Identifying z~3 Active Galactic Nuclei from the HETDEX Survey Using Machine Learning

Valentina Tardugno Poleo[1]
[1]Department of Physics, College of Natural Sciences, The University of Texas

The University of Texas at Austin
College of Natural Sciences

Abstract

We use data from the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) to probe the faint-end slope of the AGN UV luminosity function at z~3. We used optical and infrared imaging in the 24 deg^2 Spitzer HETDEX Exploratory Large Area (SHELA) survey to construct a sample of 570 potential z~3 AGN and star-forming galaxies. We extracted HETDEX spectra at the position of these sources and used machine learning methods to attempt to identify those which exhibited AGN-like features. The dimensionality of the spectra was reduced using an autoencoder and the latent space was visualized through t-distributed stochastic neighbor embedding (t-SNE). Gaussian mixture models were then employed to cluster the encoded data and a labeled dataset was used to label each cluster. The number of found AGN was used to measure an AGN fraction for different magnitude bins, allowing us to examine the faint-end slope of the AGN UV luminosity function. Combining our results with studies at other redshifts, we can explore if there is a steepening faint-end slope with increasing redshift, which would imply a potential contribution by faint AGN to the ionizing photon budget at the end of reionization. The faint-end slope remains a key parameter to investigate the role that AGN played in reionizing the intergalactic medium.

Motivation

Measuring the AGN fractions could help determine which fit more adequately describes the UV luminosity function.

