IONIZATION MAP CLASSIFICATION

Stephen†

†University of McGill

Problem Statement and The Dataset

The goal of this project was to classify simulated data of ionization maps of the universe from the epoch of re-ionization, and be able to estimate two (correlated) values: The red-shift value; and the escape fraction value - see physics section for details.

The data is stored in 3d arrays of dimensions 200^3 with values ranging from 0 to ≈ 70. Preliminary results were obtained in by cutting these blocks into 64 sub-blocks and then slicing those blocks into 50^2 image-like 2d arrays; this allowed me to play with some of the data without it being too over-whelming. For the final result, the data was prepared by directly slicing up the blocks into 200 x 200 who's dimensions were reduced by means of an auto-encoder so that the networks would be small enough to run locally.

Physics

In cosmology red-shift is used to measure distance, so correctly guessing the red-shift value tells us is how far away the regions of space are. The escape fraction is the fraction of photons emitted that make it out of their local area and start ionizing the intergalactic medium.

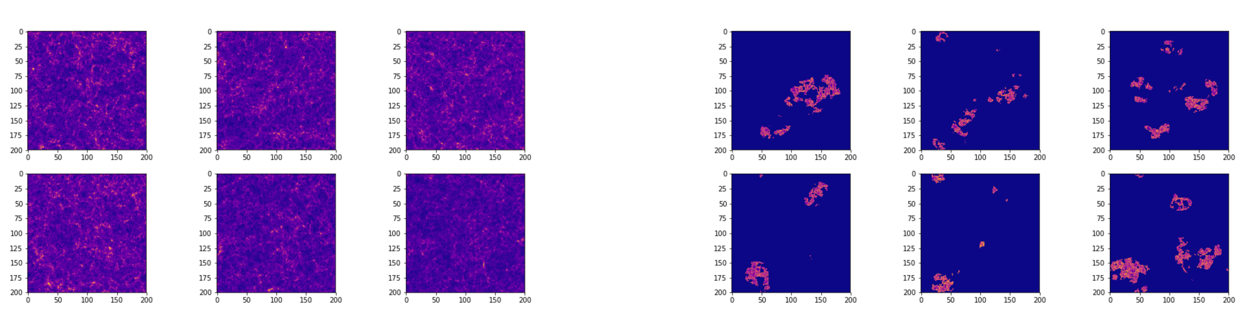

Fig. 1: Mass density maps Ionization maps (right)

The Ionization maps were generated by the mass density maps by means of a simulation [ref 1].

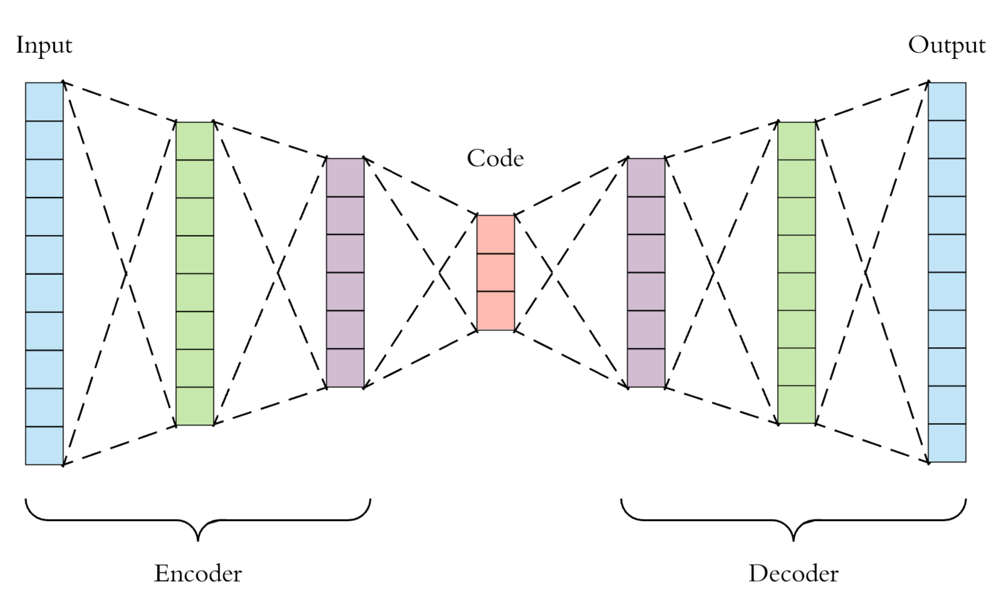

Fig. 2: Structure of autoencoder

Auto-Encoder

Multiple architectures were trained, ultimately the one which performed best was a thirteen-hidden-layered convolutional auto-encoder. No plots are included because the network was trained over several days and five epochs at a time and the training data was not saved. The encoder turned the 200×200 image into a $25 \times 25 \times 8$ array (error ≈ 0.1).

