A Machine Learning Approach to Local Energy Calibration in the ATLAS Calorimeters

Madison Howard*
Texas Tech University
(Dated: August 22, 2021)

I. Introduction

The intention of an energy calibration scheme is to provide a calorimeter signal for physics object reconstruction in ATLAS which is calibrated outside an assumption about the type of object. This is of particular importance for final-state objects with a significant hadronic signal content, such as jets.

The energy calibration method used in run 1 of The ATLAS detector was called Local Hadronic Cell Weighting (LCW) Calibration [1]. The LCW calibration aims at the cluster-by-cluster reconstruction of the calorimeter signal on the appropriate (electromagnetic or hadronic) energy scale. This method used bins and a lookup table to calibrate the energy. In this, the cluster energy resolution is expected to improve by using other information in addition to the cluster signal in the calibration. The calorimeter signal inefficiencies that calibration includes non-compensating calorimeter response, signal losses due to clustering, and signal losses due to energy lost in inactive material.

The LCW method worked well for high energy and is the current standard. However, LCW calibration had poor response for energies less than 100 GeV.

A Machine Learning (ML) approach to energy calibration could help to alleviate the issues that arise with LCW calibration. The hope of this approach is that it will be better at calibrating the energy for all energies. Also, due to the amount of data that ATLAS has, LCW calibration is computationally more expensive than a ML based model.

In tandem with such a scheme, this paper outlines a project where we calibrate energy based on Geant4 simulation data. The data from the full simulations are stored in ROOT tuples, in a tree format. Each row vector in the dataset contains the signal, moments and information related to the composition for one topo-cluster. Since this data is produced through simulation, we have access to information, such as the topo-cluster true energy, which is not available in an experimental situation. This makes such a simulated format valuable for developing robust ML enabled energy calibration schemes.

The remaining part of the paper is organized as follows. We first introduce how we devise ML frameworks that are capable of multi-dimensional regression. Then we describe the various cuts of data and learning experiments we conducted along with results and discussion. We summarize the findings and remarks on future works in the last section.

II. Methodology

Machine learning, specifically Deep Learning Neural Networks (DNN), have been used as a new energy calibration method opposed to LCW calibration for topo-clusters initiated by charged and neutral pions.

A concern we have is if the DNN is overtraining or not training enough, and finding that optimal solution is the motivation to look at the different kinds of loss functions.

The goal of this project is to improve on ML calibration below 10 GeV and apply that same calibration method to jets. However, energy levels below 200 or 300 MeV do not provide meaningful inputs (moments) for ML calibration.

The data input to our ML framework represents the response of the ATLAS calorimeters to single neutral (π^0) and charged pions ($\pi^{\pm}$) from full simulations. The pions have energies between 200

MeV and 2 TeV (logarithmically sampled), and pseudorapidity within $-5 < \eta_\pi < 5$ (full acceptance in ATLAS). Each event contains exactly one pion. There is no pile-up added. The pions emerge from the nominal vertex (0,0,0) in the detector frame of reference. The magnetic fields (both solenoid and toroid) are included in the simulations (toroidal fields not really relevant for charged pions). The calorimeter response is represented by the topo-clusters (can be more than one per pion).

The calibration sequence used for LCW calibration after classification is depicted in **Fig. 1**. We are also implementing a similar sequence for ML calibration approach. For clusters with $0 < P^{EM}_{clus} < 1$ both the electromagnetic and the hadronic calibration procedures are applied and the effective weights used to determine the final energy are the result of a weighted linear interpolation between the two calibrations. The effect of each step of the calibration procedure on the total cluster energy is measured by the ratio of the cluster energy to the energy before its application. These ratios are stored in the data, see Table 1.

