A Unified Path for Anomaly Detection in HEP

Taoli Cheng

Nov. 04 @ ML4Jets 2022

Mila

Université de Montréal

A unified path combining classification and anomaly detection

T. Cheng and A. Courville,
Invariant Representation Driven Neural Classifier for Anti-QCD Jet Tagging,
JHEP 10 (2022) 152.
arXiv: 2201.07199

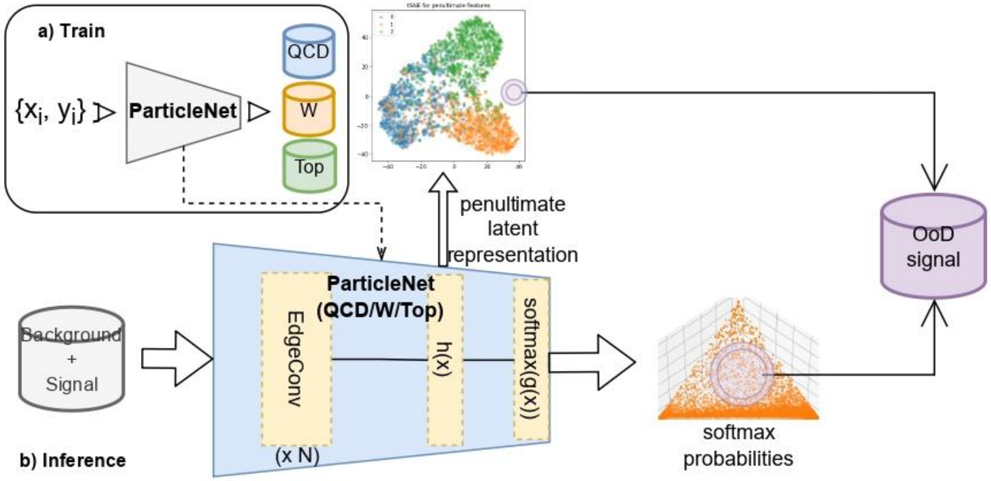
Key Messages

- A unified path for ML applications in HEP: re-utilizing Standard Model jet classifiers as generic anomalous jet taggers (event representations → classifier and anomaly detection)
- Domain knowledge → Inductive biases
- Representation-driven
- Spurious correlations and invariances: jet mass correlation

Framework

An effort in bridging SM jet taggers and general new physics searches

- In-distribution classes
- Neural feature extractor
- Anomaly detection / inference strategy
- Improved predictive uncertainty estimates



$$\mathbf{x} \mapsto \mathbf{h}(\mathbf{x});\ \mathbf{g}(\mathbf{x}) = \mathbf{W}\mathbf{h}(\mathbf{x}) + \mathbf{b};\ p_k(\mathbf{x}) = \frac{\exp(\mathbf{g}_k(\mathbf{x}))}{\sum_{k'}^{K} \exp(\mathbf{g}_{k'}(\mathbf{x}))}$$

Classifier-based Anomaly Detection – CLFAD

- A well-calibrated Standard Model jet classifier: the probability of an event being correctly classified ($\hat{y}$ = y) is equal to the predicted confidence $\hat{p}$

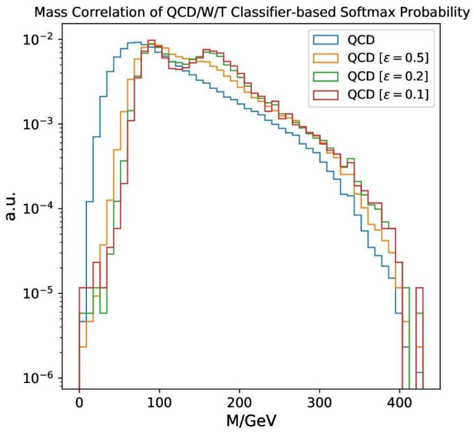
$$P(\hat{y} = y \mid \hat{p} = p) = p$$

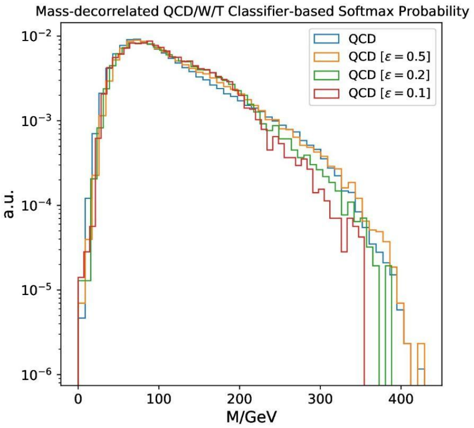
- Powerful feature extractor (architecture)
- Data-augmented mass decorrelation: populate mass distributions and match them across the in-distribution classes

$$\log p(y|\mathbf{x}) = \log p_{\cancel{M}}(y|\mathbf{x}) + \log p_M(y|\mathbf{x})$$

