# MAP-TO-MAP TRANSLATION OF SIMULATED GALAXIES WITH CONDITIONAL GANS


SKA research at
Zurich University of Applied Sciences (ZHAW)

Centre for Artificial Intelligence (CAI)
Institute for Business Information Technology (IWI)


Neuchâtel, 2024/01/22 Mon


Philipp Denzel, Frank-Peter Schilling, Elena Gavagnin

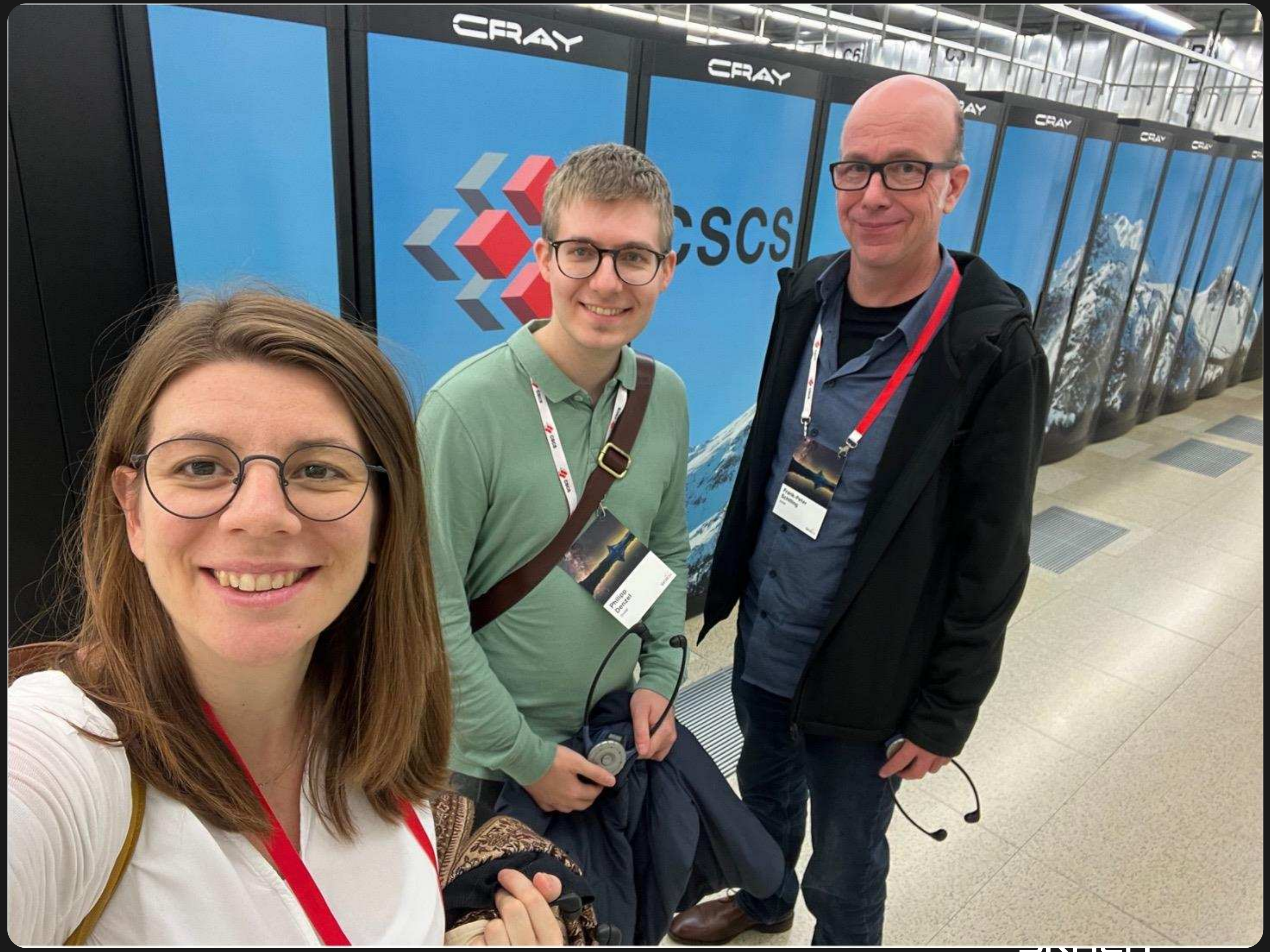
# SLIDES ON MY WEBSITE

https://phdenzel.github.io/

Link/QR code to the slides for later or to follow along

# PROJECTS AT ZHAW

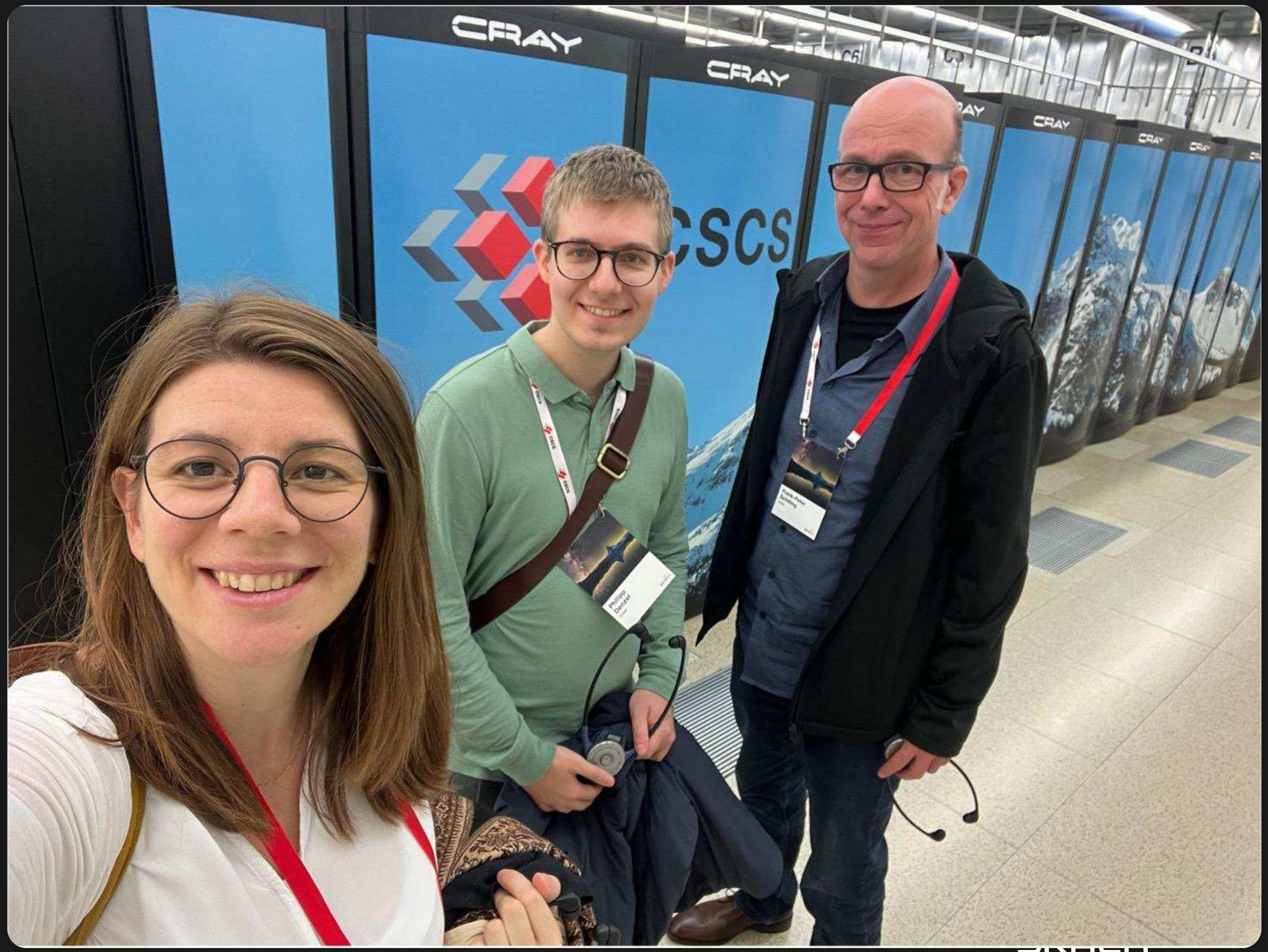
ZHAW's SKACH team at CSCS in Lugano



# PROJECTS AT ZHAW

- SKA project:
  - trained (astro)physicists, focused on ML research

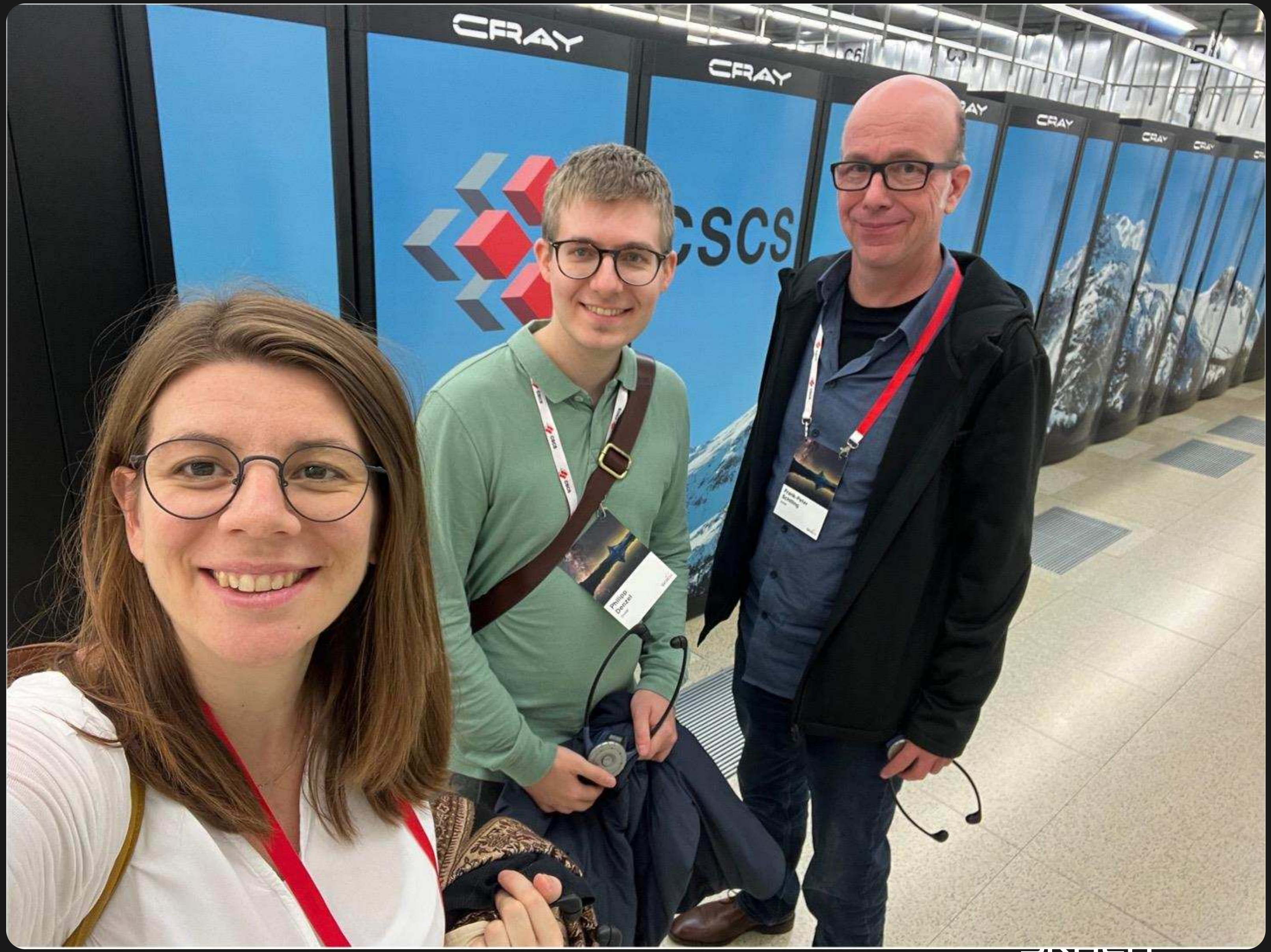