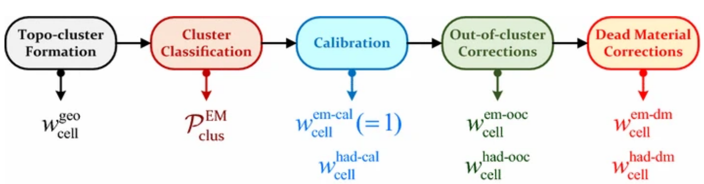

Figure 1: Calibration Sequence Flow Chart [1]

While there are many quantities produced in the simulation, the most physically significant quantity for the project is the cluster energy (target). All physical quantities used for training the network other than the target are passed as normalized quantities. We conducted a sensitivity analysis to understand which quantities of interest boosted the training regime the most. Vector sets of{energy density ($\boldsymbol{\rho\rho}\rho_{cell}$), PTD($p_T D$), distance of a cluster from calorimeter front face-- measured along the principal shower axis-- (λ_{clus}), longitudinal and lateral energy dispersion ($m^2_{lat, long}$), pseudorapidity (η_π), cluster signal significance (ς^{EM}_{clus}), energy (E^{ML}_{clus}), and second time (σ^2_t)}, {ρ_{cell}, $p_T D$, λ_{clus}, m^2_{lat}, m^2_{long}, η_π, ς^{EM}_{clus}, and E^{ML}_{clus}}, and {$p_T D$, m^2_{lat}, m^2_{long}, η_π, ς^{EM}_{clus}, and E^{ML}_{clus}} were tested with {ρ_{cell}, $p_T D$, λ_{clus}, m^2_{lat}, m^2_{long}, η_π, ς^{EM}_{clus}, E^{ML}_{clus}, and σ^2_t} having the best calibration results.

In addition to this, we have also explored various cuts on these normalized datasets. The following three cuts were the most effective:

1. Significance, ς^{EM}_{clus}, is the ratio of signal data over the background. A previous minimum significance cut of $2\sigma$was tested before, but the most effective cut for S was found to be at 0σ, just excluding negative significances.
2. A cut on $|\eta| < 0.8$ assisted with calibration at lower energies.
3. Another way to eliminate unmeaningful moments of clusters is to make the cut-off of the sum of a cluster energy for a single event number to be individually greater than 20% of the sum. For example, if a cluster has three individual moments with energy of 500, 1, and 0.5 GeV, the first 500 GeV moment will be stored and the other two will be cut out. This suppresses the shaping significantly and allows the mean to be much closer to 1 throughout the range. A major benefit to these refinement cuts is lower runtime for ML; the runtime was decreased from about 7 hours to 2 hours while keeping the most important

clusters.

Through systematic testing, we have observed the most optimal Neural Network Architecture for such a task to have 2 hidden layers each with 1024 nodes using a rectified linear activation function (ReLU). For loss metrics, we have found Mean Absolute Error (MAE) to be the most effective. And for the optimization algorithm, we used the Adaptive Moment Estimation optimizer (ADAM).

The outputs of this DNN are the corresponding weights files which gives us a checkpoint for the best signal calibration and a new data column of energy in the cluster (E_{clus}) for each signal.

III. Results

To evaluate our ML calibration scheme, we compare the ratio of the true energy and the predicted energy, $E^{dep}_{clus}/E^{ML}_{clus}$ for both our scheme and LCW as shown in **Fig 2**. The goal for calibration is that the ratio be as close to 1 as possible---meaning the calibrated energy (the output of the ML network) is as close to the true--cluster energy as possible. The ML calibration approach performs better than LCW calibration for all energies. At lower energies, the ML approach had an average error of 15% while LCW had 70%.

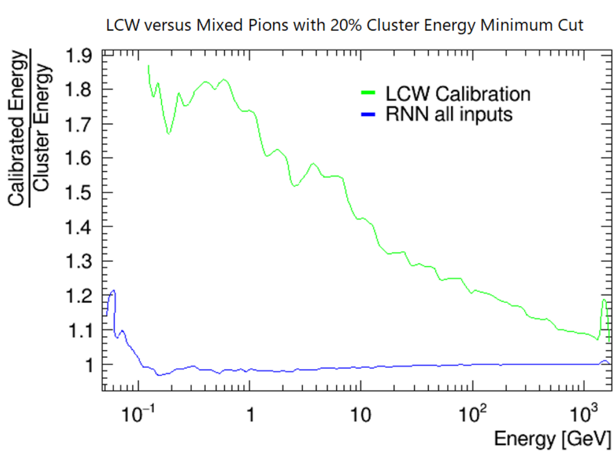

Figure 2: LCW vs DNN Calibration Comparison

In the 2D Histogram (**Fig. 3).** the color corresponds to the frequency of events, the vertical axis is the ratio E_{clus}/E^{ML}_{clus} and the horizontal axis is the, E^{ML}_{clus}. With the single particle data, when our ML calibration is applied, the average fit (red line) approximates 1. This result shows that the ML approach is successful in calibrating the energy to the desired true deposited energy.

