Classifying hadronic objects in ATLAS with ML/AI algorithms

Robert Les, Michigan State University
on behalf of the ATLAS Collaboration

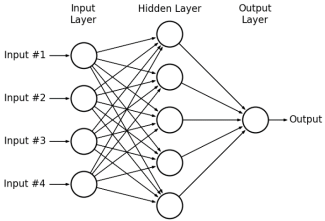
BOOST Genoa
July 29, 2024

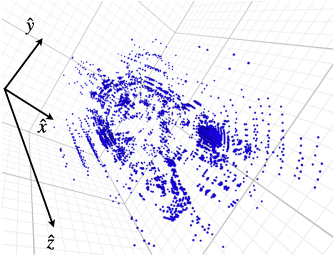
Introduction

We are well into the ML era of HEP!

ML used in all levels:

- Classification for low-level object definition
- Regression for calibration
- Density estimation for backgrounds
- Analysis specific signal/background separation
- Likelihood estimation and hypothesis testing





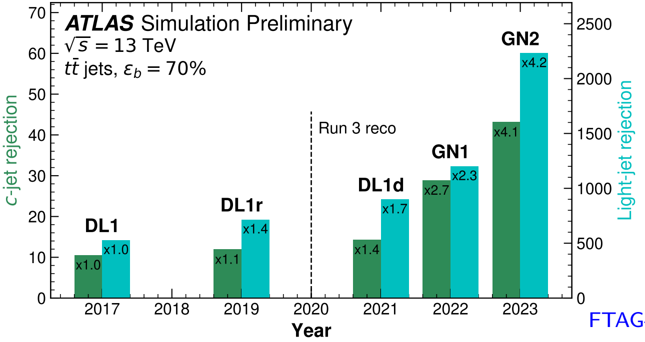
Will cover some recent results in ATLAS on hadronic object classification for

- Boosted resonance jet-tagging
- Single-pion identification
- Event missing transverse energy evaluation

Historical Perspective on Jet-Tagger Progress

The trend for boosted $W/Z/t/h \to R = 1.0$ jet identification (jet-tagging):

1) High-level taggers on jet substructure observables
   - Robust and easy to interpret at analysis/theory level
2) Machine learning taggers on jet-substructure
   - Non-trivially combine several observables for better discrimination
3) Machine learning taggers on low-level inputs
   - Lose interpretativeness for performance

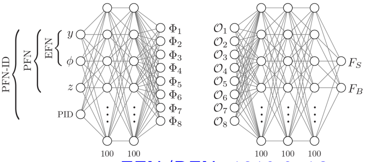


FTAG-2023-01

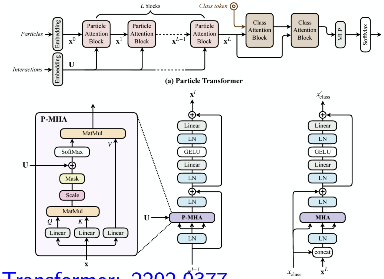
ML-era: Architectures

Machine-learning/pheno community is developing faster then we can test on data!

- Normal dense neural networks
- ResNet: CNN Architecture representing jet as image
- Energy/Particle Flow networks (EFN/PFN): General decomposition of IRC-safe observables
- ParticleNet: Graph network on point cloud
- ParticleTransformer: Transformer
- GN2X: Transformer with auxiliary tasks
- LundNet: Graph on declustering history
- PELICAN: Lorentz invariant network

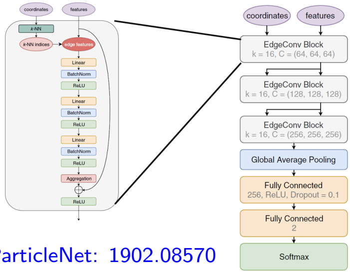


(a) EFN/PFN: 1810.05165



ParticleTransformer: 2202.0377

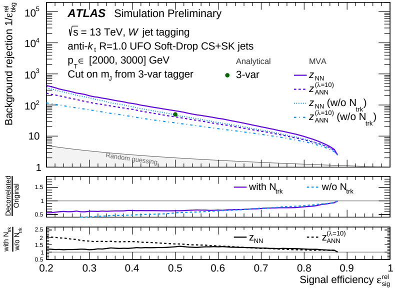
(b) Particle Attention Block.



