DESI Transient Identification Pipeline

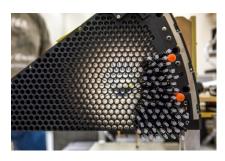
Amanda Wasserman Vashisth Tiwari Advisor: Professor Segev BenZvi

Summer Research Symposium, University of Rochester September 2020



Overview

- Dark Energy Spectroscopic Instrument (DESI) will observe over 30 million galaxy spectra.
- A small percentage of these galaxies will contain transients such as supernovae.
- Our goal is to automate the process of identifying these transients using machine learning techniques.
- Why?
 - Release the data for others to follow up
 - Ensure correct estimates of redshift
 - Identifying transient types and comparing with photometric data sets



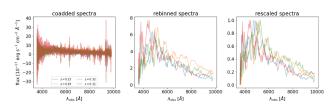
One of ten petals of the telescope's focal plane - LBNL



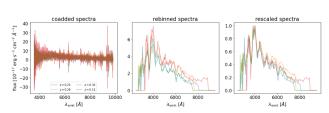


Prepossessing

- Remove NaNs/zeroes
- Weighted Rebinning
 - Removes noise
- Clipping negative flux values
 - Comes from subtracting background
- Scale between zero and one
 - Data conditioning needed in machine learning
 - Reduces training time
- De-redshift
 - Zero padding



No de-redshifting

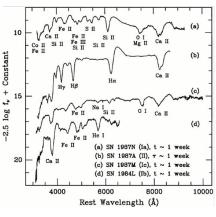


De-redshifting



CNN: Deep Neural Networks for Classification





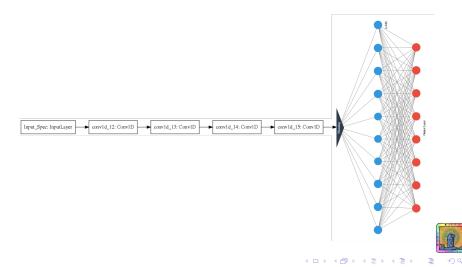
A variety of supernova spectra - Filippenko (1997)



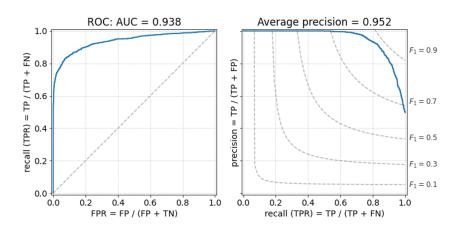
Our Deep Neural Network

Our Network's architecture:

• 5 convolutional layer • 1 dense layer (264 nodes) • 1 output



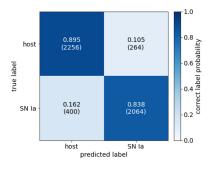
Performance: 2 classes (galaxy + SNIa)







Performance: Confusion Matrix



- •Upper left = "true negatives" (host only)
- •Lower right = "true positives" (host + SN Ia)
- •Upper right = false negatives. Confuse SN Ia for host
- •Lower left = false positives. Confuse host for SN Ia
- Ideally, the matrix is completely diagonal.
- \bullet Can improve precision, if willing to sacrifice some statistics.





Performance: Multiple Supernova Classes

 Classifying 7 different types of SN and hosts.

	Host -	0.752 (1931)	0.015 (39)	0.031 (80)	0.102 (262)	0.015 (39)	0.030 (76)	0.005 (12)	0.050 (129)
	SN Ia	0.112 (284)	0.727 (1849)	0.018 (45)	0.092 (234)	0.014 (35)	0.021 (54)	0.005 (12)	0.013 (32)
	SN Ib	0.180 (445)	0.032 (79)	0.457 (1134)	0.169 (420)	0.032 (80)	0.072 (178)	0.013 (31)	0.045 (112)
abel	SN Ib/c -	0.120 (299)	0.045 (113)	0.029 (73)	0.753 (1870)	0.014 (36)	0.006 (15)	0.002 (6)	0.029 (72)
true label	SN Ic	0.056 (115)	0.025 (50)	0.106 (216)	0.138 (280)	0.600 (1221)	0.023 (46)	0.005 (11)	0.048 (97)
;	SN IIn -	0.110 (273)	0.017 (42)	0.023 (58)	0.021 (52)	0.016 (40)	0.776 (1923)	0.008 (21)	0.027 (68)
	SN IIL/P -	0.107 (265)	0.034 (84)	0.015 (38)	0.019 (47)	0.010 (24)	0.080 (200)	0.692 (1720)	0.044 (109)
	SN IIP	0.101 (245)	0.009 (23)	0.020 (49)	0.044 (106)	0.012 (30)	0.030 (73)	0.040 (97)	0.743 (1799)
		Host	SN la	SN Ib	SN lb/c	SN Ic	SN IIn	SN IIL/P	SN IIP



predicted label

0.2

Summary / Next Steps

- Multilabel Classifier:
 - Optimize using different metrics (Precision, Accuracy, etc.)
 - Optimize the network architecture, hyper optimization, etc.
- Technical Note, DESI publication.
- Installation at the Spectro-Pipeline at the National Energy Research Scientific Computing Center (NERSC).



Acknowledgments

- This research would not be possible without the Department of Energy Office of High Energy Physics, and the University of Rochester Discover Grant.
- Much thanks for the University of Rochester DESI group and in particular our advisor, Segev BenZvi.
- This work is a continuation of the research done by Ouail Kitouni and Divyanshu Gandhi.



Any Questions?



