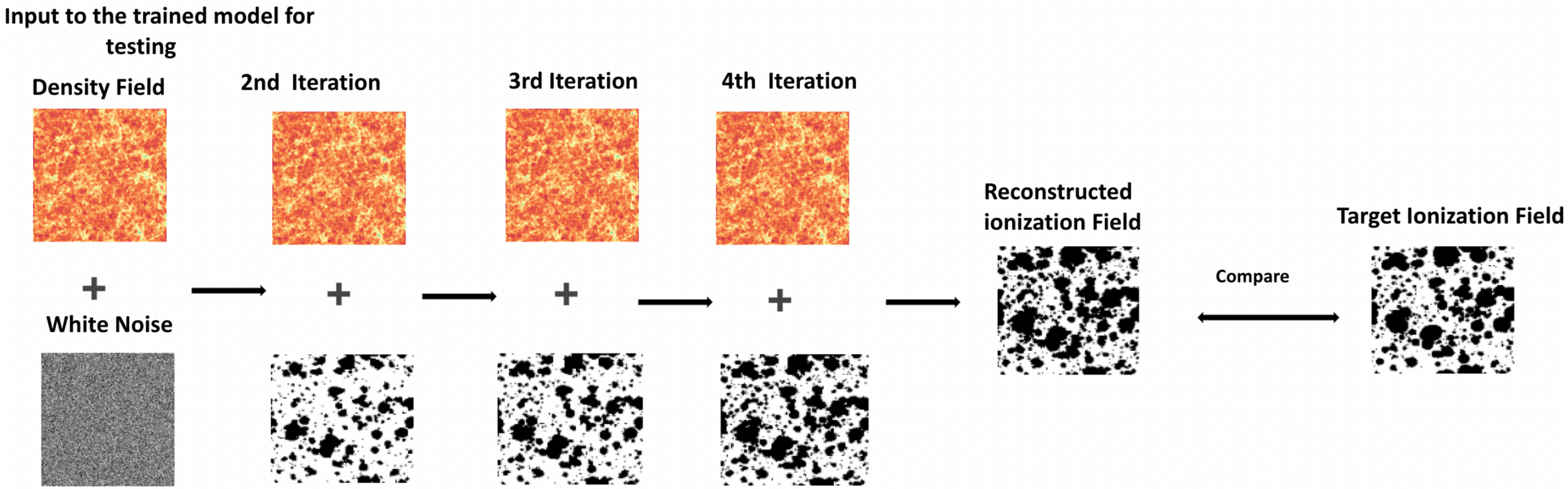


Emulating Radiation Transport on Cosmological Scales using a Denoising U-Net

Mosima Masipa, MSc student

University of the Western Cape, South Africa

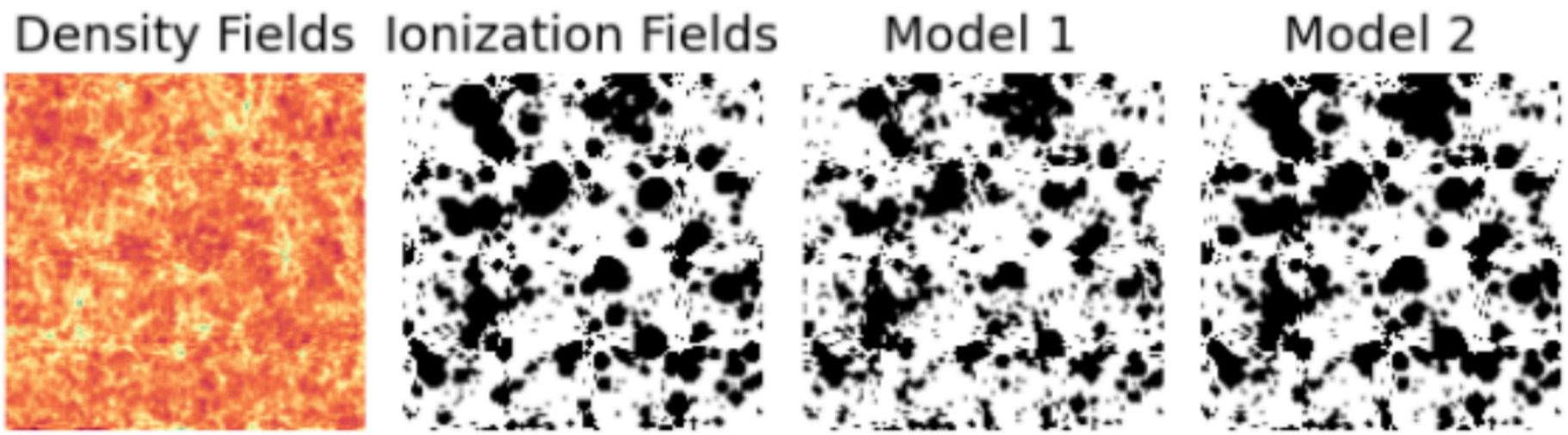
Testing Protocol?