Main Results - Mass Correlation

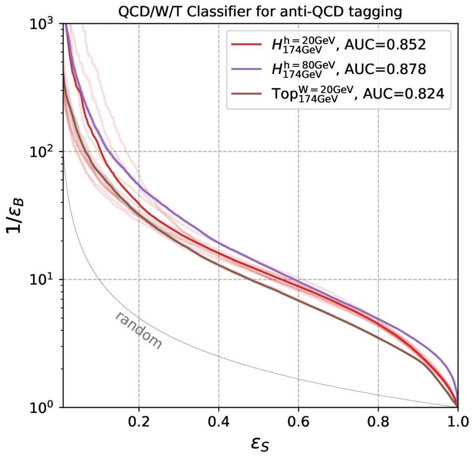
- Mass correlation is dominated by in-distribution class jet masses
- Resampling in-distribution classes to match the mass distributions

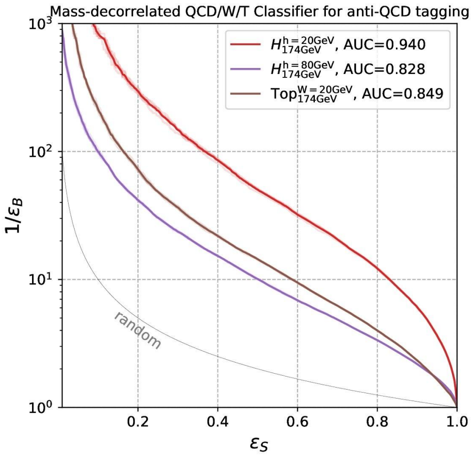




Main Results - Receiver Operating Characteristic Curves

- Mass decorrelation plays a different role here compared with the case in SM jet taggers
- Mass decorrelation eliminates the dominant factor (jet mass M) and drives powerful representation learning for Out-of-Distribution detection (**AUCs are even increased for the mass-decorrelated case**)





CLFAD: Extensions

- Datasets
  - Larger datasets
  - More jet types
- Training strategies
  - Inject signal information as an example of outlier exposure
- Simulation-to-data adaptation
  - In the same line with supervised SM jet classifiers