ZHAW's SKACH team at CSCS in Lugano



# PROJECTS AT ZHAW

- SKA project:
  - trained (astro)physicists, focused on ML research
- our expertise:
  - **deep generative modelling** of (sky) simulations
  - CV, DL, XAI, MLOps, …

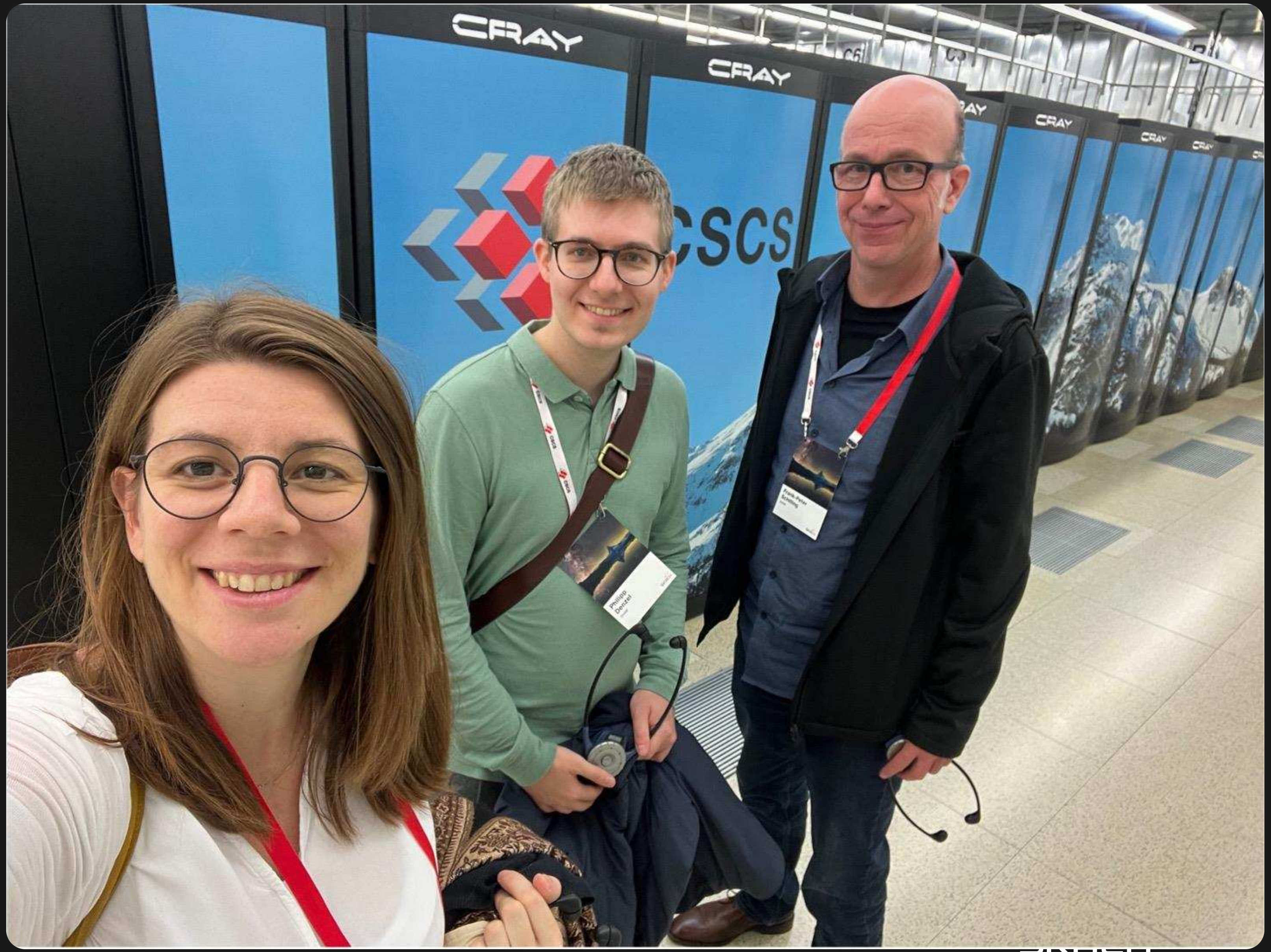
ZHAW's SKACH team at CSCS in Lugano



# PROJECTS AT ZHAW

- SKA project:
  - trained (astro)physicists, focused on ML research
- our expertise:
  - **deep generative modelling** of (sky) simulations
  - CV, DL, XAI, MLOps, …
- recently expanded efforts
  - two new projects

ZHAW's SKACH team at CSCS in Lugano



# OUTLOOK

- Map-to-map translation of simulations
- Point-cloud experiments
- Radio source classification

