Deep-Learning Jets with Uncertainties and More

Michel Luchmann

ITP, Heidelberg

May 20, 2019

based on: Deep-Learning Jets with Uncertainties and More (2019)
Sven Bollweg, Manuel Haussmann, Gregor Kasieczka, M. L., Tilman Plehn, Jennifer Thompson [arXiv:1904.10004 [hep-ph]]

Introduction

- modern Machine Learning (Deep Learning) successfully applied to LHC physics such as jet tagging[1]

  $\Rightarrow$ improved discrimination power compared to multi-variate analyses of high-level observables

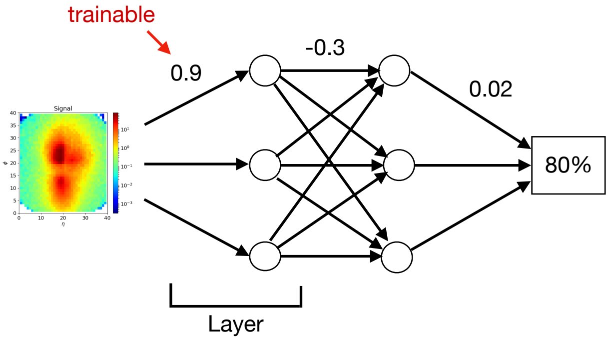
- what about uncertainties?

  $\Rightarrow$ Bayesian neural networks provide framework for capturing uncertainties

[1] G. Kasieczka et al, The Machine Learning Landscape of Top Taggers (2019) [arXiv:1902.09914 [hep-ph]]

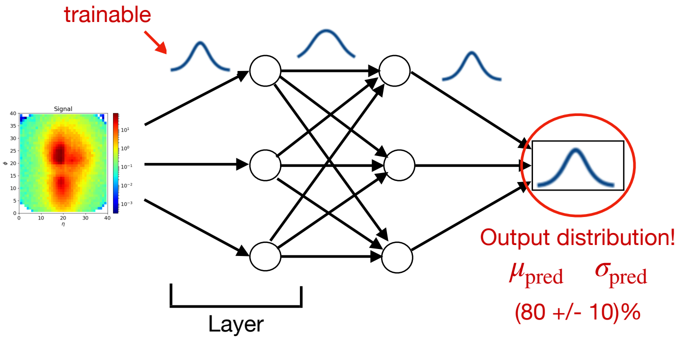
Neural networks

- neural networks can approximate arbitrary functions
- minimizing loss function $\Leftrightarrow$ maximum likelihood fit



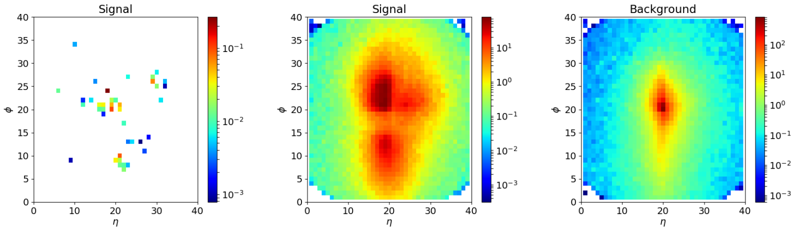
Bayesian neural networks

- replace maximum likelihood fit with Bayesian treatment
  $\Rightarrow$ distributions over weights



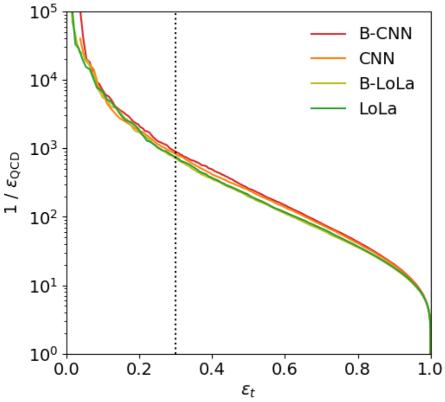
Top tagging

- MC data of pp-collisions at 14TeV
- signal: boosted top jets: $t \to qq'b \to$ fat jet
- background: QCD dijets
- input: low level observables (jet images or 4-momenta of constituents)