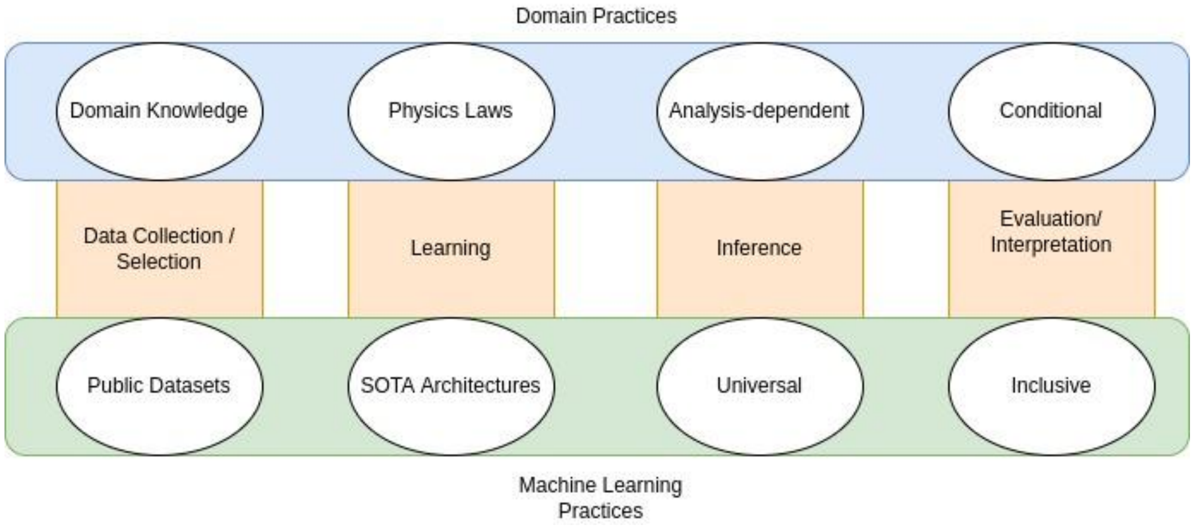
Bridging ML and HEP

Chapter: Anomaly Detection

T. Cheng, ***Bridging Machine Learning and Sciences: Opportunities and Challenges***, arXiv: 2210.13441

Interdisciplinary Research Protocols

- Datasets and benchmarks
- Learning algorithms
- Model inference
- Model evaluation
- Model generalization and robustness



Datasets and Benchmarks

- The first and foremost element in the pipeline
  - ML community: difficulties in labeling
  - HEP community: difficulties in reaching uniform settings
- Training sets:
  - Dataset generation setups
  - Input representation
- Test sets:
  - Limited test sets lead to “over-fitting”: e.g., top in anomalous jet tagging
  - Comprehensive test sets serve as a stress-tester for model development

Model Inference and Evaluation

- Evaluation procedure
  - The target class/task affects inference approaches and model selection
- Evaluation metrics
  - Area Under the ROC Curve (AUC)
  - Significance Improvement Characteristic (SIC)
- Evaluation benchmarks: beyond scalar metrics
  - Test sets
  - Metrics
  - Model examination

Model Generalization and Robustness

- Uncertainty quantification
- Distribution shifts:
    - Different event generation settings
    - Simulation-to-data adaptation
- Robustness:
    - Sensitivity over a broad range of test sets
    - Free of spurious correlations

Towards a Unified Path

- Better collaboration scheme
- Communal vocabulary between communities
- Publishing traditions (dedicated venues, open-source practices)
- …

Towards a Unified Path

- Better collaboration scheme
- Communal vocabulary between communities
- Publishing traditions (dedicated venues, open-source practices)
- …

https://ad4sd.github.io/seminars/

Thanks!

Backup

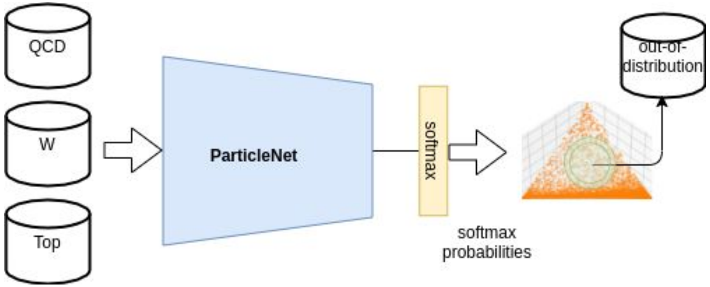
Workflow for Anti-QCD Jet Tagging

- In-distribution training data
  - Simulated large-cone QCD/W/Top jets with pT~600 GeV
  - Low-level jet constituent 4-vectors (or variants)
- Neural architecture (a decent baseline: ParticleNet) *[Huilin Qu, Loukas Gouskos. arXiv: 1902.08570]*
- Post-processing: anomaly score / “anti-QCD-ness”

$$-p_{\mathrm{QCD}}(\mathbf{x}) \text{ with } p_{\mathrm{QCD}} = p(y=0 \mid \mathbf{x})$$

$$\mathtt{MD} = (\mathbf{h}(\mathbf{x}) - \mu)^\top \Sigma^{-1} (\mathbf{h}(\mathbf{x}) - \mu)$$

- Out-of-distribution test sets
  - OoD class 1: H (174 GeV) → hh (h→ jj) with h (20 GeV)
  - OoD class 2: H (174 GeV) → hh (h→ jj) with h (80 GeV)
  - OoD class 3: “Top” (174 GeV) with W (20 GeV)



Anomaly Scores

Softmax probabilities vs Logits vs Representation Layer (Final Features)

- Softmax probabilities based scores
  - Maximum Softmax Probability: $-\max\{p_1, p_2, p_3\}$
  - Softmax Probabilistic Entropy: $-\sum_{i=1}^{k} p_i log(p_i)$
- Logits based scores
- Representation based scores
  - Distance in feature space
  - Distance-based logits: Replacing logits with feature distance for softmax

[Slide from T. Cheng, ML4Jets 2021]