# THE TIMES THEY ARE A-CHANGIN'

Zürcher Hochschule
für Angewandte Wissenschaften

# THE TIMES THEY ARE A-CHANGIN'

- the *end of the analytic era*

# THE TIMES THEY ARE A-CHANGIN'

- the *end of the analytic era*
- modern surveys: galaxies are no longer blobs

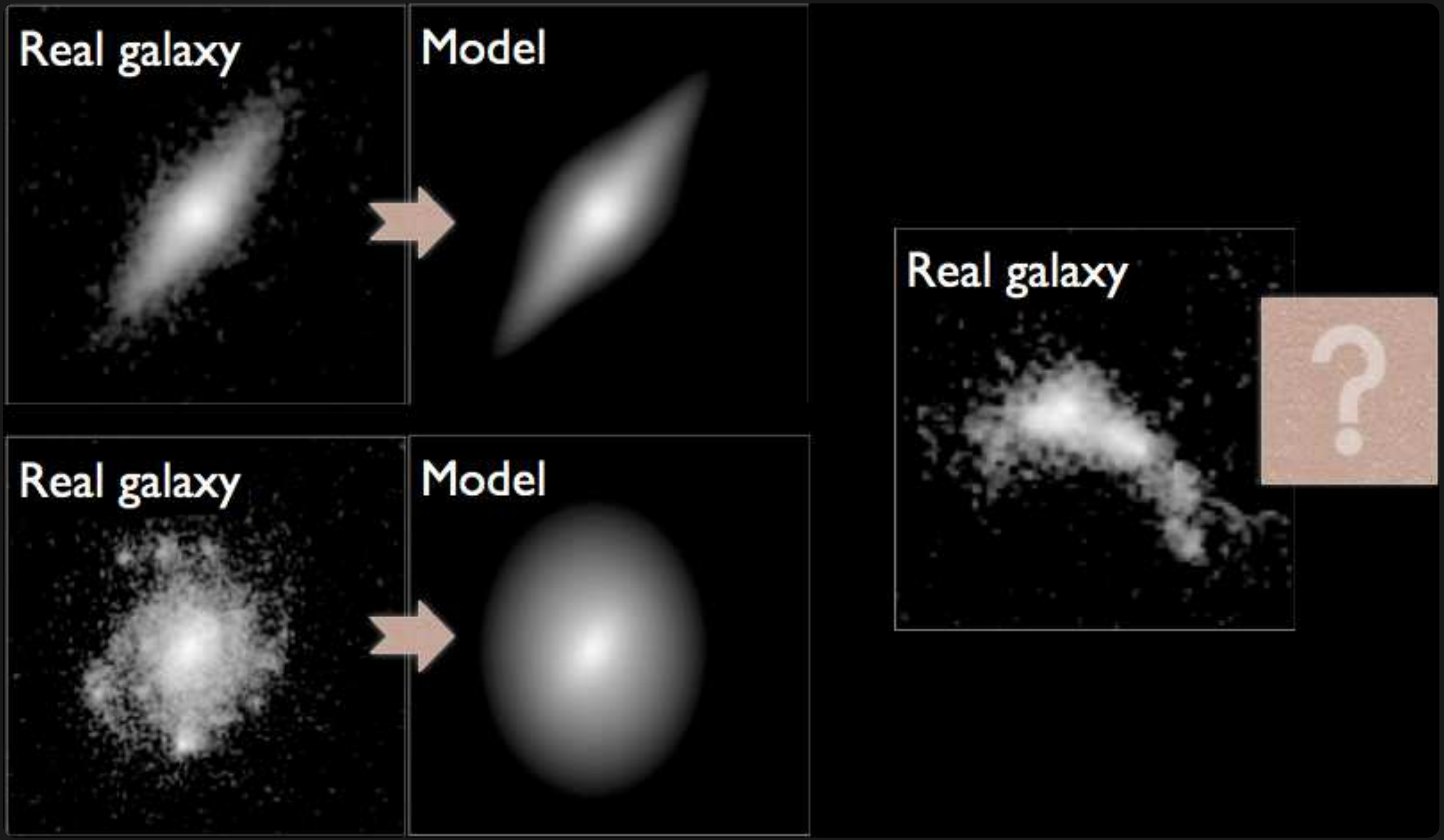
# THE TIMES THEY ARE A-CHANGIN'

- the *end of the analytic era*
- modern surveys: galaxies are no longer blobs
- rethink data analyses: analytic $\longrightarrow$ data-driven

# DEEP GENERATIVE GALAXY MODELLING

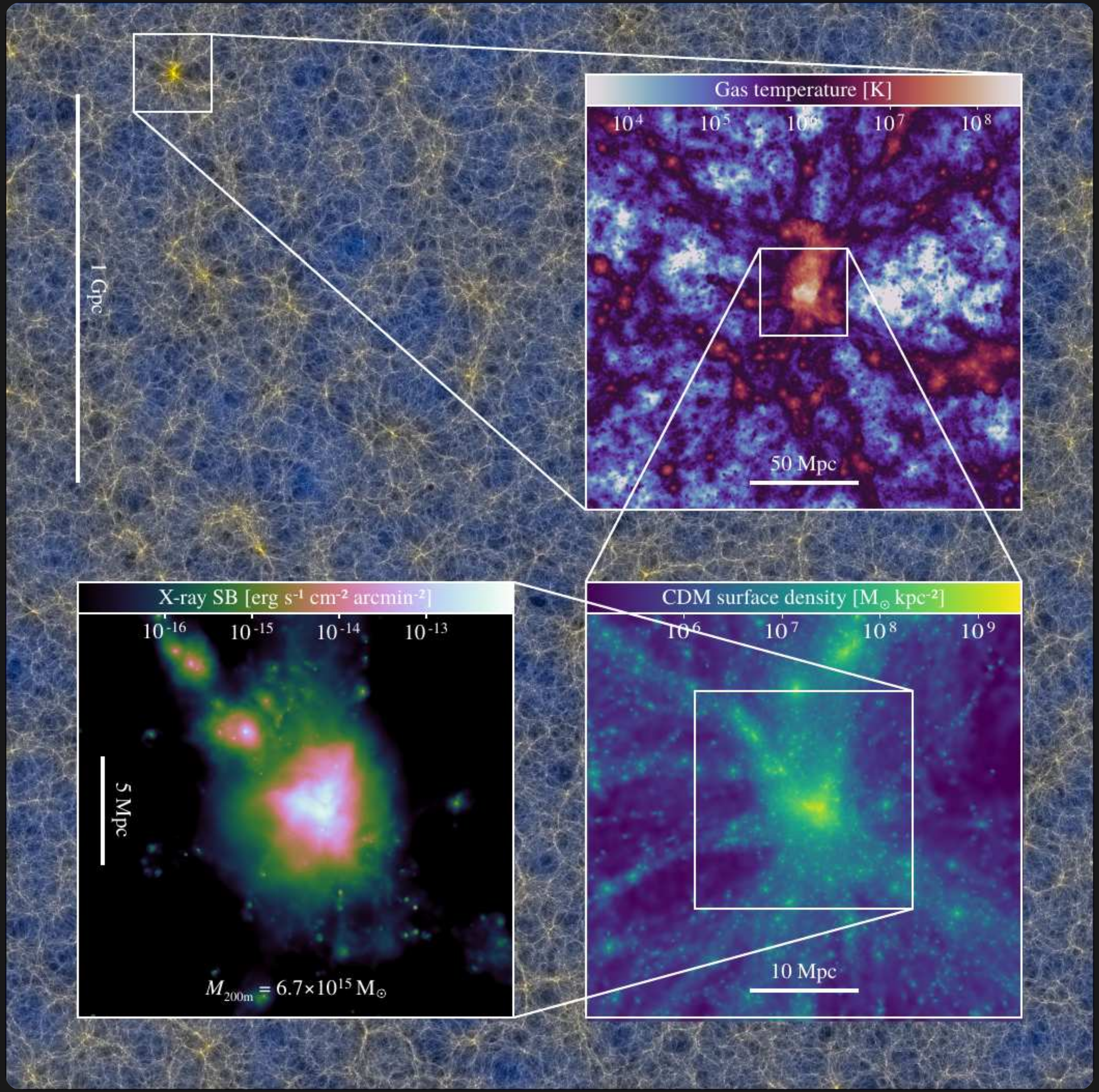
- goal is to learn an **implicit** distribution $\mathbb{P}$ from which the training set $X = \{x_0, x_1, \ldots, x_n\}$ is drawn

# Analytic models $\mathbb{P}_\theta$




Credit: Mandelbaum et al. (2014)

# Simulators $\mathbb{P}_\theta$




Credit: Schaye et al. (2023)

# Implicit distributions

- in both cases, we cannot
  - sample from (the true) $\mathbb{P}$
  - evaluate the likelihood $p_\theta(x)$
- which means: we cannot generate new **plausible** galaxies

