Resilience of Quark-Gluon Tagging

Where are the HERWIG gluons?

Lorenz Vogel

November 4, 2022

Institute for Theoretical Physics
Heidelberg University, Germany

ML4Jets Workshop 2022, Rutgers University

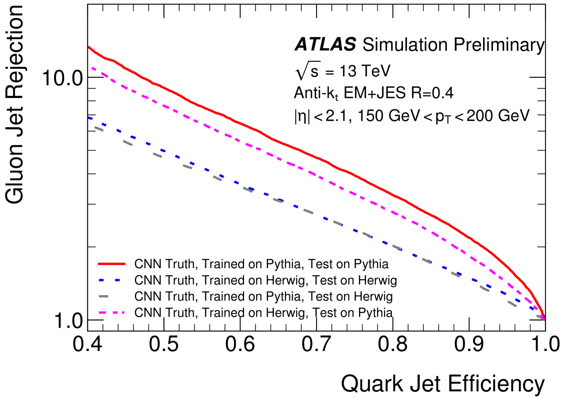
Paper in preparation — in collaboration with:
Anja Butter, Barry M. Dillon and Tilman Plehn

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Outline

Motivation

Problem of resilience in jet tagging

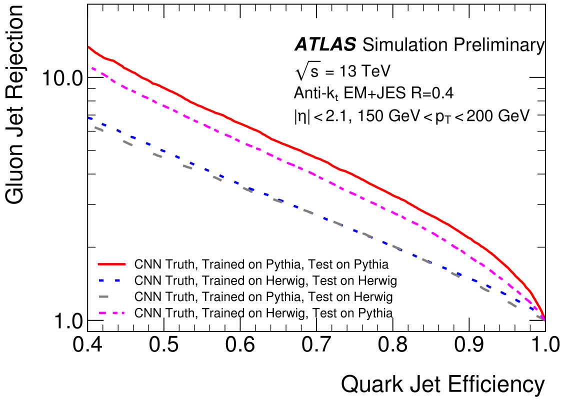


ATLAS figure showing the
PYTHIA-HERWIG discrepancy

[ATLAS Tech. Rep. and arXiv:1612.01551]

- the two standard generators, PYTHIA and HERWIG, differ substantially in their jet simulations
- gluon jets are more similar to quark jets in HERWIG than in PYTHIA
- network is worse at correctly tagging HERWIG jets than PYTHIA jets
- networks can extract robust physical information from imperfect simulations

Problem of resilience in jet tagging



This is a case study for explainable AI!

- Identifying the problematic phase-space configurations using...
  ... a Bayesian neural network
  ... reweighting (large weights)
- Building a resilient tagger...
  ... by training on mixed samples
  ... by training on reweighted data
- goal: improve jet-tagging resilience

Physics Problem

Standard sub-jet variables for quark-gluon tagging

Distinguishing quark jets from gluon jets exploits two features:

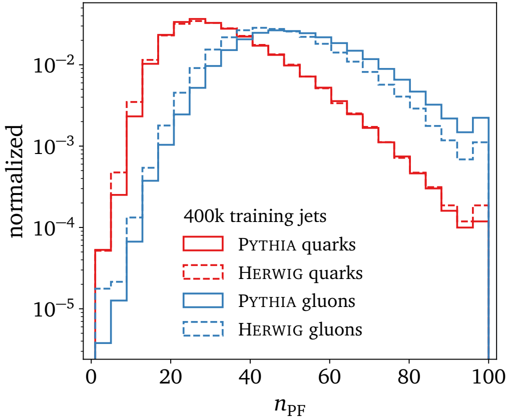
- differences in the QCD colour factors
- differences in the soft-limit splitting functions

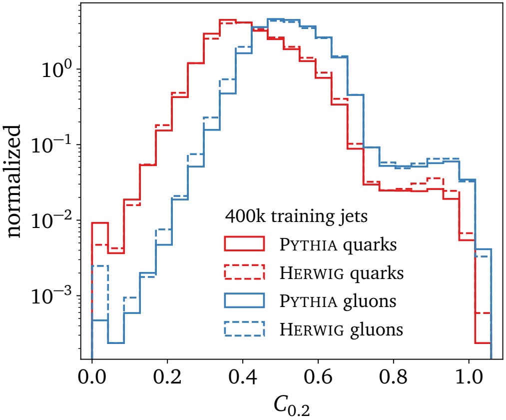
Standard kinematic sub-jet variables for quark vs. gluon discrimination expressed in terms of particle-flow (PF) objects:

$$n_{\mathrm{PF}} = \sum_{i_{\mathrm{PF}}} 1 \qquad w_{\mathrm{PF}} = \frac{\sum_{i_{\mathrm{PF}}} p_{\mathrm{T},i} \Delta R_{i,\mathrm{jet}}}{\sum_{i_{\mathrm{PF}}} p_{\mathrm{T},i}}$$

$$p_{\mathrm{T}} D = \frac{\sqrt{\sum_{i_{\mathrm{PF}}} p_{\mathrm{T},i}^2}}{\sum_{i_{\mathrm{PF}}} p_{\mathrm{T},i}} \qquad C_{0.2} = \frac{\sum_{i_{\mathrm{PF}}, j_{\mathrm{PF}}} p_{\mathrm{T},i} p_{\mathrm{T},j} (\Delta R_{ij})^{0.2}}{\left(\sum_{i_{\mathrm{PF}}} p_{\mathrm{T},i}\right)^2}$$