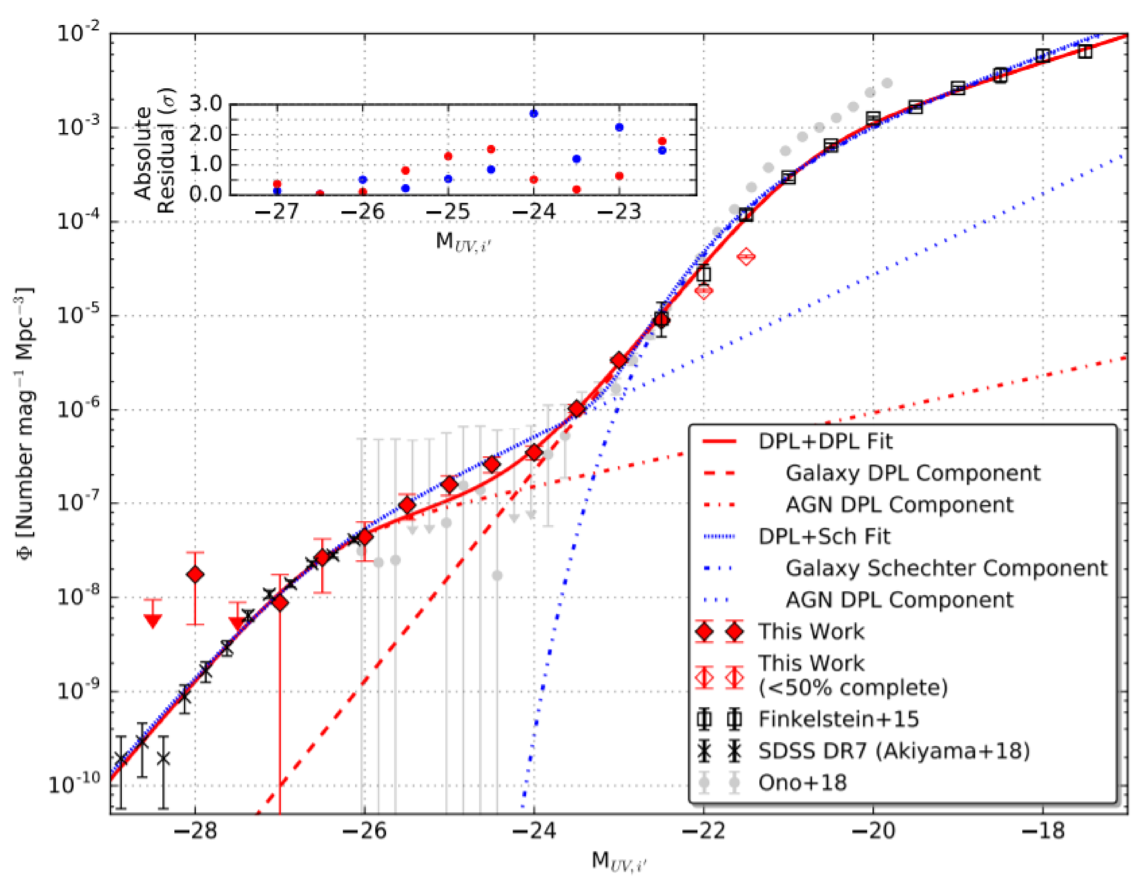

Figure 1: Rest-frame UV z = 4 luminosity function of star-forming galaxies and AGNs from the SHELA Field (Stevans et al., 2018).

Methods and Results

The 914-dimensional spectra were compressed into a 30-dimensional encoding using an autoencoder. The encoding was then visualized using t-SNE, and gaussian mixture models were employed to identify clusters.