G. Kasieczka et al., The Machine Learning Landscape of Top Taggers (2019) [arXiv:1902.09914 [hep-ph]]

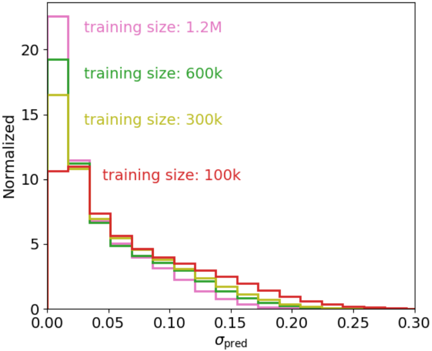
Performance



$\Longrightarrow$ Bayesian neural networks reach the same performance

Statistical uncertainties

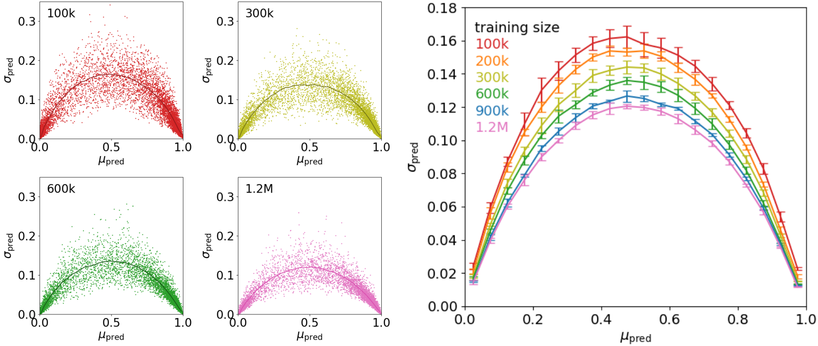
- uncertainty from finite training data
- train network several times with different training size
- histogram of σ_{pred} for 200k top jets (test sample)



Statistical uncertainties

classification: constrained output interval [0, 1]

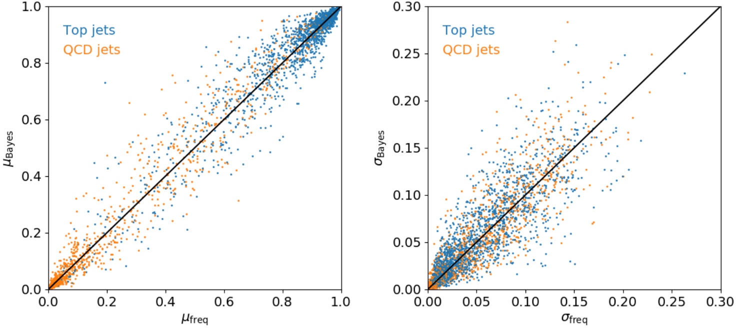
- correlation between μ_{pred} and σ_{pred}
- full picture in $[\mu_{\text{pred}}, \sigma_{\text{pred}}]$-plane



$\Longrightarrow$ BNNs capture uncertainty from finite training data!

Comparison to frequentist approach

- train deterministic network N times on statistically independent data
  $\Rightarrow$ N trained neural networks
- calculating μ_{pred} and σ_{pred} from N predictions

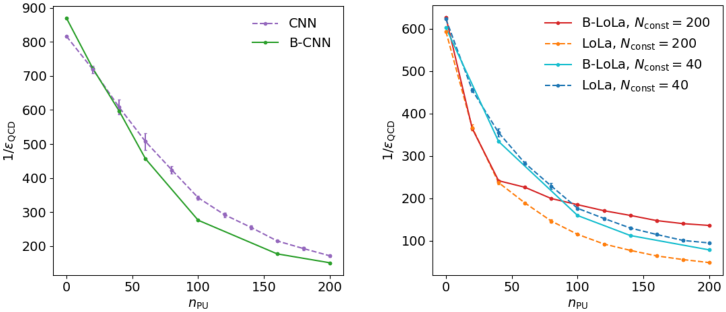


Systematic uncertainties

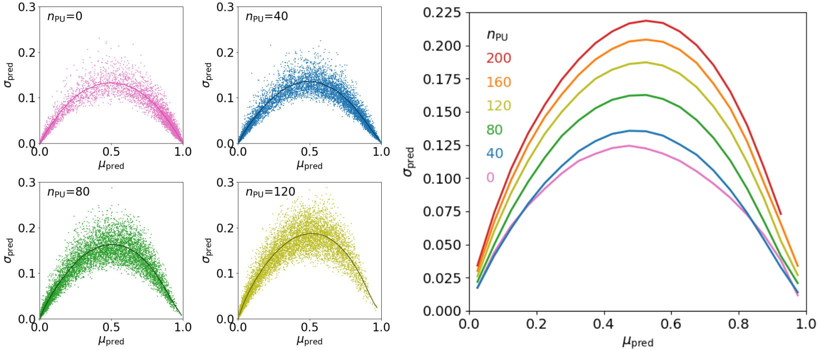
- introducing systematic uncertainty in test sample:
  - jet energy smearing ($\rightarrow$ paper)
  - pile up (technically not sys. unc.)
- sys. uncertainty not included in training sample
- pile up : several interactions per bunch crossing

