Computer Vision and Application to the Classification of the Hadronic Decay Modes of the Tau Lepton with the ATLAS detector

Torrey Saxton

August 15, 2018

Abstract

In this report, we examine the use of deep learning neural networks for the classification of hadronic decays. Through a series of two-dimensional, convolutional neural network, we use Monte-Carlo simulated tau lepton decays to train a classification algorithm. We classify the decays into five main modes: letting "p" stand for a charged pion, and "n" stand for a neutral pion, the five modes are as follows: 1p0n, 1p1n, 1pXn, 3p0n, and 3pXn.

1 Introduction

On July 4th, 2012 the discovery of the Higgs Boson from the combined experimental efforts of ATLAS and CMS was announced to the world. While the discovery of this particle resulted in an experimental confirmation of the Brout-Englert-Higgs mechanism [1, 2], there are still many questions to be answered about the Higgs. It has been established that the Higgs boson couples to the tau lepton to a precision of approximately 30% [3]. With the large datasets collected by ATLAS, we will attempt to study this coupling in greater detail. The research presented here focuses on detecting tau lepton decays in the ATLAS detector, focusing on collisions producing visible tau lepton decays in the barrel of the ATLAS detecter ($|\eta| < 1.1$) [4]. The life time of a tau lepton is quite short, and as such these particles are not directly measured by the ATLAS detector; rather their decay products are. The tau lepton has 5 main hadronic decay modes: 1p0n, 1p1n, 1pXn, 3p0n, and 3pXn, which are the focus of the study presented here. Through the use of a series of two-dimensional convolutional neural network, and Monte-Carlo simulated tau lepton decays, we have created a machine learning algorithm for use in detecting tau leptons in the ATLAS detector.

Over the course of the Summer Student Programme at CERN, my main role in this project was investigating the architecture of the neural network. This involved programming in "sanity checks" to allow us to have a visual understanding of what the network is producing after it has been trained, evaluating instances of "false-positives" - when the network has improperly identified a decay - and learning how to remove unintentional bias from the network, and overall striving to improve the accuracy of the network as a

whole. This has involved extensive use of the Keras library [5], and extensive utilization of hardware provide from TechLab [6].

2 The ATLAS Detector and Network Design

The ATLAS detector at CERN is a general purpose detector whose sole purpose is the study of fundamental particle physics. Through the use of electromagnetic calorimetry, full azimuthal angle coverage, and high detector granularity, the ATLAS detector seeks to record data for the vast number of events that occur during collisions. This requires the development of reconstruction algorithms, and working backwards through jet products by physicists to determine that an event involved a tau lepton. This is done here by implementing a series of two dimensional convolutional neural networks, to create an algorithm for classifying the hadronic decays of tau leptons within the ATLAS detector.

Deep learning algorithms today employ convolutional neural networks for efficient image recognition. A qualitative approach to a convolutional neural network, is that each neuron reads in a "patch" of an image. By increasing or decreasing the patch-size, the neuron is trained on small or large details. On average, by training the neurons on smaller details, one can then take advantage of a concept known as "max pooling", in which one now groups together neurons, and creates a new "patch" of data, which contains the maximum value contained with in each neuron in the group. This allows the network to be trained on larger details as well, which helps improve the image recognition capabilities of the network.

In the algorithm we employ, a different image is provided to the network corresponding to the different layers of the calorimeter. A convolutional network is ran separately over each of these images, and combined with max pooling. We then combine all of the two-dimensional convolutions into a single one-dimensional neural network. Furthermore, the track data is also provided to the neural network, allowing us to see how where the track algorithm correlation with energy deposit in the calorimeter influences the network's decision. The track layer is already a one-dimensional layer, which then is concatenated with the calorimeter layers of the network, providing the final structure of the network trained as seen in Figure 1.

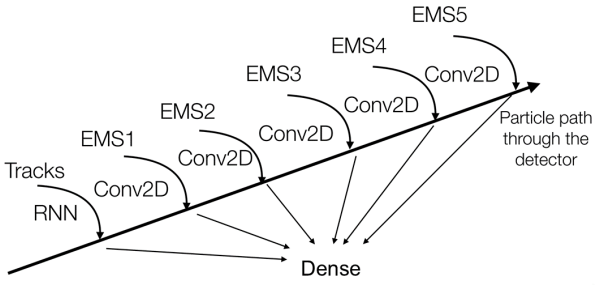
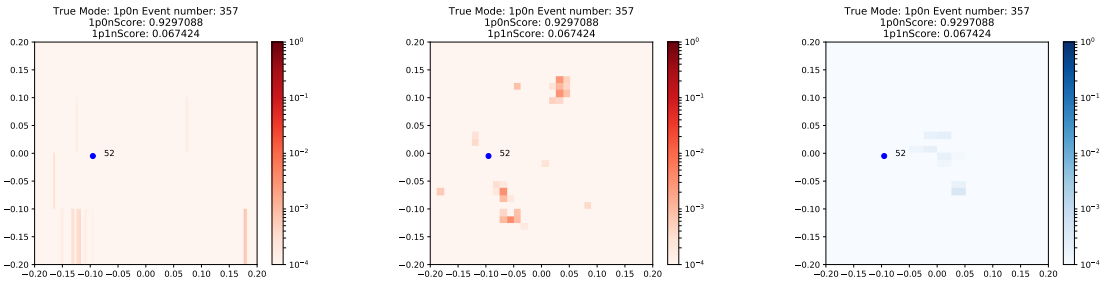