# Implicit distributions

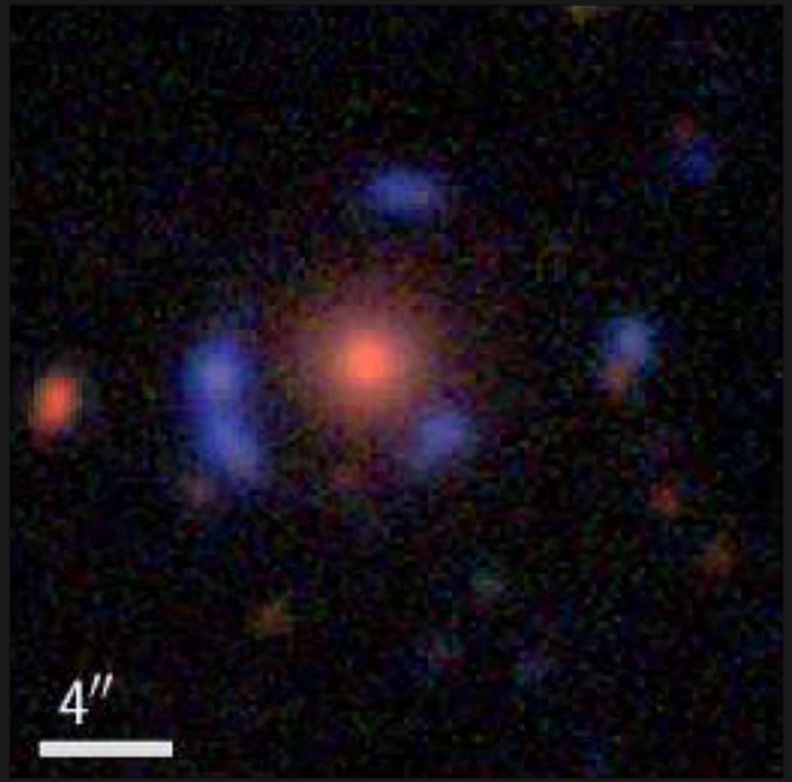
- in both cases, we cannot
  - sample from (the true) $\mathrm{P}$
  - evaluate the likelihood $p_\theta(x)$
- which means: we cannot generate new **plausible** galaxies
- what for?

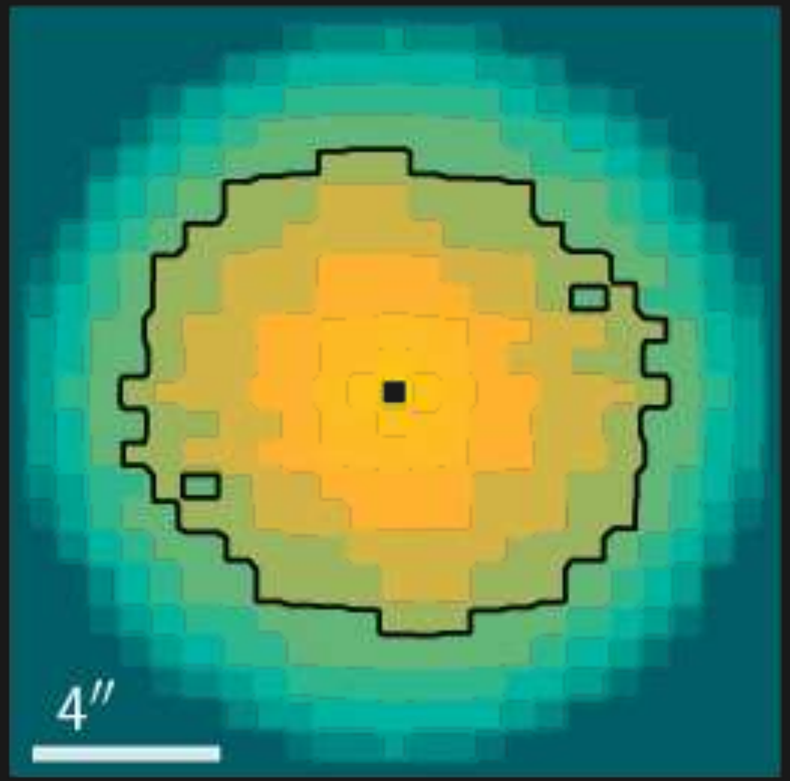
# For instance: strong lensing


2006, Credit: NASA, ESA, D. Player (STScI)

