Machine Learning for Pions

IML Forum, Particle Flow

Maximilian Swiatlowski
TRIUMF, obo of the ATLAS Collaboration
Jets at the LHC

When **quarks** or **gluons** are produced during a collision…

They “shower” into more gluons and quarks…

Which hadronize into stable (or unstable particles)

**This is a jet!**

Need to measure not just a single particle, but hundreds per event (mostly **pions**, 2/3 charged, 1/3 neutral)

Image credit: B. Nachman
Measuring jets is a huge challenge: resolution is clearly worse compared to other types of final states!

But the high branching ratio of the Higgs and possible new physics to jets still makes jets **critical** for searches and measurements

**And new ideas can have a big impact!**
Why is Jet Resolution So Bad?

• Three main effects:

  1. Fluctuations due to pileup (won’t talk much about this today)

  2. Fluctuations in fragmentation/hadronization/etc. (see backup)

  3. Fluctuations in individual particle showers (focus of this talk)
Calorimetry at the LHC

ATLAS’s and CMS’s calorimeters built to measure particles within jets: record energy and location of interactions/depositions.

However, different particles interact with the calorimeter differently: $\pi^0$ decay to $\gamma\gamma$, while $\pi^\pm$ are stable and interact directly.
Our calorimeters are calibrated to the EM scale: EM showers (from $e$, $\gamma$, $\pi^0$) are measured ‘correctly’

Resolution is good: all showers are ‘similar’

Hadrons can also interact with nuclei, produce low energy neutrons, etc.: undetectable energy

Every shower is unique: huge resolution penalty from variations
One way to improve: use inner-detector tracks to measure hadrons whenever possible: significantly improved resolution in many cases!
Huge number of studies in the literature!

Can we also improve jet inputs with ML?
• Existing expert-tuned algorithms in ATLAS for PFlow and calorimeter calibration follow a step-by-step approach

• Follow a similar strategy for ML: break the problem into pieces and solve independently

• Today: focus on hadronic/EM shower separation, and full detector energy regression
Point Cloud Deep Learning Methods for Pion Reconstruction in the ATLAS Experiment

ATL-PHYS-PUB-2022-040
• Study single $\pi^\pm, \pi^0$ datasets from simulation

• Solve single particles before going onto multi-particle “confusion” problem

• Primary calorimeter objects are “topological clusters”

• Formed from $\sim$188k calorimeter cells in the detector

• Represent topoclusters as “point clouds” formed from calorimeter cells

• Previously studied CNNs and image-recognition techniques: point clouds more adaptable to calorimeter geometry
• GNN is a natural architecture for a point cloud dataset

• Simultaneously train GNN to classify pion class ($\pi^0$: EM shower; $\pi^\pm$: hadronic shower) and calibrate energies

• Also consider simple “Deep Sets” point cloud model: no edge information
• All ML methods significantly outperform baseline $P_{EM}$ method!

• GNN and DeepSets also robust out to forward $\eta$ regions: no special tuning for geometry required

• ML can be used to identify shower types in the calorimeter
Can also compare pure calorimeter energy calibration with GNN compared to baseline EM and LCW techniques.

- GNN significantly improves calorimeter-only energy resolution!
- DeepSets improves over very basic EM calibration, worse than expert-tuned LCW
- Model was not thoroughly optimized: could potentially be improved
• Calibrating calorimeter alone has intrinsic limitations due to hadronic shower fluctuations

• Tracker is natural counterpart: better resolution at low energy

• Point cloud networks can be easily extended to include tracking information

• Track considered just another “node” in the point cloud; include binary mask to label input as track or calo cell
• GNN and DeepSets continue to be studied

• Very simple DNN (simple MLP) also implemented to compare performance

• Also explore “transformer” algorithm

• Several iterations of “Message passaging layer” (displayed on right) helps model learn relationships between nodes via attention mechanism; nodes ultimately summed in pooling layer

• Similar methods developed by Nilotpal et al (previous talk)
• All ML architectures deliver significantly improved energy reconstruction over full energy range
Here, compare resolution of ML techniques to Calo-only and tracker-only results

ML results all significantly improve over baselines

- Transformer and GNN significantly improve over DNN: graph/message layers matter!