Standard sub-jet variables for quark-gluon tagging

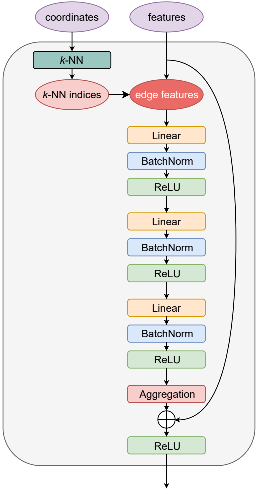


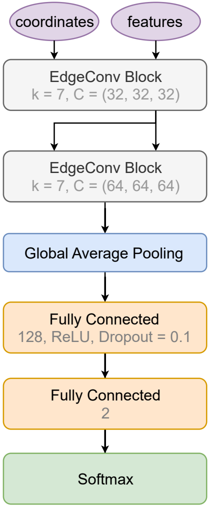


Gluon jets tend to have a higher particle multiplicity and a broader radiation pattern than quark jets and quark-jet constituents carry a larger average fraction of the jet energy (due to harder fragmentation)

Bayesian ParticleNet (BPN)

ParticleNet: jet tagging via particle clouds





graph-based jet-tagging architecture

- dynamic graph convolutional neural network (DGCNN)
- particle-cloud representation (jets as unordered sets of particles)
- explicitly respects permutation symmetry and flexibility to include arbitrary features for each particle
- edge convolution (EdgeConv)
- feature input: seven kinematic variables and experimentally realistic particle IDs
- coordinate input: $(\Delta\eta, \Delta\phi)$ coordinates
- 100 highest-p_{T} constituent particles

Basics of Bayesian neural networks

In Bayesian neural networks (BNNs), the network parameters ω are treated as random variables with corresponding probability distributions

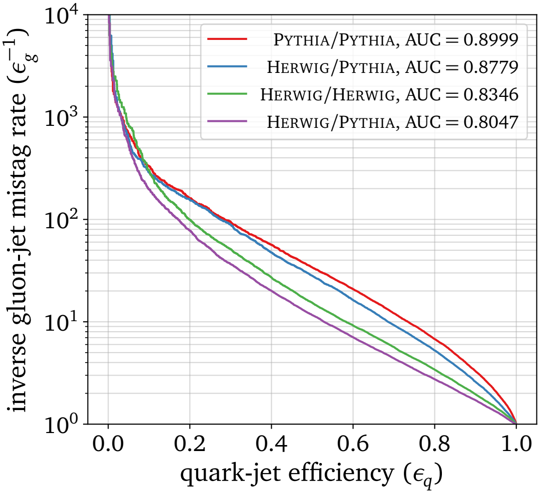
variational inference (VI): approximate intractable posterior $p(\omega|\mathcal{D})$ with a simplified and tractable family of distributions $q_\theta(\omega)$

$$p(y^*|x^*,\mathcal{D}) = \int \mathrm{d}\omega\, p(\omega|\mathcal{D}) p(y^*|x^*,\omega) \overset{\text{VI}}{\approx} \int \mathrm{d}\omega\, q_\theta(\omega) p(y^*|x^*,\omega)$$

The parameters θ of the variational posterior $q_\theta(\omega)$ are learned by minimizing the Kullback-Leibler (KL) divergence:

$$\min_\theta \mathrm{KL}[q_\theta(\omega), p(\omega|\mathcal{D})] \quad \xrightarrow{\text{Bayes}} \quad \mathcal{L}_{\text{BNN}} = \underbrace{\mathrm{KL}[q_\theta(\omega), p(\omega)]}_{\text{regularization term}} \underbrace{- \int \mathrm{d}\omega\, q_\theta(\omega) \log p(\mathcal{D}|\omega)}_{\substack{\text{negative log-likelihood term:} \\ \text{(binary) cross-entropy loss}}}$$

Bayesian ParticleNet (BPN)

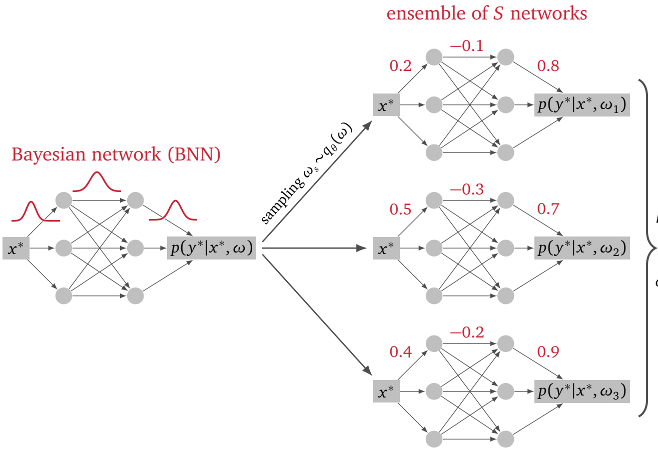


BPN trained/tested on different PYTHIA-HERWIG combinations

- we can reproduce the PYTHIA-HERWIG discrepancy
- BPN shows similar performance to the deterministic ParticleNet while providing additional information about uncertainties

Basics of Bayesian neural networks

By sampling over weight space, an ensemple of ordinary networks is created, each giving a different prediction (jet-by-jet error estimates for the tagging output)



output: predictive mean and standard deviation

$$\mu_{\text{pred}} = \frac{1}{S}\sum_{s=1}^{S} p(y^*|x^*,\omega_s)$$

$$\sigma^2_{\text{pred}} = \frac{1}{S}\sum_{s=1}^{S}\left[p(y^*|x^*,\omega_s)-\mu_{\text{pred}}\right]^2$$

