Title: Detecting Halos in the Lyman-α Forest using Neural Networks

Authors: Teng Hu, Matthew Pieri, Duarte Muñoz Santos,,Simon Weng

Abstract:
The circumgalactic medium (CGM), the interface between the interstellar medium (ISM) and the intergalactic medium (IGM), plays a crucial role in galaxy evolution, encompassing processes such as gas accretion, star formation, and feedback mechanisms. While previous studies have predominantly focused on high column density hydrogen absorbers, i.e., Damped Lyman-alpha systems (DLAs, $N_{HI} > 10^{20}$ cm^{-2}) and Lyman Limit Systems (LLSs, $N_{HI} > 10^{17}$ cm^{-2}), which are relatively rare and represent only a fraction of the CGM, absorbers with lower column density ($N_{HI} < 10^{17}$ cm^{-2}) are more prevalent but pose significant detection challenges due to their subtle and complicate spectral signatures. To address this, we employ a machine learning approach utilising a U-Net architecture, which is a convolutional neural network with an encoder-decoder structure and skip connections to detect dark matter halos within the Lyman-α forest. Trained on mock spectra generated from the TNG50 cosmological simulation, where halos are labelled in the training data, the neural network effectively captures both fine and large-scale spectral features, both of which are critical for halo identification. The network demonstrates robust performance across a wide range of halo masses. For example, it achieves an F1-score of 0.68 for halos in the mass range of 10^{11} to 10^{12} solar masses. Furthermore, we explore the potential of this framework to classify the Lyman-alpha forest within the context of the cosmic web, offering an alternative approach for IGM tomography and a novel method for building the key large-scale environmental context in which to study galaxy evolution and structure formation.