Top Jet Tagging using Graph Neural Networks and Z' Mass Regression

Group Members

Dipanwita Mondal, Sashibhusan Sahoo, Harshita Mishra, Khushi Singla, Susobhan Chattopadhyay, Hitesh Kadam, Pawan Kumar, Suprajaa D

July 11, 2026

Project Instructor : Dr. Sanmay Ganguly

Physics Motivation

Why are we looking for new heavy particles?

Although the Standard Model successfully describes elementary particles, it cannot explain

- Dark Matter
- Neutrino Masses
- Matter–Antimatter Asymmetry
- Gravity

Many Beyond Standard Model (BSM) theories predict a new heavy neutral gauge boson

$$Z'$$

which can be produced at the LHC through

$$pp \to Z' \to t\bar{t}.$$

Physics Goal :Search for a heavy resonance (Z')

Boosted Top Quarks

For a heavy resonance,

$$M_{Z'} \gg 2m_t,$$

The dominant decay for top quarks is

$$t \to Wb$$

followed by

$$W \to q\bar{q}$$

Normally,

$$b,\ q,\ \bar{q}$$

are reconstructed as three separate jets. However, at high momentum these decay products become highly collimated and merge into a single

Boosted Fat Jet

Challenge: Boosted top jets look very similar to energetic QCD jets, making their identification a challenging classification problem.

Why Machine Learning?

Each boosted jet contains detailed internal information carried by Particle Flow (PF) candidates. Each PF candidate stores

$$(px,\ py,\ pz,\ E,\ d_0,\ dz,\ \mathrm{Charge},\ \mathrm{PID},\ \mathrm{PUPPI},\ldots)$$

Instead of using handcrafted observables, modern approaches learn directly from all particles inside the jet.

$$\boxed{\text{PF Candidates}} \rightarrow \boxed{\text{GNN(ML)}} \rightarrow \boxed{\text{Top Tagging}}$$

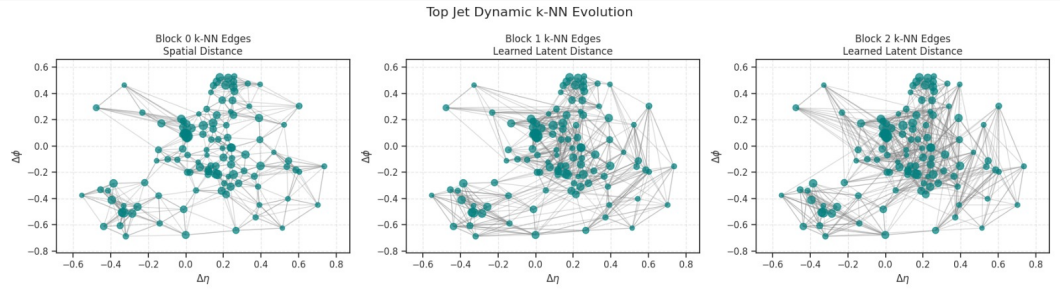
Steps

- Make the appropriate graph / hyper-graph data using given processed .h5 data.
- Train GNN / Hyper-GNN to perform top-jet tagging vs QCD-jet rejection.
- Regress Z' mass from a mixed test sample.
- Plot a ROC curve for classification comparing the architectures.
- Ploting a scatter plot for the mass regression and response curve (Predicted - True) / True

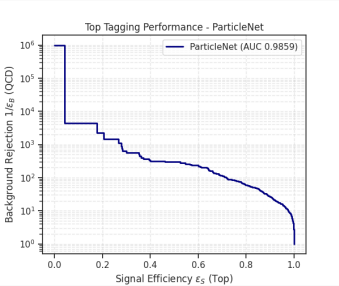
ParticleNet: Jet as a Graph

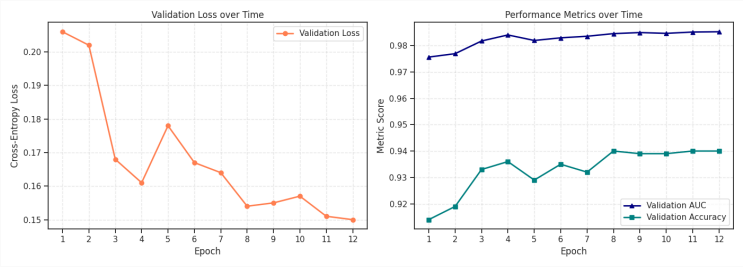
- A jet is an **unordered set of particle-flow candidates (PFCands)**, not an image or sequence.
- Each particle is represented as a **graph node** with features:
  - p_T
  - $\nabla\eta$ and $\nabla\phi$
  - $\log(p_t)$ and $\log(E)$
  - log(rel-pt) and log(rel-E)
  - dR
  - Impact Parameters (d0, dz)
- Nodes are connected to their k **nearest neighbours (k-NN)** in the learned feature space.
- The graph is **dynamic**: neighbour connections are recomputed after every EdgeConv layer, allowing the network to learn increasingly meaningful particle relationships.