Bayesian neural networks for classification

Averaging over an ensemble of S networks with different weight configurations:

$$p(y^*|x^*) \approx \int \mathrm{d}\omega\, q_\theta(\omega) p(y^*|x^*,\omega) \approx \frac{1}{S}\sum_{s=1}^{S} p(y^*|x^*,\omega_s) =: \mu_{\text{pred}}$$

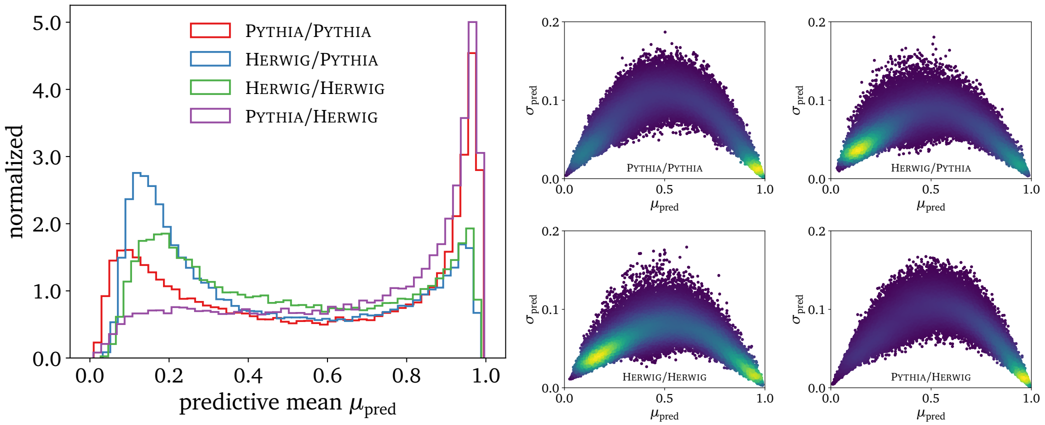
The uncertainty for each prediction is defined as:

$$\sigma^2_{\text{pred}} := \frac{1}{S}\sum_{s=1}^{S}\left[p(y^*|x^*,\omega_s) - \mu_{\text{pred}}\right]^2$$

After the final sigmoid transformation, the unconstrained BNN outputs turn into a correlated mean and standard deviation:

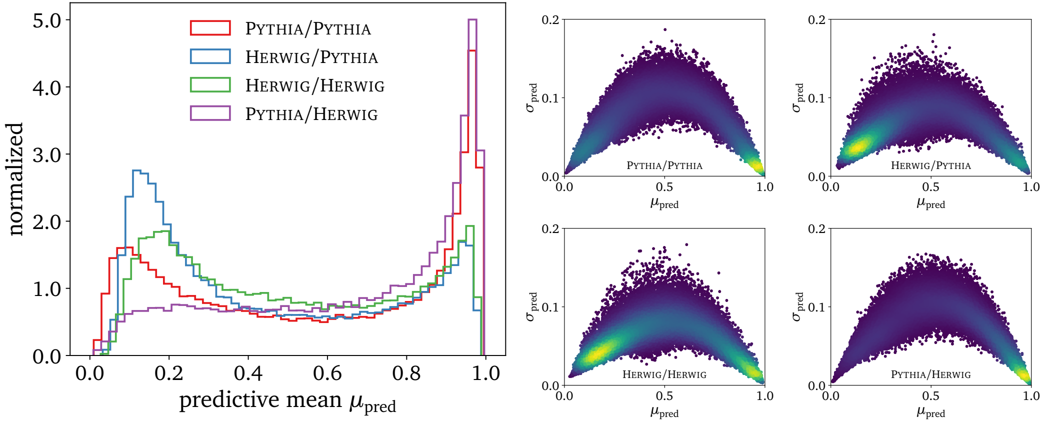
$$\sigma_{\text{pred}} \approx \mu_{\text{pred}}\left[1-\mu_{\text{pred}}\right]\sigma^{(\text{unconstr})}_{\text{pred}} \qquad \text{with} \qquad \mu_{\text{pred}} \in [0,1]$$

BPN: where are the HERWIG gluons?



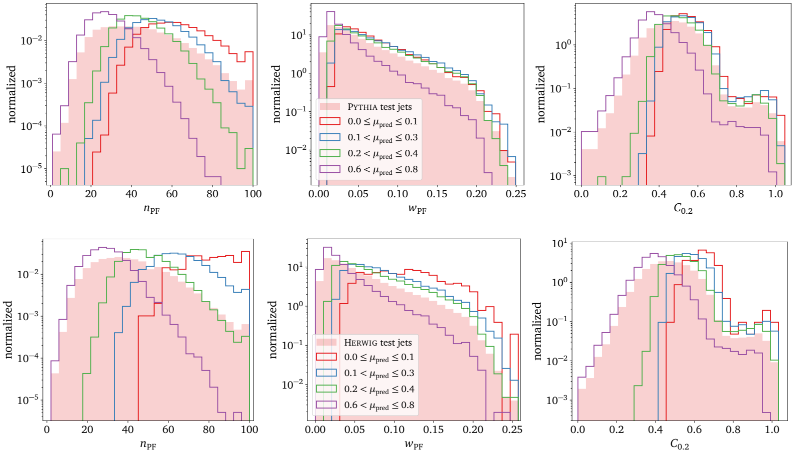
Predictive distributions of BPN trained/tested on different generator combinations:
HERWIG-trained networks have difficulties in confidently tagging gluons

BPN: where are the HERWIG gluons?