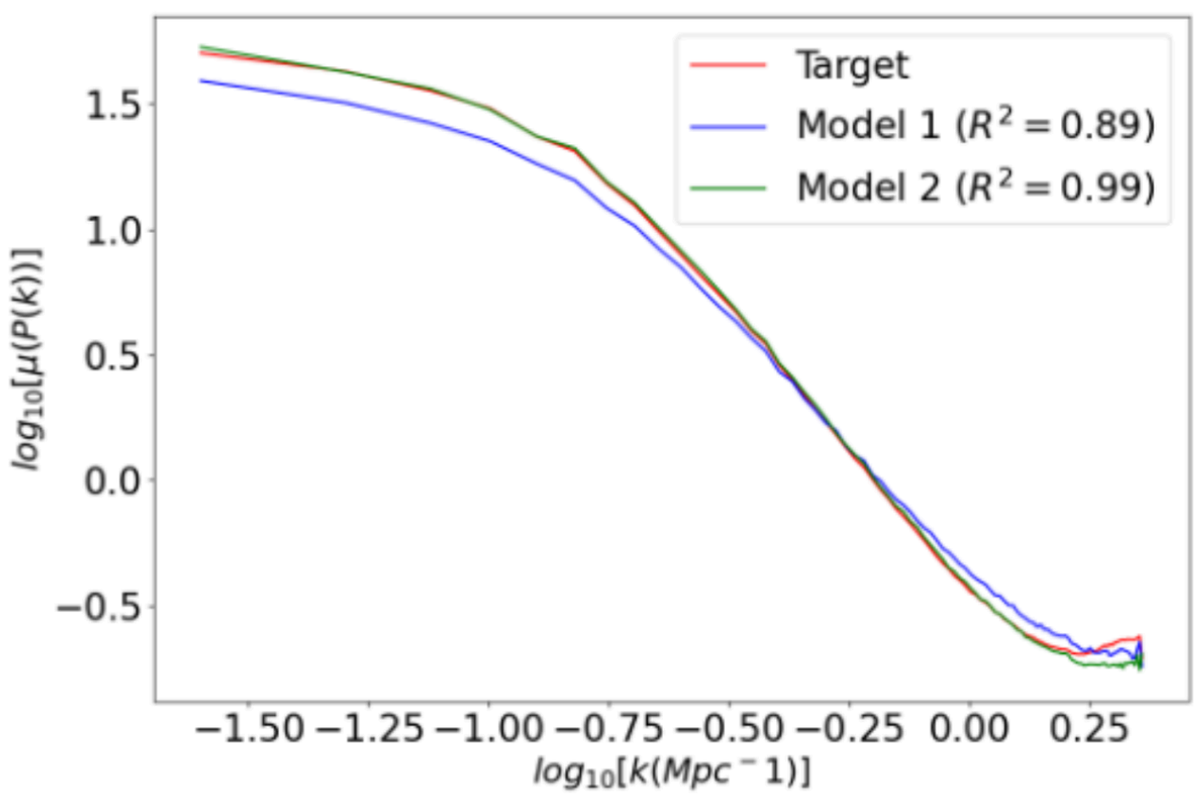


Emulating Radiation Transport on Cosmological Scales using a Denoising U-Net

Mosima Masipa, MSc student

University of the Western Cape, South Africa



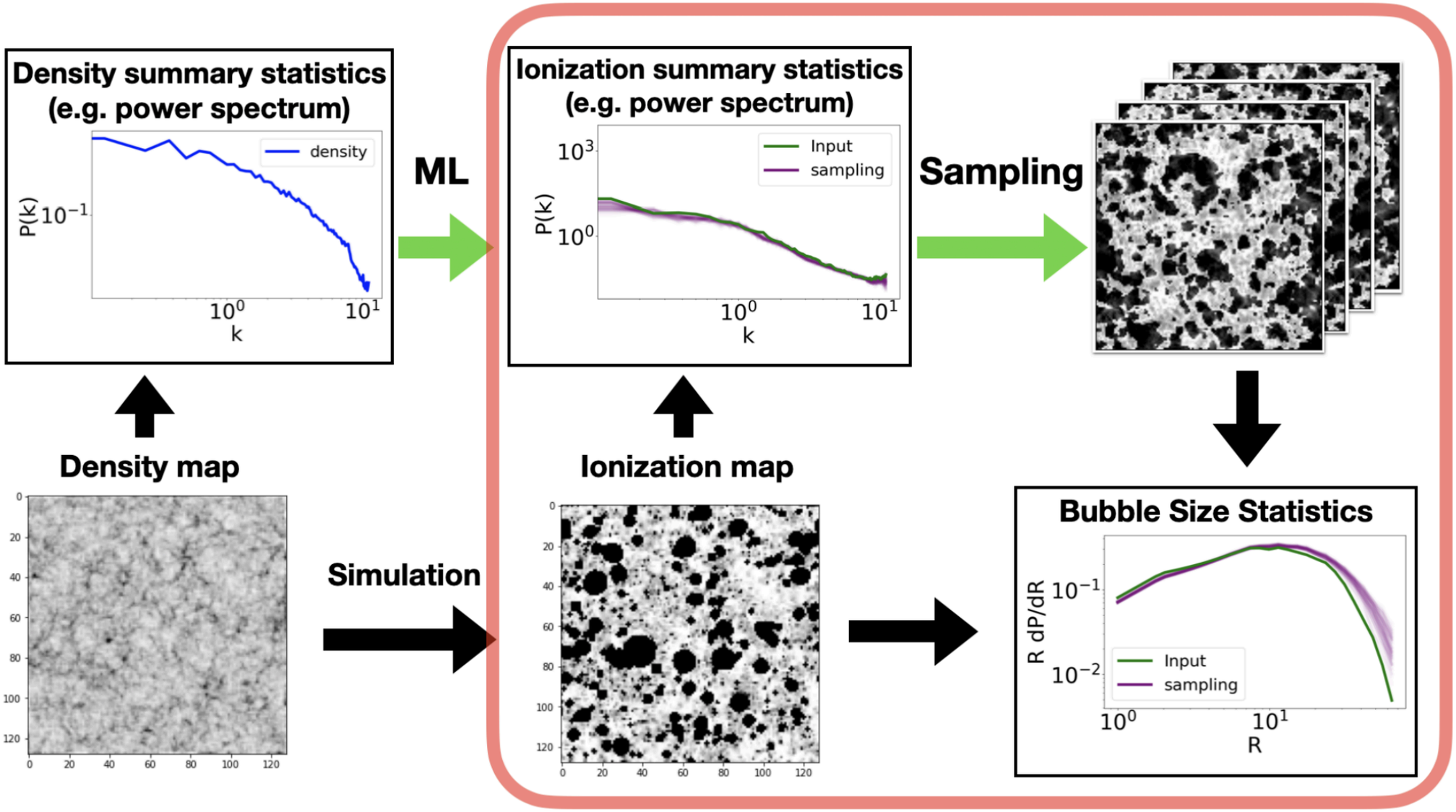


Towards a non-Gaussian generative model of large scale reionization maps (Lin, Hassan, Blancard, Eickenberg, Modi 2022, accepted to NeurIPS 2022).

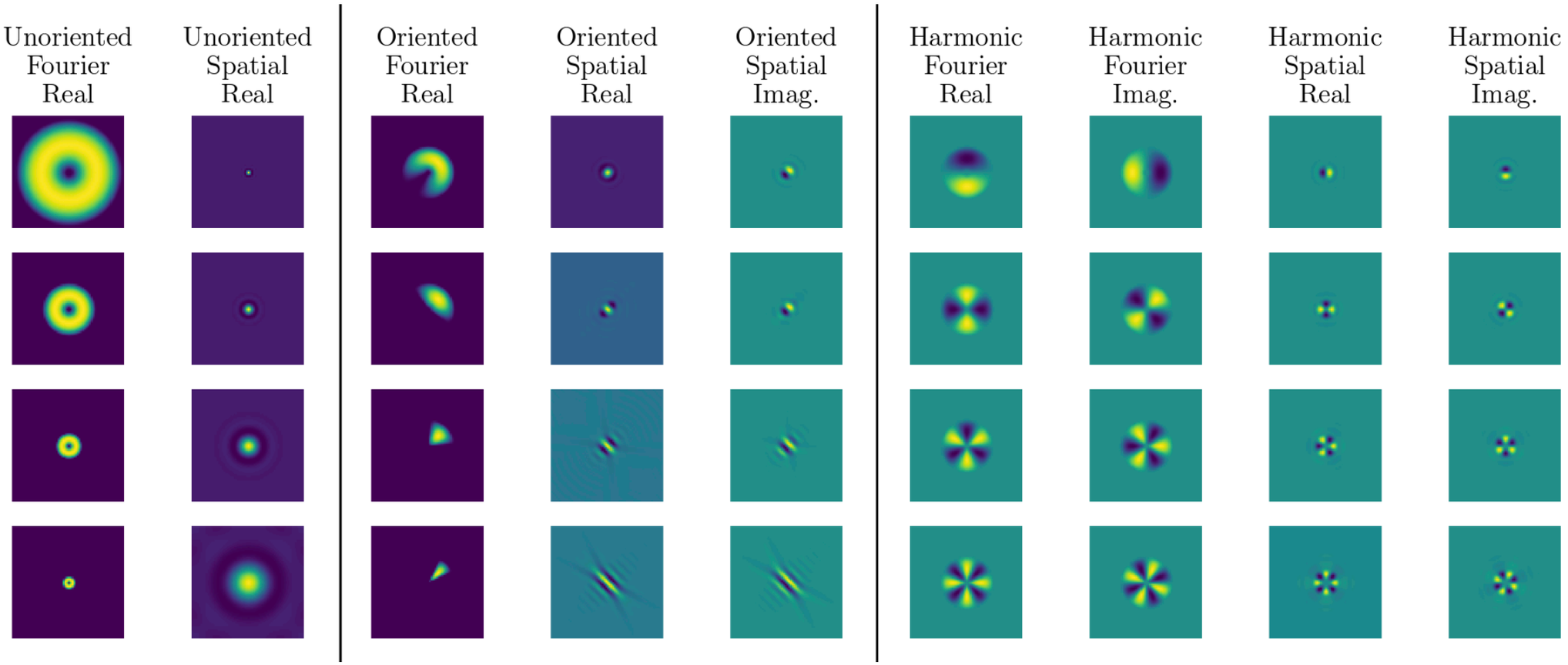
Towards a non-Gaussian Generative Model of large-scale Reionization Maps

Yu-Heng Lin, PhD student

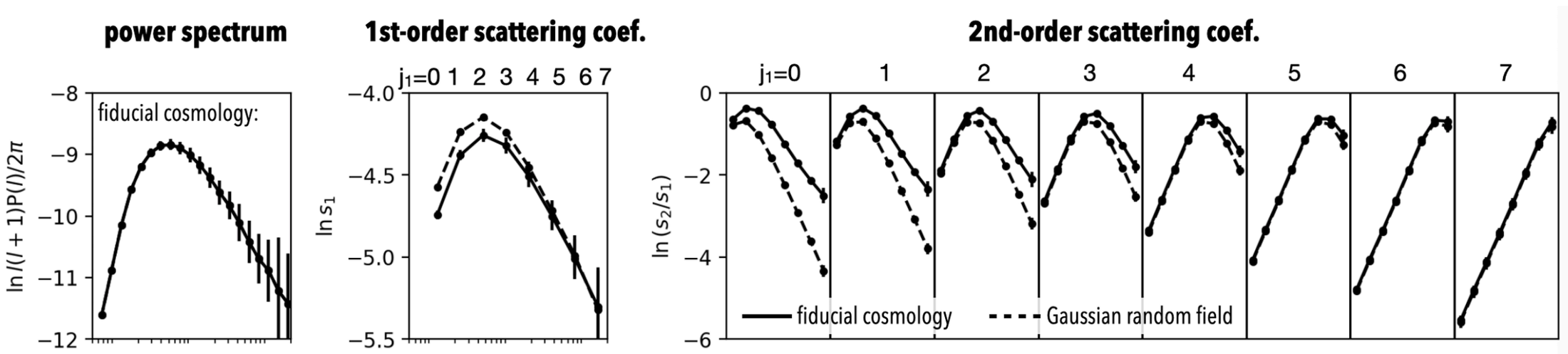
University of Minnesota



Wavelet Scattering Transforms capture non-Gaussiantiy —> optimal summary statistic



Eickenberg et al, incl. Hassan 2022



Cheng et al 2020

Towards a non-Gaussian Generative Model of large-scale Reionization Maps

Yu-Heng Lin, PhD student

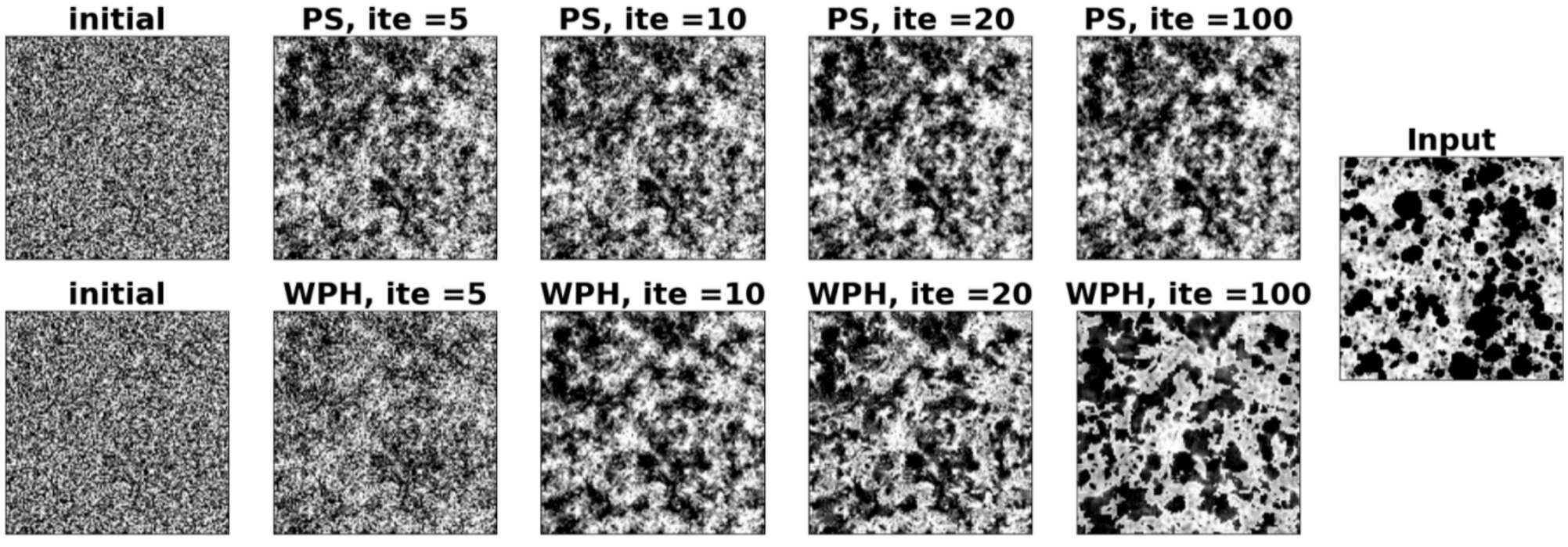
University of Minnesota

Steps:

1 - obtain your favorite summary statistics of the input!
We tried power spectrum and Wavelet Phase Harmonics!

