Master Thesis – Project Description

Using Generative Adversarial Networks to Analyse Next-Generation Astronomical Data

Ludvig Doeser – doeser@kth.se

December 22, 2020

The proposed project aims at developing a high-fidelity generative model of galaxy images using Generative Adversarial Networks (GAN's). The framework of GAN's is fairly general; it is a type of unsupervised learning, where a machine tries to learn the generative process that generates data. In addition, GAN's provide probability distributions from which new samples can be generated.

The machine learning (ML) model using GAN's aims to account for various physical properties such as the evolution of the galaxy population as a function of redshift, different morphologies, such as spiral and elliptical galaxies, metallicity, and age. We would train a GAN on Galaxy images while accounting for the telescope response, that is, finite resolution, noise, luminosity limits etc. In the simplest form, this can be achieved by using a pin-hole camera model, which can be added as a fixed element to the generative model. The GAN would then learn a prior on how galaxies appear, and one could use this prior in galaxy image analyses, with the goal to e.g.:

- Provide super-resolution galaxy images
- Use the detailed features of galaxies to do distance tri-angulation
- Measure the redshift distance relation using only galaxy images.
- Prepare to use the billion galaxies expected to be observed with next-gen galaxy surveys, which might provide powerful cosmological constraints.
- Enable fast and efficient simulations of galaxy mock data

Relevance

The next-generation of astronomical instruments, such as the Vera C. Rubin Observatory, will generate enormous amounts of image data to chart the cosmic matter distribution by taking high-resolution pictures of more than 3 billion galaxies. This image data set holds significant cosmological information that needs to be extracted using novel data analysis and machine learning techniques. Specifically, the brightness, shapes, morphologies, and alignment of galaxies provide important information on the formation of the cosmic structure as well as dark matter and dark energy phenomenology. However, standard analyses make simplifying assumptions, such as assuming galaxies to be ellipsoids or discs ignoring the actual detailed features and shapes of galaxies.

Novel machine learning techniques have recently been shown the promise of taking into account full image data to significantly improve measurements of the photometric redshifts, indicating the distance to the galaxies. This includes information on the brightness but also the size distribution of galaxies. Detailed knowledge on the intrinsic shapes of galaxies is furthermore relevant for weak gravitational lensing data since gravitational lenses tend to distort observed images and knowledge on the undistorted image allows us to measure gravitational dark matter potentials. Going beyond simple approximate models of galaxies using machine learning (ML) is a promising path forward to perform high-precision cosmology with next-generation data.