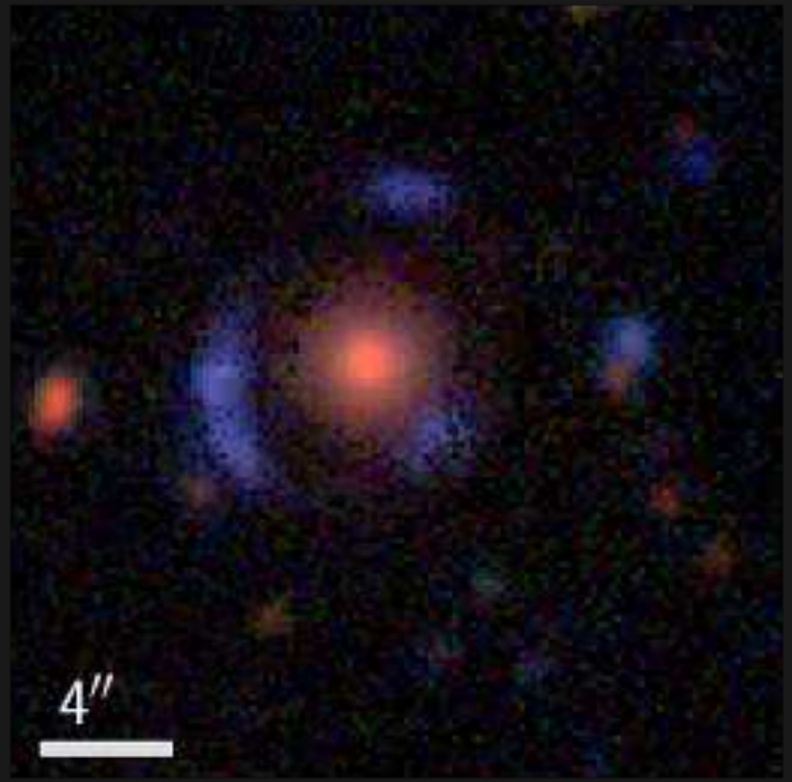
# Strong lens modelling

input data ⟶ latent representation ⟶ reconstruction



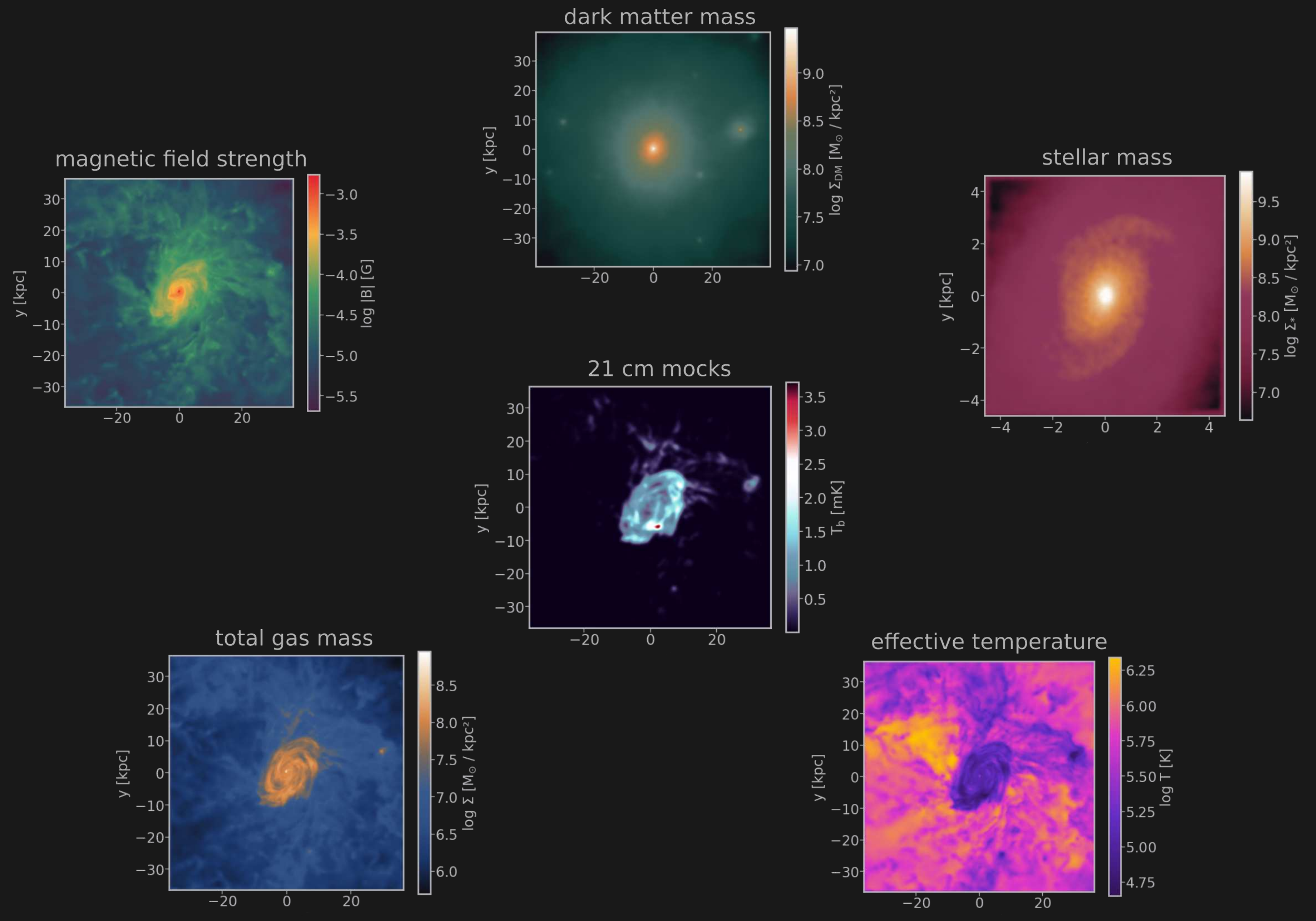

Denzel et al. (2021)

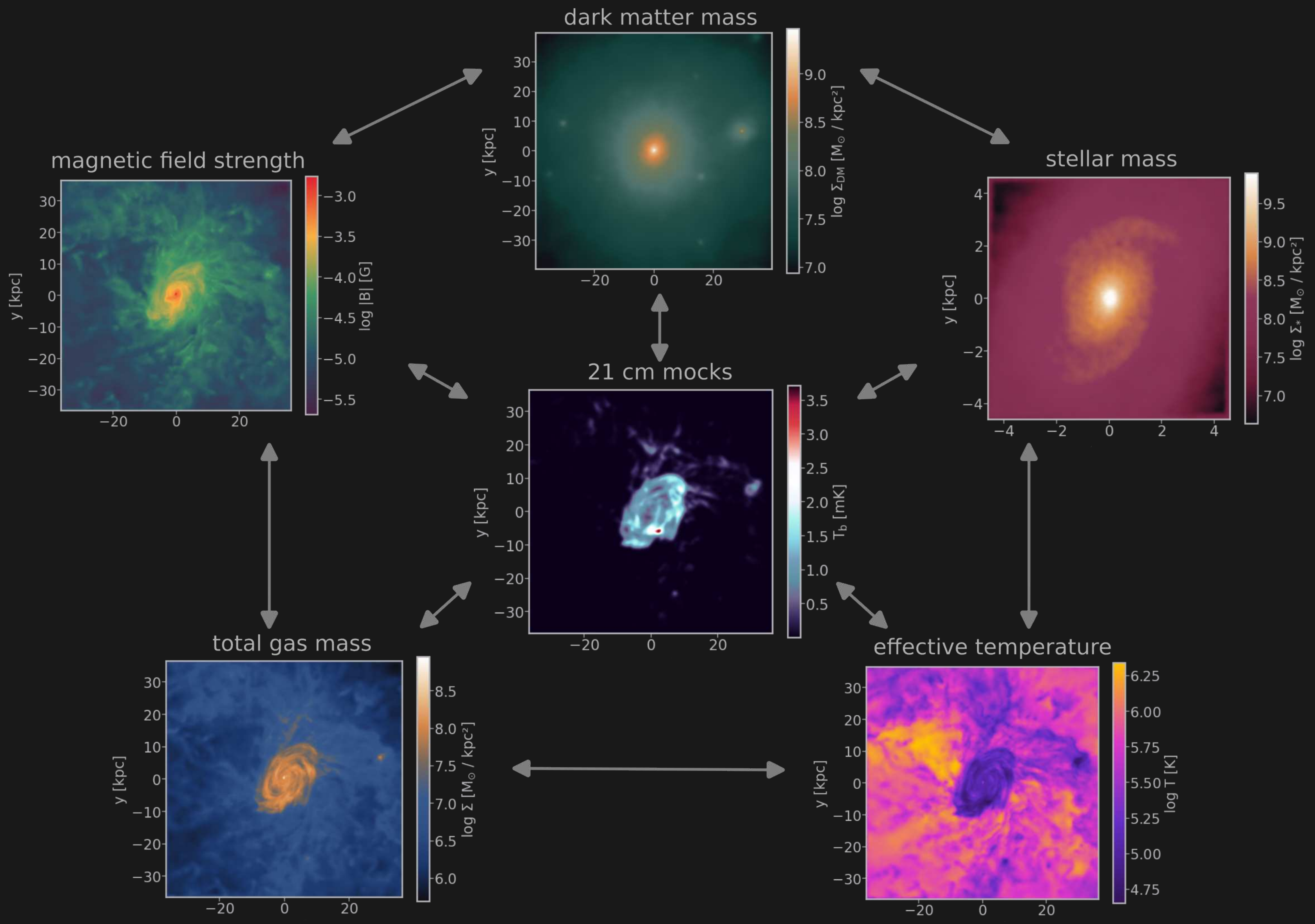


# Dataset: SPH simulations

B-field (TNG100), Credit: IllustrisTNG

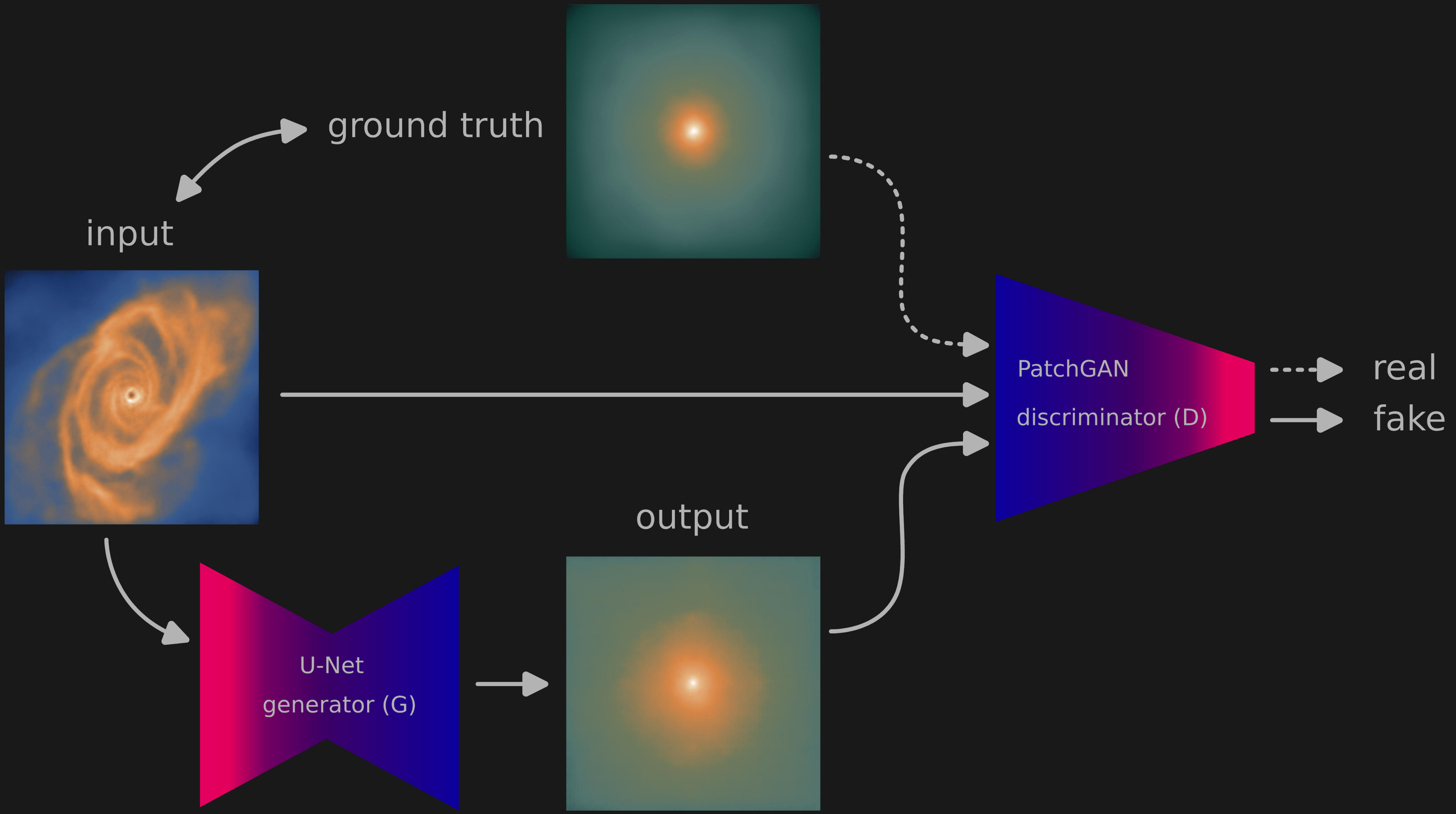
- projected IllustrisTNG galaxies
- 6 domains:
  - dark-matter, stars, gas,
    HI, temperature, magnetic field
- ~ 3000 galaxies
- ~ 10000 images / domain
- augmented:
  - up to 5x randomly rotated
- scale: 2 dark-matter half-mass radii