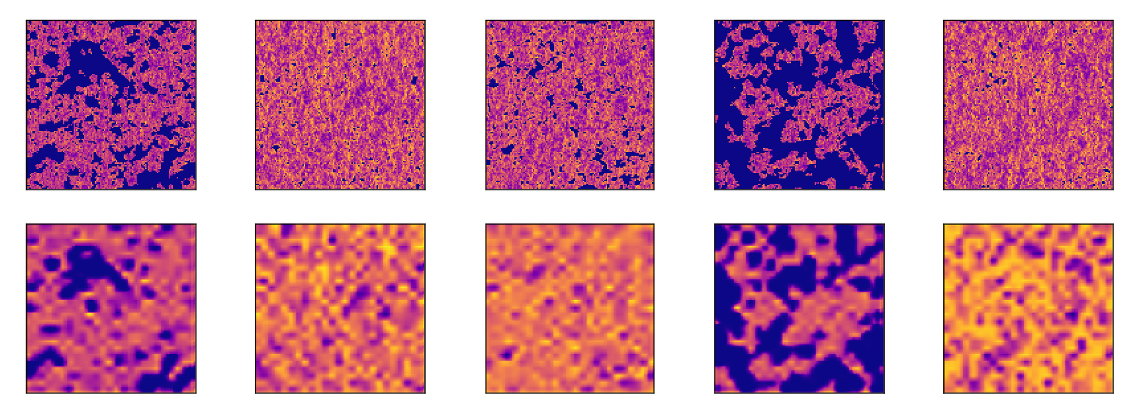

Fig. 3: Input image (top); decoded image (bottom)

asdf

Results

By guessing the parameters individually it was possible to obtain accurate results for both the escape fraction and the red-shift values, given that the other parameter was constant.

The red-shift values were easy to obtain, almost no tweaking of parameters was required to get above close to perfect accuracy. The baseline for this using a fully connected neural network was 48 percent.

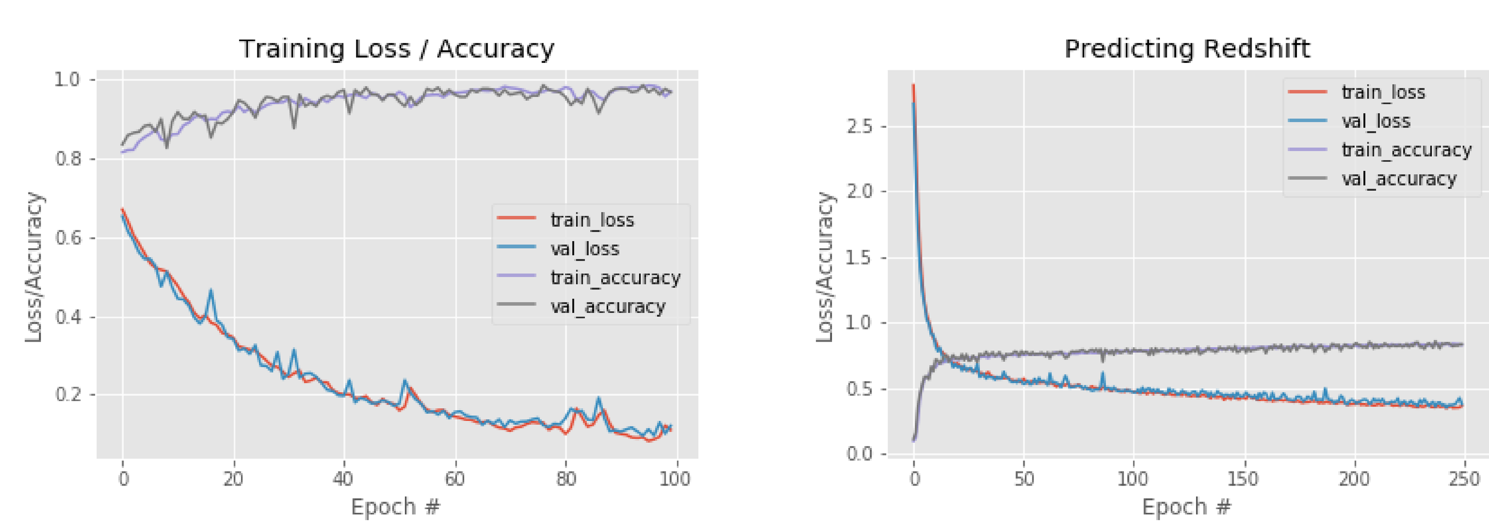

Fig. 4: Left training red-shift fixed esc frac; Right training red-shift multiple esc frac

Results

The escape fraction was not quite as easy, and much tweaking of parameters and redesigning was required.