ParticleNet: 1902.08570

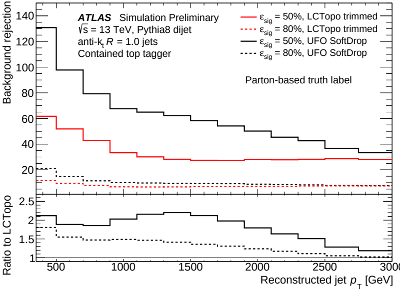
Baseline DNN top-taggers / W-tagger at $50/80\%$ signal efficiency over 15/10 substructure observables

- Factor $\sim$1.6/6 gain for W/top-tagging vs single substructure variable



Tagger performance deeply connected to jet definition!

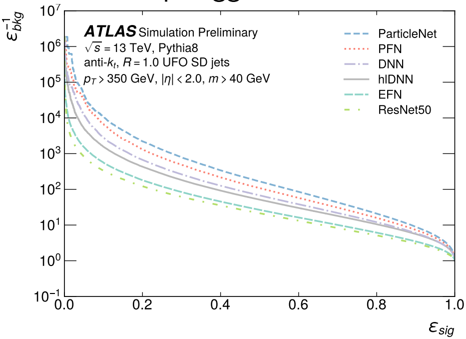
- Over Run2 moved from calorimeter $\rightarrow$ calorimeter+tracking information
- Factor 2 gains in background rejection
  - Due to better mass+substructure resolution



Constituent based top-tagger / W-tagger outperform high-level features ones:

- Provide network the lowest level information available: the jet constituents themselves
- Another factor 2-3 improvement!
- ResNet/EFN under-perform w.r.t theoretical performance
  - Real simulation studies important!

top-tagger ROC



W-tagging summary table

Model	AUC	ACC	ϵ_{bkg}^{-1} @ ϵ_{sig} = 0.5	ϵ_{bkg}^{-1} @ ϵ_{sig} = 0.8	# Params	Inference Time
EFN	0.920	0.835	35.1	7.95	56.73k	0.065 ms
PFN	0.931	0.853	44.7	9.50	57.13k	0.11 ms
ParticleNet	0.933	0.826	46.2	9.76	366.16k	0.36 ms
ParticleTransformer	0.951	0.880	77.9	14.6	2.14M	0.28 ms

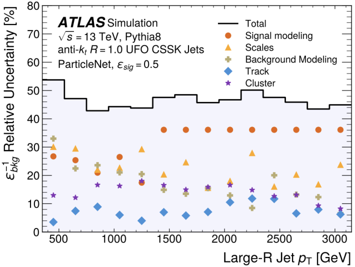
Constituent top-tagging modeling

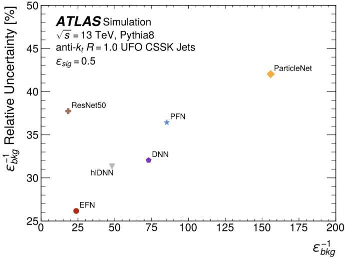
Preprint:CERN-EP-2024-159

But as we provide lower-level information the taggers can start to learn features more specific to generator/training dataset

New results evaluating **approximate** bottom-up experimental uncertainties and theory uncertainities on top-taggers

- Better taggers have higher uncertainties by almost factor 2
- Want to develop ways to break this trend
  - Hypothesis: EFN has less model dependency then other taggers since focus on IRC-safe observables?