Use image domain translation models: observations (21cm) ↔ physical properties

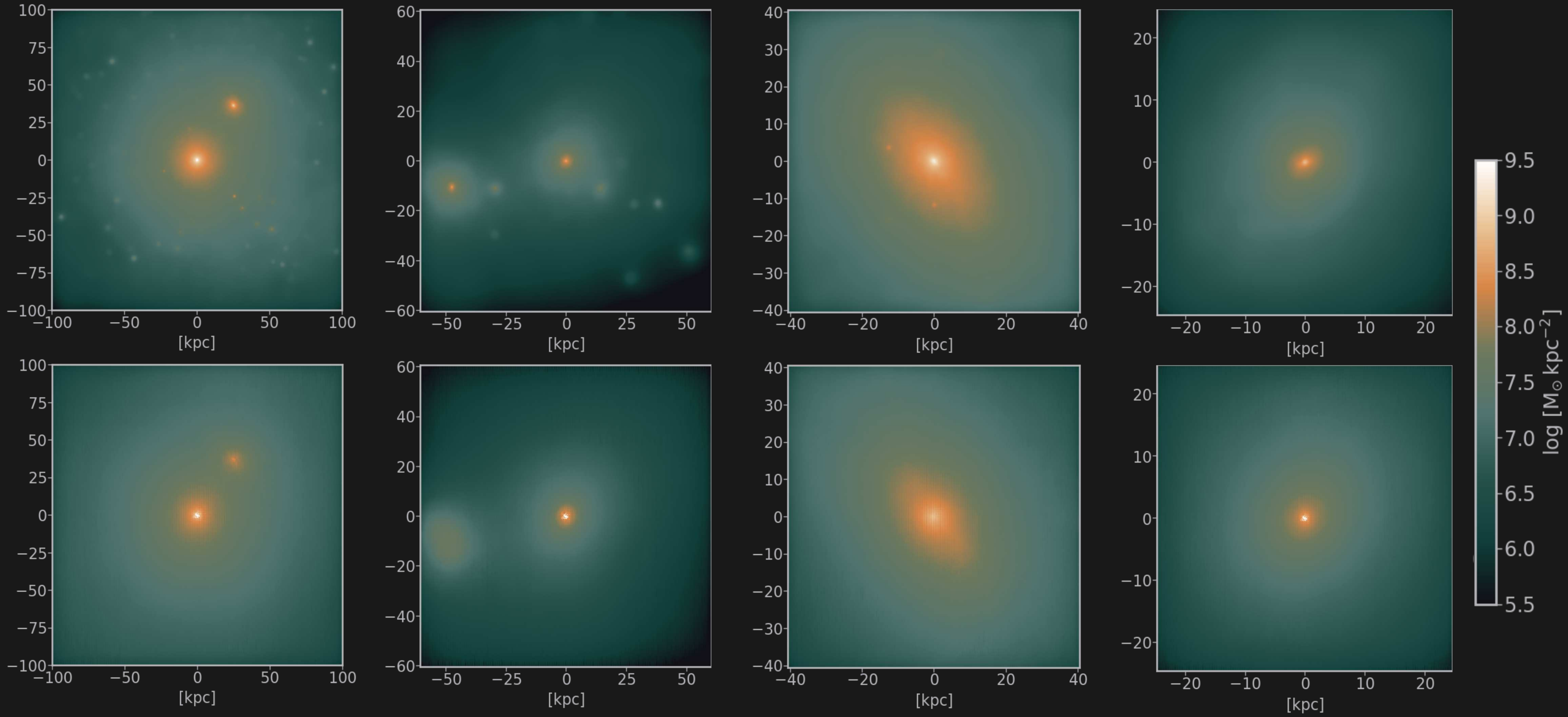
# cGANs: pix2pix schema




Use pix2pix to generate dark matter maps from mock observations

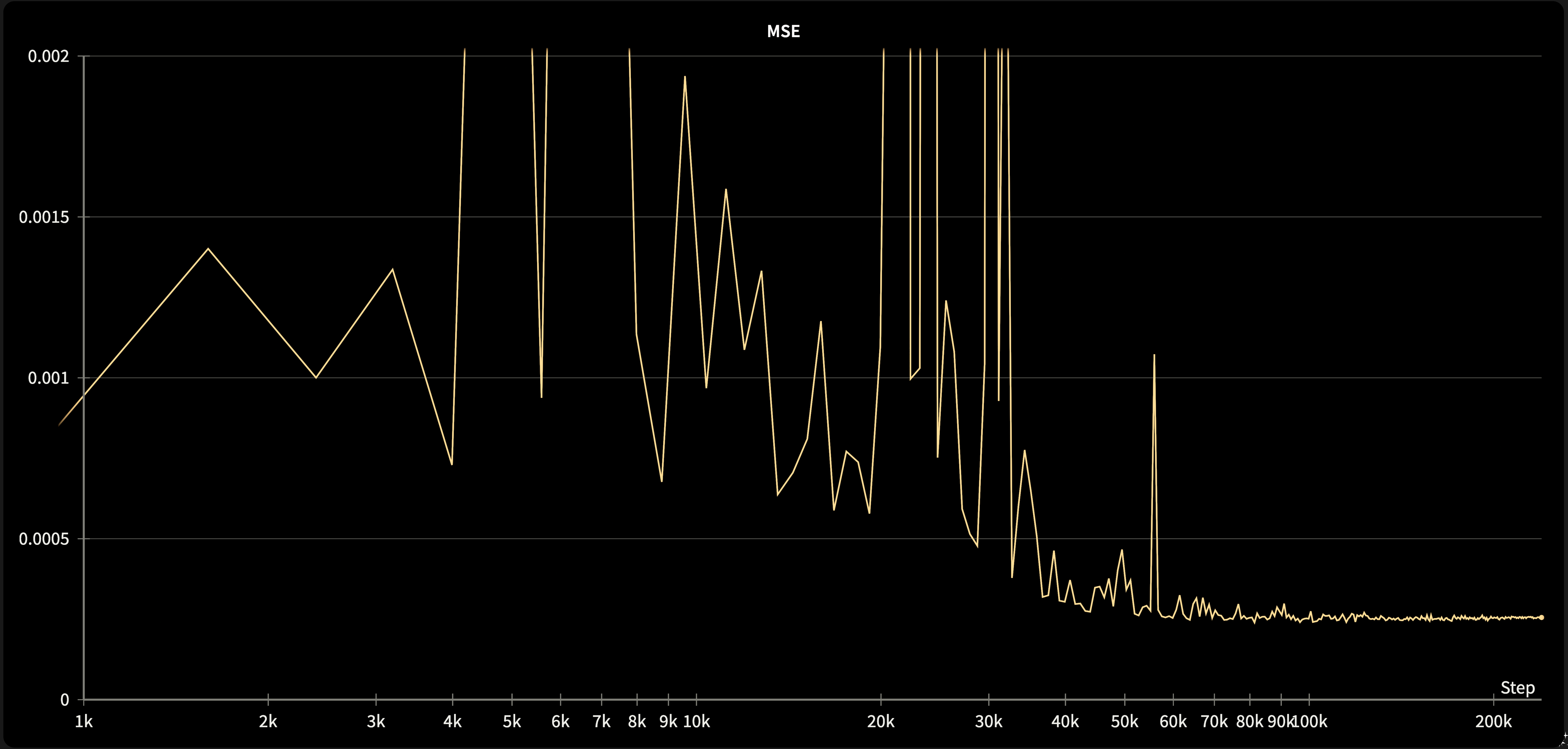
# Sampling from $\mathbb{P}_\theta$

Ground truth



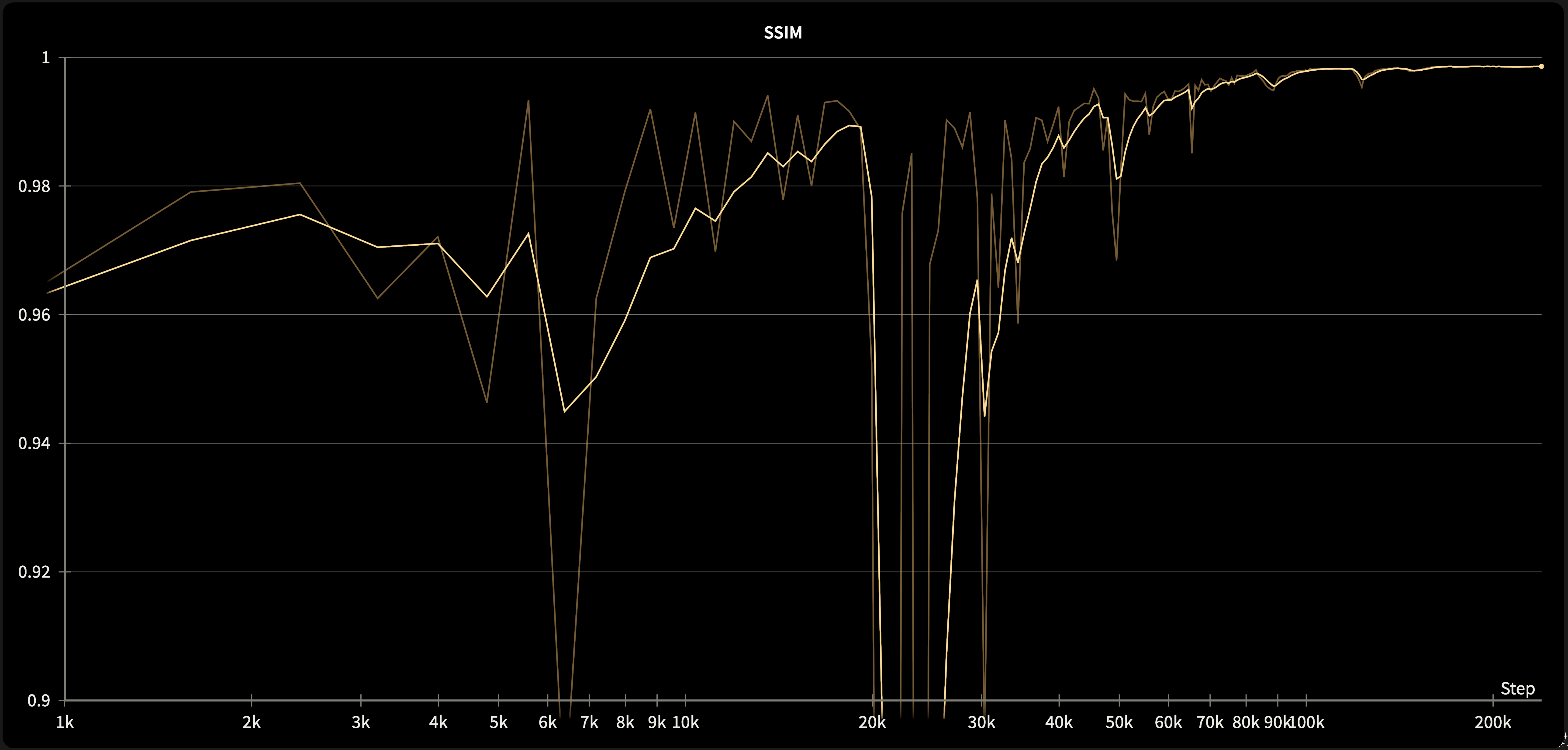

Predictions from pix2pix (from gas projections)

# Evaluation



MSE
0.002
0.0015
0.001
0.0005
0
1k
2k
3k
4k
5k
6k
7k
8k
9k
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
200k
Step


PSNR [dB]
40
35
30
25
20
Step
1k
2k
3k
4k
5k
6k
7k
8k
9k
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
200k


SSIM
1
0.98
0.96
0.94
0.92
0.9
1k
2k
3k
4k
5k
6k
7k
8k
9k
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
200k
Step

