Adoption of the Transformer Architecture for Supernova Detection in the DESI Survey

Undergraduate Senior Thesis 2023

Anthony (AJ) LaBarca
Supervised by: Professor Segev BenZvi
University of Rochester

May 9, 2023

Outline

Introduction

Deep Learning Techniques

Creation and Training of the Spectral ViT

Spectral ViT Model Performance

Conclusions

SNe Classification

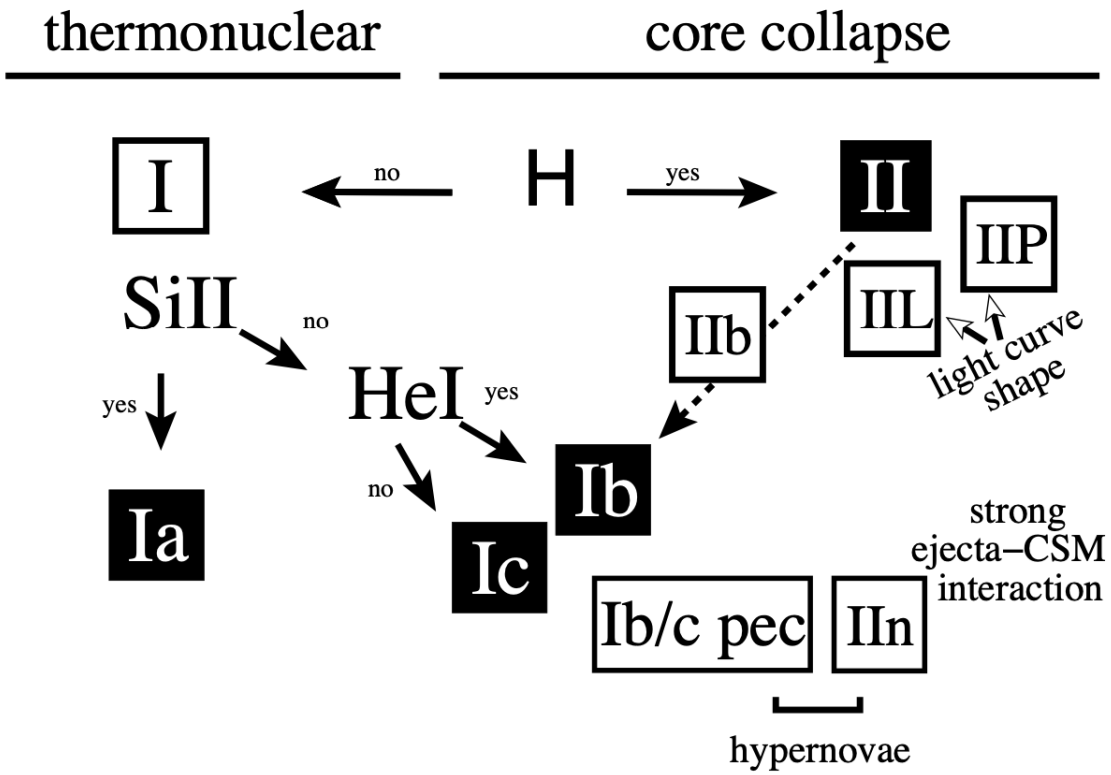


Figure: Supernova Classification Scheme (Adapted from Turatto (2003)

DESI

- Located at Kitt Peak National Observatory, Az
- Capable of resolving the [O II] $\lambda\lambda 3726, 3729$ doublet (Guy et al. 2023)
- Bright Galaxy Survey (BGS): $z < 0.5$, $\sim 10^5$ observed SNe! (DESI Collaboration et al. 2016; Hahn et al. 2022)

DESI Spectra SNe Classification

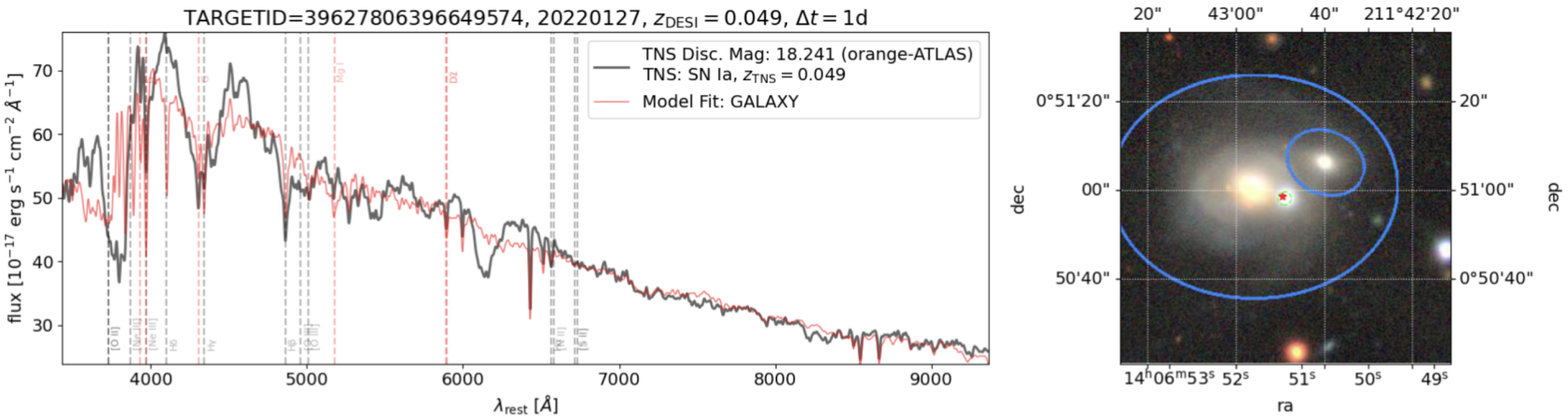


Figure: DESI Spectra of Publicly Classified SNe Type Ia

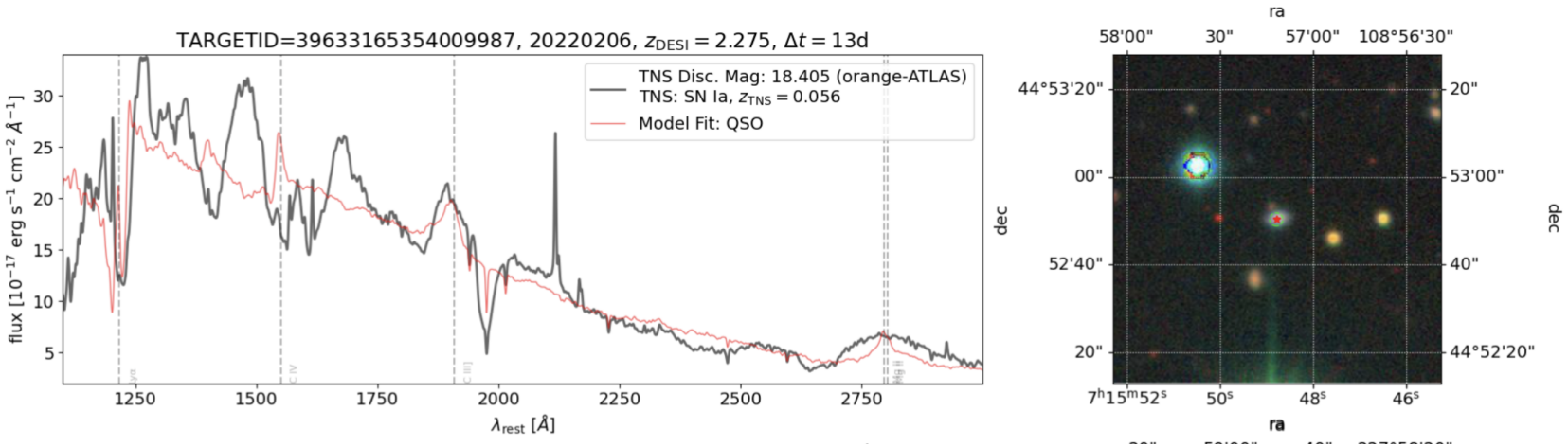


Figure: DESI Spectra of Publicly Classified SNe Type Ia – Bad Redshift Fit

SNe Template

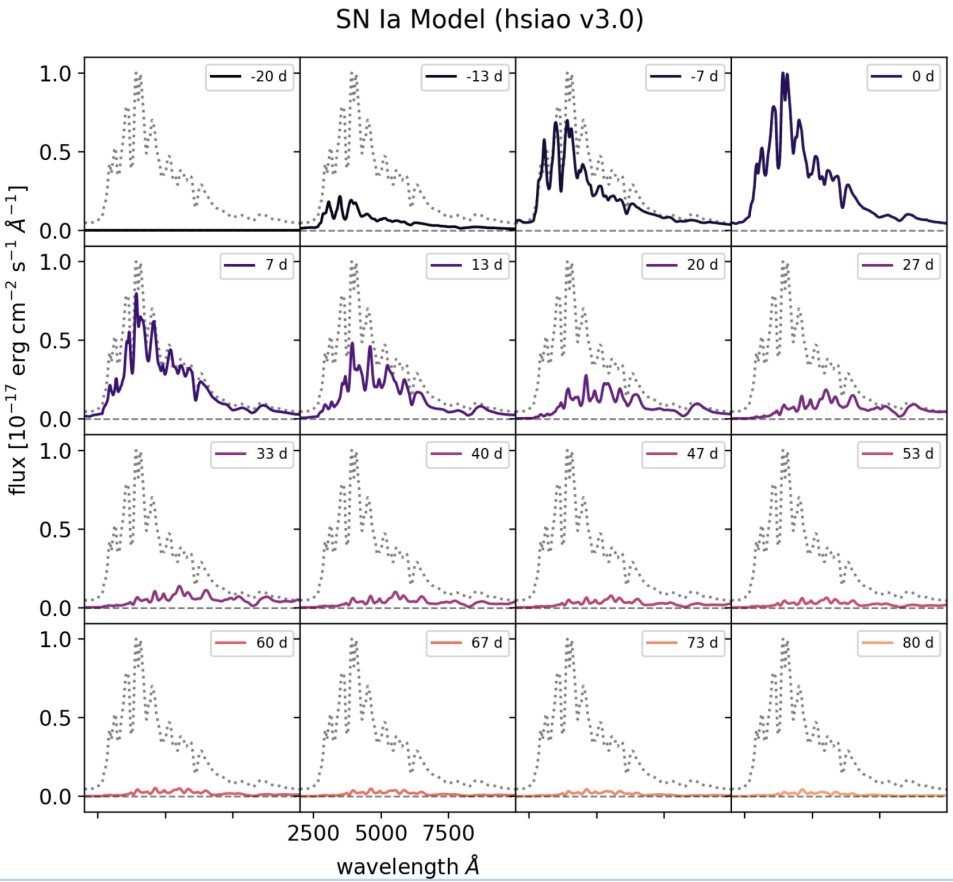


Figure: Spectral Template of SN Ia SNe (-20→ +80 days after peak) Adapted from BenZvi, Tiwari, and Wasserman (2020)