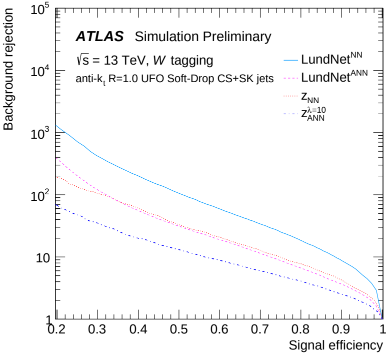
Open-data available! See Kevin Grief Poster

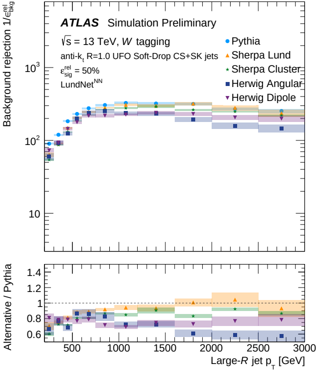
Train a network inspired by ParticleNet on the Lund-plane points:

- Graph neural-network on C/A re-clustering history of jet formation

But also see comparably modeling dependency as other networks

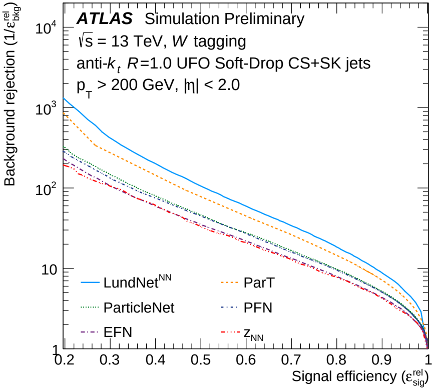
- Lund-plane point position can be suggestive if IR/collinear splitting
- Can clip out regions of Lund jet plane to remove IRC unsafe regions?

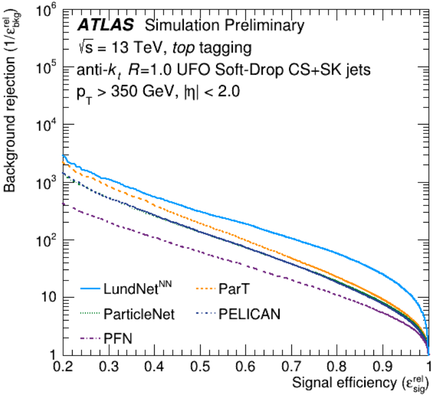




Of all explored architectures in ATLAS, the physics driven Lund-net tagger performs the best so far

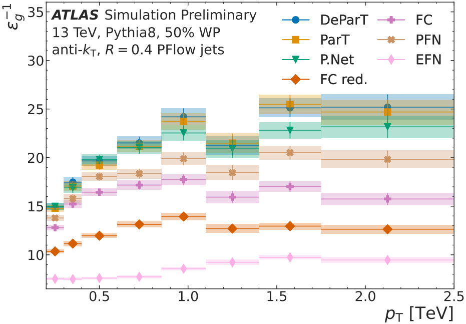
- Better than standard transformers or symmetry driven PELICAN
- Is stronger hyper-parameter optimization needed?
- How does the modeling dependence fit into this picture?
- Are detector details important?

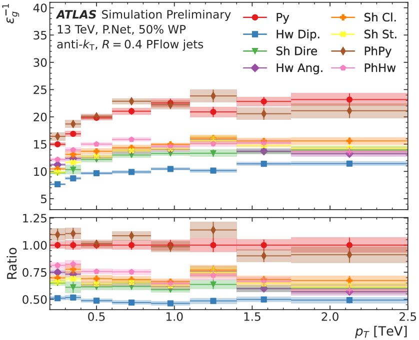




Similar studies also in $R = 0.4$ q/g-tagging!

- Interesting in the context of vector-boson scattering/fusion measurements
- Harder problem than $W/t/h$-tagging
  - Main differences: colour-factors $\rightarrow$ more radiation in gluon-jets
- All previous conclusions follow here
- Currently limited by capability to calibrate such a tagger