ParticlenNet: Jet as a Graph

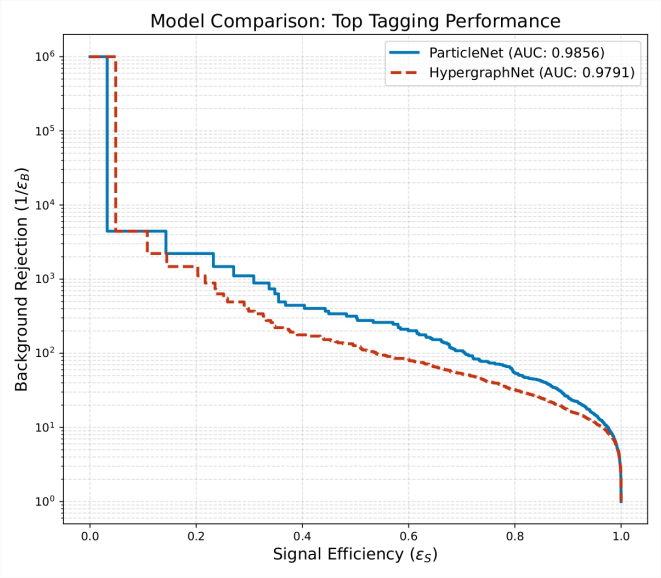


Results

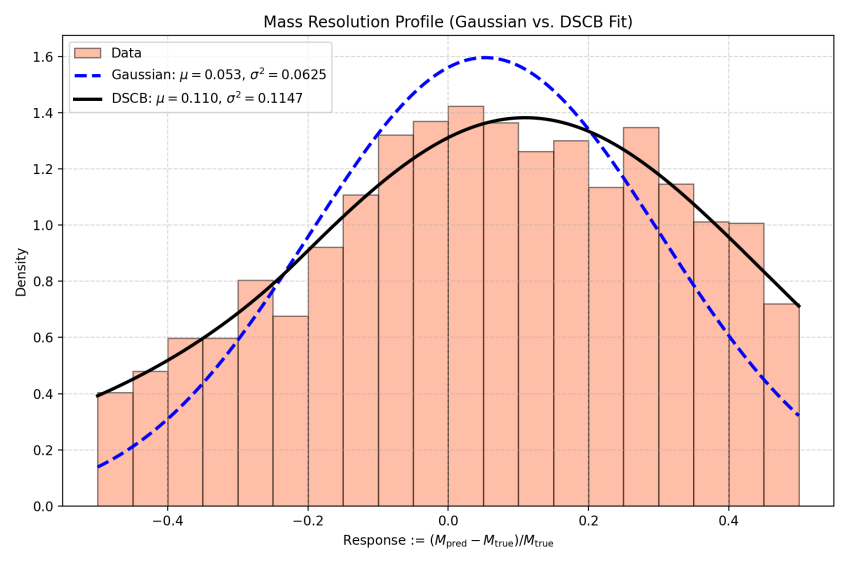

Top Tagging Performance - ParticleNet
ParticleNet (AUC 0.9859)
Background Rejection $1/\varepsilon_B$ (QCD)
Signal Efficiency ε_S (Top)


Validation Loss over Time
Validation Loss
Cross-Entropy Loss
Epoch
Performance Metrics over Time
Metric Score
Validation AUC
Validation Accuracy
Epoch

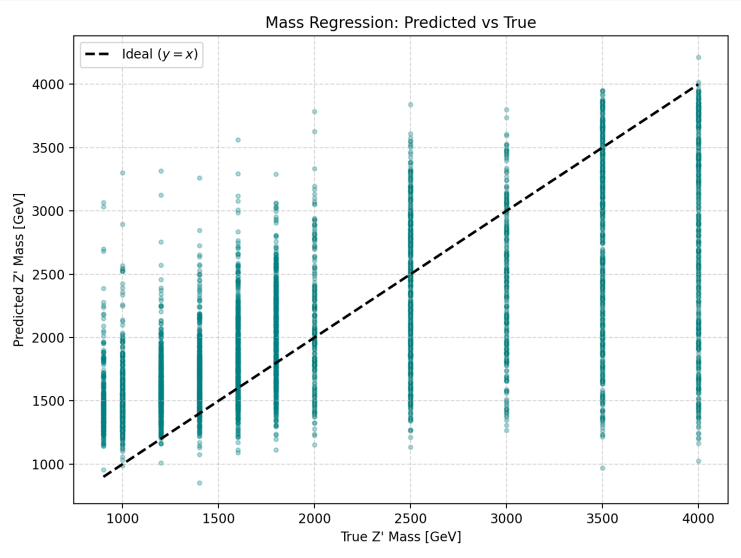
Results


Model Comparison: Top Tagging Performance
ParticleNet (AUC: 0.9856)
HypergraphNet (AUC: 0.9791)
Background Rejection ($1/\varepsilon_B$)
Signal Efficiency (ε_S)
10^0
10^1
10^2
10^3
10^4
10^5
10^6
0.0
0.2
0.4
0.6
0.8
1.0

Results


Mass Resolution Profile (Gaussian vs. DSCB Fit)
Data
Gaussian: $\mu = 0.053$, $\sigma^2 = 0.0625$
DSCB: $\mu = 0.110$, $\sigma^2 = 0.1147$
Density
Response := $(M_{\text{pred}} - M_{\text{true}})/M_{\text{true}}$

Results


Mass Regression: Predicted vs True
Ideal ($y = x$)
Predicted Z' Mass [GeV]
True Z' Mass [GeV]
1000
1500
2000
2500
3000
3500
4000

Thank You!

Questions and Discussion