DL in DESI Transient Analysis

- 1D CNN (Wasserman 2021)
- 2D CNN (Sepeku 2022)

Problem: High precision, Low recall

Multi-Layer Perceptron

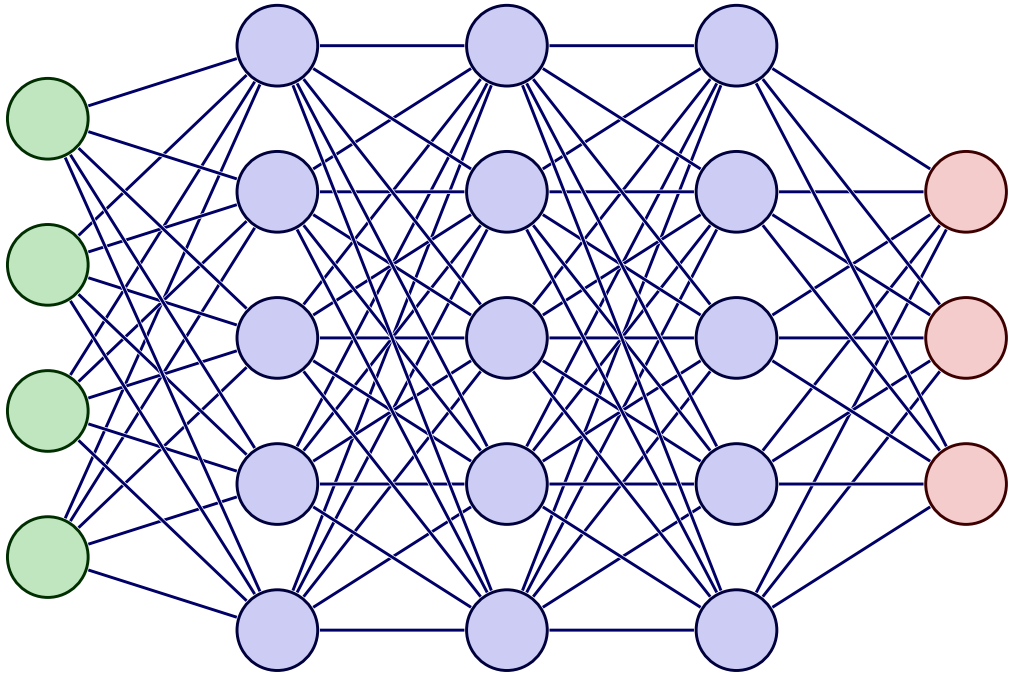


Figure: Multi-Layered Perceptron (adapted from Neutelings (2021))

CNN Architecture

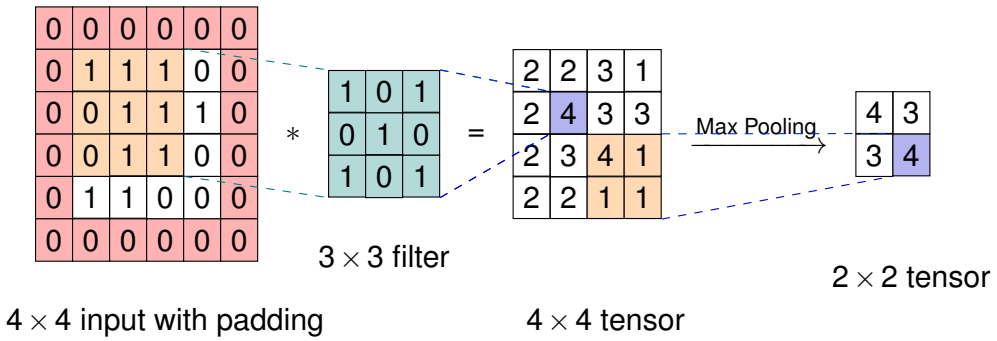


Figure: CNN Unit (adapted from Neutelings (2022))

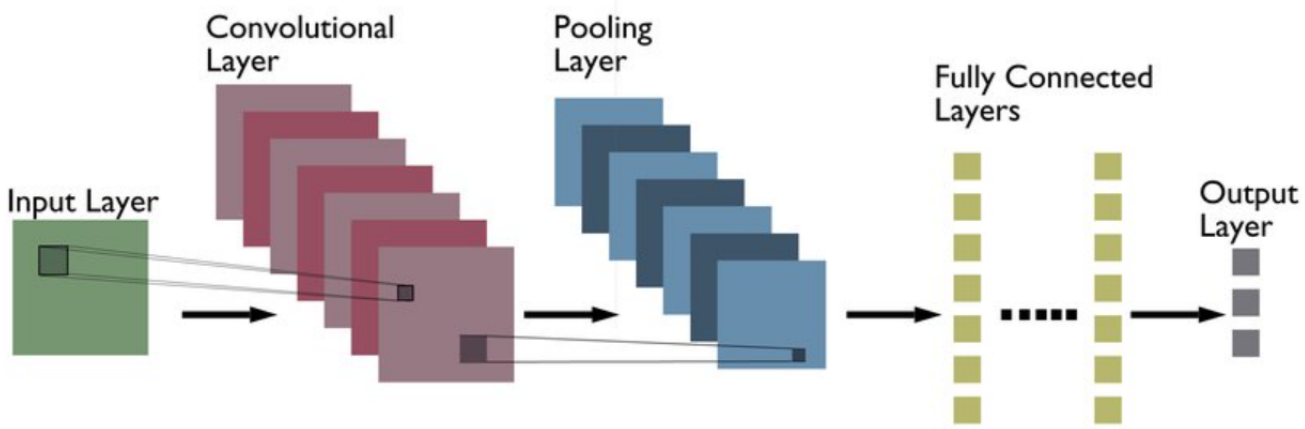


Figure: Traditional CNN Architecture for Classification (adapted from Kumar 2022)

2D CNN Performance

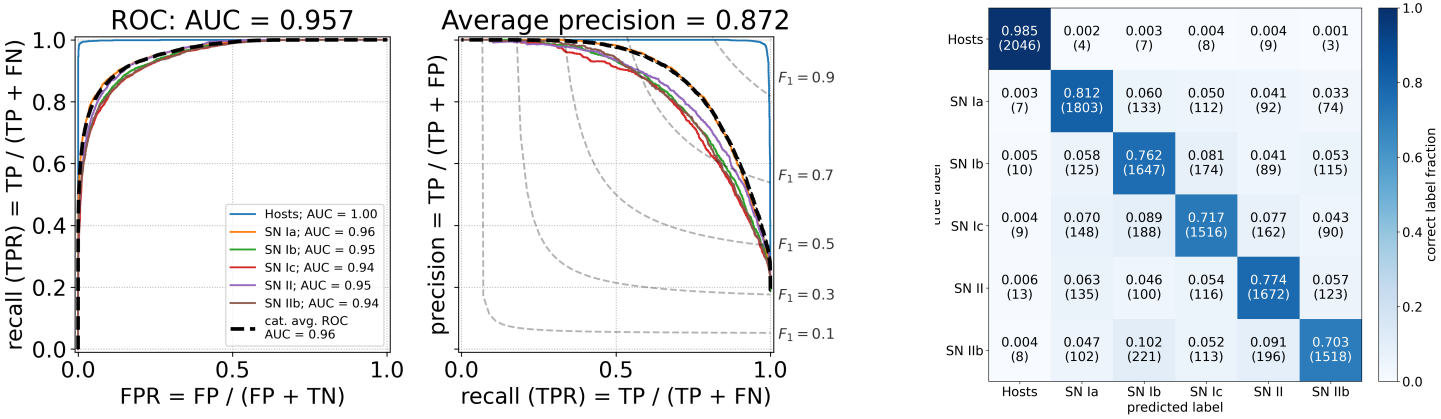


Figure: CNN Classifier

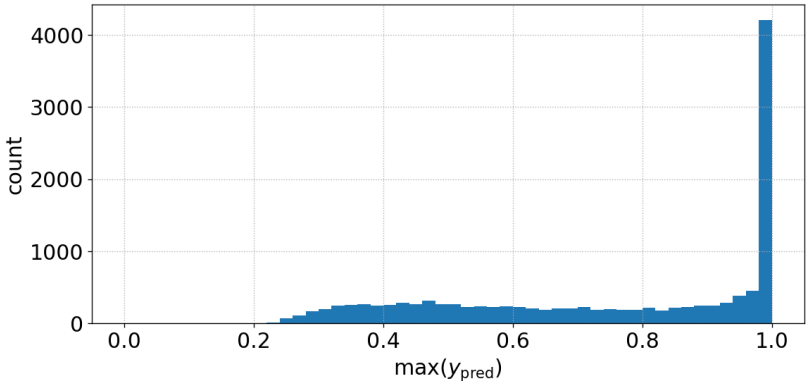


Figure: CNN's Maximum Output Vector

CNN Cut to Increase Performance

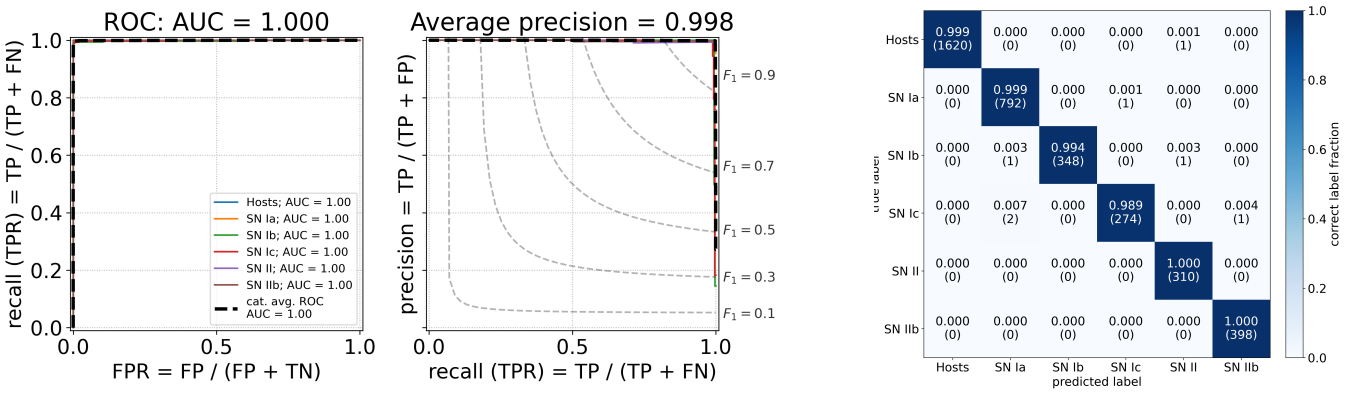


Figure: CNN Classifier 99% confidence cut (28.8% remaining)

