Pileup Mitigation at the HL-LHC Using Attention Neural Networks

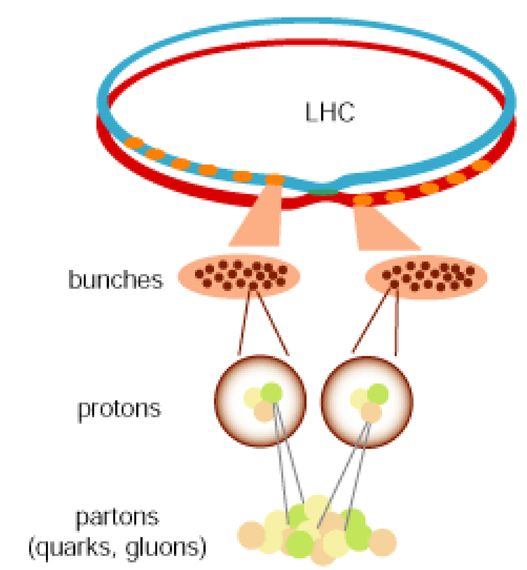
Luke Vaughan
Oklahoma State University

Abstract

Pileup mitigation at the LHC is crucial for experimental High Energy Physics analyses. Rejecting pileup background gives more sensitivity to signal and helps reduce the combinatorics needed to reconstruct invariant mass of underlying particles. As the LHC is upgraded, more pileup will be introduced requiring better pileup identification algorithms.

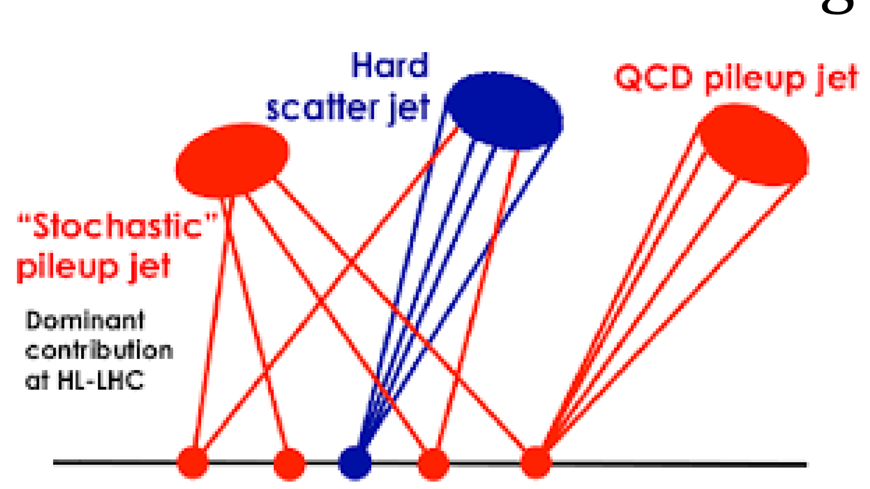
Introduction

At the Large Hadron Collider, LHC, protons are accelerated to 0.999 999 990 c at energies of 6.5 TeV (γ≅6,930). However, colliding microscopic protons isn't as easy as hitting the broad side of a barn, 1 b = 10^{-28} m^2.



Therefore, to increase the chances of a collision, protons are collided in bunches of about 100 billion protons every 25 nanoseconds.

On average, only 1 out the billion of protons interacts via deep inelastic scattering and produces collimated particles called a "hard scatter" jet. Many of the other protons interact via soft inelastic scattering and produces stochastic particles which are clustered as "pileup" jets.



The LHC will be upgraded in 2030 to the High-Luminosity LHC. The mean number of interactions per bunch crossing will increase from 60 to 200, bringing much more pileup.