2 - Generate a white noise and obtain the same summary statistics.

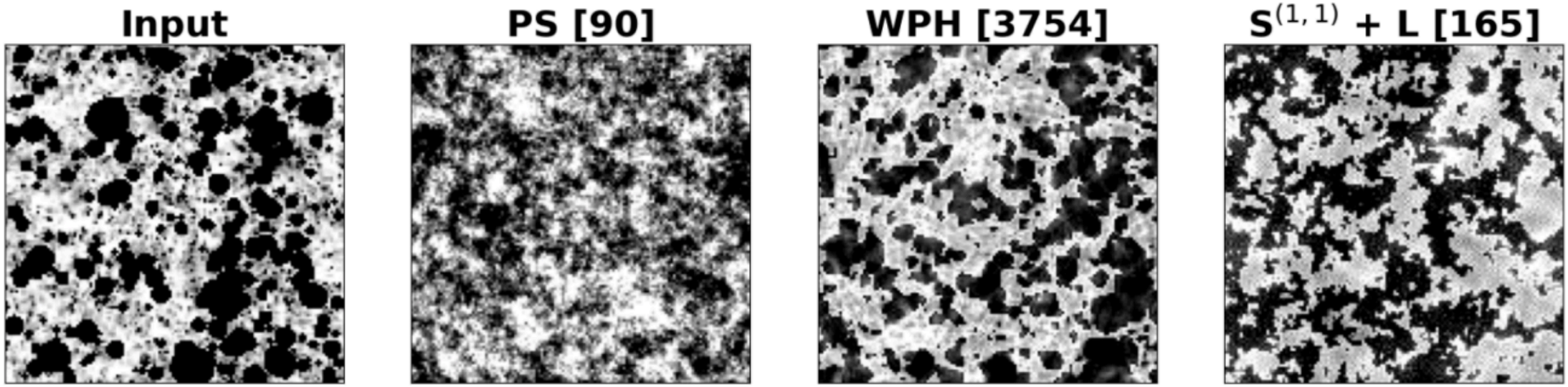
3 - Optimize the noise to minimize the following loss function: $L_i(u) = |\phi_i(u) - \phi_i(s)|^2$

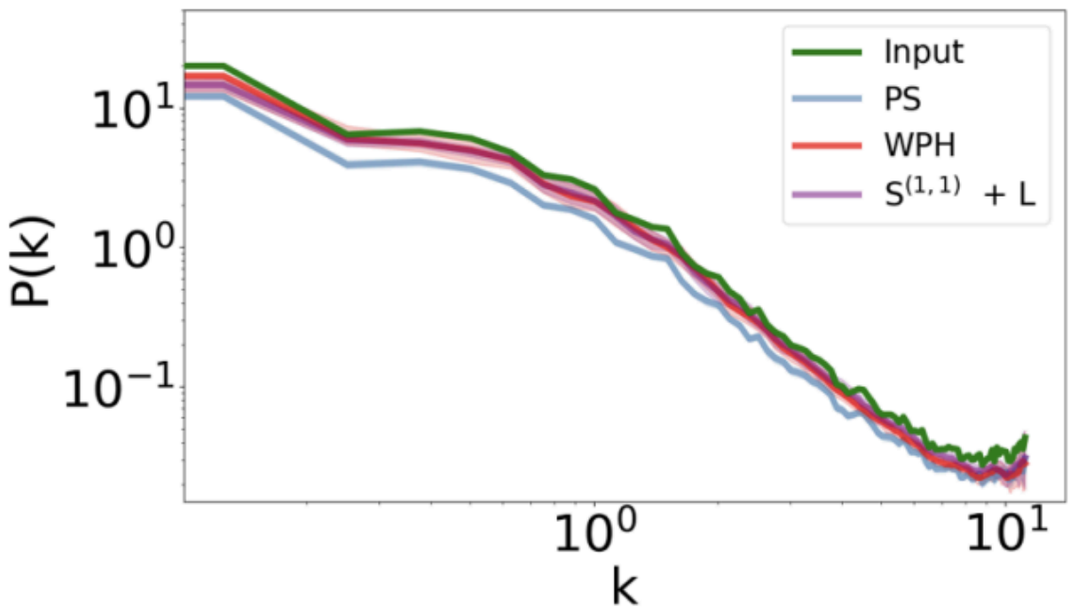


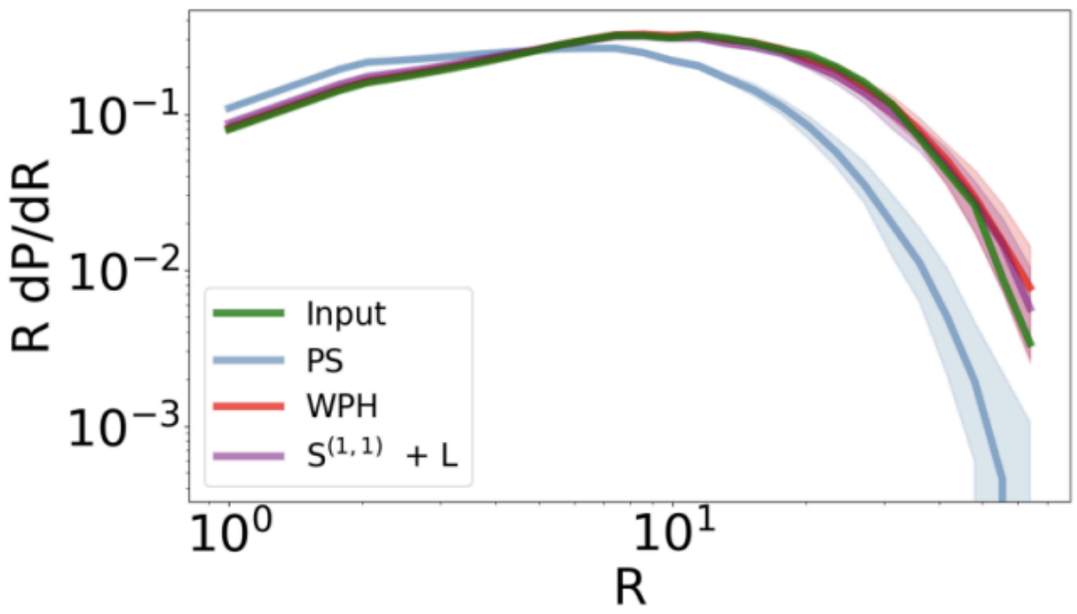
Towards a non-Gaussian Generative Model of large-scale Reionization Maps

Yu-Heng Lin, PhD student

University of Minnesota







Invertible Mapping between fields in CAMELS (Andrianomena, Hassan, Paco 2023, accepted to ICLR 2023).

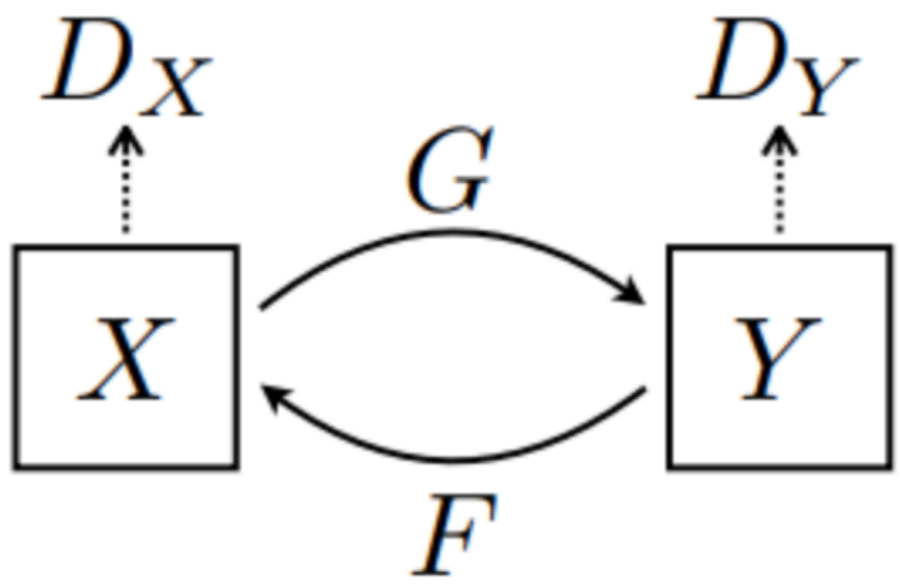
Invertible Mapping between fields in CAMELS

Sambatra Andrianomena
Operations Scientist
South African Radio Astronomy Observatory

For Multi-Messenger surveys :

The aim is to train CycleGAN to convert

- Dark matter (Mcdm) to HI (Neutral Hydrogen)
- Mcdm to B (Magnetic field)
- HI to B