The Transformer Architecture

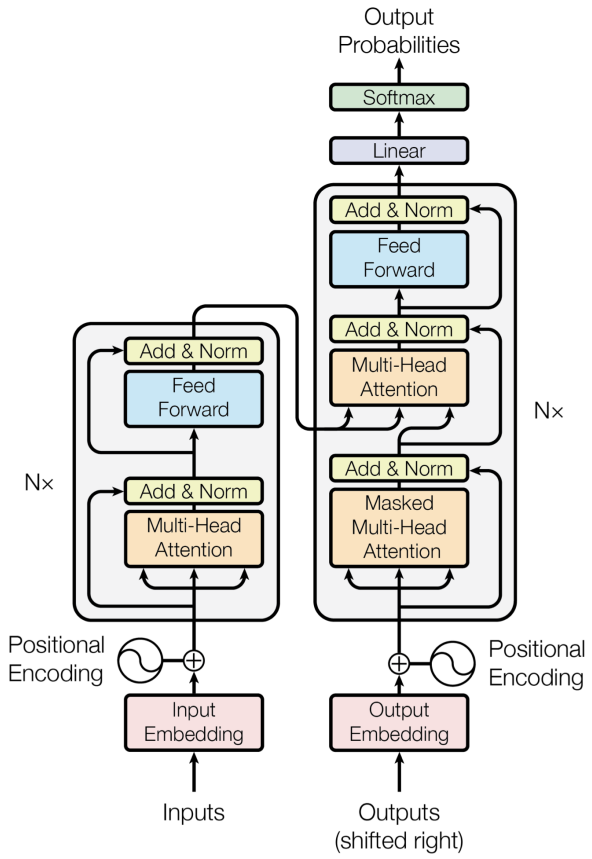


Figure: Original Transformer Architecture (Image from Vaswani et al. (2017))

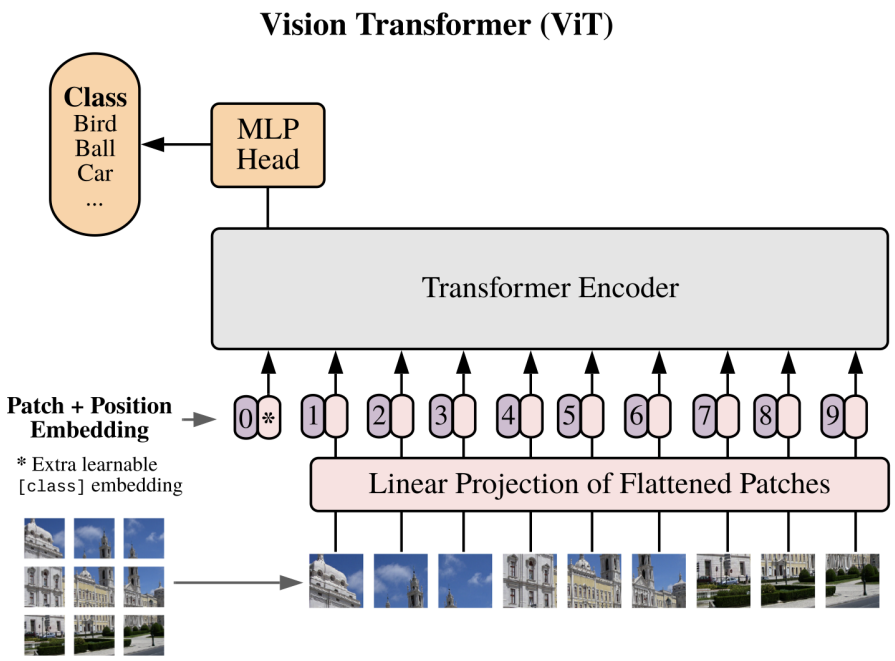
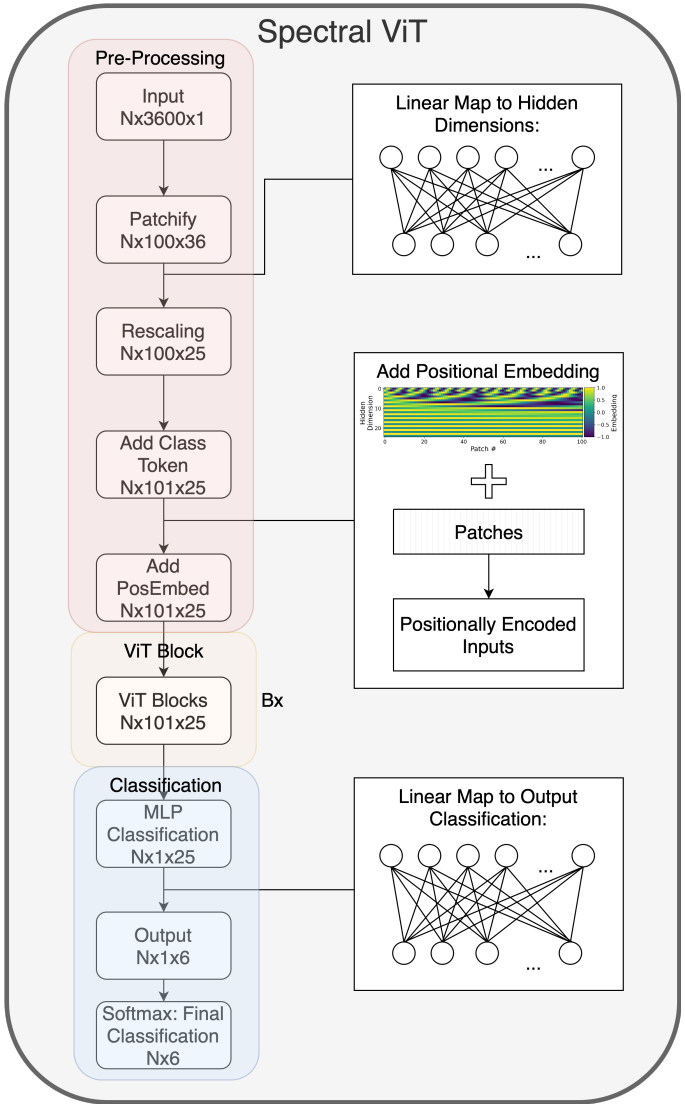


Figure: ViT Architecture (Image from Dosovitskiy et al. (2020))

Spectral ViT

Transformer architecture must satisfy

- 1D Nature of Spectra (more like original transformer)
- Cut up data into 'patches' synthetically (more like ViT)



Positional Embedding

$$\text{Embedding}_{ij} = \begin{cases} \sin\left(\frac{i}{10000^{(j/\text{patch size})}}\right) & \text{if } j \text{ is even} \\ \cos\left(\frac{i}{10000^{((j-1)/\text{patch size})}}\right) & \text{if } j \text{ is odd} \end{cases} \tag{1}$$

Positional Embedding

$$\text{Embedding}_{ij} = \begin{cases} \sin\left(\frac{i}{10000^{(j/\text{patch size})}}\right) & \text{if } j \text{ is even} \\ \cos\left(\frac{i}{10000^{((j-1)/\text{patch size})}}\right) & \text{if } j \text{ is odd} \end{cases} \quad (1)$$

