Probabilistic deep learning for cosmology and beyond

Supervisor: Jason McEwen (MSSL, UCL)

Project Outline

The current evolution of our Universe is dominated by the influence of dark energy and dark matter, which constitute 95% of its content. However, an understanding of the fundamental physics underlying the dark Universe remains critically lacking. Forthcoming experiments have the potential to revolutionalise our understanding of the dark Universe. Both the ESA Euclid satellite and the Rubin Observatory Legacy Survey of Space and Time (LSST) will come online imminently, with Euclid successfully launched in July 2023 and the Rubin LSST Observatory having recently achieved first light. Furthermore, the Simons Observatory is in advanced stages of construction. Sensitive statistical and deep learning techniques are required to extract cosmological information from weak observational signatures of dark energy and dark matter.

The classical approach of deep learning is to make single predictions. A single estimate of a quantity of interest, such as an image, is typically made. For robust scientific studies, however, single estimates are not sufficient and a principled statistical assessment is critical in order to quantify uncertainties. Bayesian inference provides a principled statistical framework in which to perform scientific analyses. In cosmology, in particular, Bayesian inference is the bedrock of most cosmological analyses. While such approaches provide a complete statistical interpretation of observations, which is critical for robust and principled scientific studies, they are typically computationally slow, in many cases prohibitively so. Furthermore, in such analyses prior information typically cannot be injected by a deep data-driven approach.

In the proposed project we will develop probabilistic deep learning approaches, where probabilistic components are incorporated as integral components of deep learning models. Similarly, we will also develop statistical analysis techniques for which deep learning components are incorporated as integral components. This deep hybrid approach, where statistical and deep learning components are tightly coupled in integrated approaches, rather than considered as add-ons, will allow us to realise the complementary strengths of these different approaches simultaneously. For some examples of related research, please see the following recent papers: McEwen et al. 2021, arXiv:2111.12720; Spurio Mancini et al. 2022, arXiv:2207.04037; Polanska et al. 2023, arXiv:2307.00048 (although in this PhD project we will go beyond the Bayesian model comparison focus of these works).

Specifically, we will develop novel probabilistic deep learning models, variational inference techniques and simulation-based inference approaches. These new methodologies will be applied to various cosmological problems and probes, focusing on the cosmic microwave background and weak gravitational lensing, and will include generative models for emulation and inference approaches for the estimation of not only the parameters of cosmological models but also to assess the most effective models and physical theories for describing our Universe.

The student should have a strong mathematical background and be proficient in coding, particularly in Python. The student will gain extensive expertise during the project in deep learning, going far beyond the straightforward application of existing deep learning techniques, instead focusing on the construction on novel probabilistic deep learning approaches and their application to novel problems in cosmology and beyond. The expertise gained in foundational deep learning will prepare the student well for a future career either in academia or industry. In particular, the emerging field of probabilistic deep learning is a speciality highly sought after in industry by many companies, such as Google/DeepMind, Facebook, Amazon and many others.