## Next steps

- physical metrics: radial/elliptical (NFW) profiles
- substructure just above the resolution limit
- still not able to evaluate $p_\theta(x)$
- GANs: average performance expected to be slightly worse compared to autoregressive and score-based methods

# POINT-CLOUD EXPERIMENTS

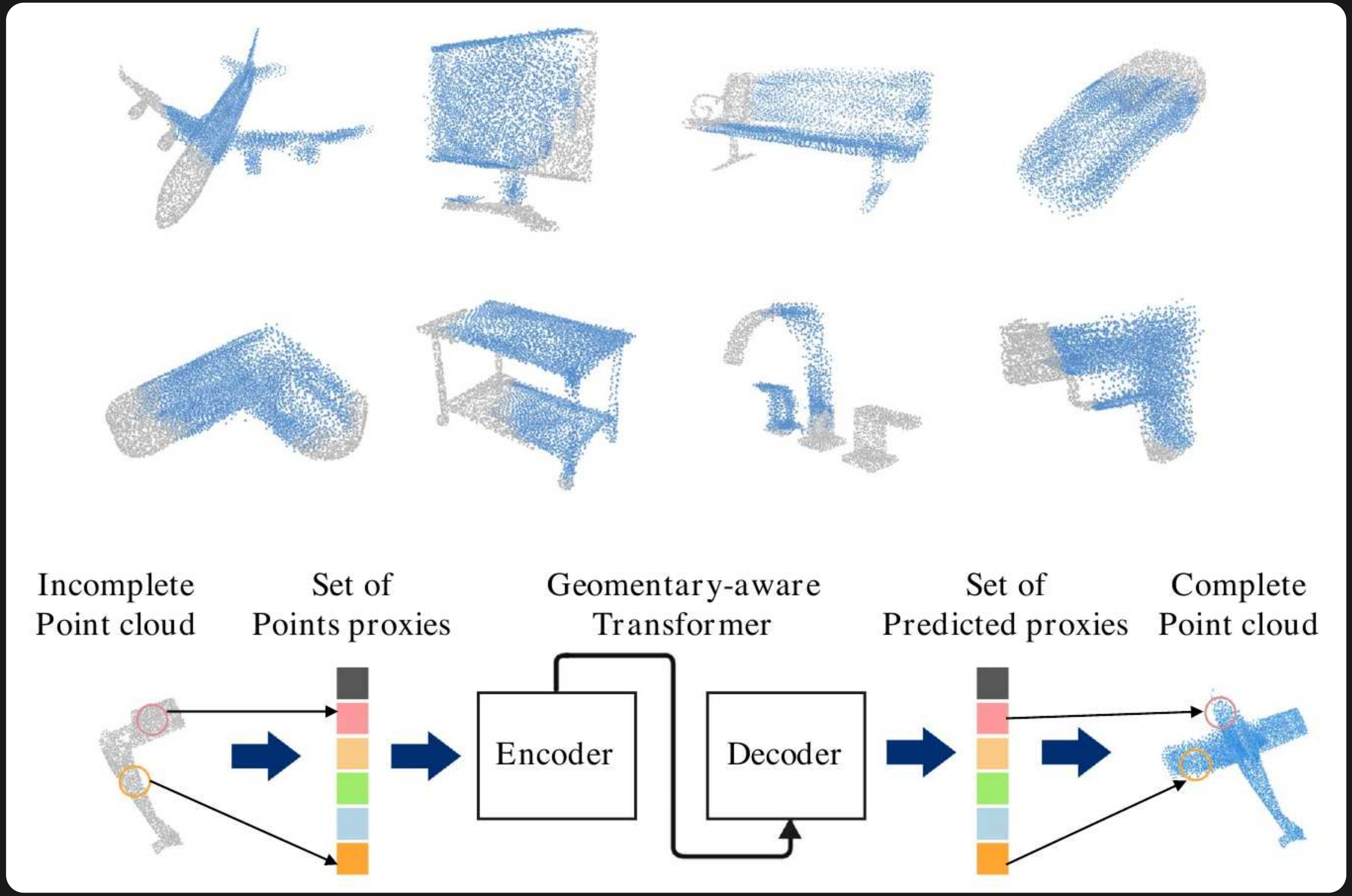
- generative models for full 3D+ simulations

| Property | SPH data | Point clouds |
|---|---|---|
| applications | hydrodynamics | 3D scanning, CAD, etc. |
| list of coordinates | ✓ | ✓ |
| unordered | ✓ | ✓ |
| invariance: vector-row perm. | ✓ | ✓ |
| invariance: geometric transf. | ✓ | ✓ |
| discrete | ~ | ✓ |
| smoothing kernel | ✓ | ✗ |

# AdaPoinTr

- initial tests indicate feasibility
- application: DM-only simulation, generate baryonic particle types (stars, gas, etc.)

