This Is My Title
Along With More Clarification

by

Anthony J La Barca

Submitted in Partial Fulfillment of the

Requirements for the Degree

Bachelor of Science in Physics and Astronomy

Supervised by Segev BenZvi, Ph.D.

Department of Physics and Astronomy
School of Arts and Sciences

University of Rochester

Rochester, New York

2023

To whomever

Acknowledgments

Lorem Ipsum

Abstract

Lorem Ipsum

Contents

Acknowledgments iii

Abstract iv

List of Tables vii

List of Figures vii

1 Introduction 1

1.1 Citation styles 1

1.2 Figures 2

1.3 Tables 3

2 Machine Learning Techniques 4

2.1 Convolutional Neural Networks 5

2.1.1 CNNs on Spectroscopic Data 5

2.2 Introduction of Transformers 7

2.2.1 ViT on Spectroscopic Data 8

Bibliography 10

A Detailed Diagrams of Various Neural Networks **12**

A.1 Fully Connected Neural Networks 12

A.2 Convolutional Neural Networks 12

A.3 Transformers 12

A.3.1 Multi-Head Attention 12

B Title of Appendix B **13**

List of Tables

1.1 This is a table 3

List of Figures

1.1 This is a black box 2

2.1 CNN Diagnostics 6

2.2 Max value of the output vector for the CNN. 6

2.3 CNN Diagnostics 7

1. Introduction

1.1 Citation styles

These are the different citation styles for author-year.

The standard `\cite` command produces the following output: **clarke1990rendezvous**.

The `\textcite` command produces the following output: **clarke1990rendezvous**.

The `\parencite` command produces the following output: (**clarke1990rendezvous**).

The `\footcite` command produces a footnote citation[1].

[1] **clarke1990rendezvous**.

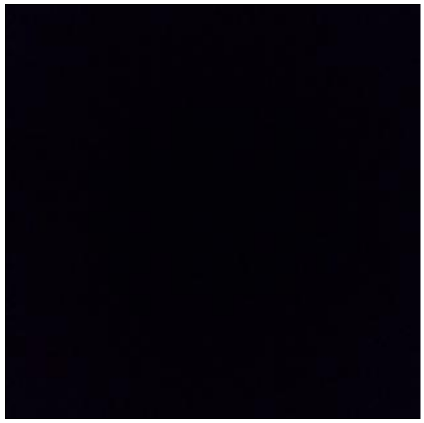


Figure 1.1: This is a black box

1.2 Figures

Always prefer to float figures to the top of the page using the [t] option in the figure environment. This is shown in Figure 1.1.

1.3 Tables

Use vertical lines sparingly in tables. They're unnecessary bloat. Write the code for tables in a separate tex file, and include it in when required. Also preferrably use sans serif font for tables (because of their information density) using `sffamily` in table definition.

Spam	Ni	Swallow	Shrubbery
A	1	2	3
E	3	4	5
C	6	9	3
M	4	1	1

Table 1.1: This is a table

2. Machine Learning Techniques

To aid in the classification of supernovae, machine learning techniques provide flexibility and scalability that have been previously unattainable by manual methods. The simplest form of machine learning algorithms are the feed-forward fully connected neural networks (FCNs). These networks apply a series of linear transformations to the input data, followed by a bias and a non-linear activation function. The output of the network is another vector tailored for whatever the purpose of the FCN is (i.e. for classification, is a vector of probabilities that the input data belongs to, see Appendix A.1).

This architecture, while simplistic, is very powerful. Cybenko (1989) showed that for any boolean function (i.e. 2 class classification), a FCN with a single hidden layer could approximate any function arbitrarily well; this is known as the universal approximation theorem. This theorem is the basis for the success of FCNs in many fields, but also the reason why they are not the best choice for all problems. The main issue with FCNs is that they are not invariant to translations, which is essential for many problems such as image recognition, time series analysis, and natural language processing. Therefore, alternative options have been proposed.