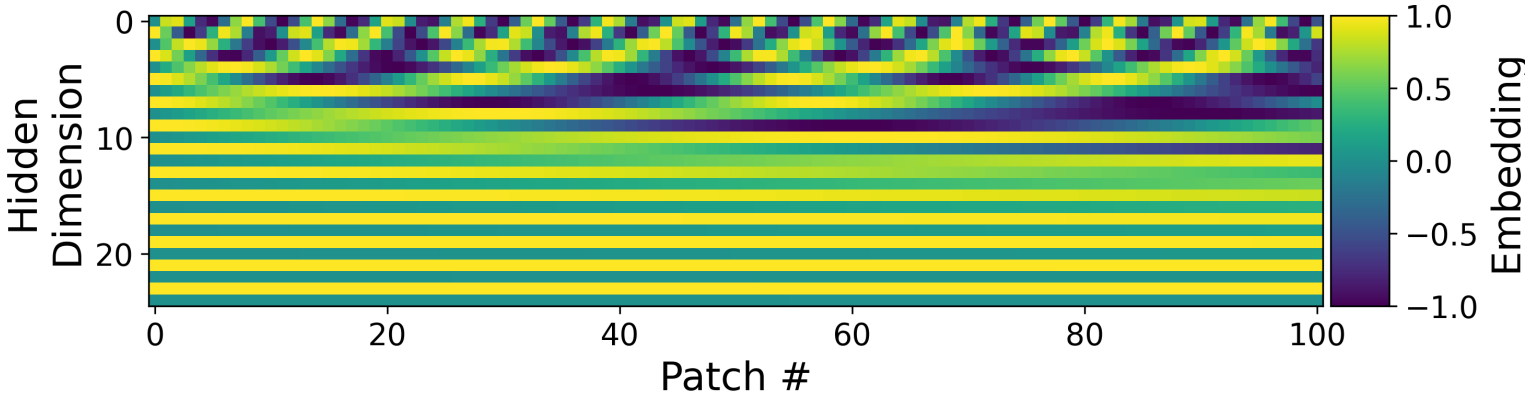


Figure: Positional Embeddings for ViT

Attention

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V \tag{2}$$

Attention

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V \tag{2}$$

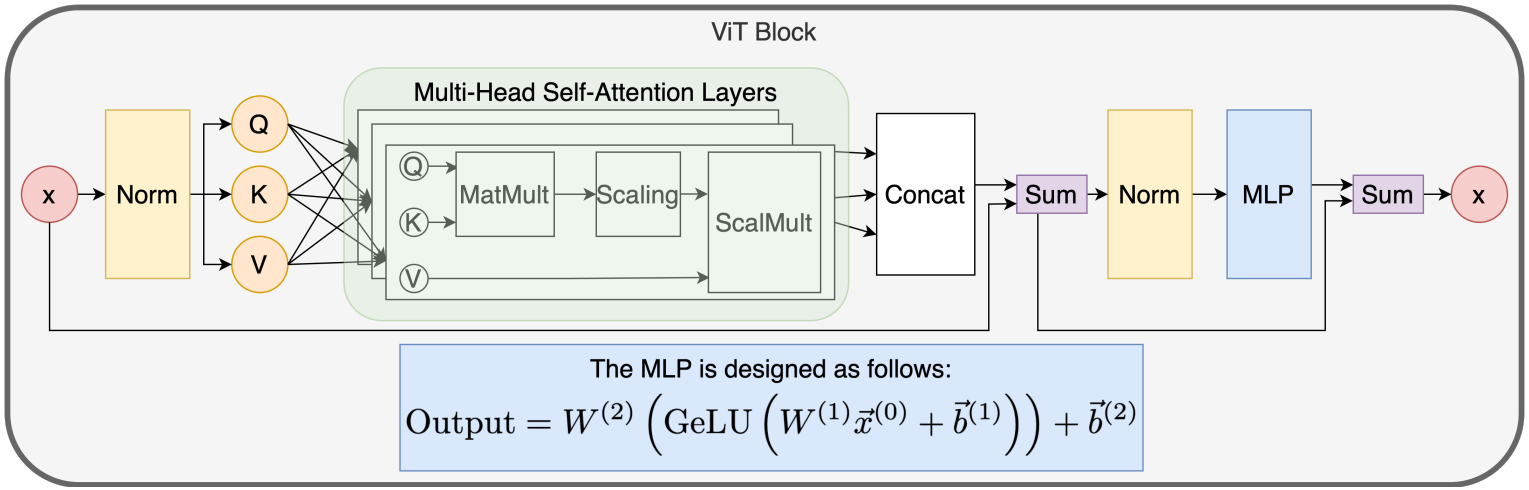
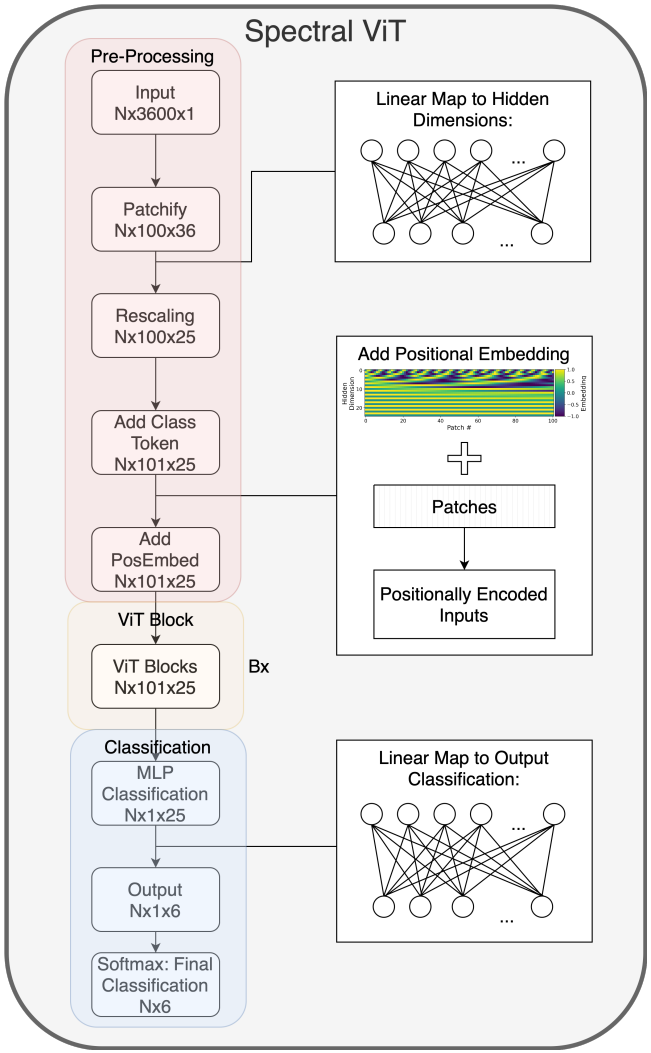
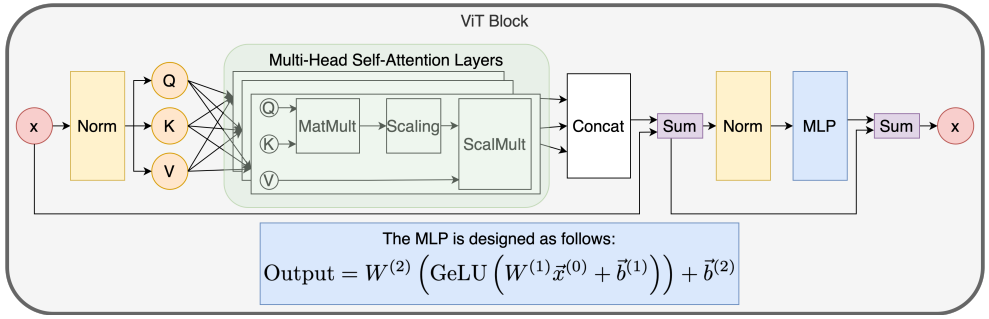