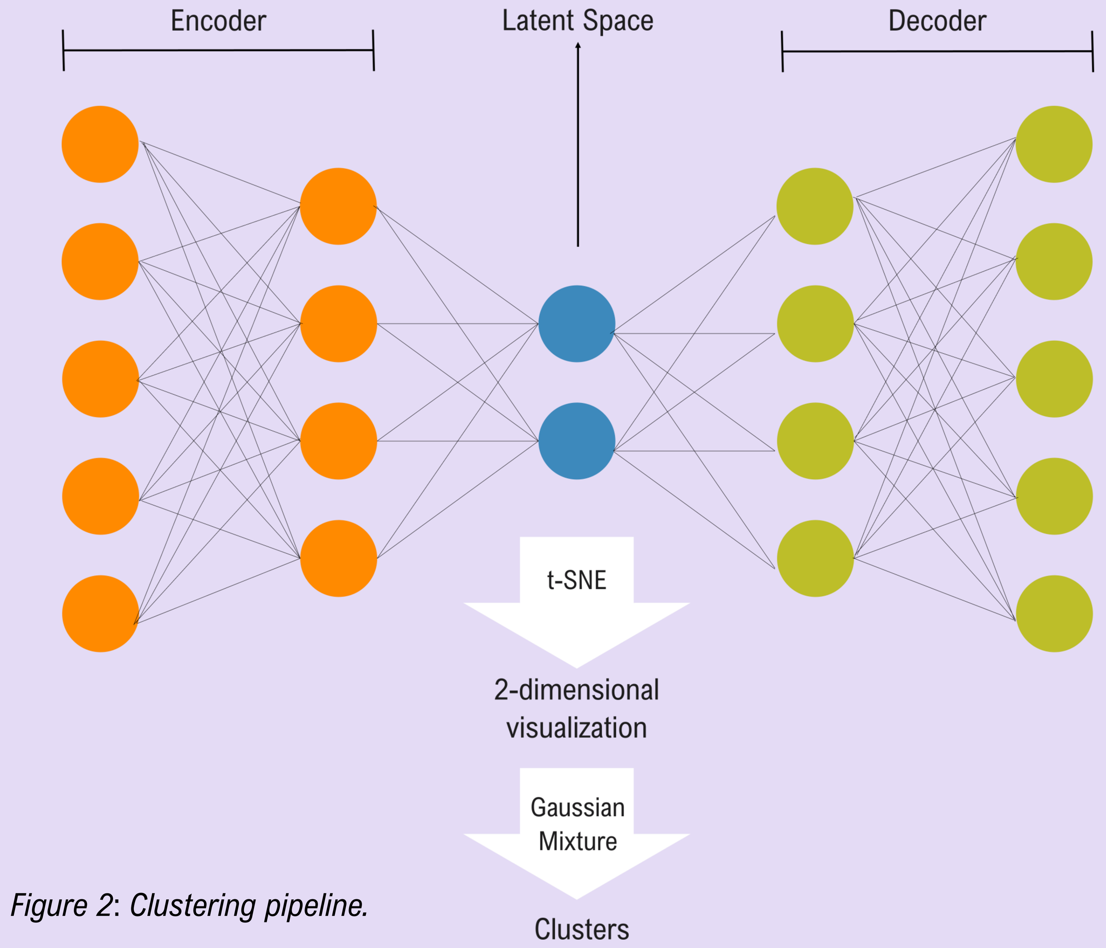

Figure 2: Clustering pipeline.

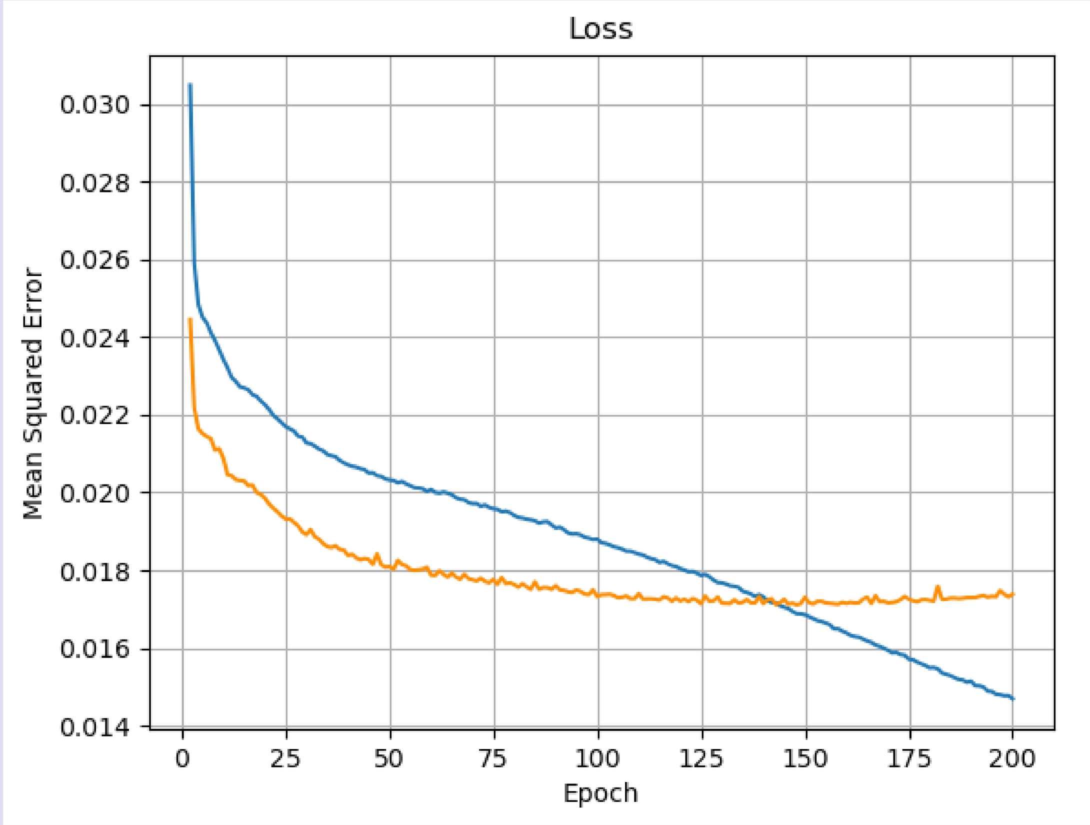

Figure 3. Training and validation loss curves for the autoencoder.

