Fall 2018 Capstone Project
Interim Update

J.a.a.S.

November 1, 2018

1 Progress Summary

The team has been working on learning the distribution the neutral hydrogen of numerical simulations of the universe. We approach the problem using Generative Adversarial Networks (DCGAN, WGAN and MMDGAN) and a vanilla Variational Autoencoder. Until now, we have implemented the data loader functions, data transformations, and initial network models. This week we are implementing different architectures for the mentioned deep learning models and deciding which ones fit best our problem.

2 Next Steps

As we continue to improve our current models we pan to implement a GAN framework that tries to replicate the 3D power spectrum of our real sample, by incorporating this to the loss function. In addition to modeling, we are also working on implementing a good 3D power spectrum function as a validation metric with the help of one of our mentors.

3 Issues

One of the main problems that we face is the sparsity of the data. The hydrogen masses and their distribution in the simulated universe is highly skewed where the clusters of galaxies are bunched up and the other parts are almost completely empty. In addition, the magnitude of the range of values is really high and scaling might adversely affect the algorithm due to more information being in the lower end of the scale. To resolve this, we have been working under memory constraints (one cube being 64GBs) and trying to do several transformations such as scaling, standardizing and box-cox transformations. Also, we are looking into GAN architectures used in High Energy Physics like LAGAN where locally connected layers with unshared weights are used. All in all, either or both of these paths should improve the results in the following days.