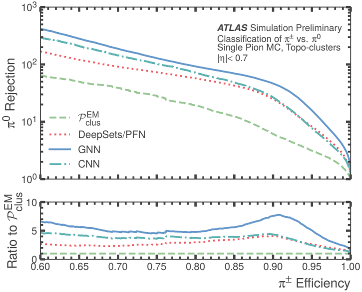


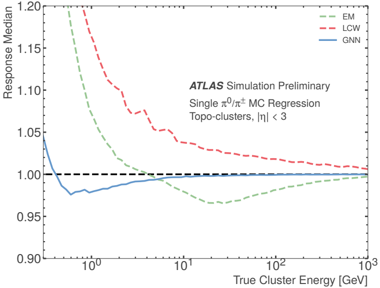


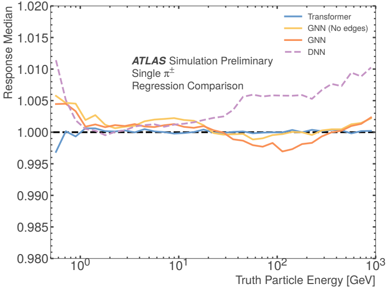
Single-pion ID and energy regression

Similar architectures studies for the classification and regression of single-pions
- Important for calibration of jet constituents

Similar improvement trends as in tagging results:
- Point-cloud/graphs better representations
- Tracking information helps







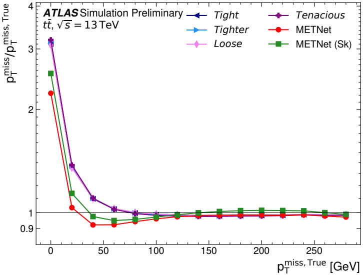
Missing transverse momentum calculation relies on well calibrated objects:

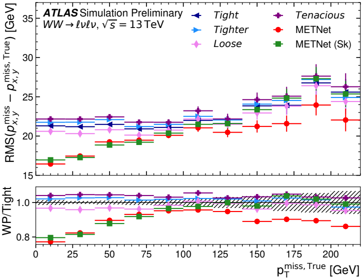
$$\underbrace{\underbrace{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{\substack{\text{selected}\\ \text{electrons}}} \mathbf{p}_{\mathrm{T}}^{e}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},e}} \underbrace{- \sum_{\substack{\text{accepted}\\ \text{photons}}} \mathbf{p}_{\mathrm{T}}^{\gamma}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\gamma}} \underbrace{- \sum_{\substack{\text{accepted}\\ \tau\text{-leptons}}} \mathbf{p}_{\mathrm{T}}^{\tau_{\mathrm{had}}}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\tau_{\mathrm{had}}}} \underbrace{- \sum_{\substack{\text{selected}\\ \text{muons}}} \mathbf{p}_{\mathrm{T}}^{\mu}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\mu}} \underbrace{- \sum_{\substack{\text{accepted}\\ \text{jets}}} \mathbf{p}_{\mathrm{T}}^{\mathrm{jet}}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,jet}}}}_{\text{hard term}} \underbrace{\underbrace{- \sum_{\substack{\text{unused}\\ \text{tracks}}} \mathbf{p}_{\mathrm{T}}^{\mathrm{track}}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,soft}}}}_{\text{soft term}} .$$

Several semi-arbitrary working points defined to define the input objects

New METNet DNN algorithm to combine these choices as a regressor to truth value

- Much better scale and resolution with respect to truth
- Can also adjust network to output a score confidence to define a METNet significance: $\mathrm{METNetSig} = |\boldsymbol{E}_T^{miss}|/\sigma(|\boldsymbol{E}_T^{miss}|)$





Summary

Presented several results on ML techniques for hadronic object identification:

- Constituent level: single-pion identification/response
- Jet-level: jet origin classification
- Event-level: MET calculation and response

Usage of ML techniques have provided improvements almost everywhere

- ML allows us to easily utilize the lowest-level information and their correlations

But it is not the end of the line:

- New ML algorithms always incoming
  - Full detector performance is important for these
- Pushing into regime where networks might be learning simulation specific features
  - Understanding the source of this model-dependence will be crucial to generalize the performance of these new taggers
  - Behavior in data is fundamental for its usability in physics
- ATLAS has top-tagging Open-data to provide inputs for the pheno/ML community to also study these

BACKUP

Calibrations

The tagger approach to substructure limited by capability to calibrate

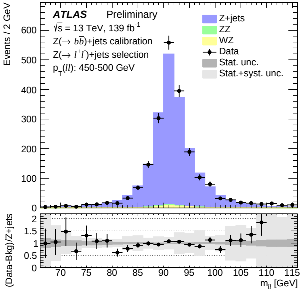
- Train on simulation, but also apply to data