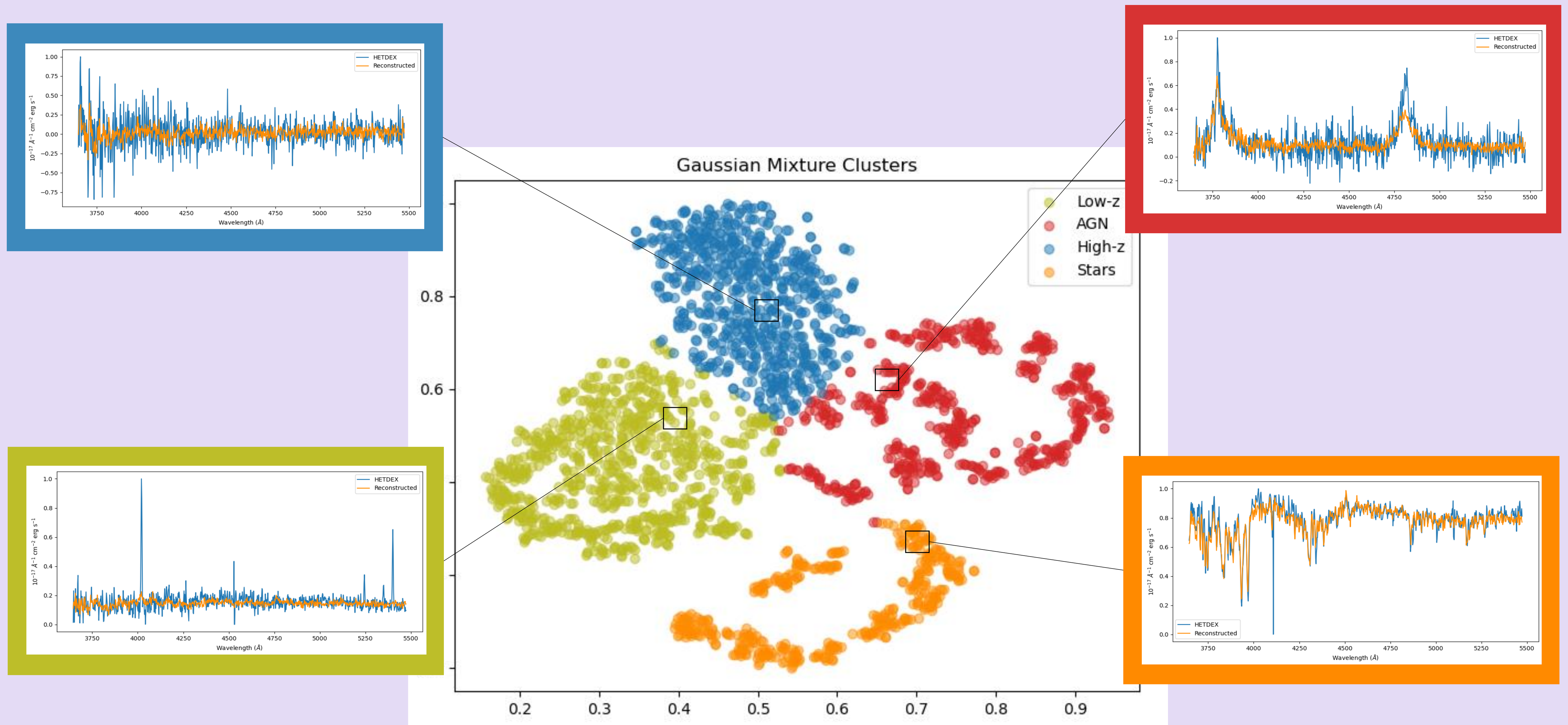

Figure 4: Gaussian mixture clusters and sample spectra. The blue spectra represent raw data extracted from the HETDEX survey. The orange spectra represent reconstructed spectra obtained from the 30-dimensional encoding.