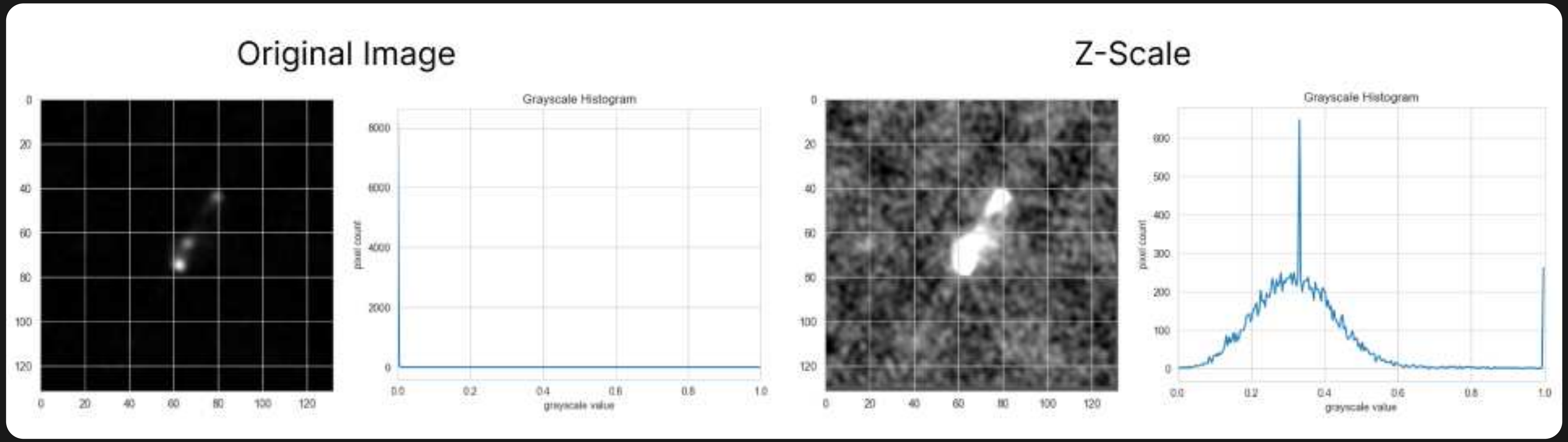
Yu et al. (2023)

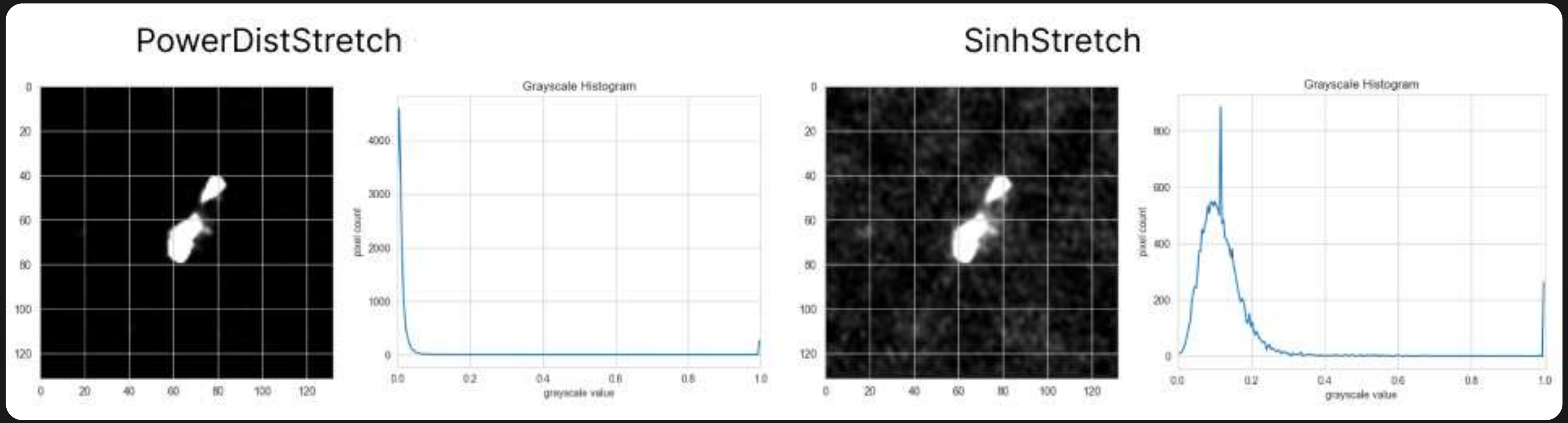


# RADIO SOURCE CLASSIFICATION

- idea developed with Michele Bianco (EPFL)
- student Manuel Weiss: tested SOTA classification & detection architectures
  - ResNet, EfficientNet, ViT, etc. / YOLOv8, DINO, etc.
- goal: testing on the GLEAM survey
- Radio Galaxy Zoo Object Detection Data Set (11’836 labelled images)

# Data preprocessing & augmentations



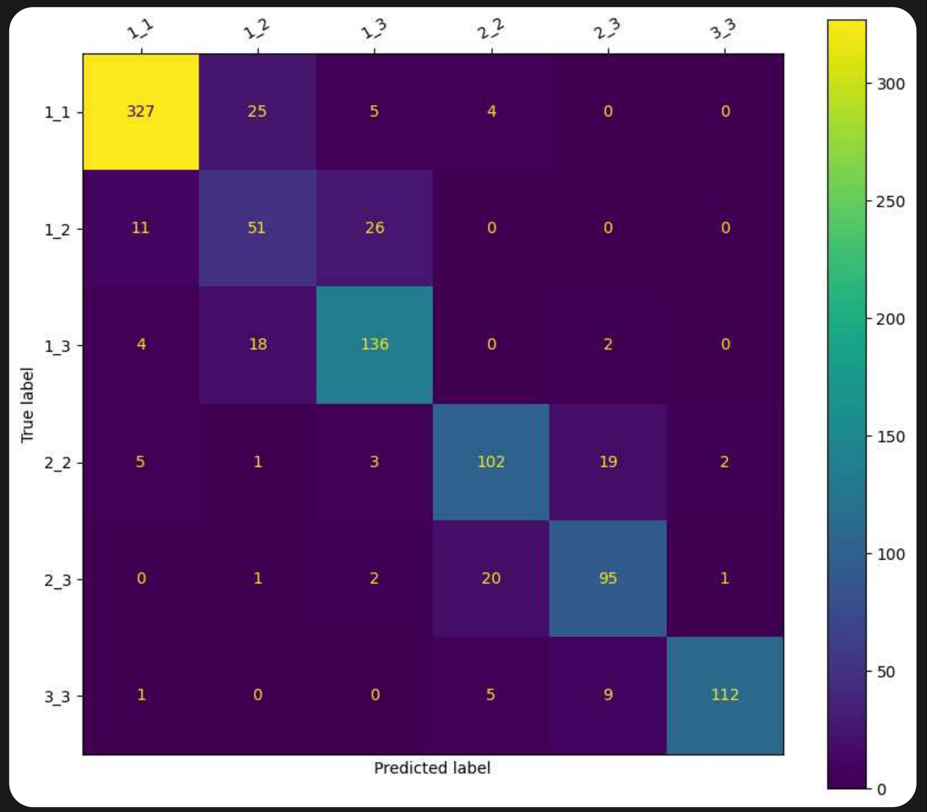



# Difficulties

| Classes | 1_ 1 | 1_ 2 | 1_ 3 | 2_ 2 | 2_ 3 | 3_ 3 |
|---|---|---|---|---|---|---|
| Samples | 5300 | 1331 | 1412 | 1251 | 1208 | 1334 |

- unbalanced dataset
- even humans have difficulties distinguishing
  - 1_ 2 vs 1_ 3 $\longrightarrow$ FR1 vs FR2
  - mislabelled samples?

Confusion matrix for the best ResNet model



# Preliminary results

- probably mislabelled data
- best model: ResNet (small, not pretrained)

| Model | Top1 [%] | Top2 [%] | F1 [%] | Precision [%] | Recall [%] | ensemble σ |
|---|---|---|---|---|---|---|
| ResNet | 89.36 | 97.57 | 86.24 | 87.40 | 85.44 | 4.7% |
| ViT | 76.60 | 89.46 | 69.64 | 70.10 | 69.38 | - |

# CONTACT

https://phdenzel.github.io/

philipp.denzel@zhaw.ch

frank-peter.schilling@zhaw.ch

elena.gavagnin@zhaw.ch

Created by phdenzel.