Figure: ViT Block Architecture for the Spectral ViT

Final Architecture





Training on Synthetic Data

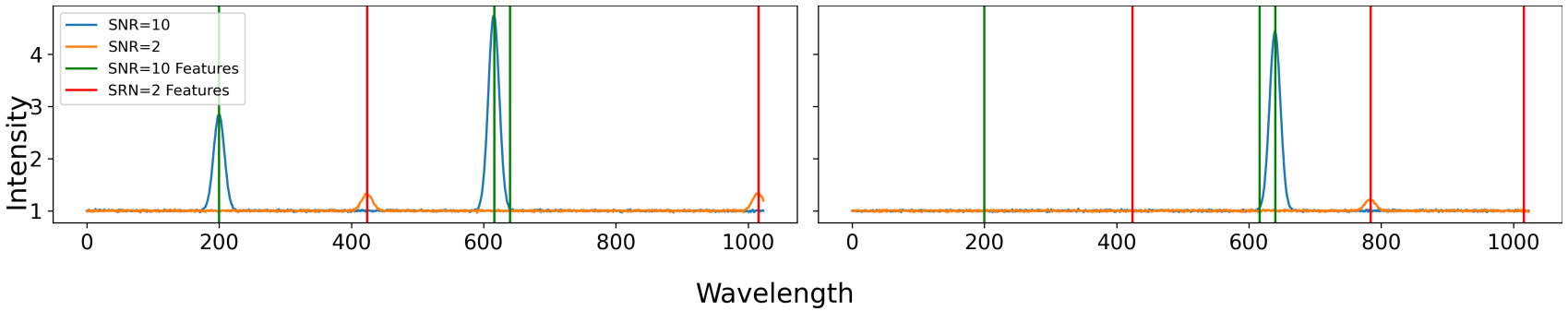


Figure: Synthetic spectra

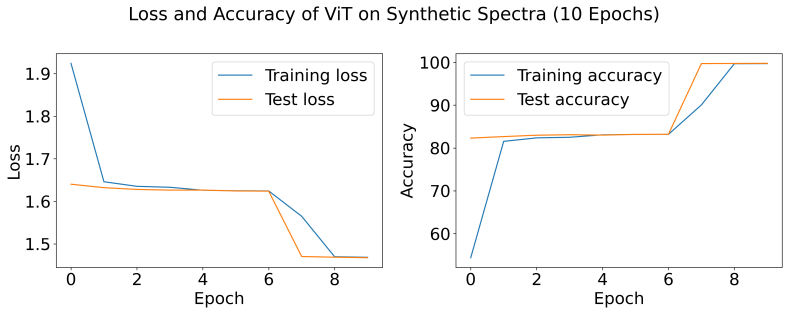


Figure: SNR=10

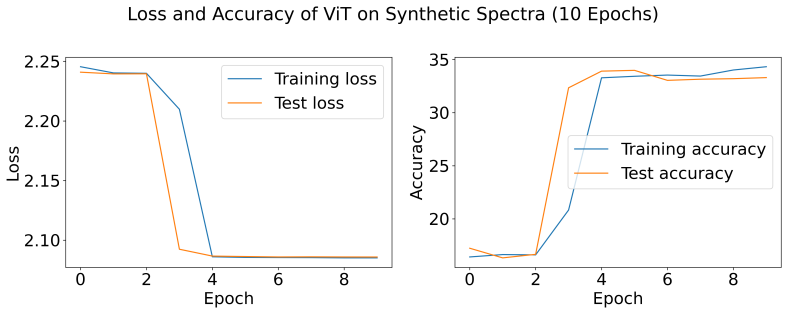


Figure: SNR=2

DESI Synthetic Data

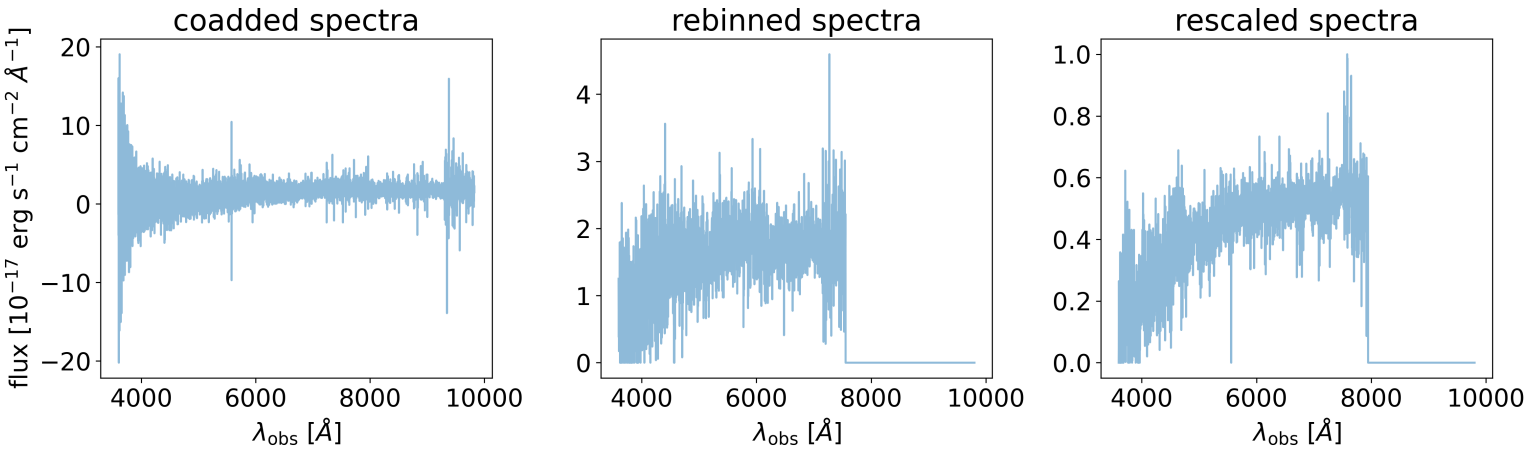


Figure: DESI Spectra Pre-Processing – Redshift Corrected

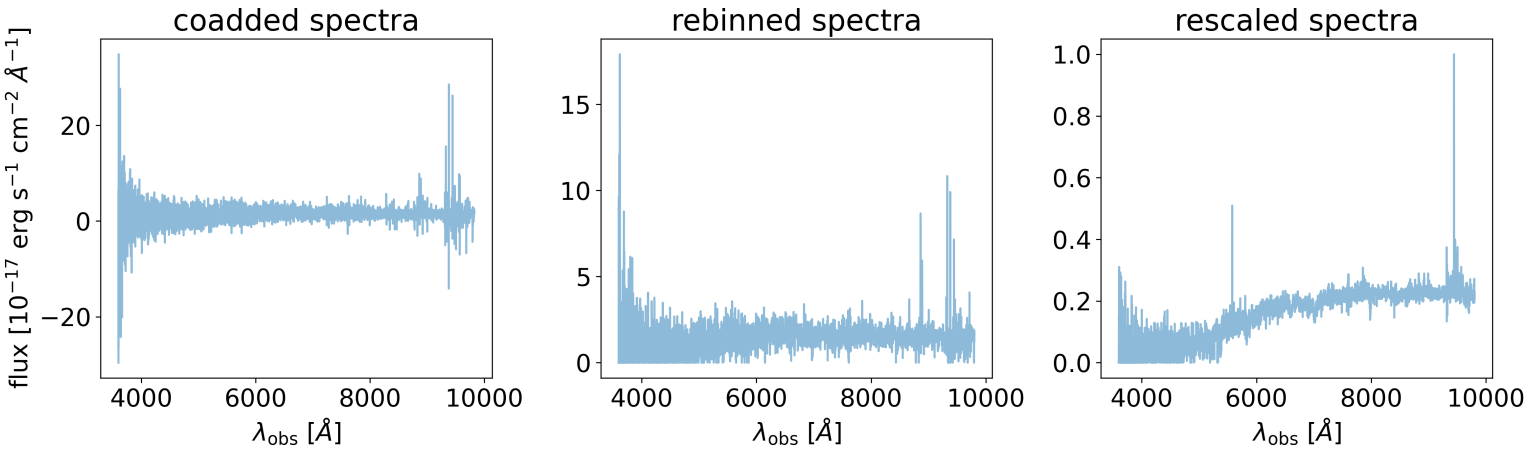


Figure: DESI Spectra Pre-Processing – Non-Redshift Corrected

CNN Training

Hyperparameter	Value
Learning Rate	0.00004135238172950965
Regularization	0.06
Dropout Rate	0.40904759925886294
Batch Size	50

Table: Hyperparameters of the CNN, model adapted from Sepeku (2022).

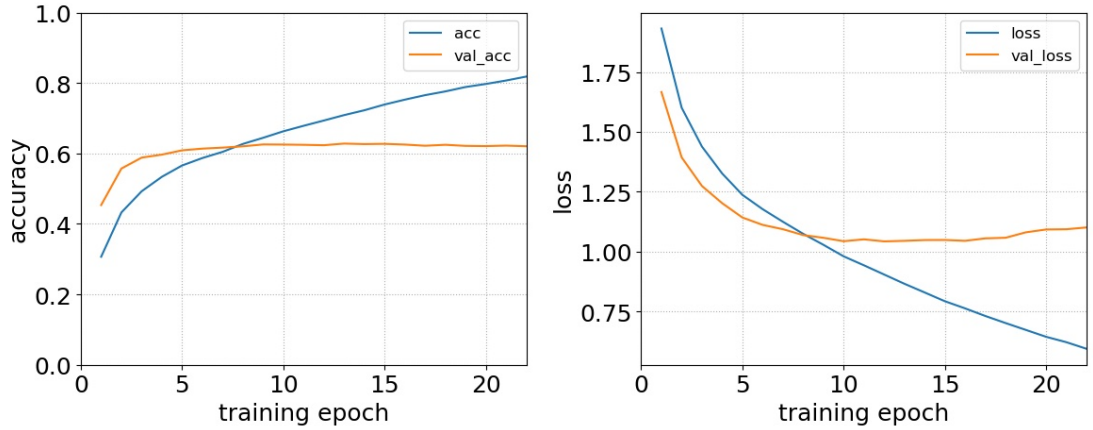


Figure: Hyperparameters used during CNN training (Sepeku 2022)

Spectral ViT Training

Hyperparameter	Smaller Architecture	Larger Architecture
Learning Rate	0.0001	0.01
Batch Size	16	16
Blocks	4	4
Heads	5	12
Hidden Dimension	25	72

Table: Hyperparameters used during Spectral ViT training

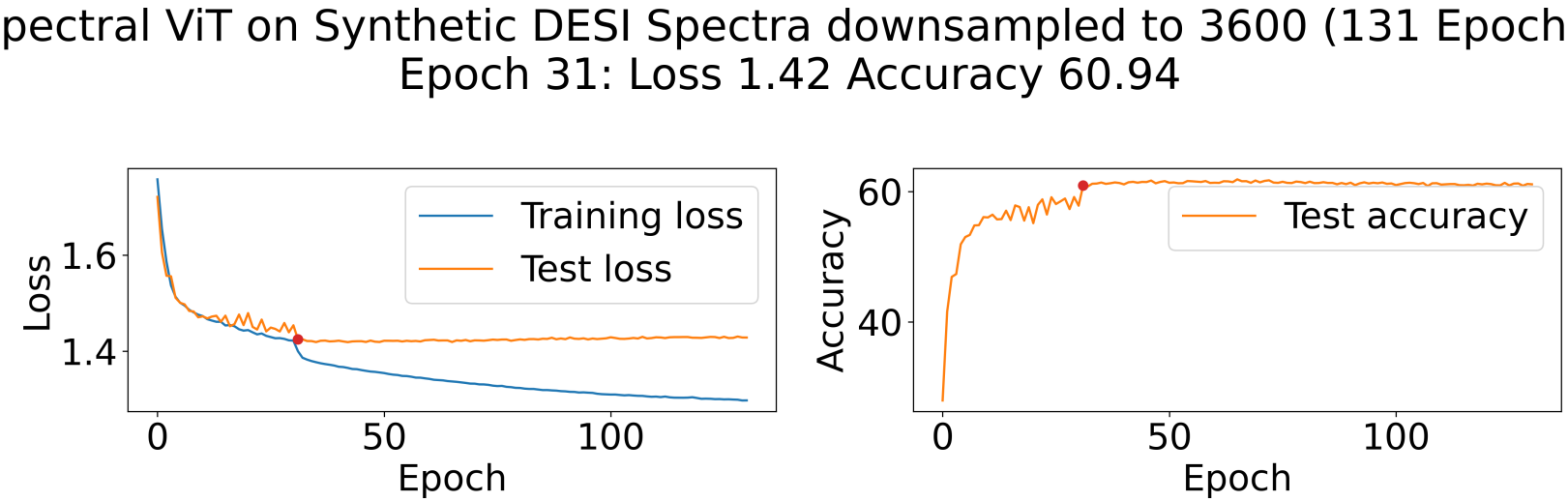


Figure: Training of Spectral ViT: V1 (3600 Bins)

Additional (Failed) Training

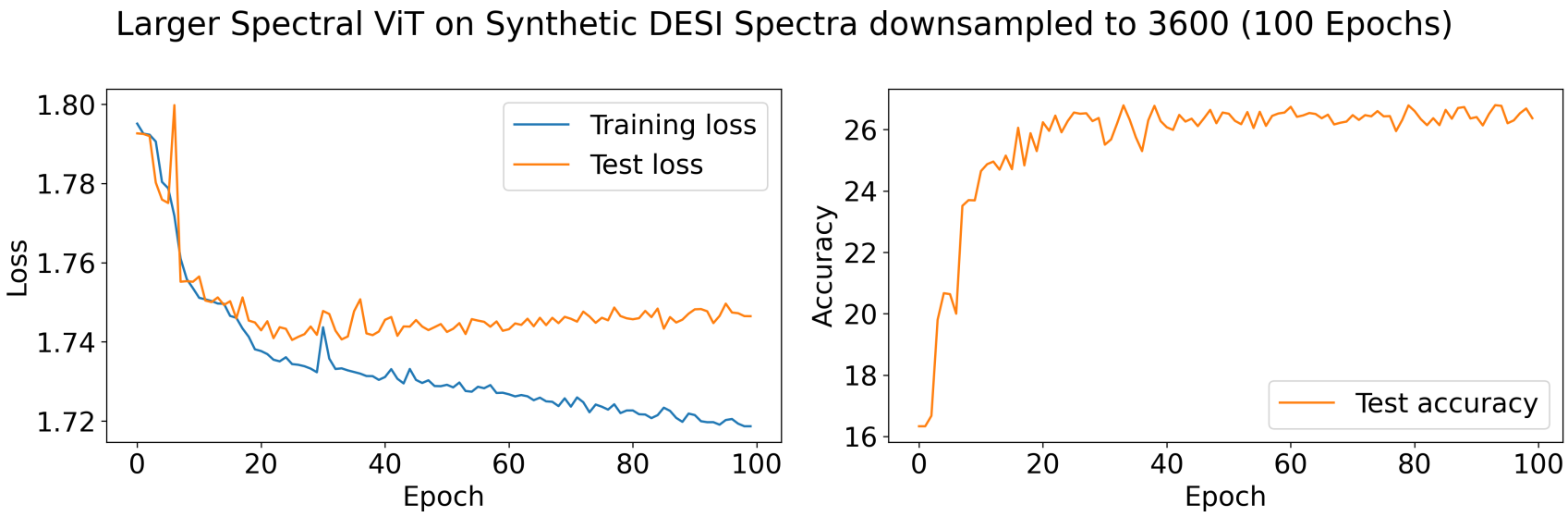


Figure: Training of Spectral ViT: V1 Big (3600 Bins)