Figure 1: Schematic outline of the neural network employed in developing the classification algorithm

This approach allows us to see how the network correlates energy deposits between different layers of the calorimeter when making classification decisions. This also allows us to observe the networks capability for handling cases where the energy deposit between layers are widely spread due to the trajectory of the particle, but again only focusing on the barrel of the detector. The distribution of data provided to the network is discussed in Table 1.

Sample	Training	Validation	Testing
1p0n	321672	40209	40209
1p1n	745511	93189	93189
1pXn	318920	39865	39866
3p0n	246244	30780	70781
3pXn	132379	16547	16548

Table 1: The data provided to the network is divided into three segments. 80% of the data is used to train the algorithm, 10% of the data is used in validation to prevent overtraining, and 10% is used in testing the accuracy of the final algorithm.



(a) First layer of the calorimeter

(b) Second layer of the calorimeter

(c) Third layer of the calorimeter

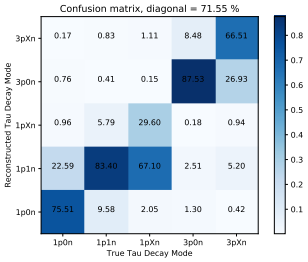
Figure 2: Shown here are energy deposits in the calorimter. This is an instance in which the energy deposits between layers have little correlation between them, but the neural network successfully identified the decay as a 1p0n decay. The x axis represents η and the y axis represents ϕ. The color represents the amount of energy deposited in the cell, according to the bar scale.

As shown in Figure 2, the algorithm is quite capable of handling cases where the energy deposits between layers are widely spread, which is an ideal feature of an algorithm being used to classify decay modes from images. As these decay modes involve charged particles, it is unlikely that the particle will travel straight through the detector without its trajectory being deflected due to the magnetic field. Using information such as the η and ϕ of the cells, we looked for location bias within the network, as to whether certain locations were more likely to result in a successful classification.

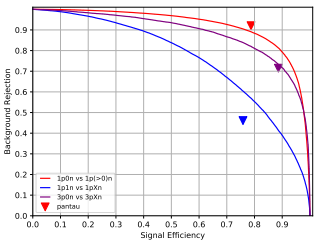
Our overall goal is to maximize the off-diagonal of the confusion matrix shown below in Figure 3; in the matrix, each column has a value of 100 when summed, as the values within the column represent probabilities. Values along the off-diagonal of the matrix represent the efficiency of the network correctly classifying a given decay mode. Values outside of the off-diagonal represent how often the network incorrectly identifies a given decay mode as another decay mode.

3 Investigation of the Network Performance

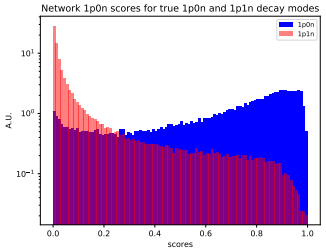
Through the use of the confusion matrices, ROC curves, and individual decay mode scores, we have investigated the performance of the network, as seen in Figure 3.



(a) Confusion matrix for a trained network.



(b) ROC curve for a trained network. The triangles represent the accuracy of algorithm Pantau.

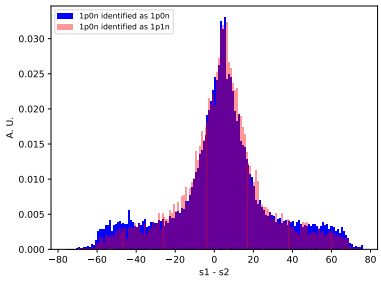


(c) 1p0n scores calculated by the classification algorithm for the true 1p0n and 1p1n decay modes.

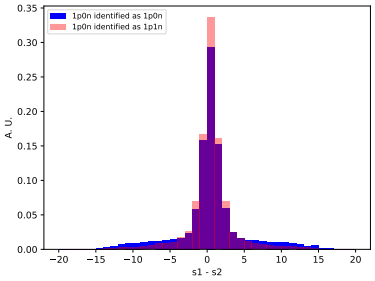
Figure 3: The figures shown above provide insight into the performance of the classification algorithm. Figure (a) shows the confusion matrix, which provides insight into how the algorithm is classifying or misclassifying particular decay modes. Figure (b) shows the ROC curve, where it is comparing efficiency to the Pantau algorithm, the current baseline for ATLAS. Finally, Figure (c) shows the algorithm scores for 1p0n for true 1p0n and 1p1n decay modes.