Correlation between the predictive mean and the predictive uncertainty:
HERWIG-trained networks have difficulties in confidently tagging gluons and seem to be less stable

BPN: standard sub-jet variables



Gluon-tagged jets tend to have a larger number of constituents
while quark-tagged jets tend to have a smaller number of constituents

Jet Reweighting

Per-jet reweighting using the BPN classifier

Ideal per-jet weight to morph the HERWIG simulation into the PYTHIA one:

$$w(x) = \frac{p_{\mathrm{PY}}(x)}{p_{\mathrm{HE}}(x)} \approx \frac{C_{\mathrm{PY/HE}}(x)}{1 - C_{\mathrm{PY/HE}}(x)}$$

- training a classifier $C_{\mathrm{PY/HE}}$ to distinguish the two simulations (PYTHIA vs. HERWIG), separately for quarks and gluons

Per-jet reweighting using the BPN classifier

Ideal per-jet weight to morph the HERWIG simulation into the PYTHIA one:

$$w(x) = \frac{p_{\mathrm{PY}}(x)}{p_{\mathrm{HE}}(x)} \approx \frac{C_{\mathrm{PY/HE}}(x)}{1 - C_{\mathrm{PY/HE}}(x)}$$

- training a classifier $C_{\mathrm{PY/HE}}$ to distinguish the two simulations (PYTHIA vs. HERWIG), separately for quarks and gluons

Using the reweighting factors w, we can then train a classifier $C_{q/g}$ to do quark vs. gluon classification on w^r-reweighted HERWIG datasets:

$$\mathcal{L}_{\mathrm{BCE}}^{q/g} = -\sum_{i \in \{q\}} w(x_i)^r \log C_{q/g}(x_i) - \sum_{i \in \{g\}} w(x_i)^r \log(1 - C_{q/g}(x_i))$$

- $r = 0$: $w^0 = 1$, i.e. classical/original HERWIG data
- $r = 1$: $w^1 = w$, i.e. HERWIG fully reweighted into PYTHIA

Per-jet reweighting using the BPN classifier

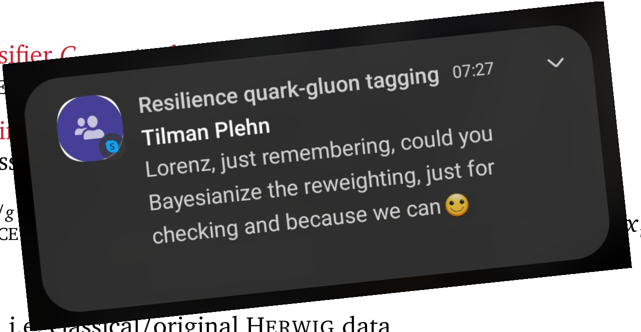
Ideal per-jet weight to morph the HERWIG simulation into the PYTHIA one:

$$w(x) = \frac{p_{\text{PY}}(x)}{p_{\text{HE}}(x)} \approx \frac{C_{\text{PY/HE}}(x)}{1 - C_{\text{PY/HE}}(x)}$$

- training a classifier C
  (PYTHIA vs. HE

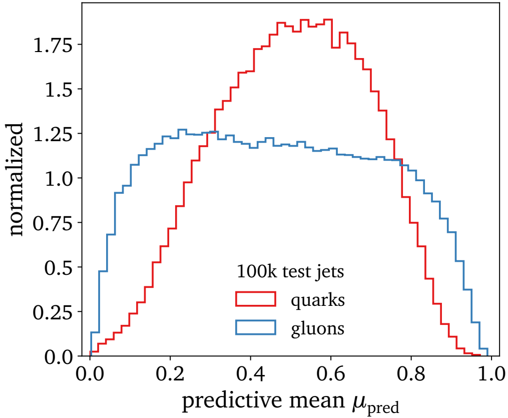
Using the reweighti
quark vs. gluon class

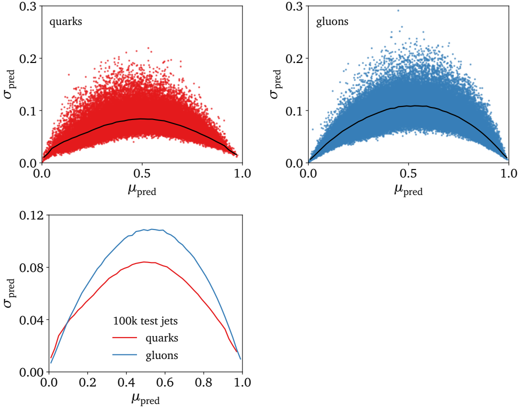
$$\mathcal{L}_{\text{BCE}}^{q/g} \quad x_i))$$



- $r = 0$: $w^0 = 1$, i.e. classical/original HERWIG data
- $r = 1$: $w^1 = w$, i.e. HERWIG fully reweighted into PYTHIA

