Map-to-map translation

for SKA mock observations and cosmological simulations

by Philipp Denzel, Frank-Peter Schilling, Elena Gavagnin

Zürcher Hochschule für Angewandte Wissenschaften

The **SKA** (Square Kilometre Array) is an international, next-generation radio astronomy facility that will lead to groundbreaking new insights in *astrophysics and cosmology*. It will be operated over three sites: *HQ in the UK*, the mid-frequency array in *South Africa (SKA-mid)*, and the low-frequency array in *Australia (SKA-low)*. The two telescopes under construction will combine signals received from thousands of dishes and small antennae to reach extremely high sensitivity and angular resolution. In particular, the sensitivity of the SKA in the **21cm hydrogen line** will map the spatial distribution of neutral hydrogen (HI) of billions of galaxies to unprecedented precision, promising new insights into **galaxy formation and evolution**.

Image credits: SKAO

Hydrodynamical simulations have been instrumental in the study of the Universe and particularly galactic dynamics. Highly tuned on a vast range of scales, they produce **realistic galaxy** models from first principles. With the latest suites of large-scale cosmological simulations, extracted mock observations have become almost indistinguishable from actual ones. The **IllustrisTNG project** produced a suite of state-of-the-art simulations counting thousands of galaxies at redshifts below 1 (see illustration on the right). Deep learning is particularly apt to scale up to such big datasets and to translate the more abstract, physical principles that underlie complex observations such as by the SKA.

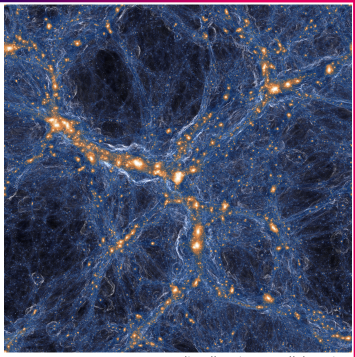
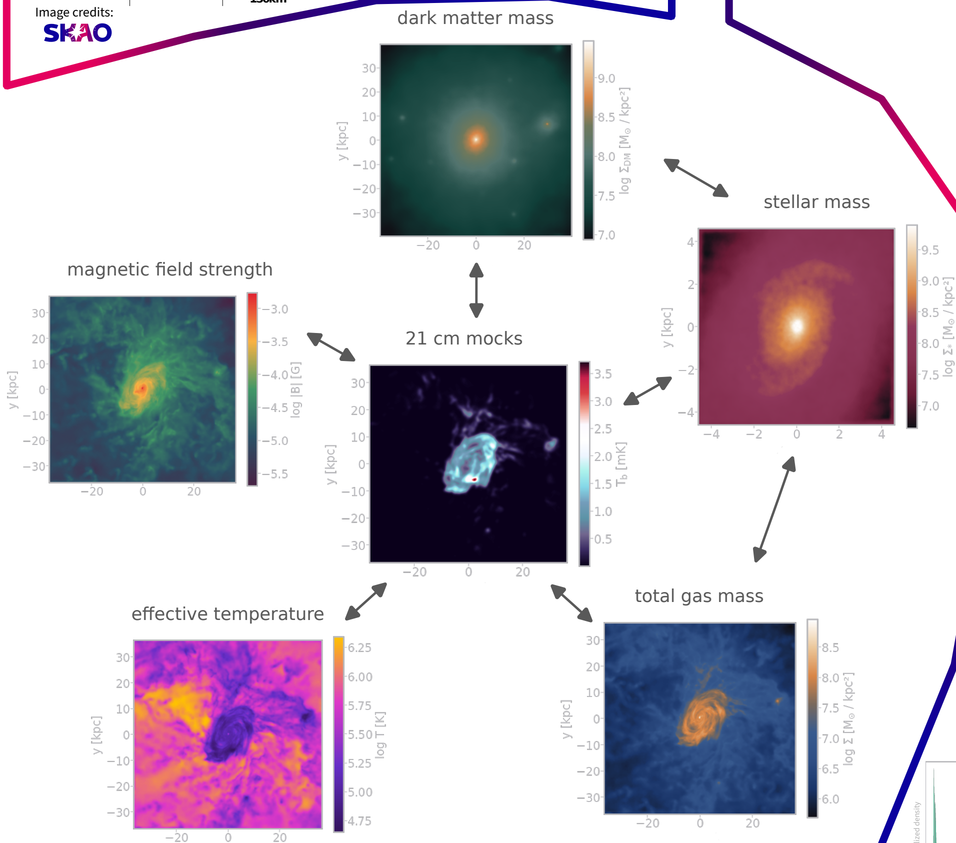