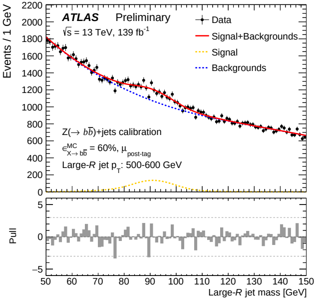
Often use top-down approach, where calibrate to a well known resonance

- $t\bar{t}$, $V+\text{jets}$, $V+\gamma$
- Can measure QCD jet rejection in multijet topologies

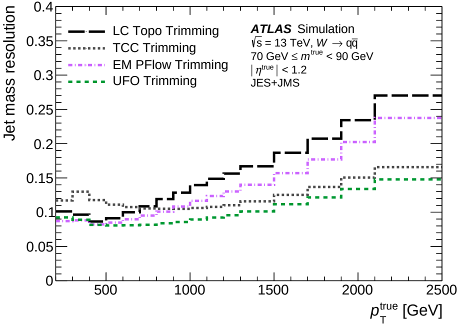
Difficulties can arise:

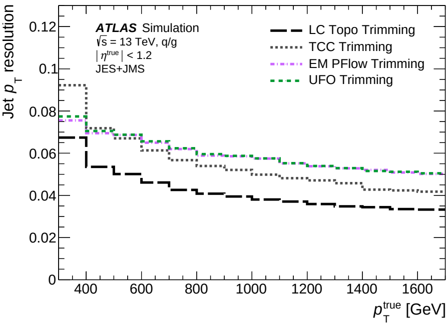
- No easy SM resonance: e.x. calibrating to h
- Limited probe statistics at high-p_{T}
- Limited trigger/selection capabilities at low-p_{T}





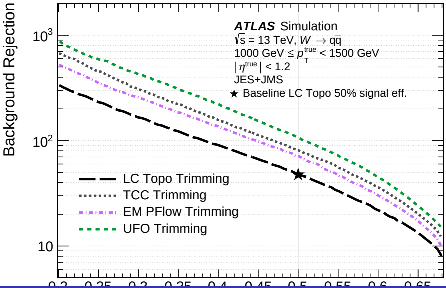
Jet collections





Tagger performance deeply connected to jet input definition!

- Over Run2 moved from Topological Clusters → Track-Calo Clusters → Unified Flow-Objects
- Better mass+substructure resolution



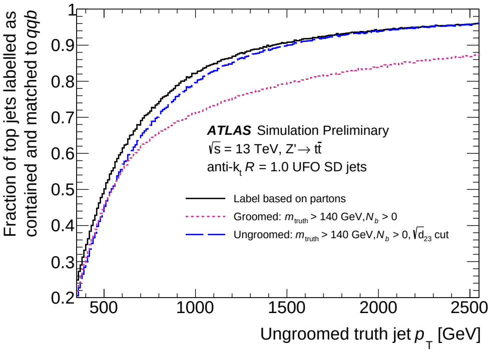
Taggers Definitions

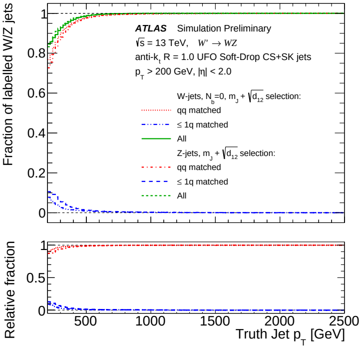
Over several publications iterated on top and W truth label defintions

- Move from parton history to ghost-associated matching between truth and reco jets