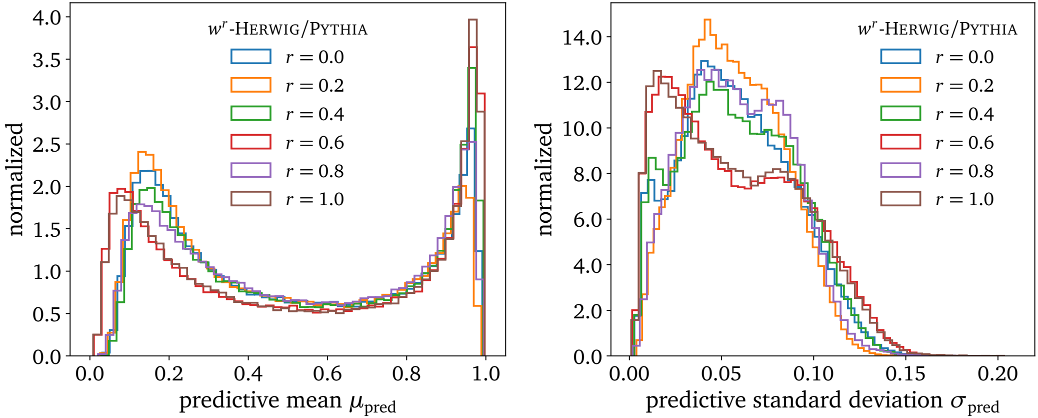
HERWIG → PYTHIA reweighting: predictive distributions





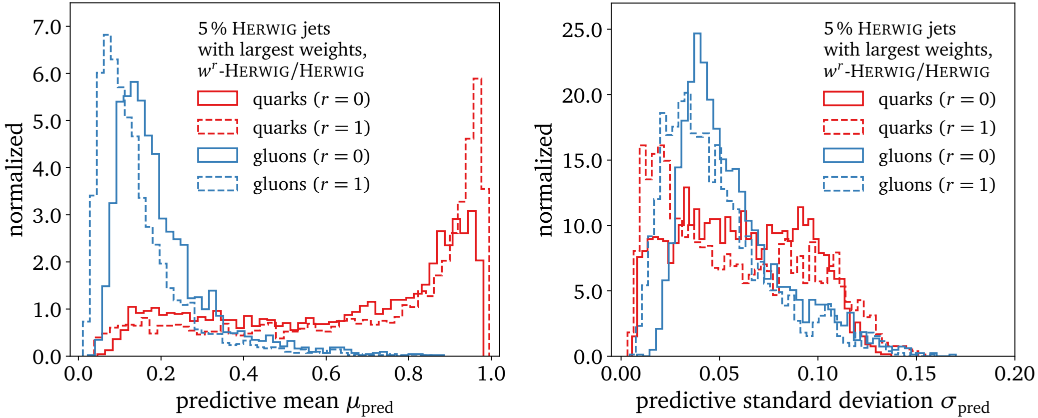
The quark classifier shows some instability because it's harder to distinguish PYTHIA and HERWIG quark jets (they are more similar) – the gluon-classifier training seems to be more stable

HERWIG → PYTHIA reweighting: continuous reweighting scheme



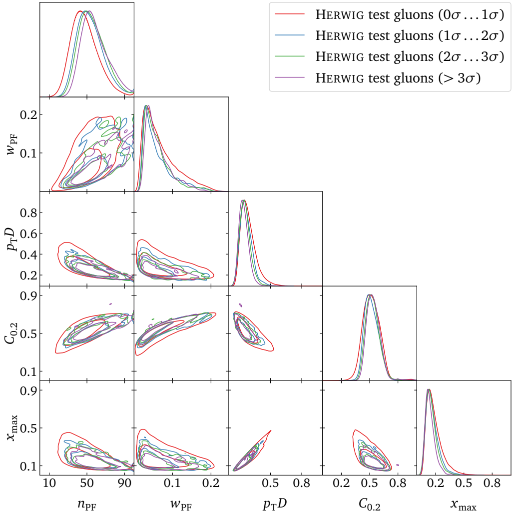
The instability of HERWIG-trained networks is removed when training on PYTHIA-like data and reweighting allows the network to confidently classify gluon jets

HERWIG → PYTHIA reweighting: large reweighting factors



How do HERWIG gluons with large reweighting factors migrate from the less confident "classical HERWIG" classifier to the more confident "PYTHIA-like" classifier?

Herwig → Pythia reweighting: large reweighting factors

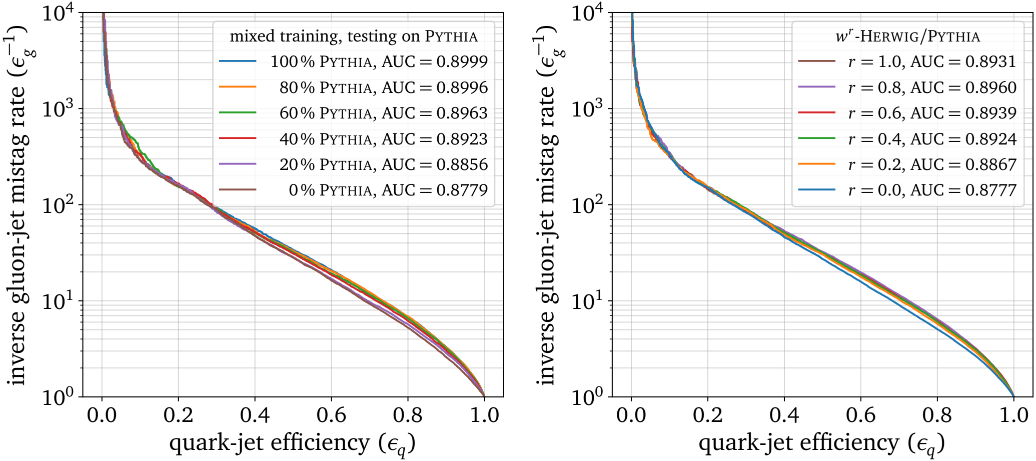


two-dimensional correlation plots between high-level observables

- interesting things happen in the particle multiplicity n_{PF} and the radiation distribution w_{PF}

Jet-Tagging Resilience

Resilience through mixed training and reweighting



Conclusions

Conclusions

Using ML to understand the difference between two datasets:

- Bayesian networks provide a powerful framework to include uncertainties in deep learning
- Bayesian ParticleNet (BPN) shows similar performance to the deterministic ParticleNet while providing additional information about uncertainties
- exploring phase-space regions that HERWIG gluons don't fill
- particle multiplicity n_{PF} seems to be most important for the tagging output: predictive distributions are strongly correlated to the number of jet constituents
- Bayesian networks are a really nice tool if things go wrong!