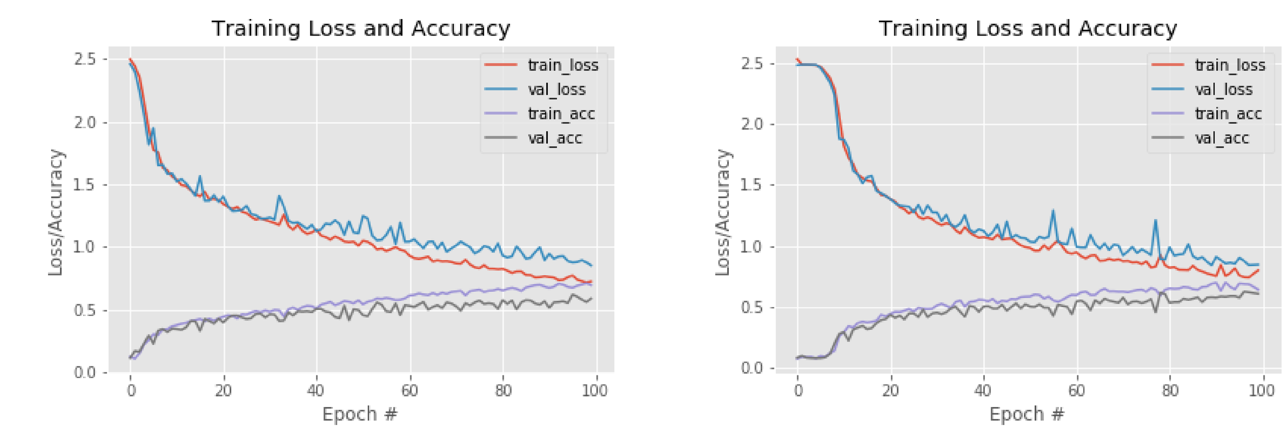

Fig. 5: Training to recognise the escape fraction

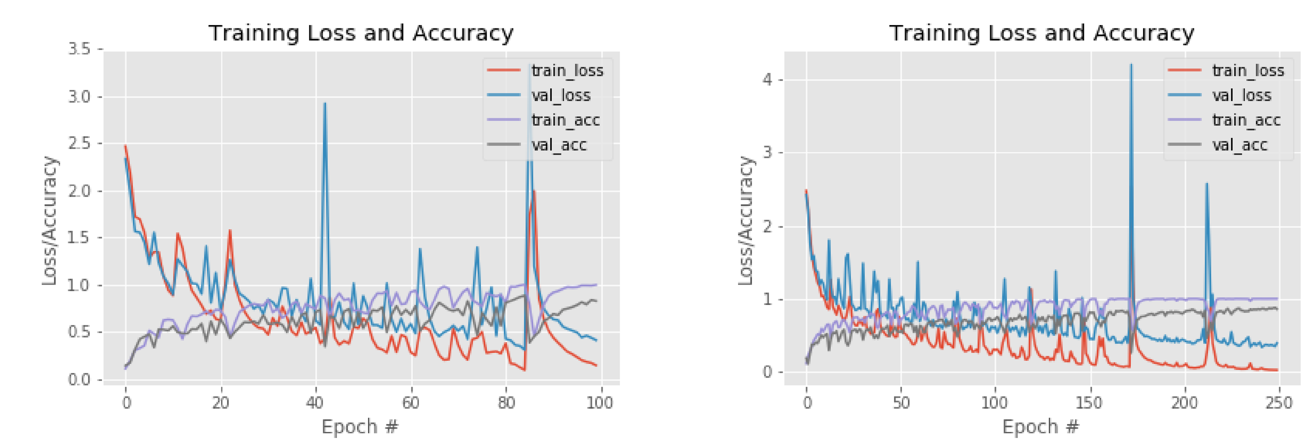

Fig. 6: Training to recognise the escape fraction

When the network was fed random slices as apposed to just slices corresponding to a single esc frac value

Future Work

Learn to classify all the parameters (e.g. the Turnover mass). Successfully learn the escape fraction, somehow I think down-sampling is drowning these out. Learn make the classifications inter-act with each other : aka, use the information from the red-shift classifier to help the esc-frac classifier make better predictions.

References

1. Emulating Simulations of Cosmic Dawn for 21cm Power Spectrum Constraints on Cosmology, Reionization, and X-Ray Heating - Nicholas Kern, Adrian Liu et al. https://iopscience.iop.org/article/10.3847/1538-4357/aa8bb4/pdf
2. Extensive use of online guids and resources, especially towards-datascience.com - multiple tutorials followed. As well as the keras documentation.
3. Sam's summer research project - https://github.com/samgagnon/remove-wedge