Pile up: Impact on performance

$1/\epsilon_{\mathrm{QCD}}$ for reference efficiency $\epsilon_t = 0.3$

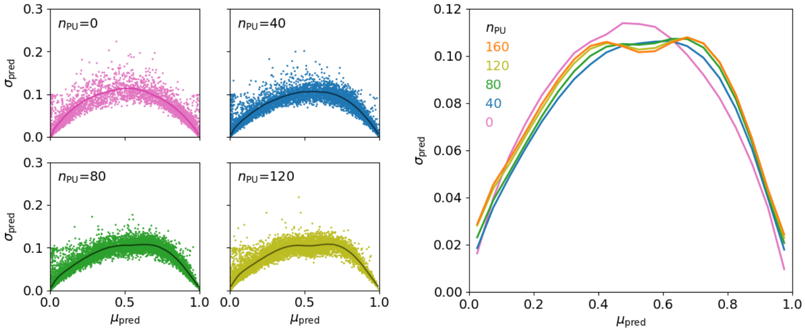


Pile up: LoLa



$n_{\rm PU}$=0
$n_{\rm PU}$=40
$n_{\rm PU}$=80
$n_{\rm PU}$=120
$\sigma_{\rm pred}$
$\mu_{\rm pred}$
$n_{\rm PU}$
200
160
120
80
40
0

Pile up: CNN



$\Longrightarrow$ no increased standard deviation, but correlation curve gives us hint about instability!

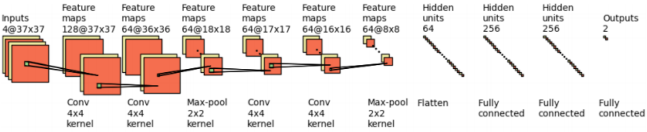
Summary

Bayesian neural networks:

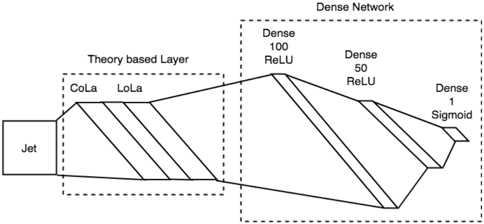
- reach same performance
- capture uncertainty from limited data (statistical uncertainty)
- give us additional information about stability (pile up)
- frequentist interpretation possible
- require more time for training and testing compared to single deterministic tagger

Architectures:

- CNN[2] (Convolutional neural network): Image based



- LoLa (Lorentz Layer)[3] 4-momenta based



[2] S. Macaluso and D. Shih, JHEP 1810, 121 (2018) doi:10.1007/JHEP10(2018)121 [arXiv:1803.00107 [hep-ph]].

[3] Anja Butter et al, SciPost Phys. 5, 028 (2018), DOI: 10.21468/SciPostPhys.5.3.028, [arXiv:1707.08966 [hep-ph]]

Dataset and pile up

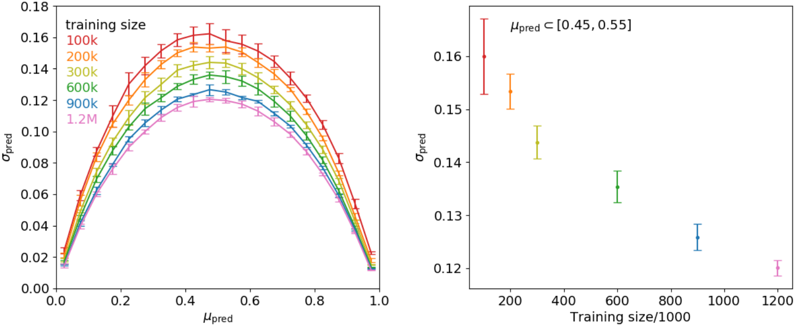
dataset:

- 14tev, hadronic tops, qcd diets with pythia
- Delphes Atlas card for detector simulation
- anti-kT 0.8 jets in pT range [550,650]
- $|\eta_{\mathrm{jet}}| < 2$
- available: here

pile up events:

- min-bias events generated with pythia
- added to signal and background jets and reclustered

Statistical uncertainty



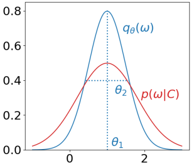
Bayesian neural networks

- Bayes theorem:

$$p(\omega|C) = \frac{p(C|\omega)\,p(\omega)}{p(C)} \quad \Rightarrow \quad p(c^*|C) = \int d\omega\, p(c^*|\omega)\, p(\omega|C)$$

- variational interference: $p(\omega|C) \approx q_\theta(\omega)$ (Gauss)

  $\longrightarrow$ minimize $\mathrm{KL}(q_\theta(\omega), p(\omega|C))$



- approximate via Monte Carlo:

$$\mu_{\text{pred}} = \frac{1}{N}\sum_{i=1}^{N} p(c^*|\omega_i) \qquad \sigma^2_{\text{pred}} = \frac{1}{N}\sum_{i=1}^{N} \left(p(c^*|\omega_i) - \mu_{pred}\right)^2$$

$\omega_i \in q_\theta(\omega)$
N: number of Monte Carlo samples