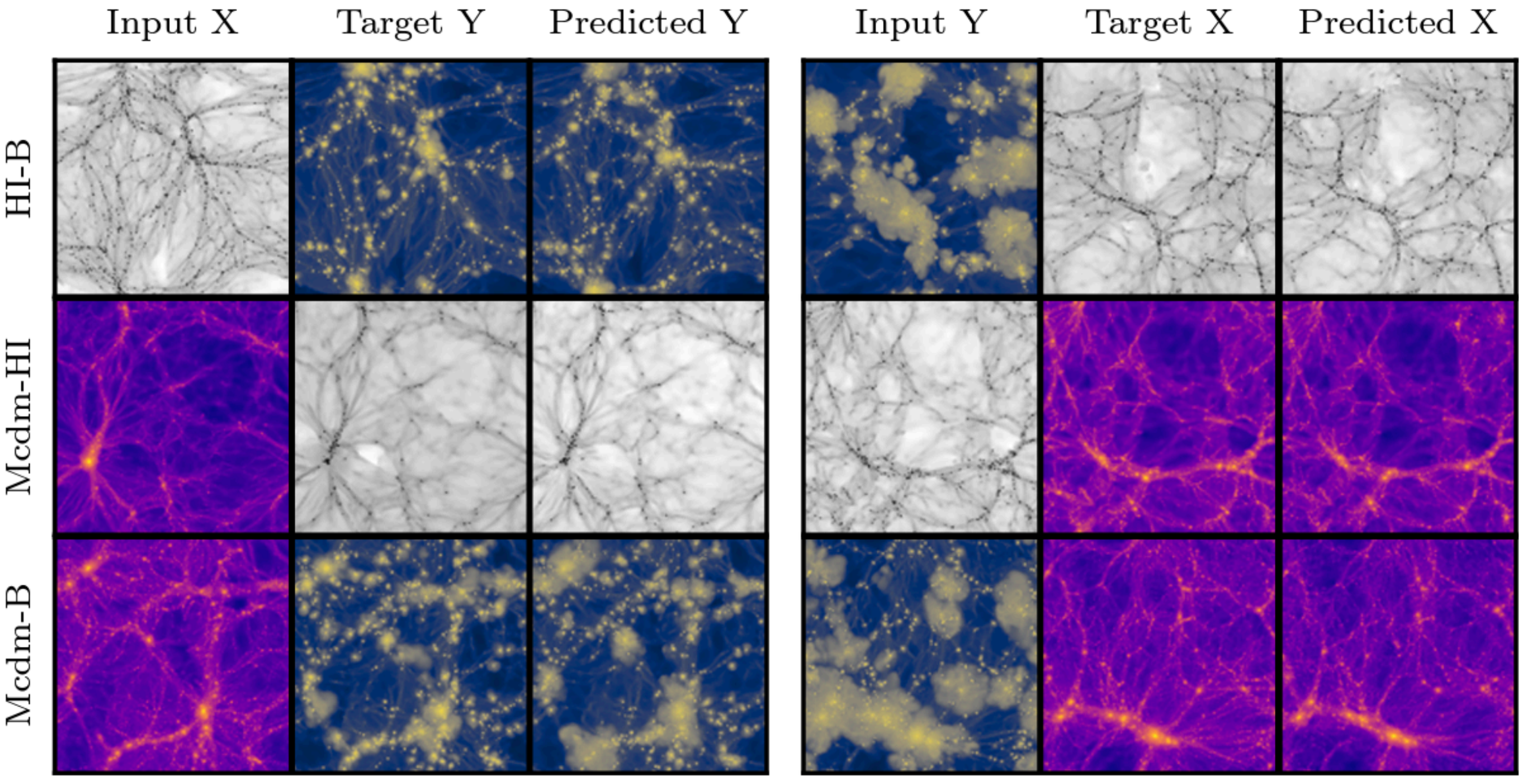


$$\mathcal{L}_{\text{tot}} = \mathcal{L}_{\text{GAN}}(G_X, D_X, Y, X) + \mathcal{L}_{\text{GAN}}(G_X, D_X, Y, X) + \lambda_{\text{cycle}}\mathcal{L}_{\text{cycle}} + \lambda_{\text{id}}\mathcal{L}_{\text{id}}$$

Diagram credit: https://www.tensorflow.org/tutorials/generative/cyclegan

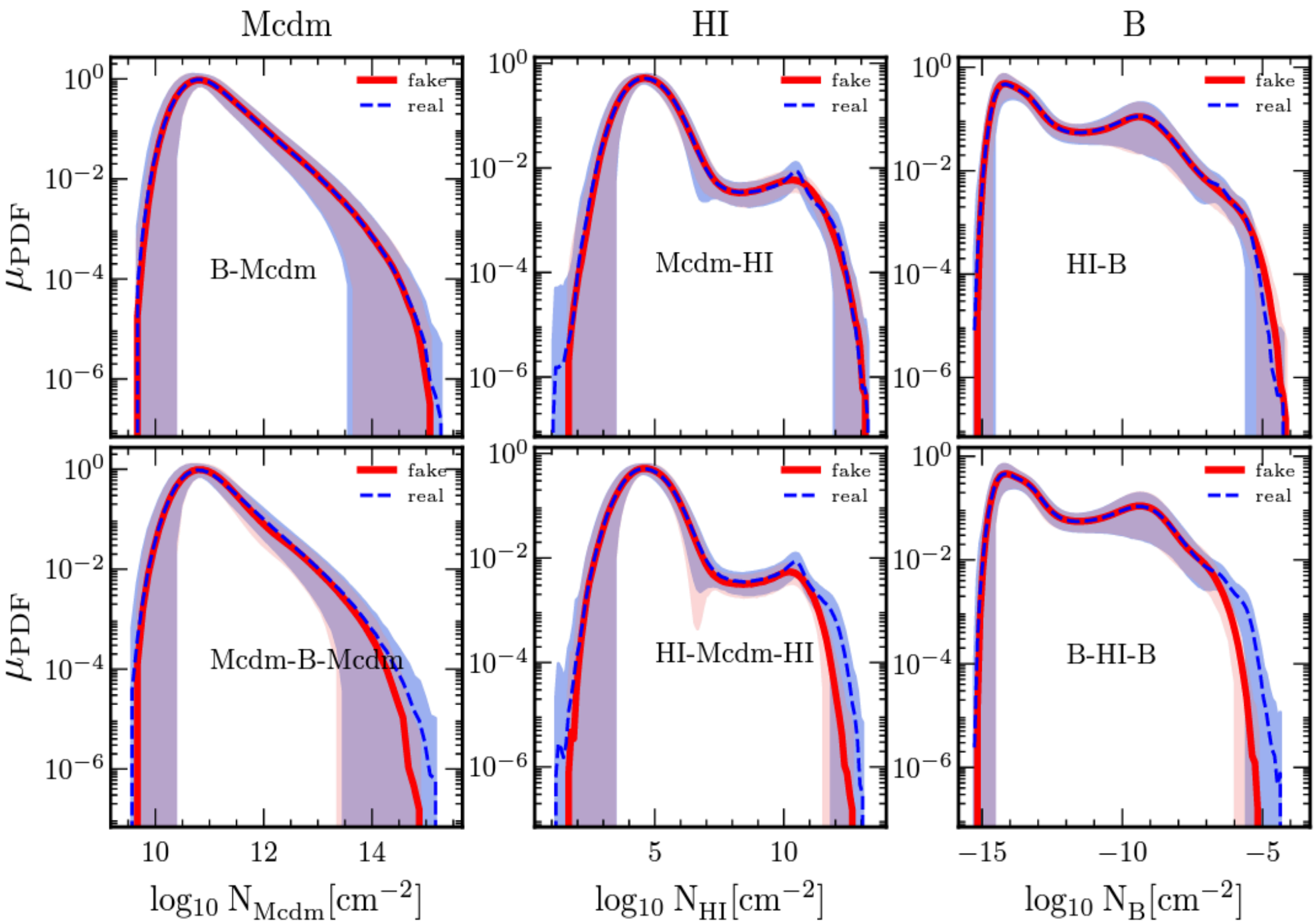
Invertible Mapping between fields in CAMELS

Sambatra Andrianomena
Operations Scientist
South African Radio
Astronomy Observatory

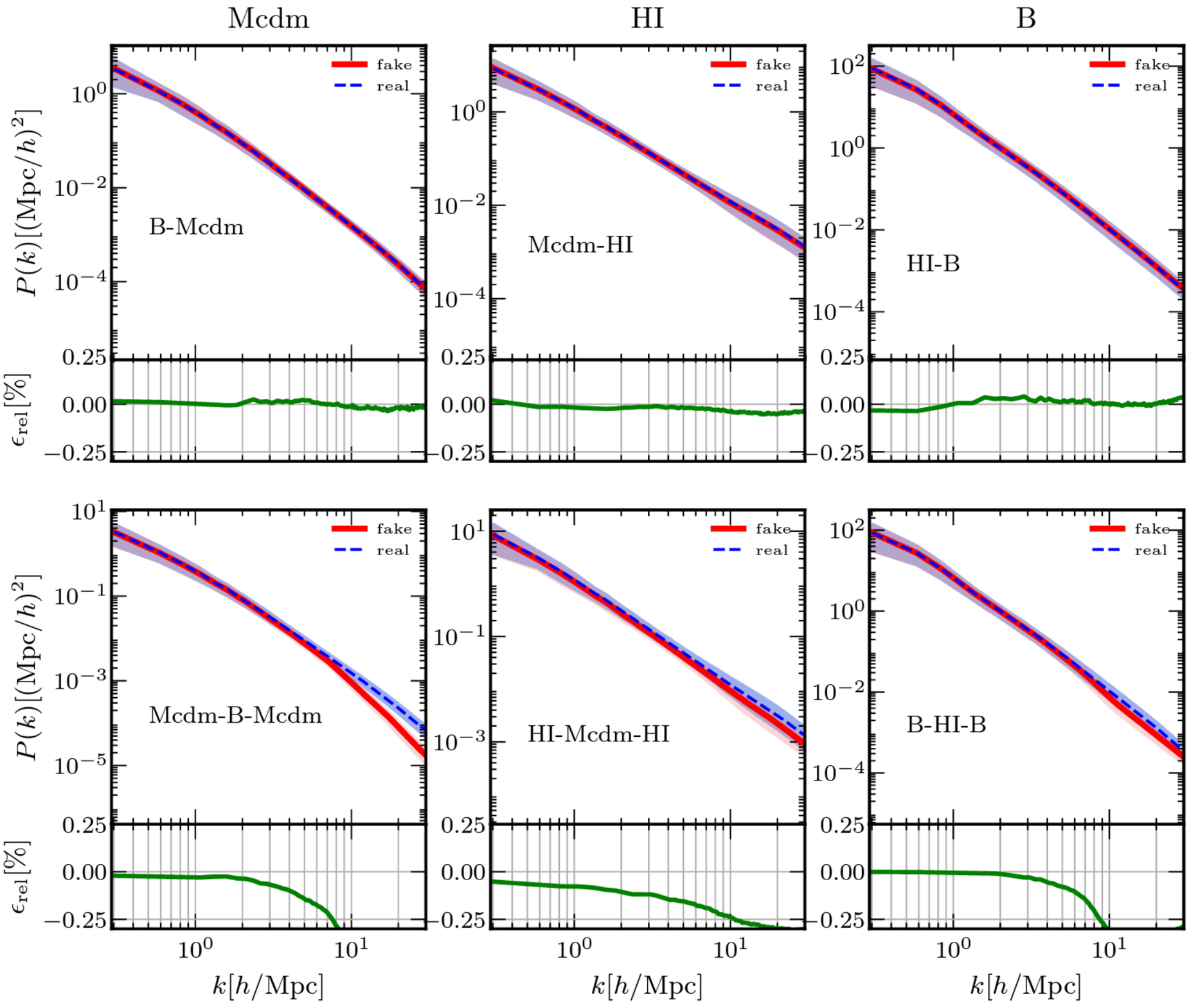


Invertible Mapping between fields in CAMELS

Sambatra Andrianomena
Operations Scientist
South African Radio Astronomy Observatory