Additional (Failed) Training

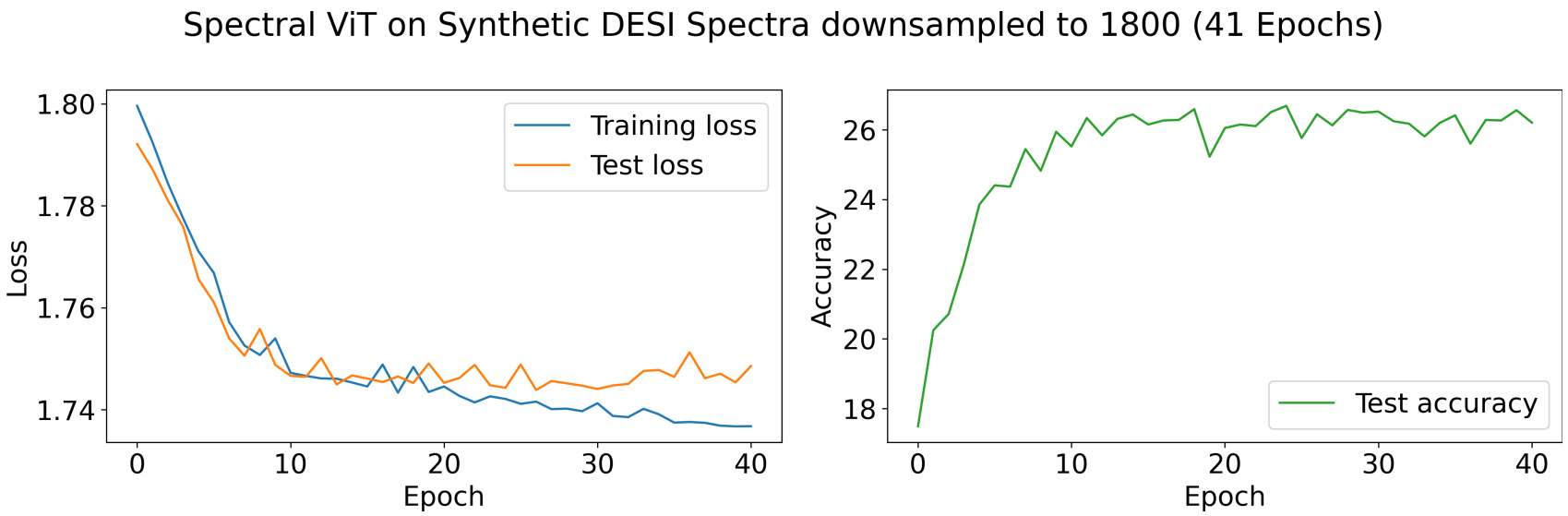


Figure: Training of Spectral ViT: V1.2 (1800 Bins)

Additional (Failed) Training

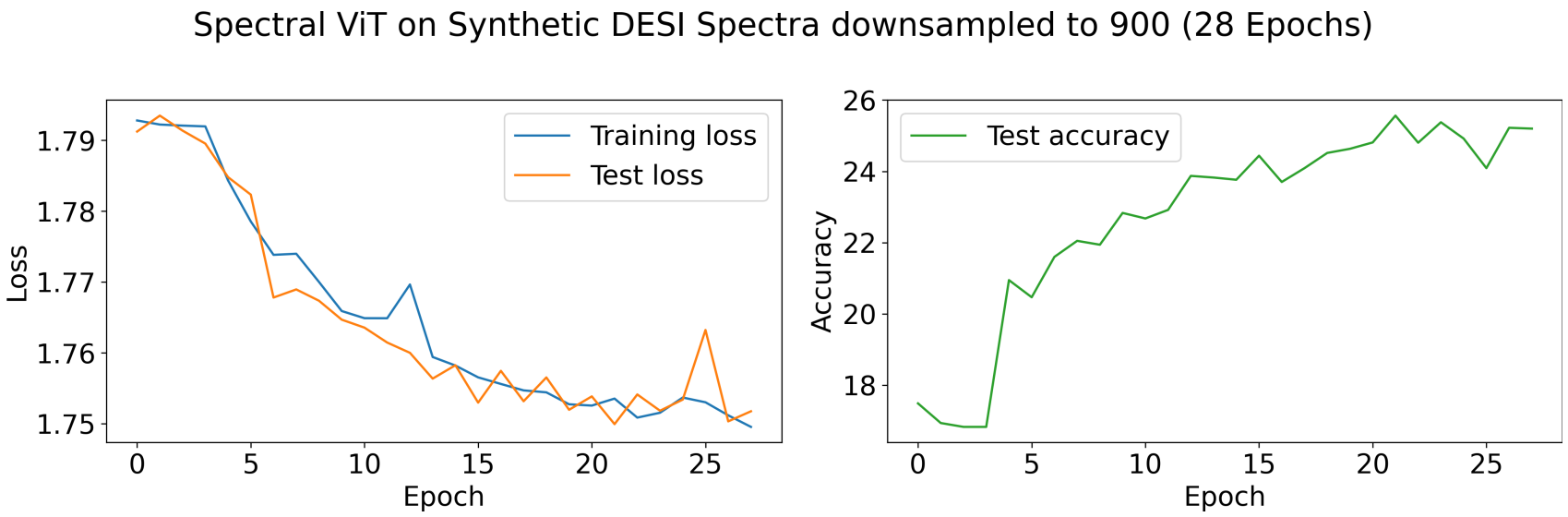
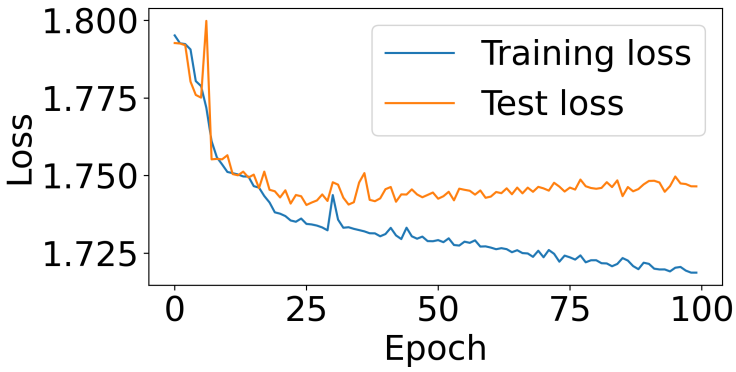
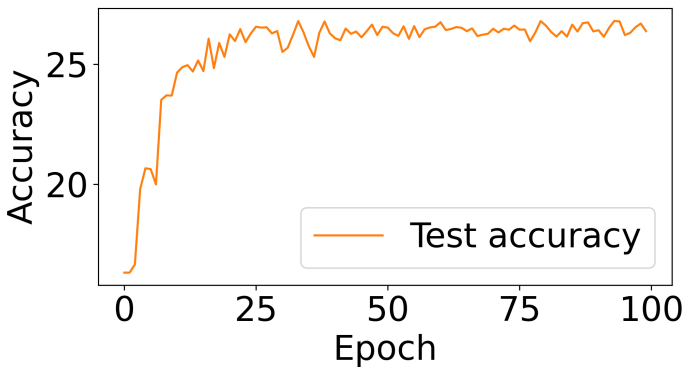


Figure: Training of Spectral ViT: V1.3 (900 Bins)

Training for Non-Redshift Corrected Data

Larger Spectral ViT on Synthetic DESI Spectra downsampled to 3600 (100 Epochs)