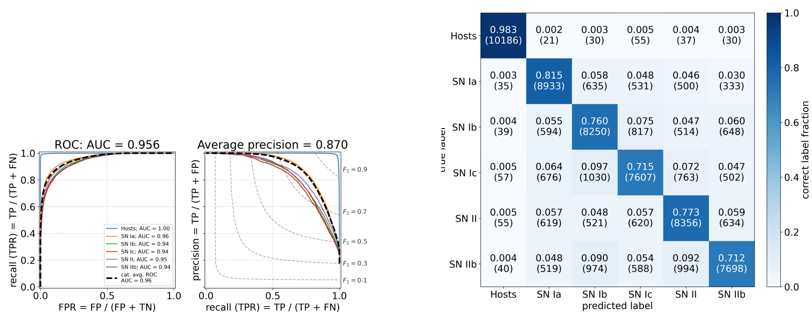
2.1 Convolutional Neural Networks

Convolutional neural networks (CNNs) were originally developed as a solution to positional differences in the input data (Fukushima 1979). This architecture was shown to be effective in image recognition tasks by LeCun, Huang **and** Bottou (2004), and was popularized after the success of AlexNet in the ImageNet challenge Krizhevsky, Sutskever **and** Hinton (2012), and have since been applied for use in many other vision tasks, as well as other fields such as NLP and audio processing. CNNs are composed of a series of convolutional layers, usually followed by a pooling layer, and then a fully connected layer (Appendix A.2). The convolutional layers are composed of a series of convolutional filters (usually referred to as feature maps) which are applied across the entire input space. The application of the same filter across the entire input space is what gives the CNN its invariance to translations.

2.1.1 CNNs on Spectroscopic Data

Due to their affinity for disregarding spacial invariance, CNNs theoretically should translate to spectral classification quite well. In fact, CNNs have been applied to spectroscopic data in the past, even on the DESI dataset. Parks **andothers** (2018) used a CNN to detect strong emission lines in DESI spectra with great success. * Talk about 1D CNN currently running, and preprocessing used *

This work provides a baseline for the use of CNNs on supernovae classification, but leaves more to be desired. Therefore, an alternate architecture was proposed by * Cite Eddies work * which augments the preprocessing of the spectra with a conversion to a 2D image. This 2D image was then fed to a more traditional vision CNN architecture (Appendix A.2). This architecture was shown to train quickly



((a)) ROC curve **((b))** Confusion Matrix

Figure 2.1: CNN Diagnostics

(less than 1 hour on a single GPU), but was not able to achieve astonishingly high accuracy. Figure 2.1 shows the ROC curve and confusion matrix for the CNN trained on synthetic spectra.

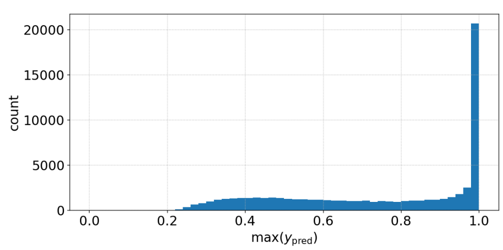
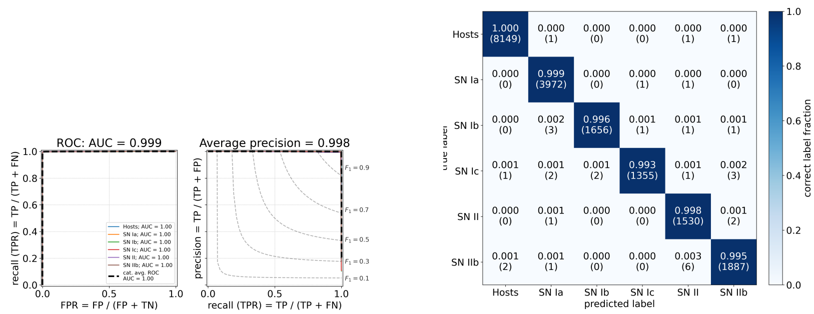


Figure 2.2: Max value of the output vector for the CNN.

The maximum value of the output vector was used to determine the predicted class, which may not be the best choice for this problem. Figure 2.2 shows the distribution of the maximum value found. As shown, there are approximately 20 thousand spectra that are classified very confidently, but there are many more that are not.



((a)) ROC curve **((b))** Confusion Matrix

Figure 2.3: CNN Diagnostics

Taking a look at the diagnostics again, but only for highly confident classifications, a much better ROC curve and confusion matrix is produced (Figure 2.3). Therefore, it is clear that the CNN, when confident in its classification, produces accurate results. This cut, however, is not ideal, removing *insert percent*% of the data.

The next step in training would be to either increase the confidence of the CNN, or to move to a different architecture, with the hopes of increasing not only the overall accuracy of the networks, but also the number of confident classifications.