Invertible Mapping between fields in CAMELS



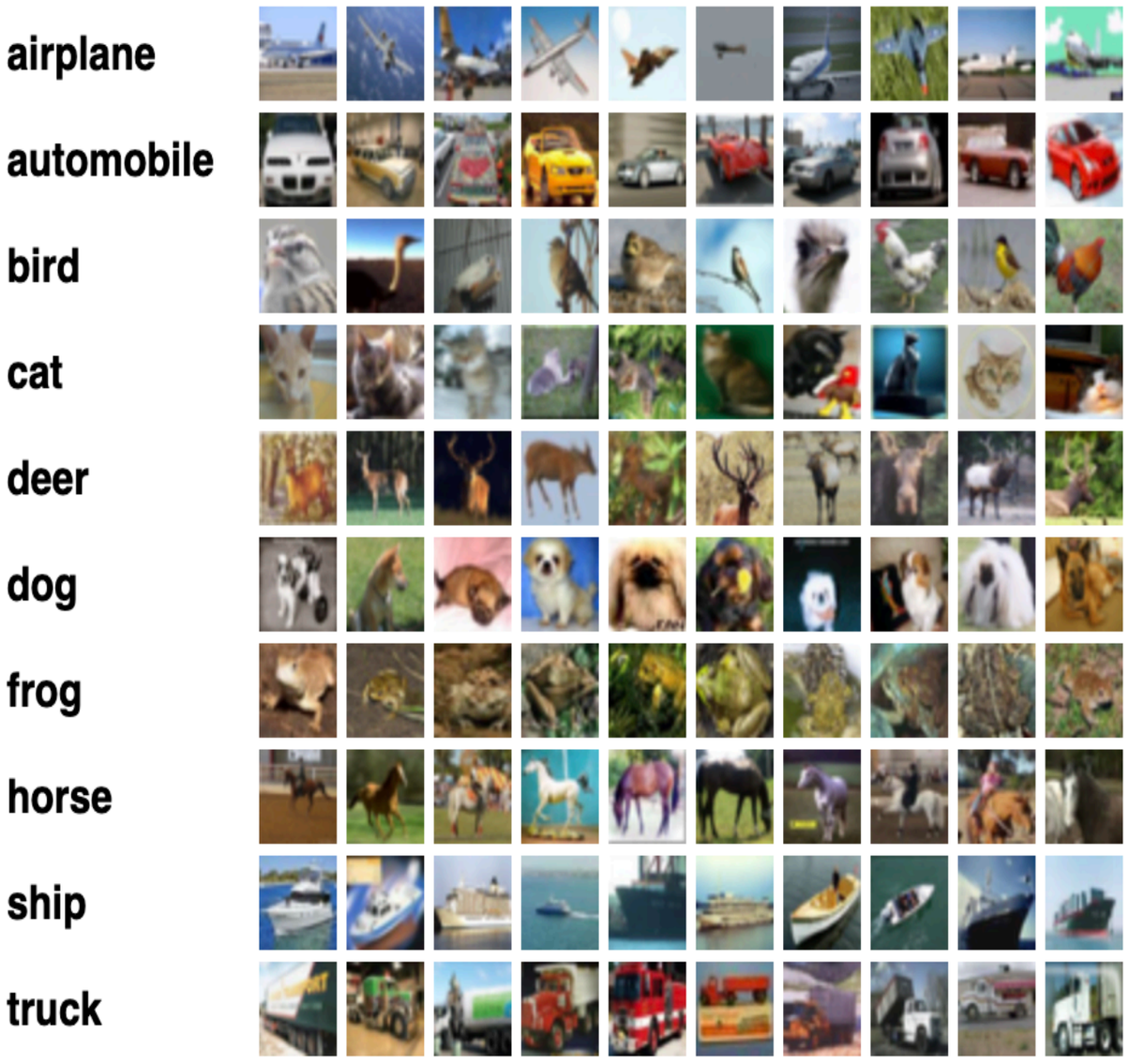
Sambatra Andrianomena
Operations Scientist
South African Radio Astronomy Observatory

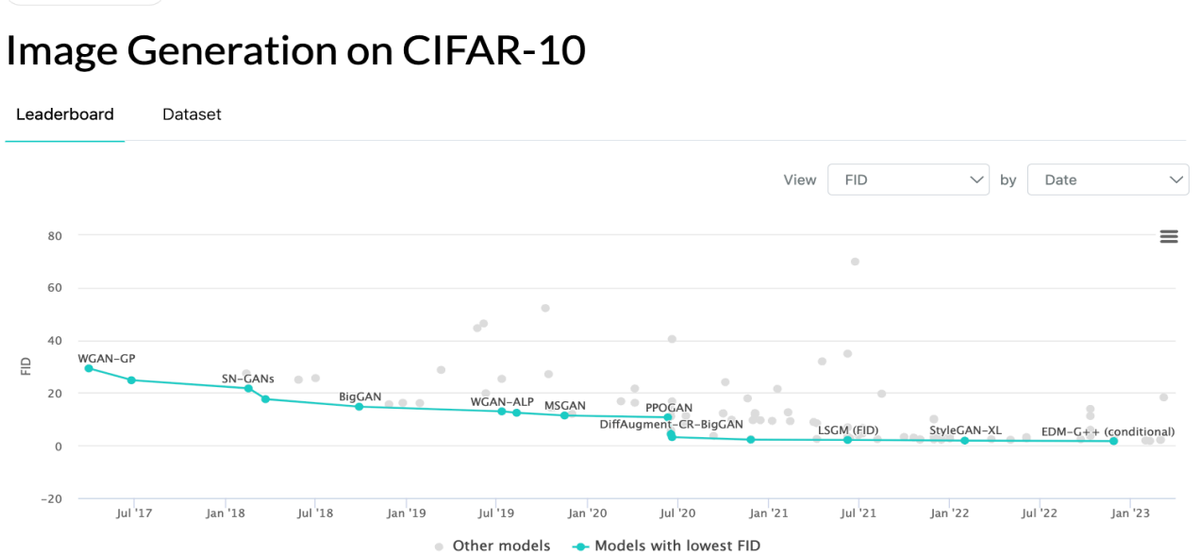
Generating high-fidelity HI maps using score-based diffusion models (Started here, Hassan, Wu, Lovell, Cooray + in prep, join us in slack @ galevo23-p11)

Why diffusion models?

On CIFAR-10 alone:

- Best ever model.
- 3 out of the first 5.
- 6 out of the first 10.





Filter: GAN | Score-based | VAE | Flow-based | Diffusion | Transformer | PFGM | unconditional | ResNet | Optimal Transport | EBM | untagged

Edit Leaderboard

Rank	Model	FID	NFE	Inception score	bits/dimension	Paper	Code	Result	Year	Tags
1	**EDM-G++** (conditional)	1.64	35			Refining Generative Process with Discriminator Guidance in Score-based Diffusion Models			2022	Diffusion, Score-based
2	**PFGM++**	1.74				PFGM++: Unlocking the Potential of Physics-Inspired Generative Models			2023	PFGM
3	**EDM-G++** (unconditional)	1.77	35		2.55	Refining Generative Process with Discriminator Guidance in Score-based Diffusion Models			2022	Diffusion, Score-based
4	**StyleGAN-XL**	1.85				StyleGAN-XL: Scaling StyleGAN to Large Diverse Datasets			2022	GAN
5	**STF** (unconditional)	1.90				Stable Target Field for Reduced Variance Score Estimation in Diffusion Models			2023	Score-based