The confusion matrix provides us with a broad overview of the network performance as a whole. The ROC curve provides us with the accuracy of the classification algorithm for specific types of decay modes, and allows us to Pantau, the tau lepton classification algorithm currently used in ATLAS. Finally, the plot of the algorithm scores provides us insight into how the algorithm occasionally fails to correctly classify the decay mode, and provides an avenue for further investigation.

To ensure that the network was unbiased, my work focused on performing multiple different analyses of the network. One such analysis was the plotting of the difference in the η and ϕ components between calorimeter layers, and observing how these differences might possibly influence the classification algorithm's accuracy. Shown in Figure 4 below are the differences in the ϕ and η components between the second and first layer of the calorimeter.



(a) Difference in the η component between the first and second layers of the calorimeter.

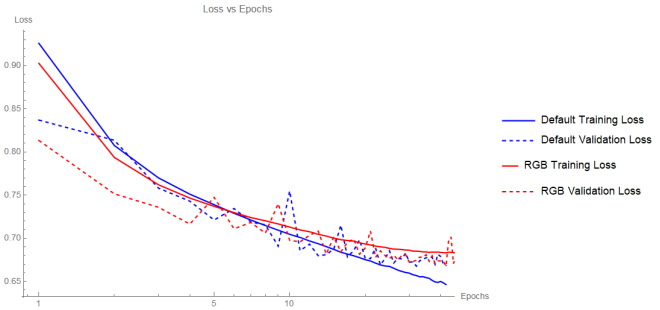


(b) Difference in the ϕ component between the first and second layers of the calorimeter.

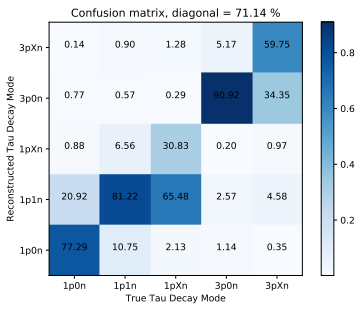
Figure 4: Shown here is a histogram of the difference in the location of the maximal energy deposit in the η and ϕ components of the first and second layers of the calorimeter. We look at the cases where the network correctly classified the event as a 1p0n as "true positive" and where it incorrectly identified the event as a 1p1n as "false positive".

As the shape for our "false positive" and "true positive" classifications are quite similar, we concluded that differences in the ϕ and η components were not influencing that accuracy of the classification algorithm.

Further investigation of this classification problem has led us to try "RGB" and "RGB-like" neural networks. These networks give us the capability of adding an ersatz depth to the network, thus allowing us to more directly teach the network the importance of the correlation between layers. The culmination of my project the construction of the "RGB-like" network. In the RGB-like network a convolutional network is ran and pooled over each layer of the calorimeter. These networks are then concatenated and further convolution is ran, and the network is then converted to a 1 dimensional dense network.



(a) The loss function of the network versus the number of epochs ran.



(b) The confusion matrix for the "RGB-Like" network.

Figure 5: As seen in Figure (a), the result of the "RGB-like" network leveling out at a higher value for the loss function than the previous architecture directly results in the off diagonal of the confusion matrix being less than it was before, as shown in Figure (b).

This network, while showing promise for being able to learn quickly, currently appears to level out much sooner than previous network architectures resulting in a lower level of accuracy than previously as shown in Figure 5.

4 Conclusion

Through the use of neural networks and the development of classification algorithms, a better resolutions of tau lepton decay products within the ATLAS detector barrel can be achieved. This will ultimately lead to a better precision on the coupling of the Higgs boson to the tau lepton, and allow further study of this particle. The convolutional neural network described is quite promising in this endeavor. It is outperforming the current ATLAS algorithm for tau leptons decaying in the barrel. Furthermore, it can be extend to reject QCD jets and estimate the 4-vector of the decay products. It will be crucial to understand how well it performs on data from the ATLAS detector itself though, and if the simulation reproduces the behavior on data accurately. This work is a first step towards a complete redesign of the reconstruction of hardonic tau leptons at ATLAS.

References

[1] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. `https://arxiv.org/pdf/1207.7214.pdf`, 2012.

[2] The CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. `https://arxiv.org/pdf/1207.7235.pdf`, 2012.

[3] The ATLAS Collaboration and the CMS Collaboration. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at *sqrts*=7 and 8 TeV. `https://arxiv.org/pdf/1207.7235.pdf`, 2012.

[4] The ATLAS Collaboration. The ATLAS experiment at the CERN large hadron collider. `http://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003/pdf`, 2016.

[5] François Chollet et al. Keras. `https://keras.io`, 2015.

[6] Aritz Brosa Iartza. `http://techlab.web.cern.ch/`.

Acknowledgements

I would like to acknowledge the University of Michigan CERN Research Experiences for Undergraduates program funded by National Science Foundation Grant Number 1659393. I would also like to thank my advisers Drs Pier-Olivier DeViveiros and Quentin Buat for their patience and guidance through this project over the course of the summer.