References and further reading

- ATLAS Collaboration
  Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector
  Technical Report (2017)
- P. T. Komiske, E. M. Metodiev and M. D. Schwartz
  Deep learning in color: towards automated quark/gluon jet discrimination
  JHEP 01, 110 (2017), arXiv:1612.01551 [hep-ph]
- H. Qu and L. Gouskos
  ParticleNet: Jet Tagging via Particle Clouds
  Phys. Rev. D 101, 056019 (2020), arXiv:1902.08570 [hep-ph]
- S. Bollweg, M. Haußmann, G. Kasieczka, M. Luchmann, T. Plehn and J. Thompson
  Deep-Learning Jets with Uncertainties and More
  SciPost Phys. 8, 006 (2020), arXiv:1904.10004 [hep-ph]
- Quark-gluon datasets:
  Pythia8 and Herwig7.1 Quark and Gluon Jets
  doi:10.5281/zenodo.3164691 and doi:10.5281/zenodo.3066475

Quark-gluon dataset details

Datasets of quark and gluon jets from the `EnergyFlow` package:

- PYTHIA 8.226 and HERWIG 7.1.4 quark (uds) and gluon jets
- generated from $q\bar{q} \to Z(\to \nu\bar{\nu}) + g$ and $qg \to Z(\to \nu\bar{\nu}) + (uds)$ processes in proton-proton collisions using a center-of-mass energy of 14 TeV
- jets are defined through the anti-k_T algorithm with a radius of $R = 0.4$
- note that no detector simulation is performed
- $p_{\mathrm{T,jet}} = 500 \ldots 550\,\mathrm{GeV}$ and $|\eta_{\mathrm{jet}}| < 1.7$
- each jet is defined by its constituent coordinates $x_i = \{(p_\mathrm{T}, \eta, \phi)_k\}$ with $k = 1, \ldots, n_{\mathrm{PF}}$ (padded with zero-particles)

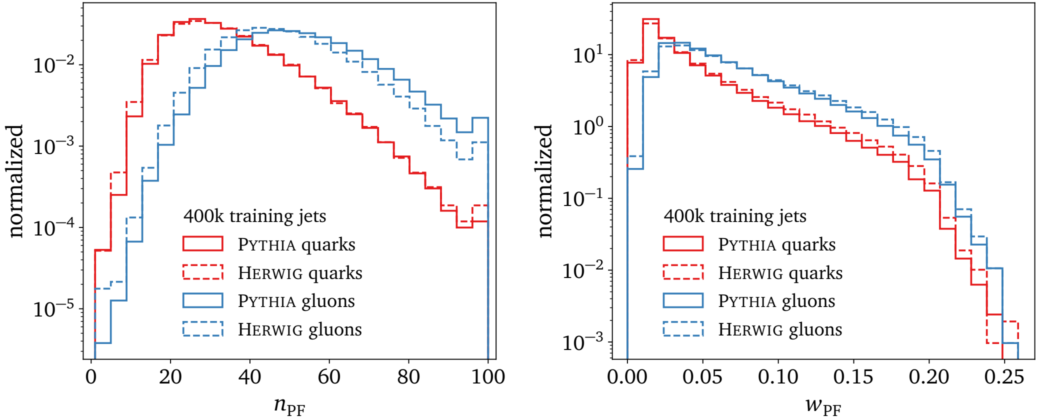
Bayesian ParticleNet implementation details

variable	definition
$\Delta\eta$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi$	difference in azimuthal angle between the particle and the jet axis
$\log p_\mathrm{T}$	logarithm of the particle's transverse momentum p_T
$\log E$	logarithm of the particle's energy E
$\log \frac{p_\mathrm{T}}{p_{\mathrm{T,jet}}}$	logarithm of the particle's p_T relative to the jet $p_{\mathrm{T,jet}}$
$\log \frac{E}{E_\mathrm{jet}}$	logarithm of the particle's energy E relative to the jet energy E_jet
ΔR	angular separation between the particle and the jet axis ($\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$)

Table 1: Kinematic input variables used in quark-gluon tagging with ParticleNet ("features")