https://paperswithcode.com/sota/image-generation-on-cifar-10

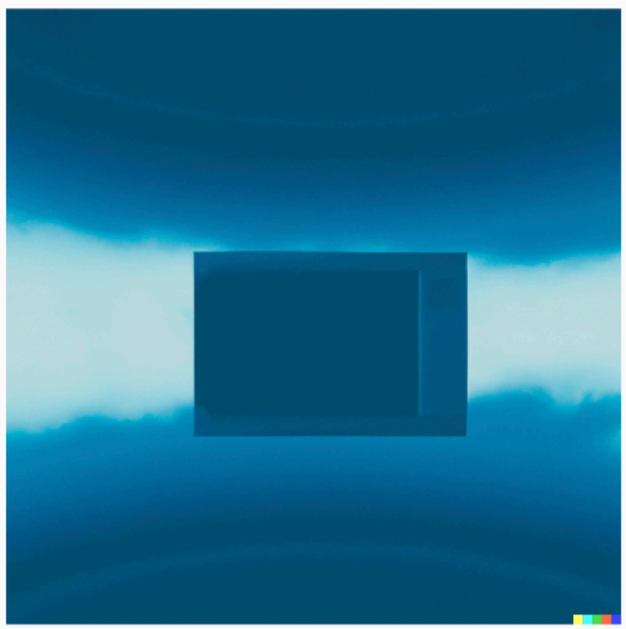
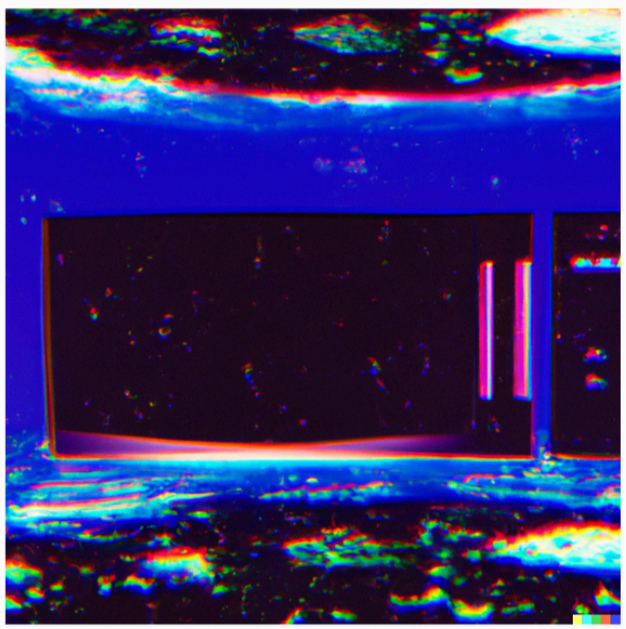
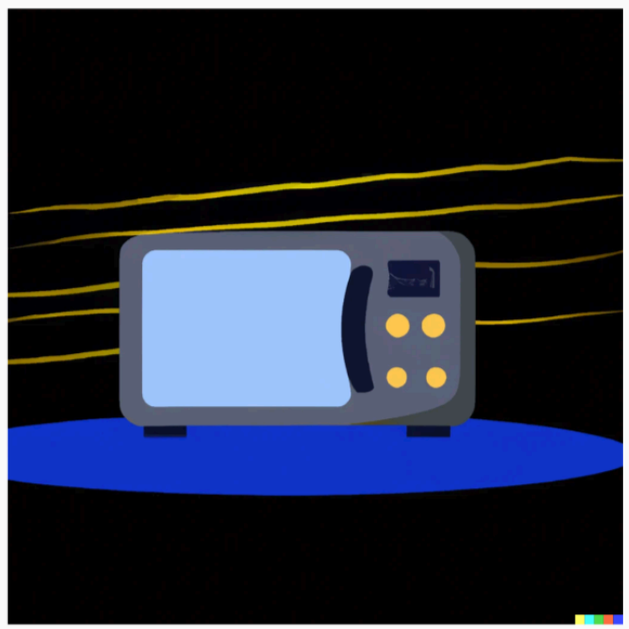
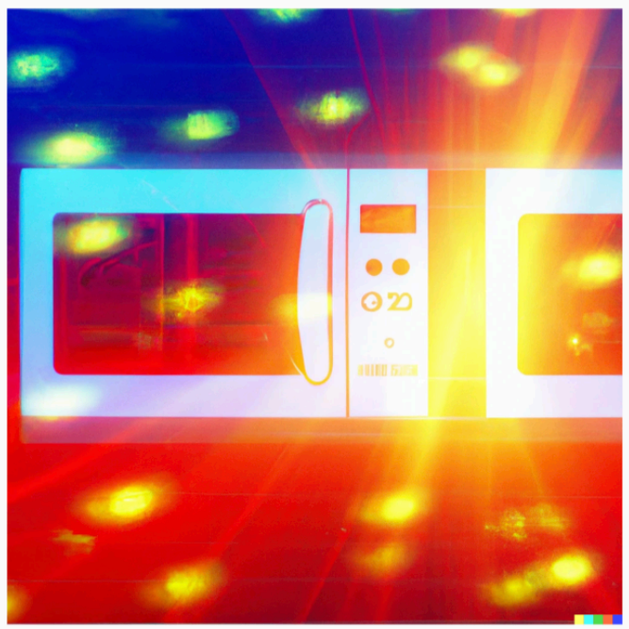
Why diffusion models?
DALL.E and ChatGPT (ask Jo! Its addictive!)

Why diffusion models?
DALL.E and ChatGPT (ask Jo! Its addictive!)

I asked DALL.E to generate the Cosmic Microwave Background (CMB), and

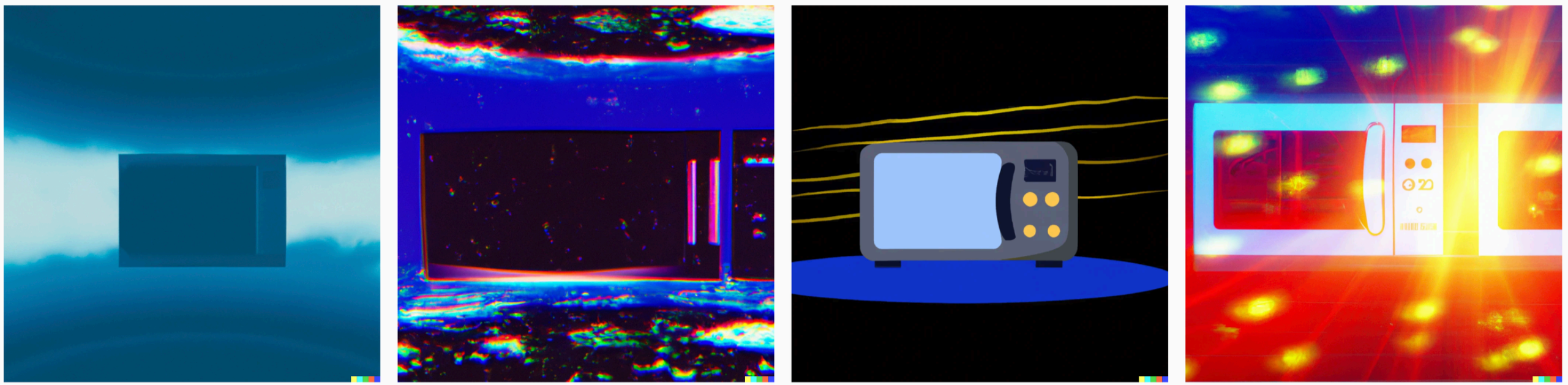
Why diffusion models?
DALL.E and ChatGPT (ask Jo! Its addictive!)

I asked DALL.E to generate the Cosmic Microwave Background (CMB), and BOOM!

Why diffusion models?
DALL.E and ChatGPT (ask Jo! Its addictive!)

I asked DALL.E to generate the Cosmic Microwave Background (CMB), and BOOM!

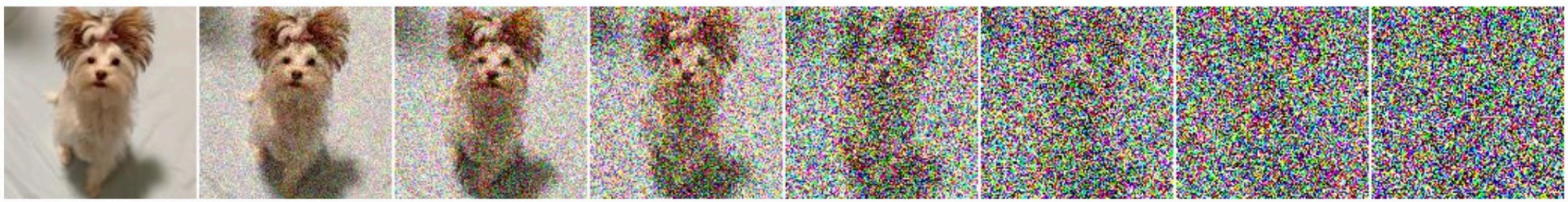


But an artist with zero physics can draw the same!
If only given ‘‘CMB’’ as a text.

Score-based diffusion models

Forward SDE (data → noise)

$\mathbf{x}(0)$ — $\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$ → $\mathbf{x}(T)$