2.2 Introduction of Transformers

Transformers are a relatively new architecture, first introduced in 2017 by Vaswani **andothers** (2017) for natural language processing (NLP) tasks. The original architecture consists of an encoder-decoder system. The encoder accepts a series of tokens, and via a series of self-attention layers and feed-forward layers, produces a series of vectors representing the input data. The decoder then takes the output of the encoder, and produces a series of tokens, one for each input token. The revolutionary aspect of this architecture is the presence of the multi-head attention layers. These layers (explained in more detail in Appendix A.3.1) allow the model to learn not only the importance of certain tokens, but there meaning in relation to other tokens. This allows for a single token to take on different meanings, essential for NLP tasks.

Once transformers were shown to have remarkable success in NLP tasks, they were quickly adapted to other fields, such as vision. Dosovitskiy **andothers** (2020) developed a vision transformer (ViT) architecture that differed from the original transformer by replacing the token input with a more involved preprocessing step. In short, the input image is broken into a series of patches, which are then flattened into a vector. These vectors, along with positional encodings, are then fed into the transformer architecture. For classification tasks, the first input token is replaced with a class token. After passing through the transformer, the class token is then run through a fully connected layer to produce the final probabilities. Further details on the ViT architecture can be found in Appendix A.3.

2.2.1 ViT on Spectroscopic Data

Previous implementations of transformers have been shown to have characteristics beneficial to the classification of spectral data. A transformer's ability to learn contextual information is essential in spectral classification. A broad absorption line, for example, may be indicative of a Type Ia supernova if in one part of the spectrum, but may be indicative of a Type II supernova if in another part of the spectrum. This contextual information is not easily learned by a CNN, as the convolutional layers are not able to learn the importance of certain parts of the input space. Attention can also play a role in identifying the purpose of features that are not in a standard location. For example, a non-k corrected spectrum might have a continuum pattern at different locations in the spectrum, but the overall shape would be similar. This change in sizing would be difficult to learn with fixed filters in a CNN, but would be identified based on their positional importance by a transformer. In addition to this, ViTs have been shown to outperform CNNs on vision tasks, which shows they are capable of focusing on learned features.

* Include caveat about the fact that ViTs take longer to train than CNNs *

Bibliography

Cybenko, G. (1989). “Approximation by superpositions of a sigmoidal function”. **in***Mathematics of Control, Signals and Systems*: 2.4, **pages** 303–314. DOI: 10.1007/BF02551274. URL: https://doi.org/10.1007/BF02551274.

Dosovitskiy, Alexey **andothers** (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.

Fukushima, Kunihiko (1979). “Self-Organization of a Neural Network Which Gives Position-Invariant Response”. **in***Proceedings of the 6th International Joint Conference on Artificial Intelligence - Volume 1*: IJCAI’79. Tokyo, Japan: Morgan Kaufmann Publishers Inc., **pages** 291–293. ISBN: 0934613478. DOI: 10.5555/1624861.1624928.

Krizhevsky, Alex, Ilya Sutskever **and** Geoffrey E Hinton (2012). “ImageNet Classification with Deep Convolutional Neural Networks”. **in***Advances in Neural Information Processing Systems*: **byeditor**F. Pereira **andothers**. **volume** 25. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

LeCun, Y., Fu Jie Huang **and** L. Bottou (2004). “Learning methods for generic object recognition with invariance to pose and lighting”. **in***Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* **volume** 2, II–104 Vol.2. DOI: 10.1109/CVPR.2004.1315150.

Parks, David **andothers** (**january** 2018). “Deep learning of quasar spectra to discover and characterize damped Ly systems”. **in***Monthly Notices of the Royal Astronomical Society*: 476.1, **pages** 1151–1168. ISSN: 0035-8711. DOI: 10.1093/mnras/sty196. eprint: https://academic.oup.com/mnras/article-pdf/476/1/1151/24261051/sty196.pdf. URL: https://doi.org/10.1093/mnras/sty196.

Vaswani, Ashish **andothers** (2017). "Attention is All you Need". **in***Advances in Neural Information Processing Systems*: **byeditor**I. Guyon **andothers**. **volume** 30. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

A. Detailed Diagrams of Various Neural Networks

A.1 Fully Connected Neural Networks

A.2 Convolutional Neural Networks

A.3 Transformers

A.3.1 Multi-Head Attention

B. Title of Appendix B