Jet Vertex Tagger

Pileup jet identification is performed by the ATLAS experiment using the Jet Vertex Tagger, JVT [1]. Two high-level variables, corrJVF and RpT, are constructed using track, vertex, and jet information. These two variables are used as input to a k-Nearest Neighbors classifier. Using a Pythia8 [2] simulation of 10k $pp \to t\bar{t}$ events:

$$\text{corrJVF} = \frac{\sum_k p_T^{\text{trk}_k}(\text{PV}_0)}{\sum_l p_T^{\text{trk}_l}(\text{PV}_0) + \frac{\sum_{n\geq 1}\sum_l p_T^{\text{trk}_l}(\text{PV}_n)}{(k \cdot n_{\text{trk}}^{\text{PU}})}} \qquad R_{pT} = \frac{\sum_k p_T^{\text{trk}_k}(\text{PV}_0)}{p_T^{jet}}$$

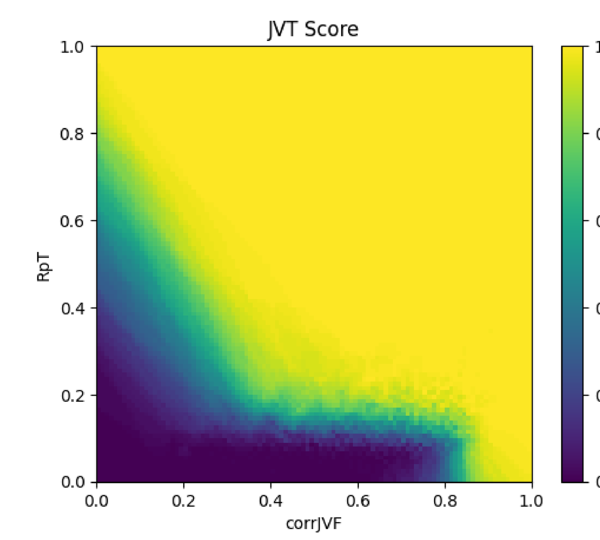
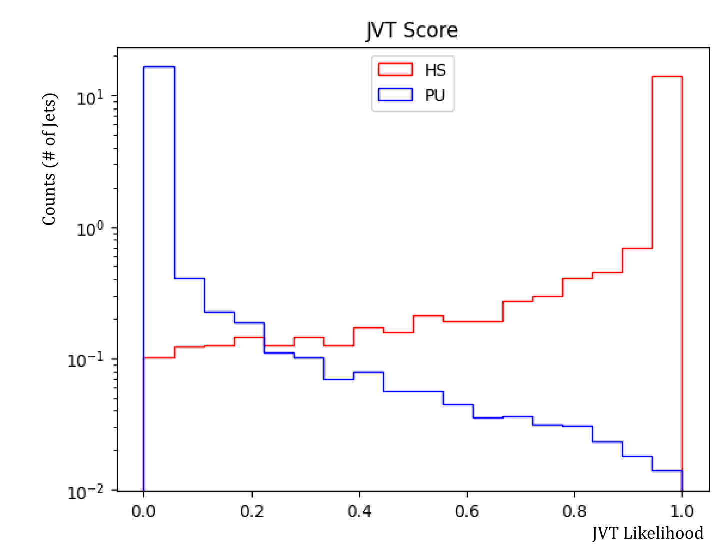


Fig (a) The landscape of JVT model in the corrJVF-RpT plane.



Fig(b) JVT shows good separation in likelihood between HS and PU jets.

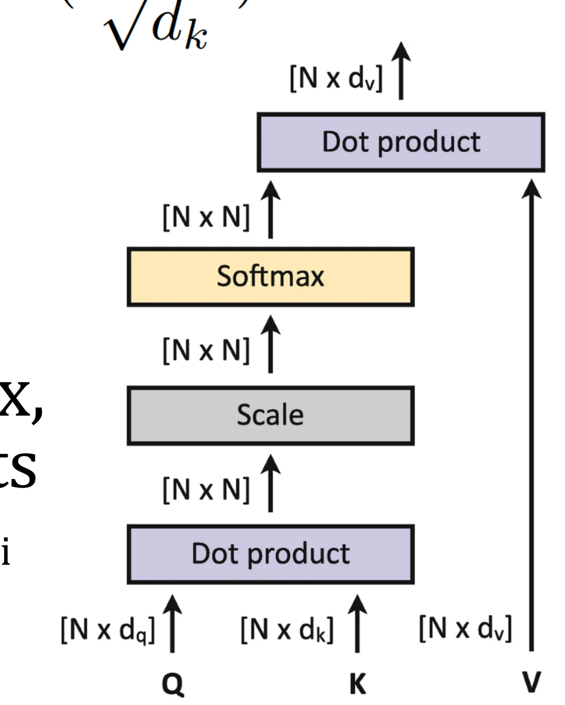
The Attention Mechanism

The attention mechanism[3] uses a scaled dot product operation on a Query and Key-Value pair to produce a context matrix or "Attention Weights" for an input set, $\mathbb{J} \in \mathbb{R}^{N,F}$.