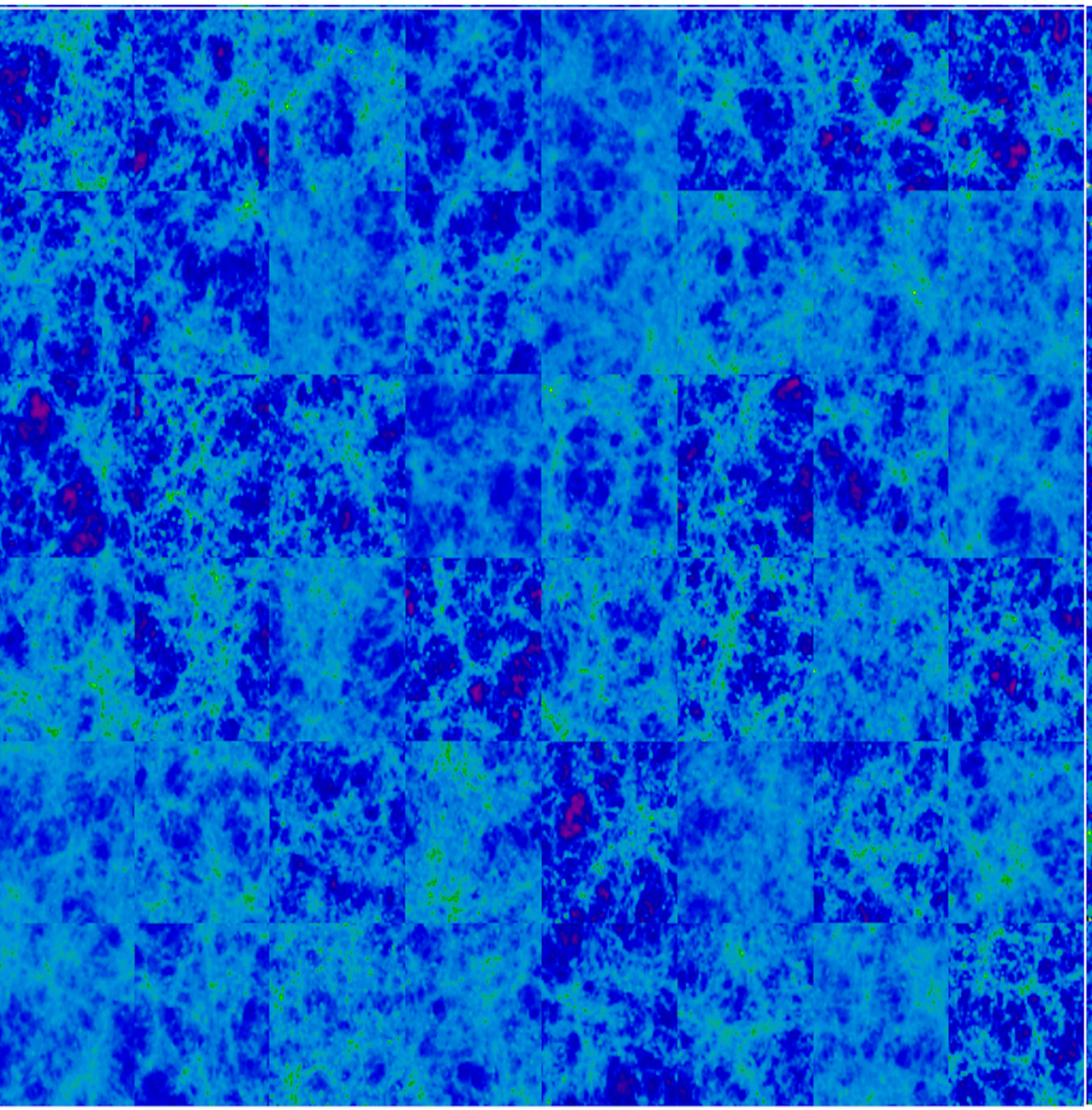
score function

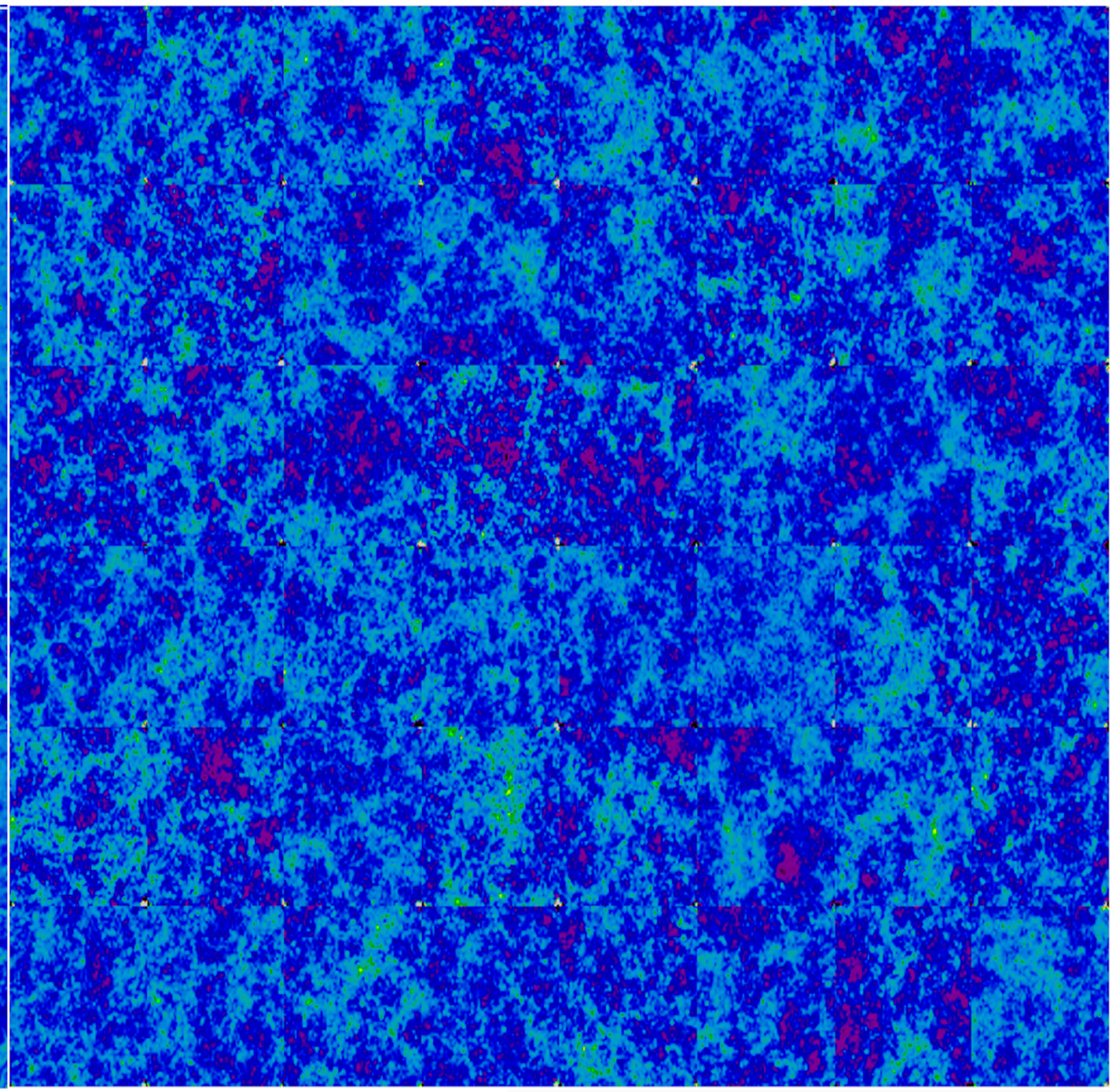
$\mathbf{x}(0)$ ← $\mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g^2(t)\nabla_{\mathbf{x}} \log p_t(\mathbf{x})\right] \mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}$ — $\mathbf{x}(T)$

Reverse SDE (noise → data)

See tutorials here: https://github.com/yang-song/score_sde_pytorch

Getting there! Started 10 days back!
Which one is CAMELS?