More robust truth-labeling helps with cleaner classification labels

- Less generator dependence

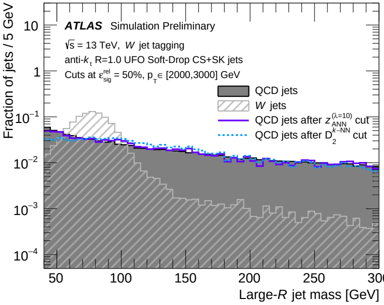


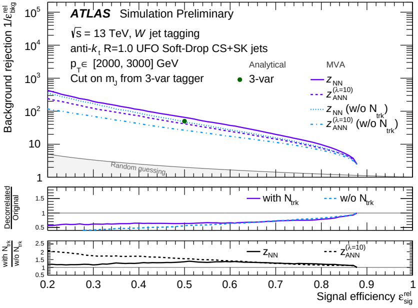


Mass Decorrelation

Similar to top-tagger, developed a DNN W-tagger over 10 substructure variables

- But obviously the DNN learns that the mass is very good discriminant
  - Sculpts background to look like signal!
  - Not an input to the network
  - Difficult to use sideband regions in analysis
- Train against a second "adversarial" network to force network to decorrelate the feature: Loss function $L = L_{classifier} - \lambda L_{adv}$
- Decorrelated tagger similar to cut-based.
- Correlated tagger has 10x better background rejection



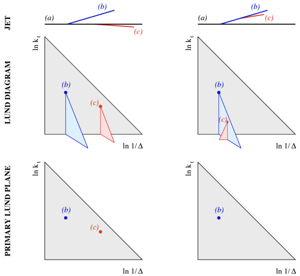


Recluster jet with C/A algorithm and each split represented on lund jet plane based on:

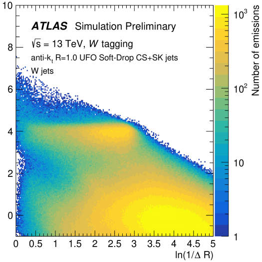
transverse momentum k_T, angle Δ, and momentum fraction z

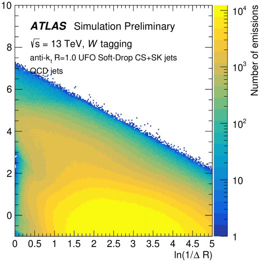
Variables can distinguish splitting type:

- ISR: low $\ln(1/\Delta)$
- Non-perturbative: low $\ln(k_T)$
- hard colinear: high $z \rightarrow$ top graph edge



arXiv:1807.04758

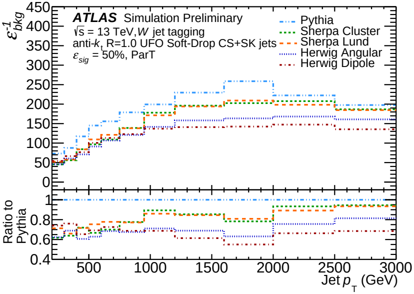




ATL-PHYS-PUB-2023-017

Similar results for constituent based W-tagger

- Transformers performing best on this dataset
- But still large modeling uncertainty



Model	AUC	ACC	ε_{bkg}^{-1} @ ε_{sig} = 0.5	ε_{bkg}^{-1} @ ε_{sig} = 0.8	# Params	Inference Time
EFN	0.920	0.835	35.1	7.95	56.73k	0.065 ms
PFN	0.931	0.853	44.7	9.50	57.13k	0.11 ms
ParticleNet	0.933	0.826	46.2	9.76	366.16k	0.36 ms
ParticleTransformer	0.951	0.880	77.9	14.6	2.14M	0.28 ms

Constituent based top taggers outperform high-level features ones:

- Provide network the lowest level information available: the jet constituents themselves
- Another factor 2-3 improvement!
- ResNet/EFN under-perform w.r.t theoretical performance
  - Real simulation studies important!

Model	AUC	ACC	ε_{bkg}^{-1} @ $\varepsilon_{sig} = 0.5$	ε_{bkg}^{-1} @ $\varepsilon_{sig} = 0.8$	# Params	Inference Time
ResNet 50	0.885	0.803	21.4	5.13	1,486,209	9 ms
EFN	0.901	0.819	26.6	6.12	1,670,451	4 ms
hlDNN	0.938	0.863	51.5	10.5	93,151	3 ms
DNN	0.942	0.868	67.7	12.0	876,641	3 ms
PFN	0.954	0.882	108.0	15.9	689,801	4 ms
ParticleNet	0.961	0.894	153.7	20.4	764,887	38 ms