Analyzing Various Machine Learning Models for Use in the Large Hadron Collider ATLAS Detector

Antonio Lobaccaro[1], Dylan Rankin[2], Max Cohen[2]

University of Pennsylvania (CAS '27)[1], University of Pennsylvania Department of Physics & Astronomy[2]

PURM Penn Undergraduate Research Mentoring Program

Background

- Large Hadron Collider: accelerates protons close to the speed of light and collides them
- ATLAS Detector: detects 1 billion particle collisions per second
  - Way too much data to store, would require a quantity of CDs that would stack up to the moon and back twice every year
  - Most collisions produce jets of low energy hadrons with a low likelihood of leading to the discovery of new particles
  - Composed of 2 "triggers" that filter out data using machine learning models that only save collisions that appear anomalous to the low-energy collisions the model is trained on
    - Level 1 Trigger (L1): filters 1B to 100,000 collisions per second
    - High-Level Trigger (HLT): filters 100,000 collisions to 3000 per second

The Dataset

- The dataset was composed of data from the LHC, as well as monte carlo simulations of various collisions.
  - The hope is that model classifies types of collisions that are rarer and of more high energy as anomalous, while classifying the dijets, which make up the majority of collisions, as background.
- Each datafile is composed of over tens of thousands collisions, which are made up of 60 numbers representing the momentum (pt) and angles (eta and phi) of 10 jets, 3 electrons, 3 muons, 3 photons, and the missing transverse momentum (MET).

Methodology

- The ATLAS Trigger System is exploring machine learning methods of anomaly detection programmed onto FPGAs to filter out data to save
- This project analyzes the performance of various autoencoders, a type of neural network that compresses data into fewer dimensions and then attempts to reconstruct it.
- For each collision an anomaly scores is calculated by taking the Mean Squared Error between the inputs and outputs.
- The algorithm classifies a collision as anomalous if the anomaly score is greater than a threshold determined by the True Positive and False Positive rates of the ROC Curve

The Model

- The model began as a generic autoencoder of dimensions 60 → 32 → 8 → 2 → 8 → 32 with pt normalized to 1 and trained over the EB_data and dijet files.
- Different iterations were tried, involving changes to normalization of pt, size of the network, zero padding in the loss function, eb weights, slicing or zeroing out parts of the dataset based on the pt values, batch normalization, regularization, changing the bases of the pt, eta, phi values, and Variational Autoencoders.

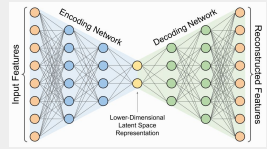


Figure 1: Autoencoder Diagram

Results

Figure 2: Base Trainer HLT ROC Curve

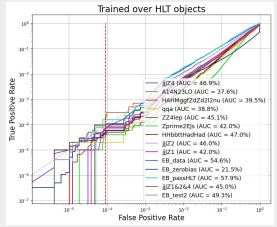


Figure 5: Improved Trainer HLT ROC Curve

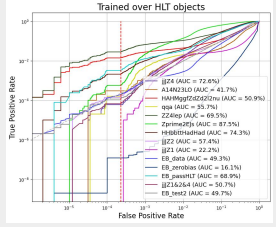


Figure 3: Base Trainer L1 ROC Curve

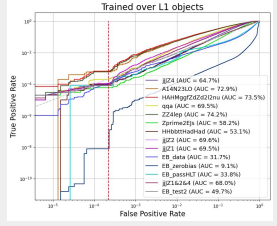


Figure 6: Improved Trainer L1 ROC Curve

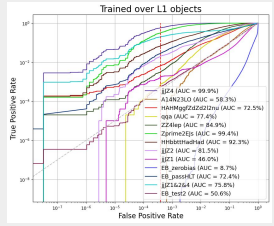


Figure 9: Improved Trainer HLT Efficiencies

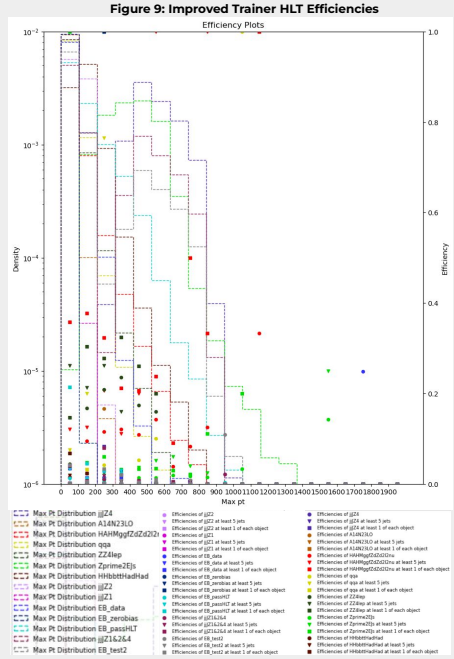


Figure 4: Sample Base Trainer Max pt vs. AD Score 2D Histogram

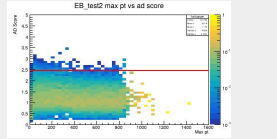


Figure 7: Sample Improved Trainer Max pt vs. AD Score 2D Histogram

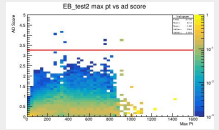


Figure 4: Sample Base Trainer Object Mult vs. AD Score 2D Histogram

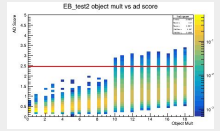


Figure 8: Sample Improved Trainer Object Mult vs. AD Score 2D Histogram

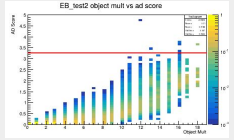
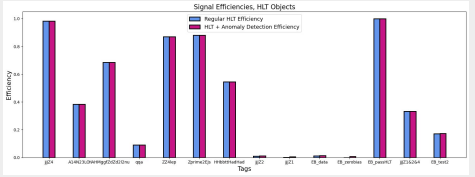


Figure 10: Difference in Signal Acceptance in HLT Trigger



Conclusion

- The base autoencoders seemed to be picking up more on patterns in the data as opposed to actual physics. For example, 2D histograms comparing the anomaly score of a signal with the number of objects or jets in each event showed much more correlation than the histograms comparing max pt to the anomaly scores.
- Removing zero padding led in the loss function reduced some of the bias with regards to the number of objects. (See Figures 4 & 8)
- Low energy jets, especially in the trigger, have a lot of noise and uncertainty, because there tends to be more pile up (particles and jets originating from a different proton-proton collision) at low energies. To reduce the effect of this noise, all jets below 50 pt GeV were zeroed out.

Supplemental Graphs of More Variations

Future Directions

- Plot more histograms and find patterns in data to find better training data combinations and transformations
- Try more variations to the traditional autoencoder and consider other types of autoencoders
- Expand upon efficiency plots by slicing the datasets in more ways and comparing models not just signals within one model

Acknowledgements

I would like to thank Professor Dylan Rankin and Max Cohen for their guidance and support on this project.