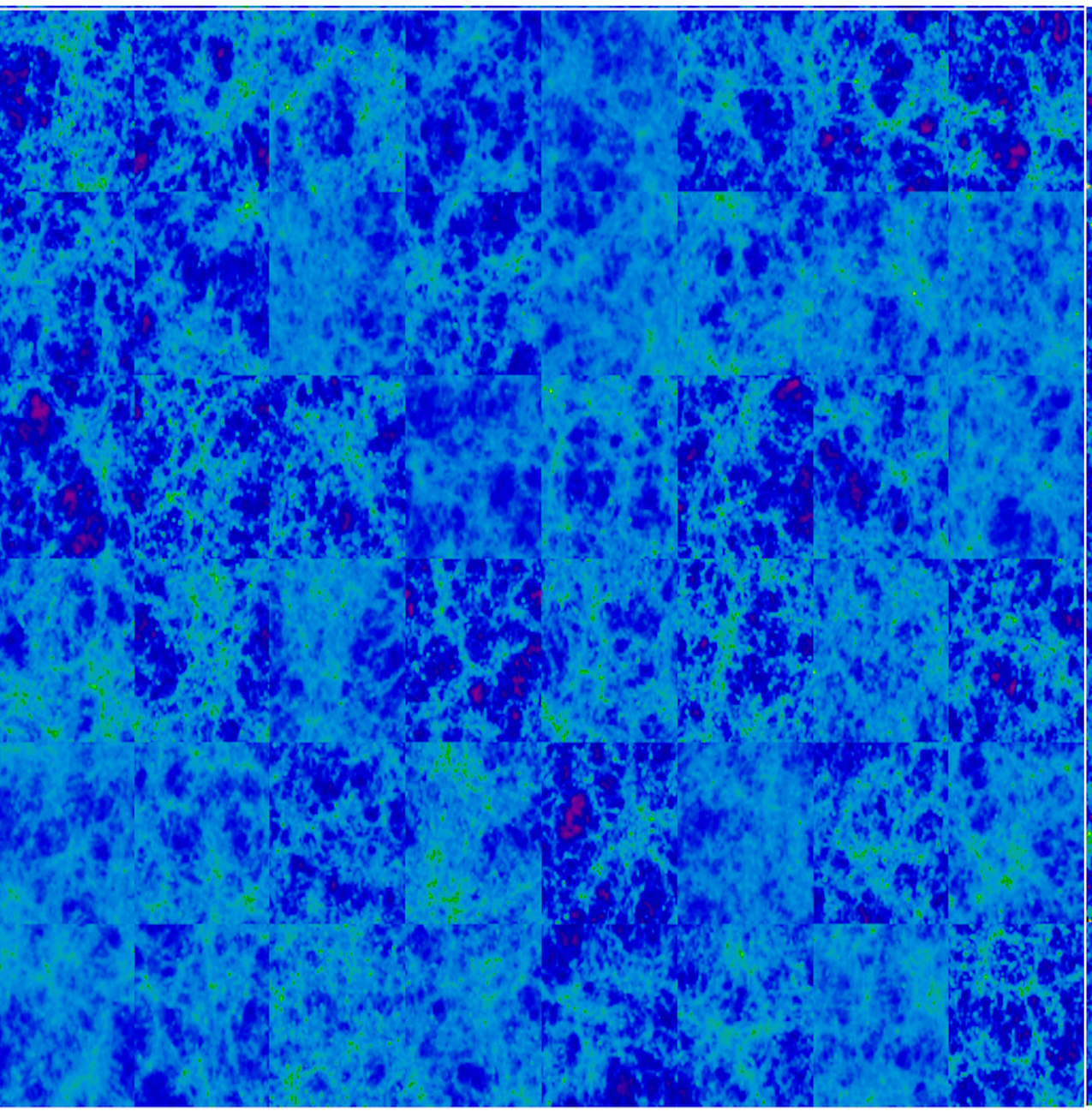




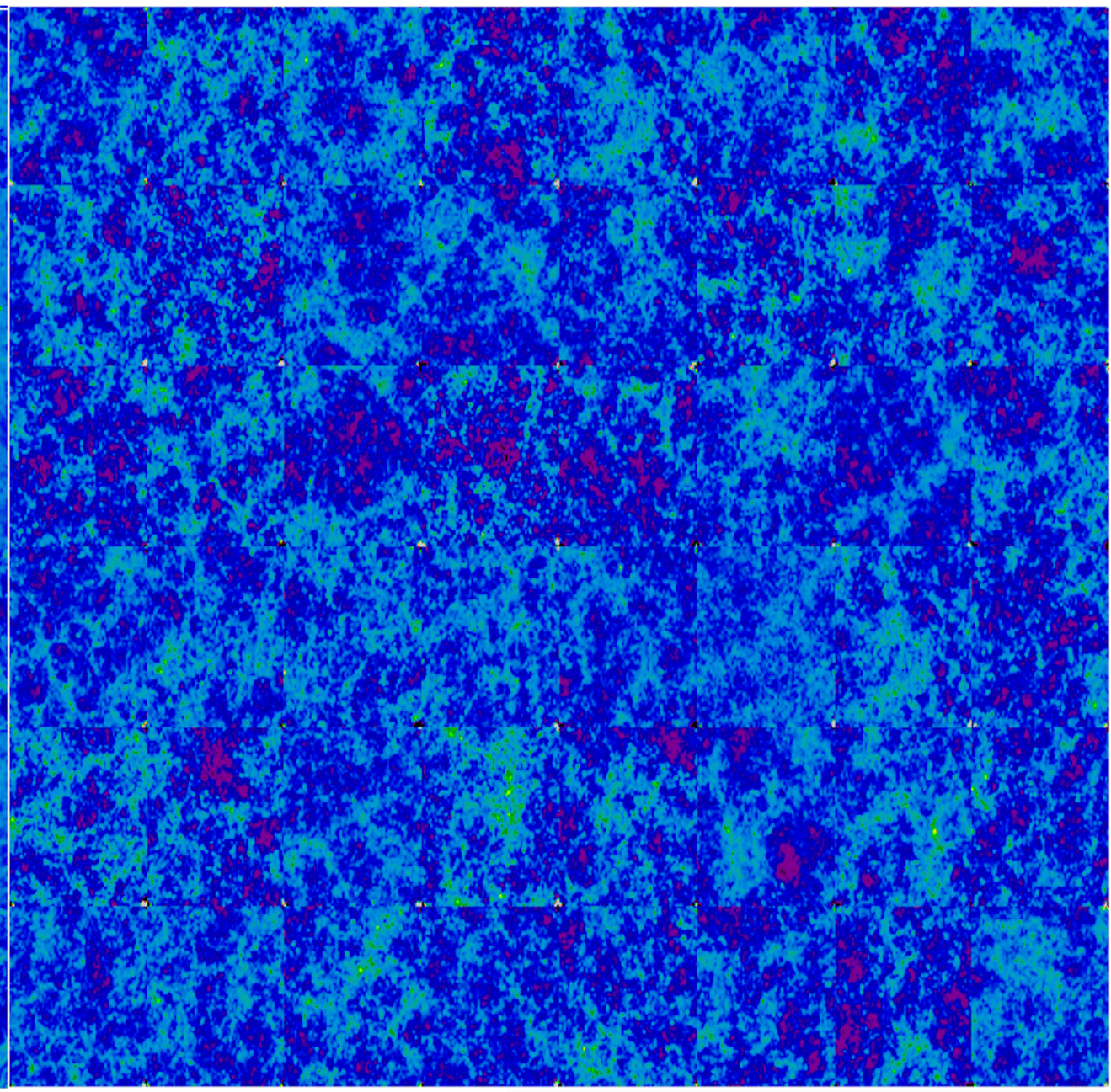
Getting there! Started 10 days back!

Which one is CAMELS?

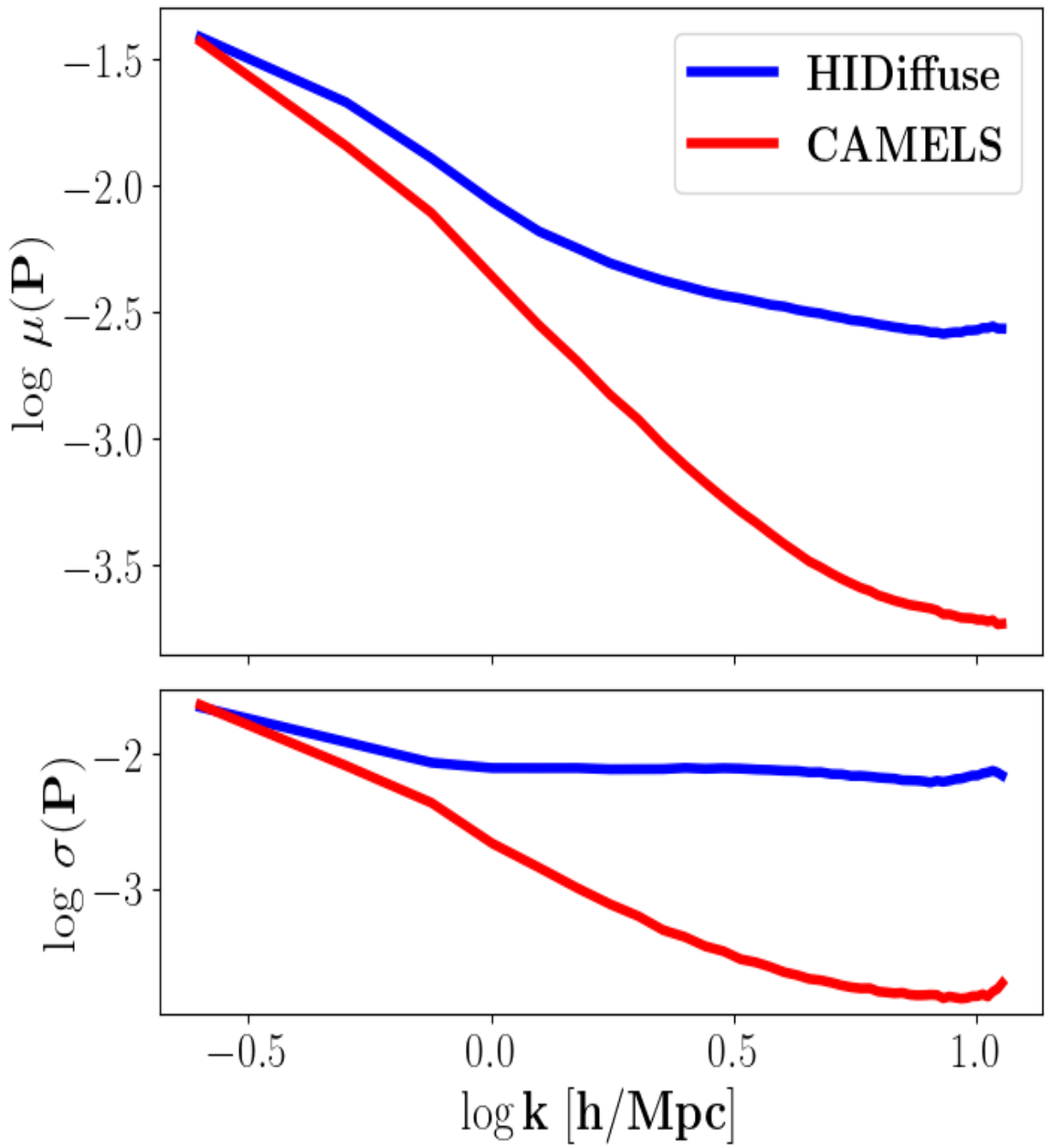
CAMELS

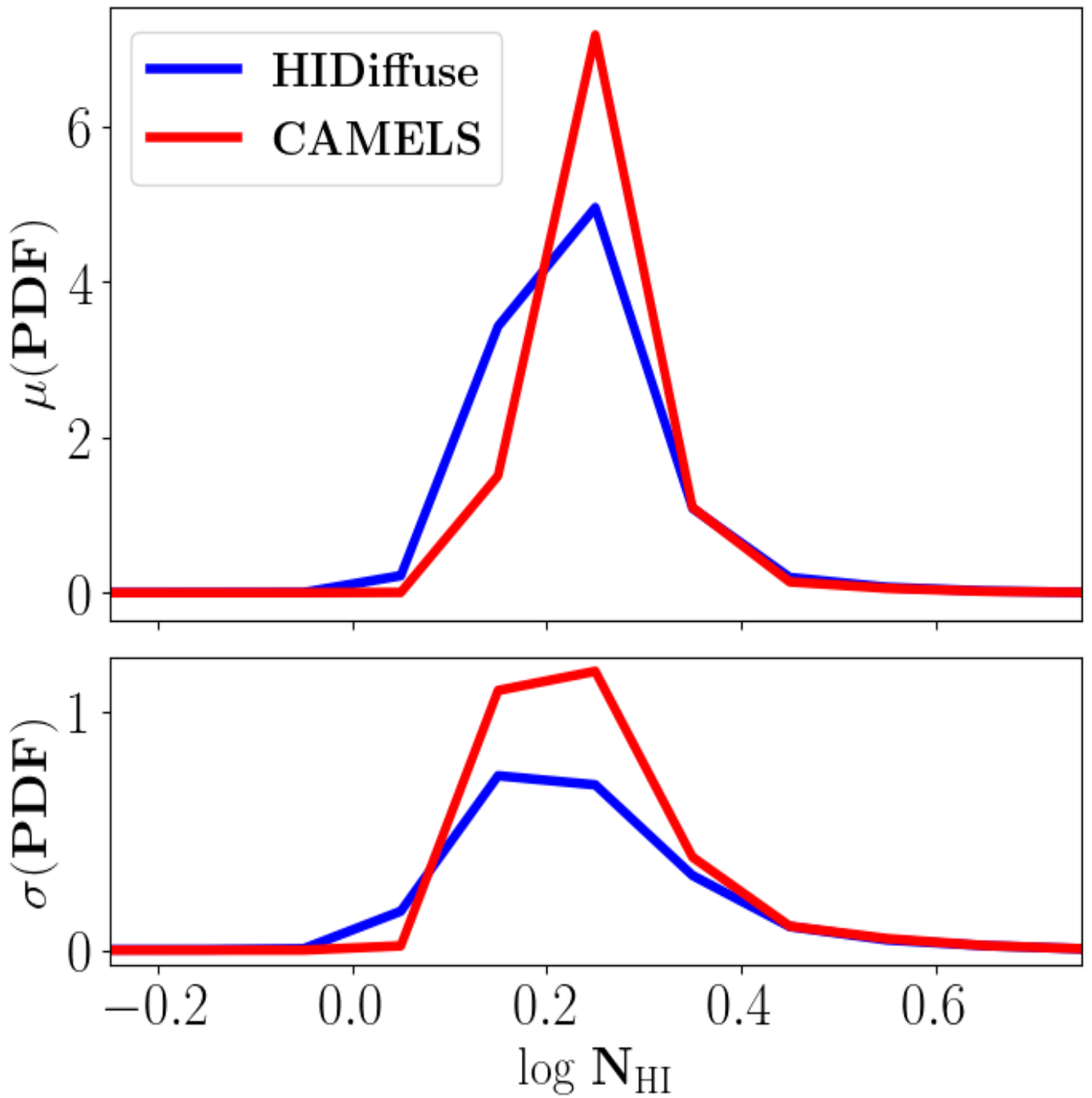


HIDiffuse



So far, somewhat recovering the PDFs, but not the powers!





Generative models are great

Questions/Comments welcome :)