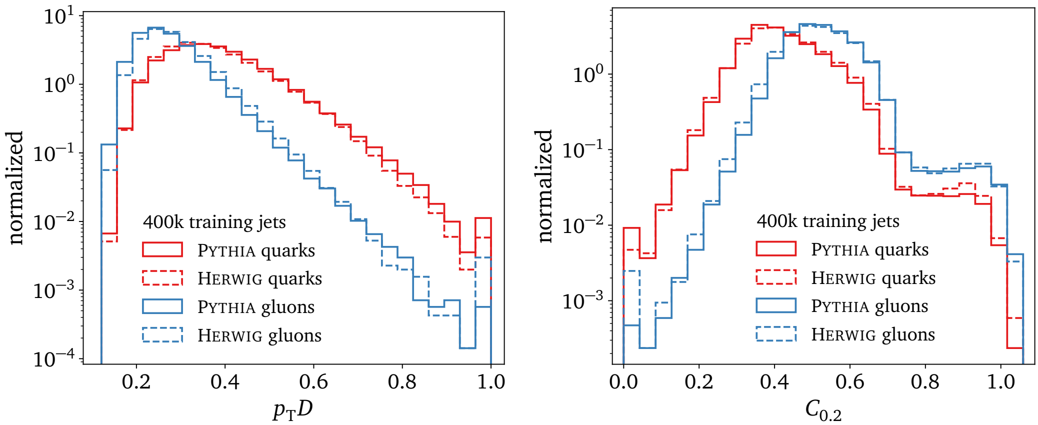
- one-hot encoded experimentally realistic particle IDs (PIDs):
  only five particle types (electrons $e^\pm$, muons $\mu^\pm$, charged hadrons $h^\pm = \pi^\pm/K^\pm/p/\bar{p}$, neutral hadrons $h^0 = K_L/n/\bar{n}$, and photons γ) and electric charge q

Standard sub-jet variables for quark-gluon tagging



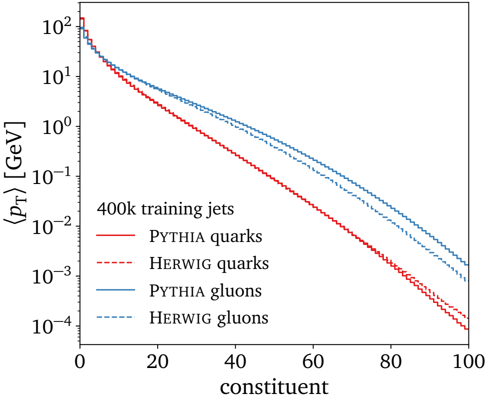
Gluon jets tend to have a higher particle multiplicity and a broader radiation pattern than quark jets

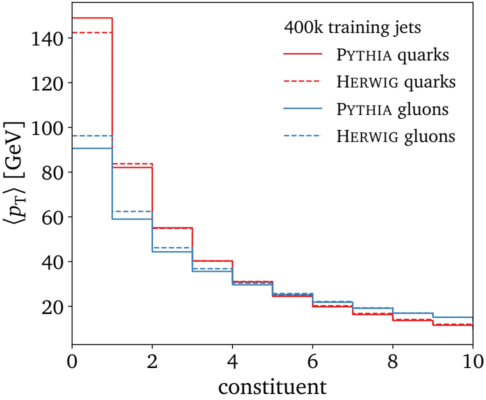
Standard sub-jet variables for quark-gluon tagging



The harder fragmentation for quarks makes quark-jet constituents carry a larger average fraction of the jet energy

Standard sub-jet variables for quark-gluon tagging





Mean transverse momentum of the p_T-ordered jet constituents:
Overall, the gluon distributions show larger differences than the quark distributions
HERWIG gluons tend to have a smaller particle multiplicity than PYTHIA gluons

Basics of Bayesian neural networks

Minimizing the Kullback-Leibler (KL) divergence:

$$\min_\theta \mathrm{KL}[q_\theta(\omega), p(\omega|\mathcal{D})] \xrightarrow{\text{Bayes}} \mathcal{L}_{\mathrm{BNN}} = \underbrace{\mathrm{KL}[q_\theta(\omega), p(\omega)]}_{\text{regularization term}} \underbrace{- \int d\omega\, q_\theta(\omega) \log p(\mathcal{D}|\omega)}_{\text{negative log-likelihood term}}$$

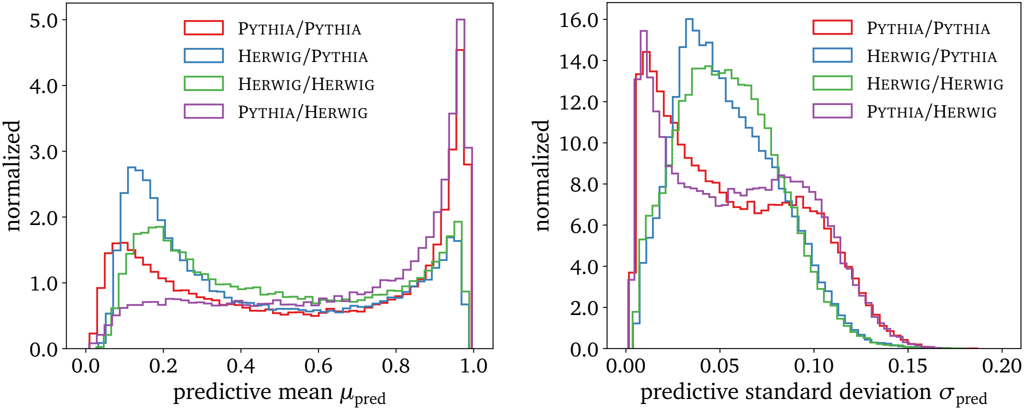
Bayes' theorem:

$$p(\omega|\mathcal{D}) = \frac{p(\mathcal{D}|\omega)p(\omega)}{p(\mathcal{D})} \quad \text{with} \quad p(\mathcal{D}|\omega) = \prod_{n=1}^{N} p(y_n|x_n, \omega)$$

Bayesian ParticleNet (BPN) loss:

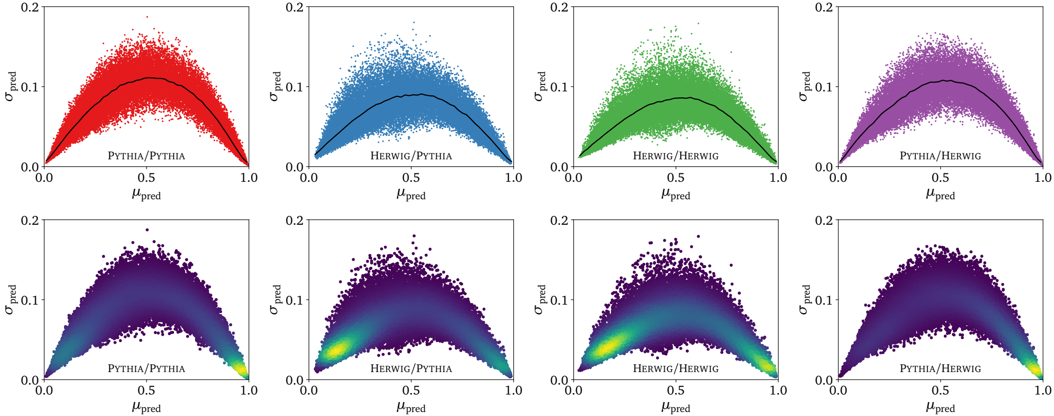
$$\mathcal{L}_{\mathrm{BPN}} = \underbrace{\frac{1}{N} \sum_{i\in\{\omega\}} \frac{1}{2} \left\{ \mu_i^2 + \sigma_i^2 - \log \sigma_i^2 - 1 \right\}}_{\substack{\text{KL-term for Gaussian priors } \mathcal{N}(0,1) \text{ and} \\ \text{Gaussian weight distributions } \theta = (\mu, \sigma)}} \underbrace{- \frac{1}{S} \frac{1}{M} \sum_{s=1}^{S} \sum_{m=1}^{M} \log p(y_m|x_m, \omega_s)}_{\substack{\text{negative log-likelihood term:} \\ \text{(binary) cross-entropy loss}}}$$

BPN: where are the HERWIG gluons?



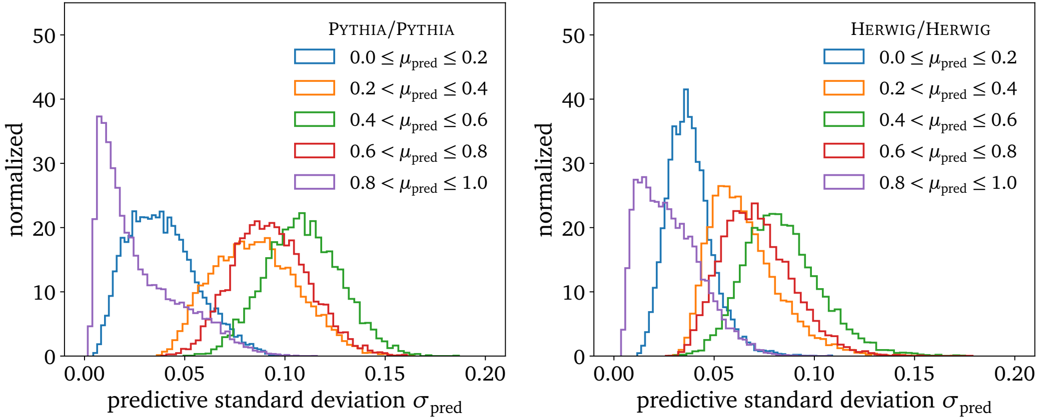
Predictive distributions of BPN trained/tested on different generator combinations:
HERWIG-trained networks have difficulties in confidently tagging gluons

BPN: training instability



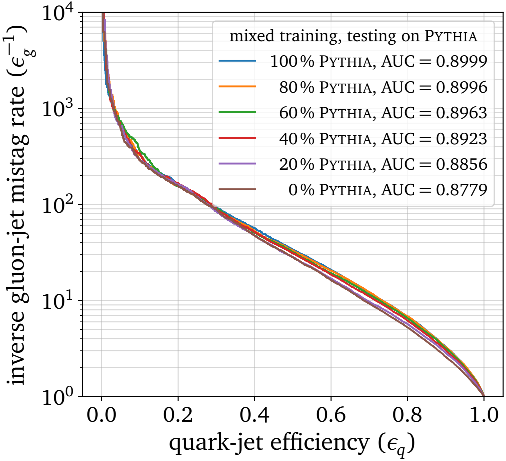
Correlation between the predictive mean and the predictive uncertainty:
HERWIG-trained networks have difficulties in confidently tagging gluons and seem to be less stable

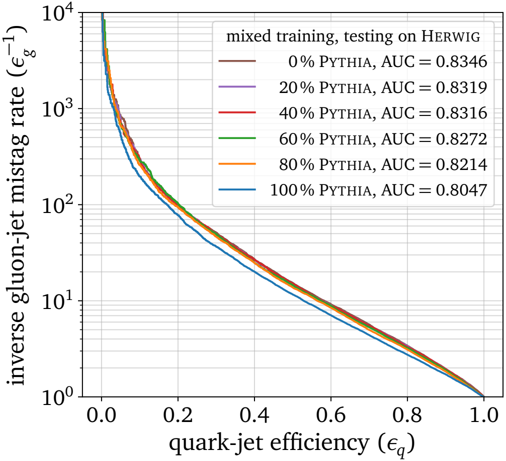
BPN: predictive standard deviation



Predictive standard deviation for different predictive-mean bins:
Herwig-trained networks have difficulties in confidently tagging gluons

Resilience through mixed training





Resilience through reweighting