Spectral ViT V1

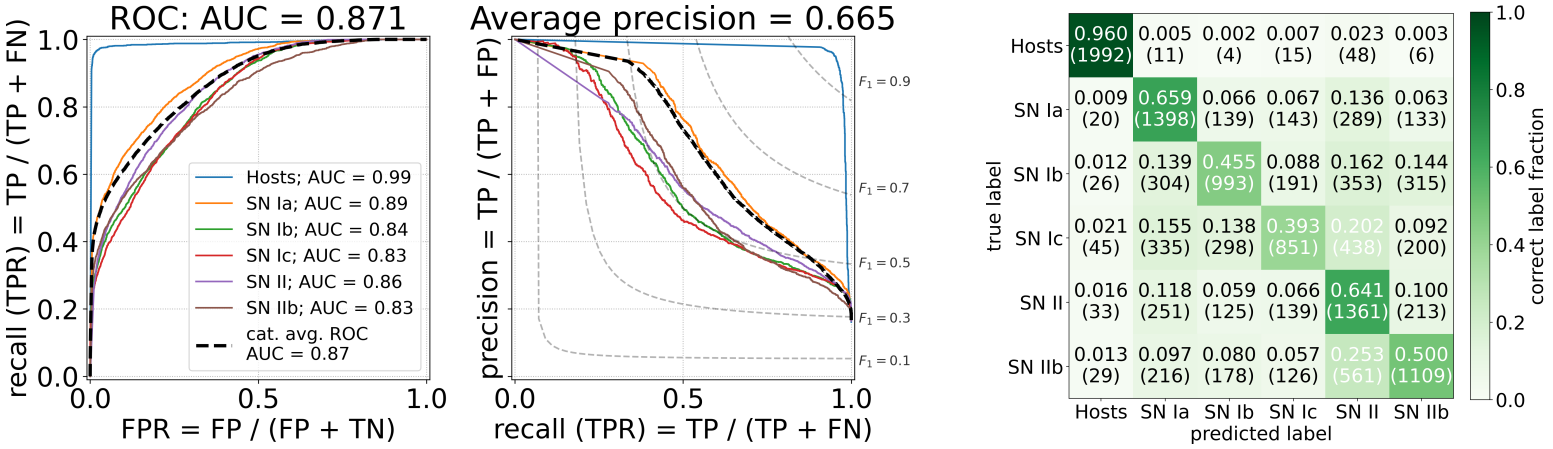


Figure: Spectral ViT V1 Classifier

Spectral ViT V1

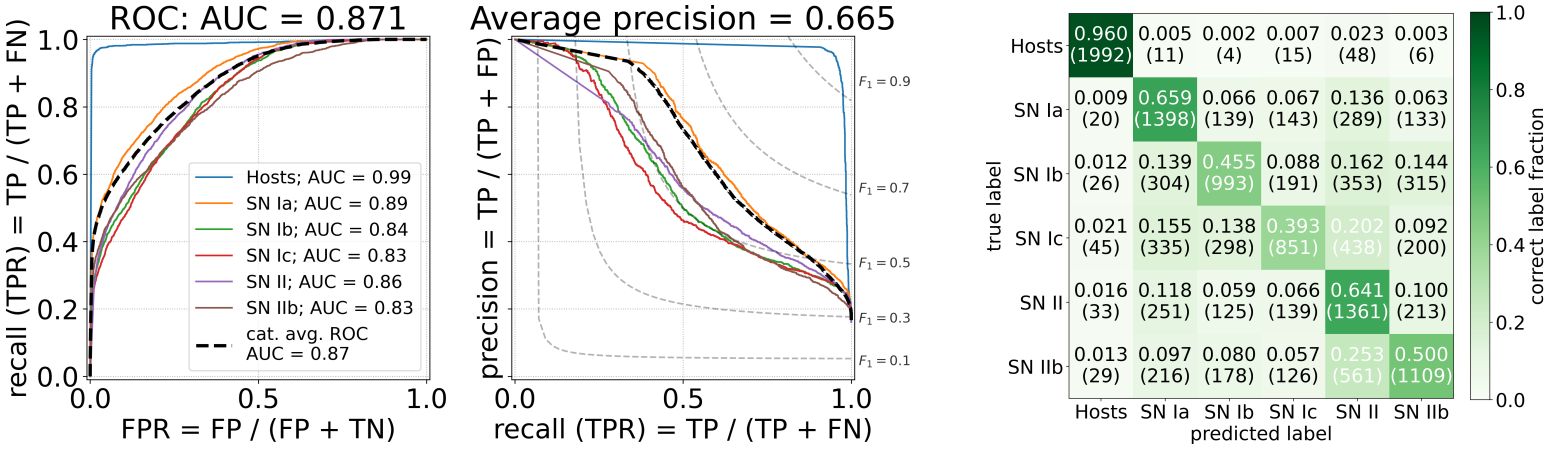


Figure: Spectral ViT V1 Classifier

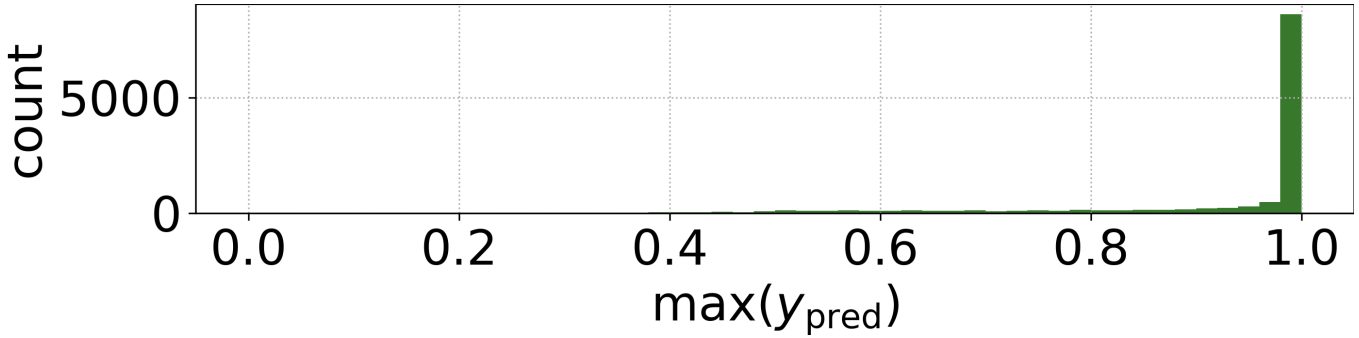


Figure: Spectral ViT V1 Maximum Output Vector Value

Spectral ViT V1 Cuts

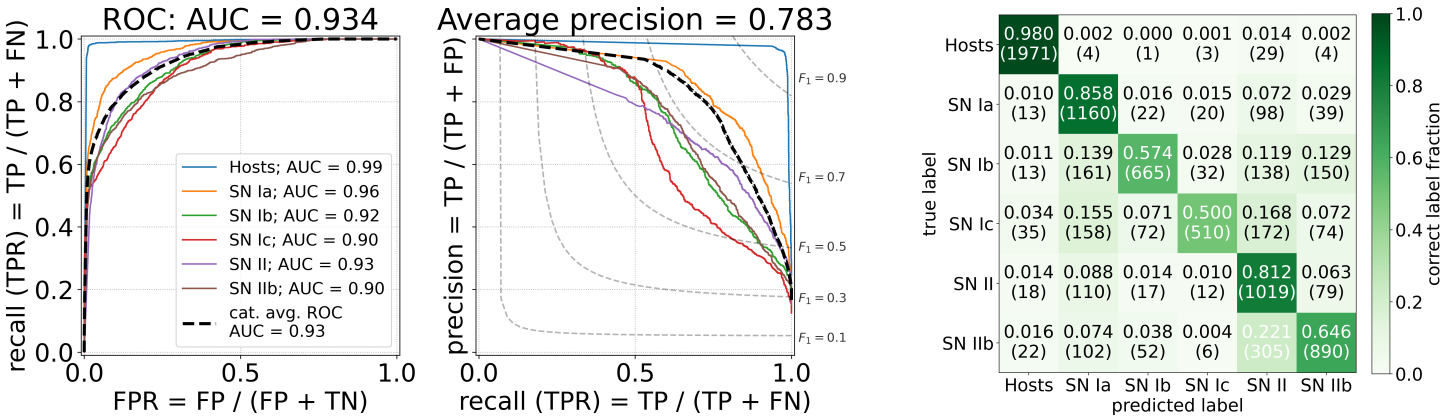


Figure: Spectral ViT V1 Classifier: 99% confidence cut (63.4% remaining)

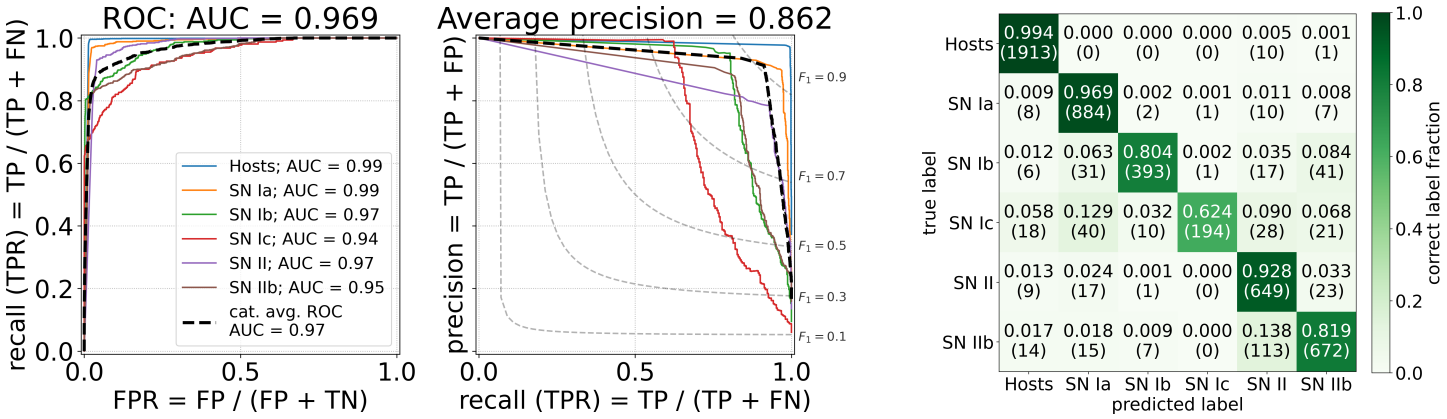


Figure: Spectral ViT V1 Classifier: 99.9999% confidence cut (40% remaining)

