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

Valentina Tardugno Poleo[1] , Steven Finkelstein[2], Gene Leung[2], Daniel Mock[3]
[1]Department of Physics, College of Natural Sciences, The University of Texas
[2]Department of Astronomy, College of Natural Sciences, The University of Texas
[3]Florida State University

The University of Texas at Austin
College of Natural Sciences

GEVIP

HETDEX
Hobby-Eberly Telescope Dark Energy Experiment
Illuminating the Darkness

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

- The UV luminosity function is consistent with both a shallow and steep AGN faint-end slope.
- Measuring the AGN fraction could help determine which fit more adequately describes the luminosity function.
- Although HETDEX does not probe z~4 galaxies, the described methodology serves as a proof-of-concept.

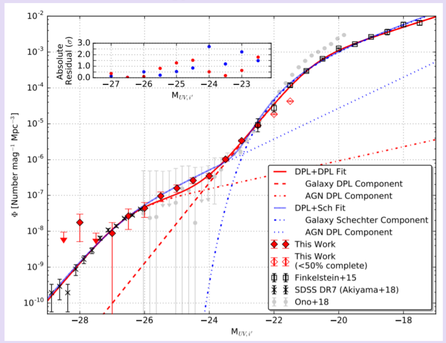

Figure 1: Rest-frame UV z = 4 luminosity function of star-forming galaxies and AGN from the SHELA Field (Stevans et al., 2018).

Methods and Results

Data Selection

1. Selected potential z~3 AGN and star-forming galaxies from the SHELA catalog
2. Extracted HETDEX spectra at the sources' position. The spectra covered wavelengths between 3645-5475 Å
3. Normalized the resulting 570 spectra by dividing the flux by the maximum flux value of each spectrum

Data Reduction and Clustering

4. Compressed the 914-dimensional spectra into a 30-dimensional space using the encoder network of an autoencoder
5. Visualized the encoding using t-SNE
6. Employed Gaussian Mixture models on the resulting two-dimensional embedding to identify clusters.

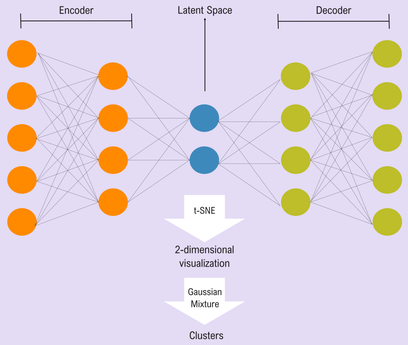

Figure 2: Schematic of analysis pipeline. The data reduction portion is composed of the encoder network and t-SNE. The clustering portion consists of the Gaussian Mixture models

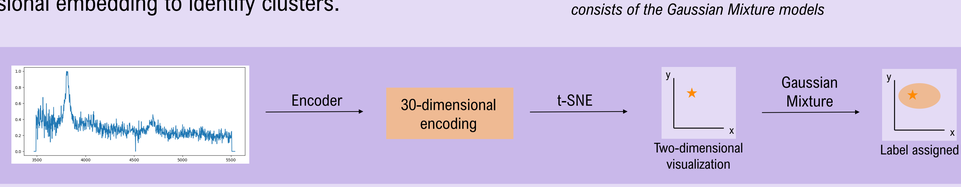

Figure 3: Path of spectra through data reduction and clustering pipeline

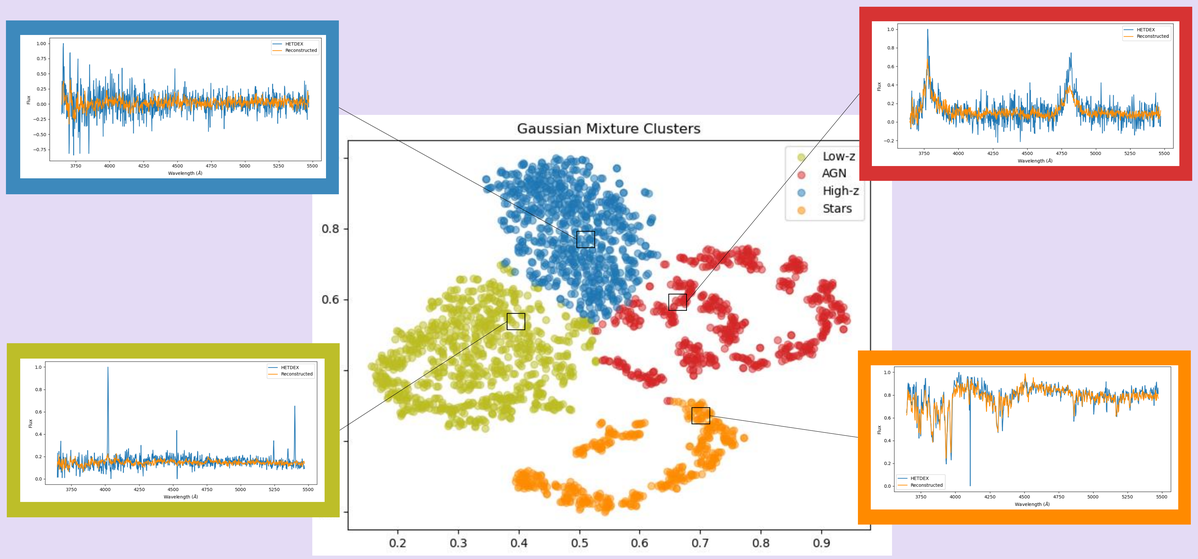

Figure 4: Gaussian mixture clusters and sample spectra. The blue spectra represent raw data extracted from the HETDEX survey. The orange spectra represent reconstructed spectra obtained from running the 30-dimensional encoding through the decoder network. The found clusters were used to remove stars and low redshift galaxies from the sample, and identify AGN and star-forming galaxies

Strengths

- Adequate reconstruction of spectral trends
- 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

Results

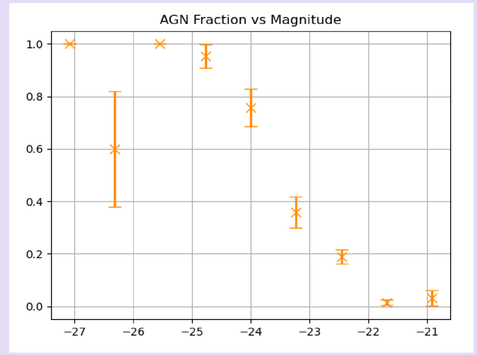

Figure 5: AGN fraction vs r-band absolute magnitude plot.

	AGN	High-z	Low-z	Stars
AGN	409	16	15	4
High-z	3	247	4	0
Low-z	9	57	768	3
Stars	21	6	28	378

Figure 6: Confusion matrix of the training data. Rows and columns represent true and predicted labels respectively

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.