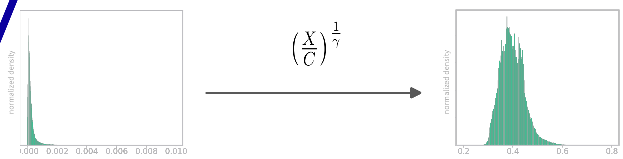
Image credits: IllustrisTNG Collaboration

For the purpose of exploring the mapping between various astrophysical properties to observational quantities, e.g. dark matter halo substructure, or stellar mass functions, we have extracted a **multi-domain dataset of over 30'000 galaxy maps** for each domain (see Figure on the left) from the IllustrisTNG *TNG50-1* box. Furthermore, **21 cm mock observations of the HI gas distribution** were modelled after *Villaescusa-Navarro et al. (2018)*. The galaxies have been selected according to their mass resolution, and were projected onto 512x512 images, extending two half-mass radii outwards from the galaxy center.

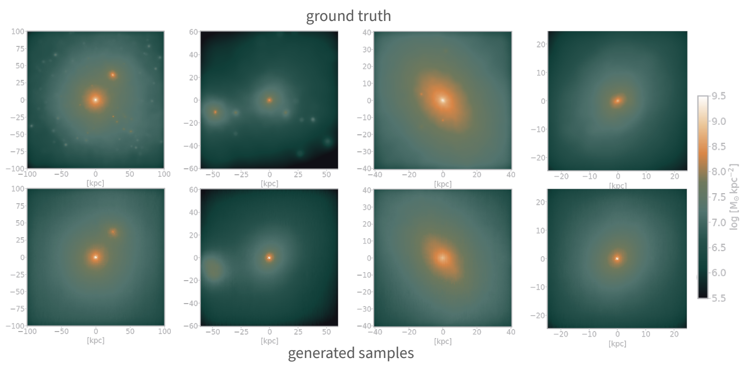


Log Σ is the corresponding 2D matter density (projected from the simulation TNG50-1) in units of solar masses per square kilo-parsecs. The projection were calculated using the *nearest-grid-point* (NGP) mass assignment scheme.

Isola et al. (2017) proposed the **pix2pix** model as a general-purpose solution to image-to-image translation problems. These conditional generative adversarial networks (**cGANs**) find a mapping from an input image of one domain to an output image of another domain by learning a loss function corresponding to the task at hand.
Since then, conditional GANs have been established as a staple of **generative models for various tasks** such as inpainting, colorization, generation from labels. However, while these models achieve high-quality results with **natural images**, **scientifc inverse problems** have so far been less explored.
While the **dynamic range** of natural images is easily grasped by neural networks, the large difference in scales for astrophysical data (cf. the range of scales for the images on the left) often hides the important features in the data. Therefore, we experimented with various scalings and found an *inverse-power scaling* to be the most flexible and suited for our models (see below).

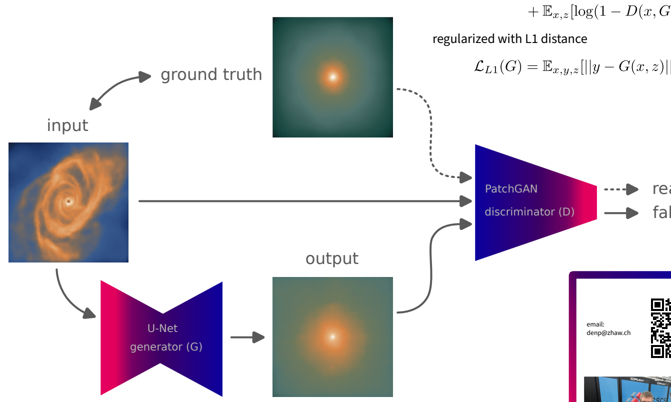


Our latest trained model (preliminary results on test dataset; *top row: ground truth, bottom row: generated samples*) yields plausible dark-matter maps **at various scales** given mock observations of projected gas maps from cosmological simulations. However, we observe that the generated maps can tend to be **too cuspy**, having not enough mass in the outskirts of the galaxies, while roughly preserving the total halo mass. Furthermore, less **prominent satellites** of high-mass galaxies above 10^{13} $M_{\odot}$ are often smoothed out or completely missing.



In future, we would like to test and compare other types of generative deep-learning models such as score or flow-based models. Moreover, to test generated maps against observational findings, we are investigating the integration of our generative models in lens modelling frameworks for testing against strong gravitational galaxy-lens observations.

pix2pix schema

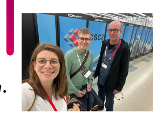


Adversarial training objective

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))]$$

regularized with L1 distance

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[||y - G(x, z)||_1]$$



Acknowlegements: we acknowledge support from SERI and the SKACH consortium.