Spectral ViT V2

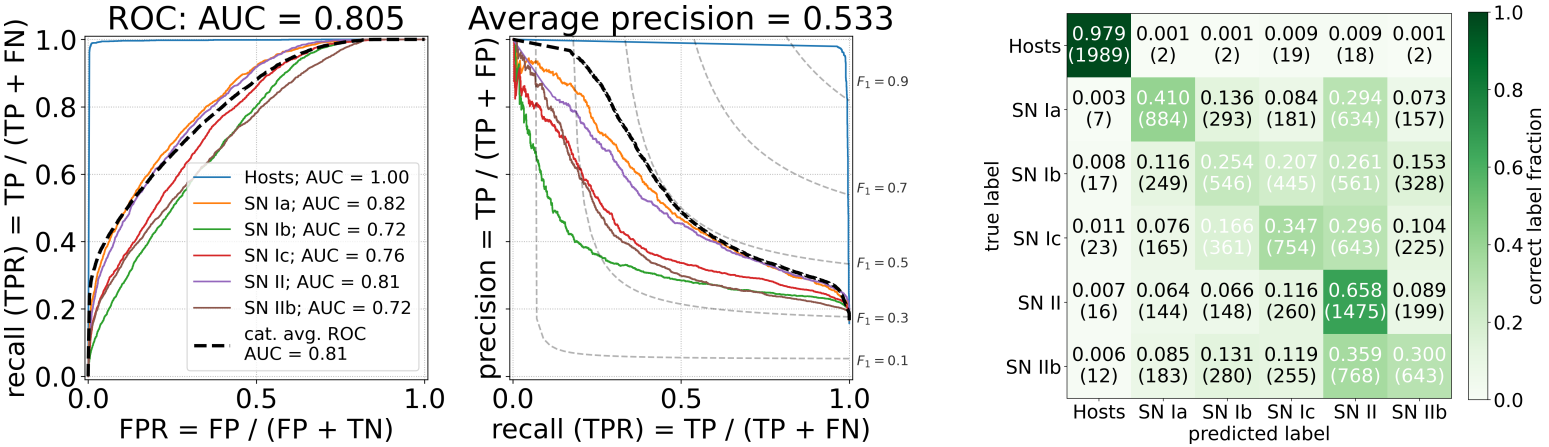


Figure: Spectral ViT V2 Classifier

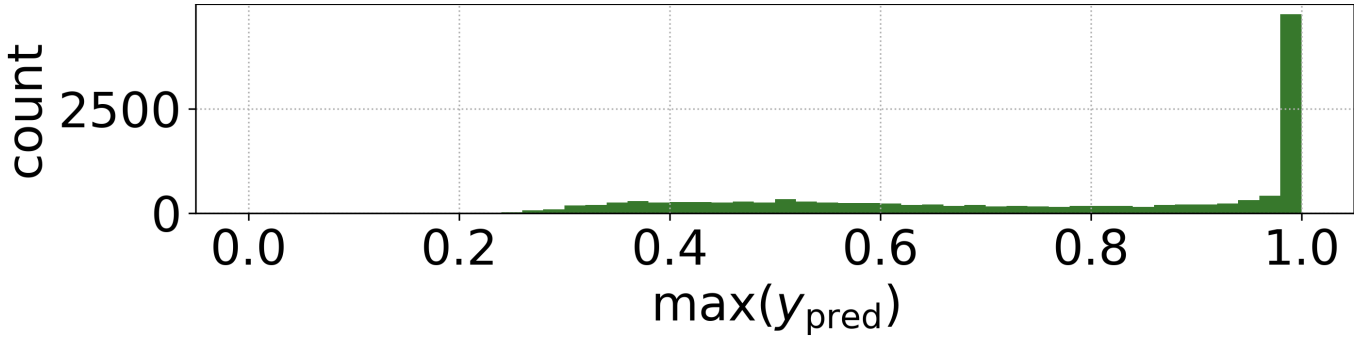


Figure: Spectral ViT V2 Maximum Output Vector Value

Spectral ViT V2 Cuts

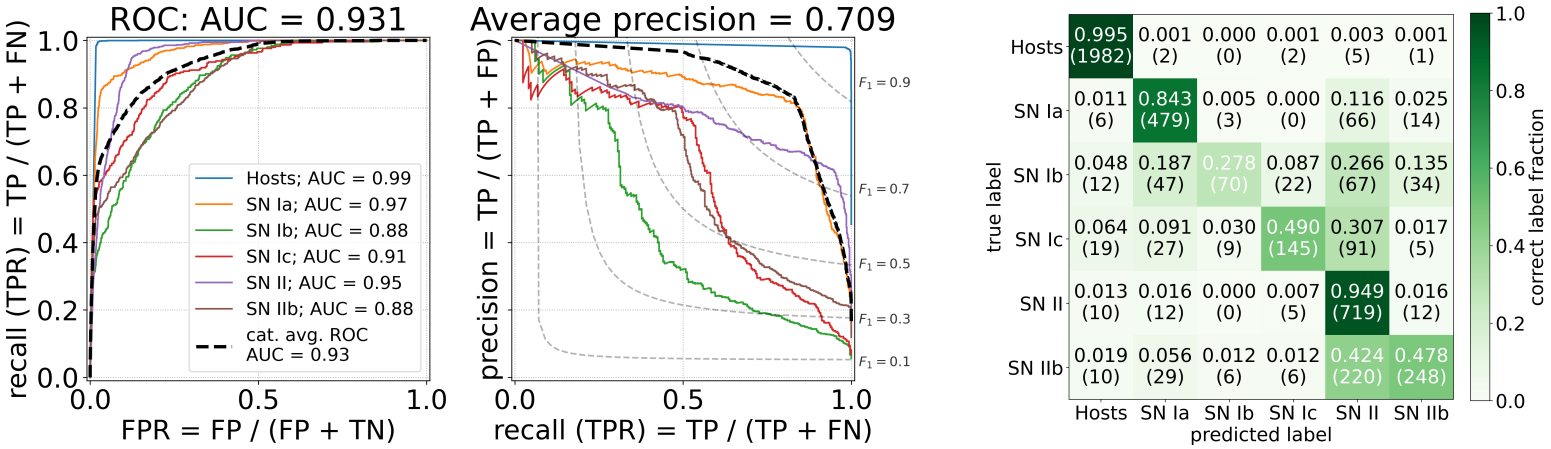


Figure: Spectral ViT V2 Classifier: 99% confidence cut (34% remaining)

Binary Classifier

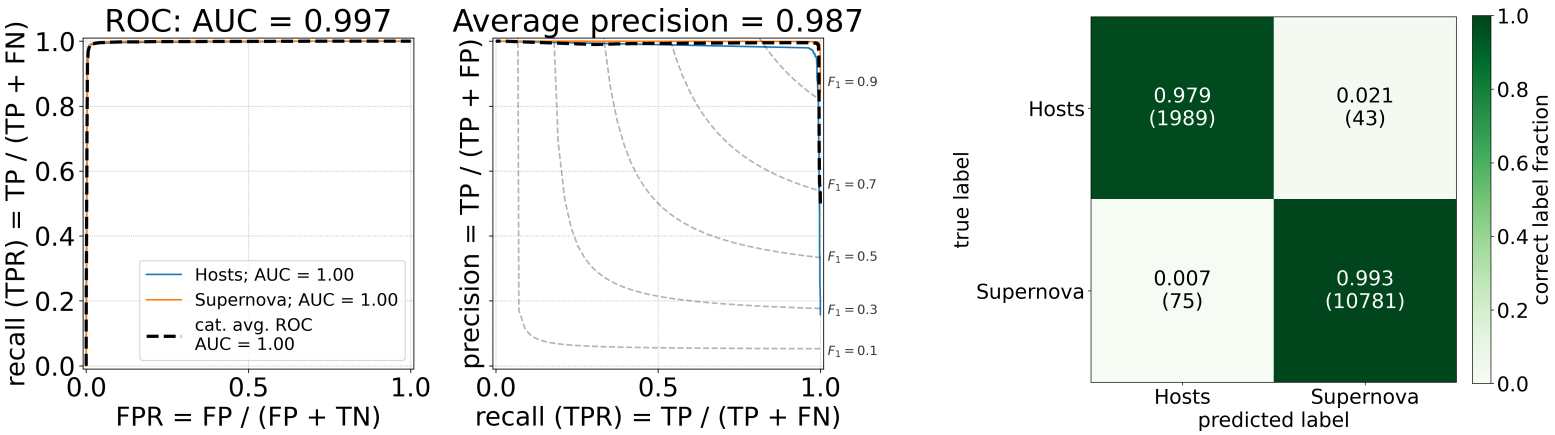


Figure: Spectral ViT V2 Binary Classifier

Other Classifiers

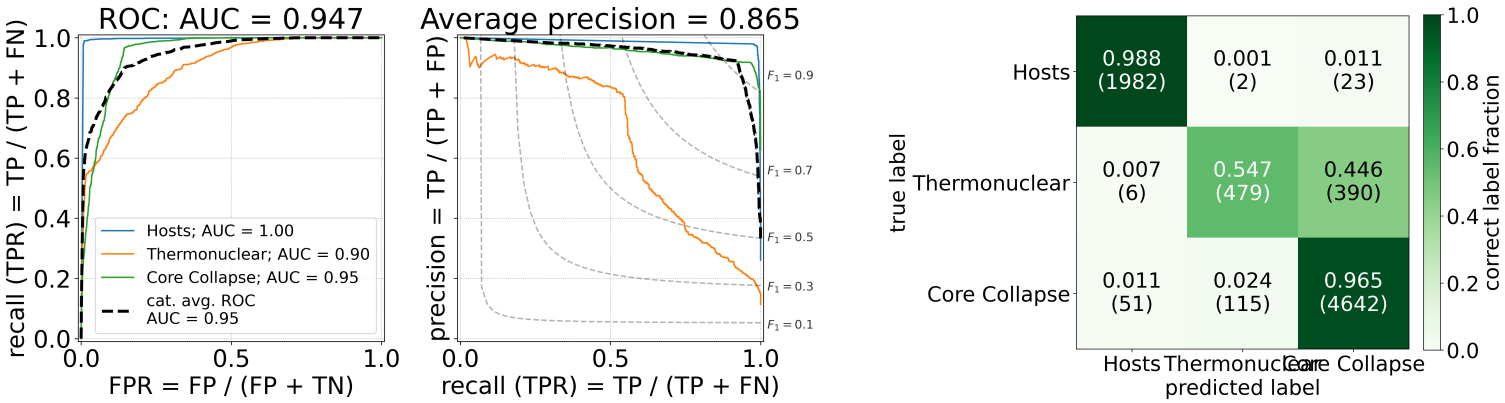


Figure: Spectral ViT V2 Progenitor Classifier: 99% confidence cut (60% remaining)

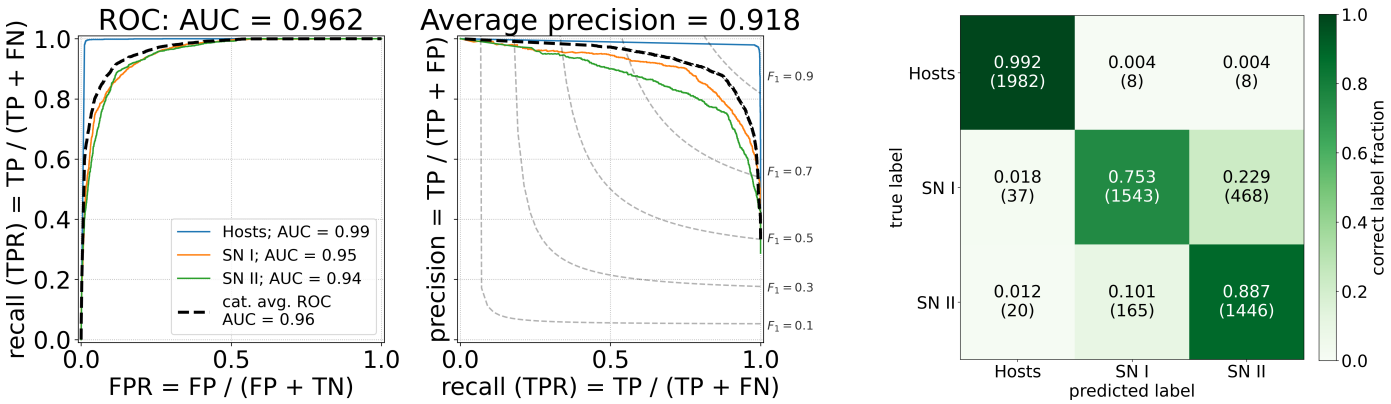


Figure: Spectral ViT V2 Type Classifier: 99% confidence cut (44% remaining)

Conclusions

Models Presented:

- V1 shows high precision AND recall rate for subtype classification
- V2 shows high precision AND recall rate for binary, type, and progenitor classification

Conclusions

Models Presented:

- V1 shows high precision AND recall rate for subtype classification
- V2 shows high precision AND recall rate for binary, type, and progenitor classification

Future Research:

- On the current model: Heat Maps, authentic DESI spectral authentication
- New Models: Train on more data, DL Training Tricks (variable learning rate, pruning, etc)
- Other Applications: Spectral ViT could be applied to other transients (Gravitational Lensing)

Acknowledgements

Professor Segev BenZvi

References I

- BenZvi, S., V. Tiwari, and A. Wasserman (Dec. 2020). *Discovering Transients (& other interesting things) in DESI Spectra*. Presentation.
- DESI Collaboration et al. (2016). *The DESI Experiment Part I: Science,Targeting, and Survey Design*. arXiv: 1611.00036 [astro-ph.IM].
- Dosovitskiy, Alexey et al. (2020). "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: *CoRR* abs/2010.11929. arXiv: 2010.11929. URL: https://arxiv.org/abs/2010.11929.
- Guy, J. et al. (Mar. 2023). "The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument". In: *The Astronomical Journal* 165.4, p. 144. DOI: 10.3847/1538-3881/acb212. URL: https://doi.org/10.3847/1538-3881/acb212.

References II

- Hahn, ChangHoon et al. (2022). *DESI Bright Galaxy Survey: Final Target Selection, Design, and Validation*. arXiv: 2208.08512 [astro-ph.CO].
- Kumar, Ajitesh (Apr. 2022). *Different types of CNN Architectures explained: Examples*. URL: https://vitalflux.com/different-types-of-cnn-architectures-explained-examples/.
- Neutelings, Izaak (Sept. 26, 2021). *Neural networks*. URL: https://tikz.net/neural_networks/.
- — (Apr. 9, 2022). *Neural networks*. URL: https://tikz.net/conv2d/.
- Sepeku, E. (May 2022). "Spectroscopic Identification of Supernova In DESI". Bachelor's Thesis. Rochester, NY: University of Rochester.
- Turatto, M. (2003). "Classification of Supernovae". In: *Supernovae and Gamma-Ray Bursters*. Ed. by K. Weiler. Vol. 598, pp. 21–36. DOI: 10.1007/3-540-45863-8_3.

References III

- Vaswani, Ashish et al. (2017). "Attention is All you Need". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- Wasserman, A. (May 2021). "Using Machine Learning to Identify Transients in the DESI Survey". Bachelor's Thesis. Rochester, NY: University of Rochester.