- ML can successfully combine information from different detectors
Conclusions
Conclusions

Better resolution can enable better physics at the LHC

We are tackling low-level inputs to jets: can we use our exquisite detector granularity to help jets catch up to other final states?

Our step-by-step approach has shown improvements over baseline in several key metrics already

Results on more complete “PFlow” to come soon!
Backup
Calorimeter-Only Pion Reconstruction With Deep Learning

ATLAS-PHYS-PUB-2020-018
Traditionally, ATLAS has used the “Local Cell Weighting” technique for calorimeter-only jet reconstruction.

1. Classify topo-clusters as “EM-like” or “Hadronic-like”
2. Apply a calibration appropriate for EM or Hadronic pions

Features used are quite simple: depth and density. Can we do more with the high granularity calorimeter information?
Average Pions

Here, compare $\pi^0$ and $\pi^\pm$ in the first layer of the calorimeter:

- Treat energy in each ‘cell’ of topocluster as pixel intensity
- Use simulated samples of pure $\pi$’s

See expected differences: $\pi^\pm$ are ‘broader’
Different calorimeter layers have different granularity

Here, show $\pi^+$ in first three calorimeter layers

Three additional layers also available

Can consider these as ‘RGB channels’ in NNs
Differences Between Pions

By just subtracting $\pi^+$ from $\pi^0$ images, can already visualize differences between EM and hadronic showers

Can deep learning classifiers use this information?
Classification
Architectures

Three general classes of NN architecture studied

DNN: Large, deep networks with cells as direct inputs

CNN: use convolutions to extract useful features from different portions of the image

DenseNet: Industry-designed, sophisticated CNN with information propagation
Classification ROC

Compare $\pi^+$ efficiency vs $\pi^0$ rejection ($1/$efficiency) for different algorithm

Compared to baseline, see huge performance improvement: factor of $12x$!
Classifier Correlation: $\pi^0$

Here, calculate the correlation coefficient between each pixel and the classifier.

Can visualize (very roughly) what the CNN is learning.

Can see the physics we expected from the images!
Energy Regressions
Use simulated data
To test energy calibrations:
Know ‘truth’ from simulation

Compare reconstructed
energy to true energy,
as a function of
True energy

Ideally: close to 1, and
narrow distributions
Correcting $\pi^+$

At ‘EM’ scale, can see energy reconstruction issues for hadronic particles, like $\pi^+$

Energy is *missed* due to non-measured nuclear interactions

Feature-based method corrects for this: see ‘correct’ energy scale for wide range of true energy, but over-compensation at low energy
Train regressions on pure $\pi^0$ and $\pi^+$ samples

Target: ‘true’ energy from simulation

Use similar NN architectures as for the classifier, but also include ‘raw’ energy
Comparing Results

**ATLAS** Simulation Preliminary
Single $\pi^+$, All Clusters

Scale goal: get close to 1
DNN outperforms default, and ‘feature-based’ correction!

Resolution goal: get close to 0
DNN again outperforms other methods
Thank you!
Also show the CNN (best architecture) performance in bins of energy

And compare to LC in bins of energy

Results are encouraging: good performance over all energies!

Factor of 100x improvement for 10-50 GeV π^0 rejection!
Combining in a Mixed Sample

- So far, evaluated only in pure samples
- Can also mix charged and neutral pions in 2:1 ratio to mimic jets
- Apply classifier at ~95% $\pi^+$ efficiency, and then apply chosen regression
- Good performance! Better median and resolution than defaults
Comparing Results

Scale goal: get close to 1

DNN outperforms default, and ‘feature-based’ correction!

Resolution goal: get close to 0

DNN again outperforms other methods
Individual particle showers fluctuate, but so do the jets themselves: the number of hadrons can vary, the types of hadrons, and so on…

Lots of techniques, both with and without machine learning, can correct for some of these fluctuations and improve resolution.
GNN vs $\eta$

ATLAS Simulation Preliminary
GNN Classification of $\pi^\pm$ vs. $\pi^0$
Single Pion MC, Topo-clusters, $|\eta|<3$

- $|\eta| \leq 0.8$
- $0.8 < |\eta| \leq 1.37$
- $1.37 < |\eta| \leq 2.5$
- $|\eta| > 2.5$