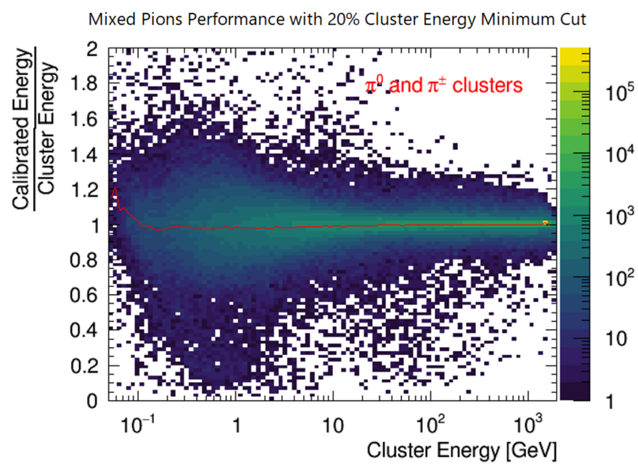

Figure 3: DNN Mixed Pions Calibration

A. Cluster Calibration on Jets

As previously mentioned, the expectation was to be able to use single particle pion data to train and clusters in jet data to test, similarly to how the LCW calibration had. Unfortunately, this was not the result because the weights produced for the single particle pion data did not apply well to the jets.

The same cut conditions mentioned in Section II were applied to jet training and jet testing data. However, when single particle training calibration was applied to jet data the histograms showed asymmetric error bars meaning there is significant systematic error. In this case, the same 20% energy minimum cut was applied to the jet data and the error bars became more symmetrical, but, because this was a significant energy cut to the jet data, it did not apply very well with the calibration. Further investigation will need to be conducted on this issue.

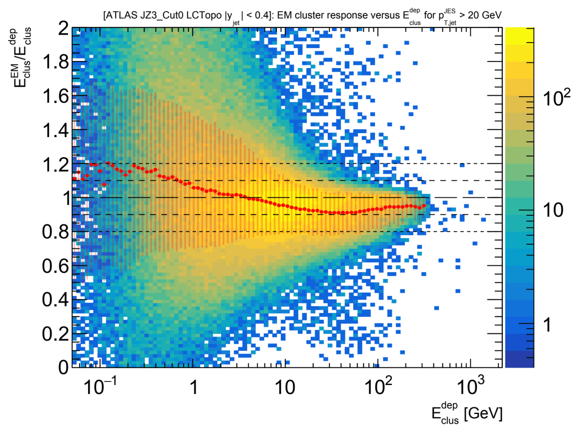

Figure 4: Jet Data DNN Calibration

IV. Conclusion and Further Work

The goal of this project was to create a Deep Neural Network to help with energy calibration to help remedy the weak points of the LCW calibration technique. This ML approach shows a vast improvement for energy calibration, especially for energies below 10 GeV. The energy level is also within 10% for energies greater than 1 GeV. Another advantage of this calibration method is the low computational resources. This project concludes that using machine learning is a viable and promising way to calibrate the energy levels of the topo-clusters

However, our attempt at applying the training of the topo-cluster data calibration to the jet data to test the calibration, as LCW has applied their calibration, has proven to be less effective and requires further investigation. A possible reason why this may be is due to the fact that single particle datasets and jet datasets had different columns of information. More work will need to be done on jet substructure and the meaning behind the asymmetry of the error bars in calculating the fit of the clusters and jets.

Altogether, ML for energy calibration shows promise for its accuracy, speed and applicability from as low as 200 MeV and up in both jets and single particle data samples. We believe that with more data and with more advanced networks this problem can be resolved.

References:

[1] Aad, G., Abbott, B., Abdallah, J. *et al.* Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C* 77, 490 (2017).

[2] N. Young. August 26, 2020. *Evaluation of Calorimeter Signal Calibration Using Machine Learning Approaches.* University of Arizona. August 19, 2021.