PACIS: Precision and Accuracy for Cosmological Imaging Surveys

The Universe on the largest scales is organized into the cosmic web: stars gather into galaxies, galaxies form galaxy clusters, which themselves are interconnected by huge walls and filaments, with immense cosmic voids in between. Such detailed information about the large scale structure of the Universe is obtained thanks to sky surveys, in which we measure source positions on the celestial sphere, their brightness in various wavelengths, and if possible, also their detailed energy distribution, which is known from their electromagnetic spectra. For most of the building blocks of the cosmic web – galaxies – their distance from us is measured from the *redshift* of the spectrum: galaxies seem the redder, the farther from us they are.

Measuring spectra for all the galaxies is however an immense challenge, requiring major time and money investments. For that reason, already today most of the galaxies that are listed in the largest catalogs – counting already hundreds of millions objects and growing – do not have spectra measured. This means that astronomers cannot directly access the key information about these galaxies, such as their distances. At the absence of direct measurement of the spectrum, redshift must be estimated in some other way.

A solution to such problems comes from modern computer science techniques. Novel *machine learning algorithms* allow astronomers to quickly and efficiently extract distance information about far-away astronomical objects only from their observed brightness at various wavelengths, sometimes using additionally angular sizes or other measured properties of galaxies. Such a method to estimate galaxy distances without measuring spectra is generally called *photometric redshifts*. A great advantage of such machine-learning approaches is their ability to rapidly work with complicated multi-dimensional data in today's largest catalogs – which would be beyond reach for standard human analysis. Within our project, we will be applying this type of novel machine-learning schemes to estimate redshifts in state-of-the-art astronomical data.

We will first focus on the data from the Kilo-Degree Survey (KiDS), one of the largest available imaging datasets with excellent brightness measurements for galaxies at several wavelengths. We will employ several sophisticated machine learning methods to extract new information from KiDS imaging and produce new catalogs of galaxies. One of them will be state-of-the-art deep learning – a 'hot topic' in today's computer science, which we will use to decode distances to galaxies directly from their multi-color images. We will subsequently adapt the machine-learning software tools developed in this project to the next-generation very big data from the Legacy Survey of Space and Time (LSST). The new galaxy catalogs obtained in our project will be used for cosmological studies with the eventual aim to resolve the mystery of dark matter and dark energy, that seem to rule the fate of the Universe.