*Self

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

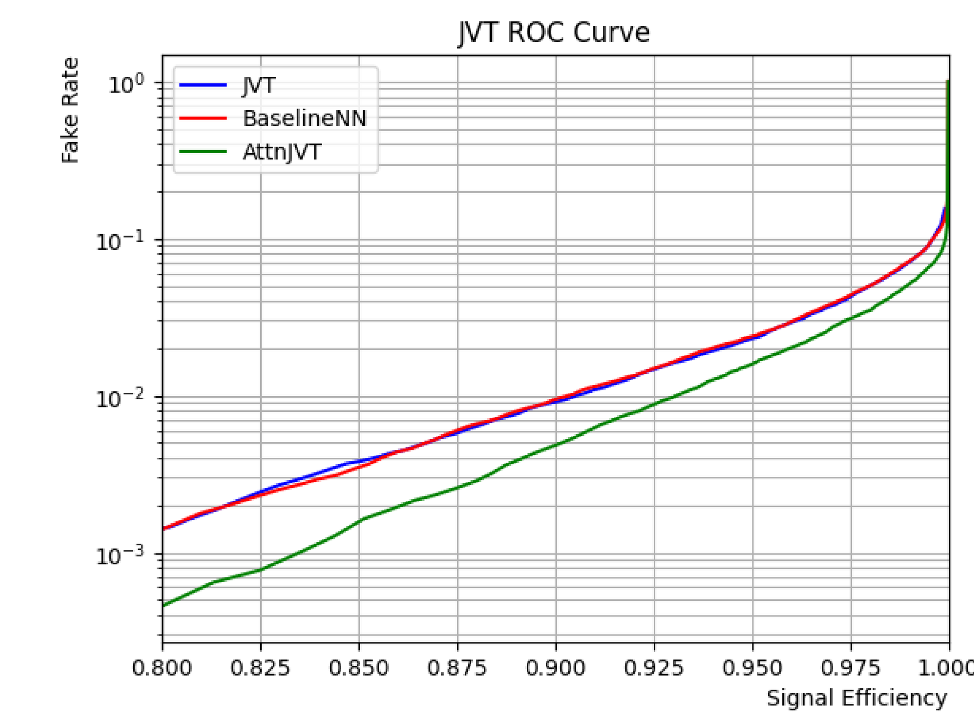
The set $\mathbb{J} = \{J_1, J_2, ..., J_N\}$ represents n jets in an event. Each jet is described by its 4-vector, p_T, η, ϕ, m. Attention produces a (N,N) mixing-matrix, M. Each element, M_{ij}, represents the dot product of the vectors J_i and J_j in the latent space. This allows jets to have a "context window" of an entire event.



Fig(c) Tensor Operations for a single Attention Layer

Results

Using Python and PyTorch, the kNN JVT model, a baseline deep NN model, and the AttnJVT model were trained using Pythia8 $pp \to t\bar{t}$ events:



Fig(d) Benchmarked Receiver Operating Characteristic Curves

The AttnJVT model is able to achieve a lower false positive rate than the kNN JVT and baseline NN. This improvement is an effect of the attention layer allowing jets to update their vector representation in the context of an event.

Conclusion

The increase in pileup for the HL-LHC requires optimized pileup mitigation algorithms. Attention provides a computationally efficient, highly parallelizable algorithm for a model to distinguish between a set of correlated HS jets from a jet that was formed from stochastic PU clustering.

References

[1] The ATLAS Experiment. *Tagging and Suppression of Pileup Jets with the ATLAS Detector*. ATLAS-CONF-2014-018.
[2] Pythia8. *A comprehensive guide to the physics and usage of PYTHIA 8.3. https://arxiv.org/abs/2203.11601*
[3] Vaswani et at. 2017. *Attention Is All You Need*. 31st Conference on Neural Information Processing Systems.