Explaining Dark Matter Halo Abundance with Interpretable Deep Learning

Ningyuan (Lillian) Guo[a], Hiranya V. Peiris[a,b], Andrew Pontzen[a], Luisa Lucie-Smith[c]

[a]Department of Physics & Astronomy, University College London, [b]The Oskar Klein Centre for Cosmoparticle Physics, Stockholm University, [c]Max-Planck-Institut für Astrophysik

Introduction

- Dark Matter Halos are building blocks of cosmic structure that galaxies form within.
- **Halo mass function** describes the abundance of dark matter halos as a function of halo mass and depends sensitively on cosmological parameters, e.g. matter density Ω_m, dark energy equation of state w_0, neutrino mass M_ν.
- Accurately modelling the halo mass function for a range of cosmological models will enable forthcoming surveys such as the Rubin Observatory Legacy Survey of Space and Time (LSST) to place tight constraints on cosmological parameters.

Problem

- Halo formation is a highly non-linear process and developing a physical understanding is difficult:

  Incomplete knowledge of which factors are needed to accurately determine the halo abundance beyond $\sigma(M)$, the rms variance of the density field smoothed on the halo mass scale.

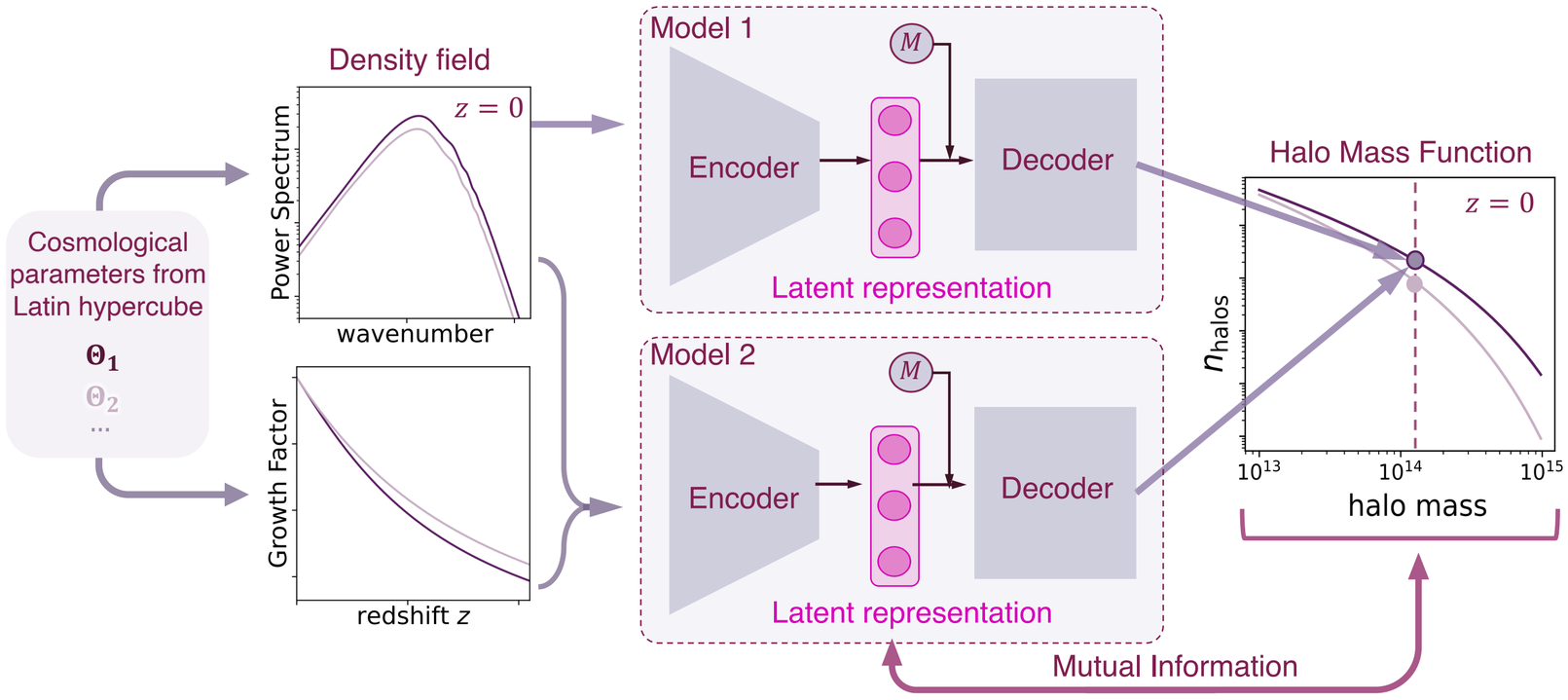
[1]Lucie-Smith, L. *et al.* Preprint at arXiv:2203.08827 (2022)
Title banner image: Boylan-Kolchin, M. *et al. MNRAS*, 398, 1150 (2009)

ningyuan.guo.20@ucl.ac.uk

Interpretable Deep Learning Framework for Knowledge Extraction

- Train an interpretable variational encoder[1] model to learn the halo mass function.
- Latent representation contains all relevant information needed to predict halo number density.
- Knowledge on which factors are relevant is extracted by interpreting the latent representation using mutual information.

Application: is growth history required to accurately model the halo mass function?



- Compare for Model 1 and Model 2 the mutual information (MI) between latent representations and the ground truth halo number density.
- If growth history is required in addition to information from the power spectrum, the total MI between Model 2's latent representation and ground truths should be higher. We also expect Model 2 to require a higher dimensional latent representation to encode growth history information.