Strengths:
- Adequate reconstruction of the general structure of the spectra
- Partial smoothing of spectra
- Effective reconstruction of spectral features such as broad emission lines and stellar absorption lines

Weaknesses:
- Potential loss of significant but more inconspicuous spectral features
- Inability to reconstruct narrow emission lines
- No identification method for pure noise spectra

Results

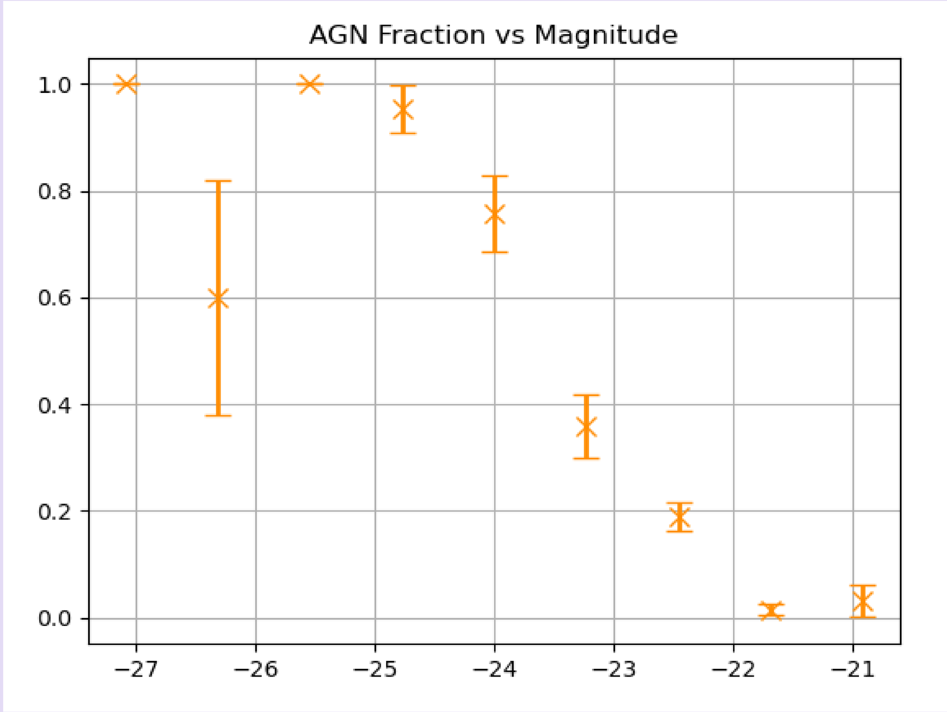

Figure 5: AGN fraction vs absolute magnitude plot.

	AGN	High-z	Low-z	Stars
AGN				
High-z				
Low-z				
Stars				

Figure 6: Confusion matrix.

Conclusion

We employed machine learning techniques to distinguish AGN and star-forming galaxies at z~3 and measured the AGN fraction at different magnitudes. First, we reduced the dimensionality of the spectra using an autoencoder. Applying t-SNE on the reduced data allowed us to visualize a separation between astronomical objects, which we identified using gaussian mixture algorithms. After removing identified contaminants, the AGN fraction was calculated for nine magnitude bins. Extending the described methodology to other redshifts could allow us to explore the evolution of the faint-end slope of the AGN UV luminosity function and assess the potential contribution of AGN to reionization.

References

Stevans, M. L., Finkelstein, S. L., Wold, I., Kawinwanichakij, L., Papovich, C., Sherman, S., Ciardullo, R., Florez, J., Gronwall, C., Jogee, S., Somerville, R. S., Yung, A. (2018). Bridging star-forming galaxy and AGN ultraviolet luminosity functions at z=4 with the SHELA wide-field survey. ArXiv. .

Acknowledgements

Thank you to Dr. Steven Finkelstein for supervising my project, as well as to Dr. Gene Leung and Oscar Chavez for their support and feedback throughout my research.