Jet Classification

Tilman Plehn

Universität Heidelberg

Nagoya 2/2020

Physics story: Nothing is ever new

LHC visionaries

– 1991: NN-based quark-gluon tagger [visionary: Lönnblad, Peterson, Rögnvaldsson]

USING NEURAL NETWORKS TO IDENTIFY JETS

Leif LÖNNBLAD*, Carsten PETERSON and Thorsteinn RÖGNVALDSSON*****

Department of Theoretical Physics, University of Lund, Sölvegatan 14A, S-22362 Lund, Sweden

Received 29 June 1990

A neural network method for identifying the ancestor of a hadron jet is presented. The idea is to find an efficient mapping between certain observed hadronic kinematical variables and the quark-gluon identity. This is done with a neuronic expansion in terms of a network of sigmoidal functions using a gradient descent procedure, where the errors are back-propagated through the network. With this method we are able to separate gluon from quark jets originating from Monte Carlo generated e^+e^- events with $\sim 85\%$ approach. The result is independent of the MC model used. This approach for isolating the gluon jet is then used to study the so-called string effect.

In addition, heavy quarks (b and c) in e^+e^- reactions can be identified on the 50% level by just observing the hadrons. In particular we are able to separate b-quarks with an efficiency and purity, which is comparable with what is expected from vertex detectors. We also speculate on how the neural network method can be used to disentangle different hadronization schemes by compressing the dimensionality of the state space of hadrons.

Physics story: Nothing is ever new

LHC visionaries

– 1991: NN-based quark-gluon tagger [visionary: Lönnblad, Peterson, Rögnvaldsson]

– 1994: jet-algorithm W/top-tagger [Seymour]

USING NEURAL NETWORKS TO IDENTIFY JETS

Leif LÖNNBLAD*, Carsten PETERSON** and Thorsteinn RÖGNVALDSSON***

Department of Theoretical Physics, University of Lund, Sölvegatan 14A, S-22362 Lund, Sweden

Received 29 June 1990

A neural network method for identifying the ancestor of a hadron jet is presented. The idea is to find an efficient mapping between certain observed hadronic kinematical variables and the quark-gluon identity. This is done with a neuronic expansion in terms of a network of sigmoidal functions using a gradient descen
network. With this method we ar
Carlo generated e^+e^- events w
model used. This approach for i
effect.

In addition, heavy quarks (b a
just observing the hadrons. In pa
purity, which is comparable with
how the neural network method
compressing the dimensionality c

Searches for new particles using cone and cluster jet algorithms: a comparative study

Michael H. Seymour

Department of Theoretical Physics, University of Lund, Sölvegatan 14A, S-22362 Lund, Sweden

Received 18 June 1993; in revised form 16 September 1993

Abstract. We discuss the reconstruction of the hadronic decays of heavy particles using jet algorithms. The ability to reconstruct the mass of the decaying particle is compared between a traditional cone-type algorithm and a recently proposed cluster-type algorithm. The specific examples considered are the semileptonic decays of a heavy Higgs boson at $\sqrt{s}=16$ TeV, and of top quark–antiquark pairs at $\sqrt{s}=1.8$ TeV. We find that the cluster algorithm offers considerable advantages in the former case, and a slight advantage in the latter. We briefly discuss the effects of calorimeter energy resolution, and show that a typical resolution dilutes these advantages, but does not remove them entirely.

except that the invariant mass of a pair is replaced by the transverse momentum of the softer particle relative to the other.

More recently, this algorithm was extended to collisions with incoming hadrons [5], and a longitudinally-invariant $k_\perp$-clustering algorithm for hadron-hadron collisions was proposed [6]. This algorithm has been compared with the more commonly used cone algorithm from the viewpoints of a parton-shower Monte Carlo program [6, 7], and a fixed-order matrix-element calculation [8], and advantages of the cluster algorithm were reported in both cases. This paper is concerned with a comparison between the algorithms for the task of reconstructing the hadronic decays of heavy particles, which was also studied in a preliminary way in [9].

The only as-yet unobserved particles of the minimal Standard Model are the top quark and Higgs boson. The search for, and study of, these particles are among the most important goals of current and planned hadron-

Fat jet taggers

Look what makes jets [Pre-LHC, jets were just annoying]

– top jets from $t \to bq\bar{q}'$ vs QCD jets

– top decays well-defined in theory

– labelled sample: semileptonic $t\bar{t}$ events

$\Rightarrow$ Fat jets as LHC physics playground

Simple top tagging [ask Michi]

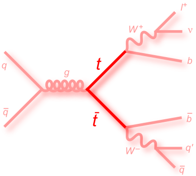
1– fat jet with $p_T > 200$ GeV

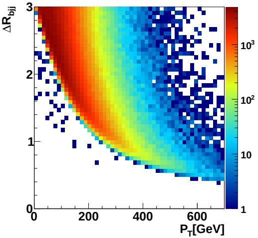
2– filtering defining 3-5 decay jets

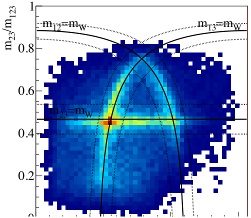
3– top mass window $m_{123} = [150, 200]$ GeV

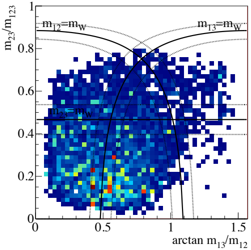
4– mass plane cuts extracting $m_{ij} \approx m_W$

$\Rightarrow$ Not rocket science, but crucial to build trust









Multi-variate taggers

Developing the benchmark

– multivariate analysis generally old news
  multivariate tagger to keep up with shower deconstruction [Soper, Spannowsky]

– optimal fat jet size R_opt [large to decay jets, small to avoid combinatorics, compute from kinematics]

$$|m_{123} - m_{123}^{(R_\text{max})}| < 0.2\, m_{123}^{(R_\text{max})} \quad \Rightarrow \quad R_\text{opt}$$

– add N-subjettiness [Thaler, van Tilburg]

– $\{m_{123}, f_W, R_\text{opt} - R_\text{opt}^\text{(calc)}, \tau_j, \tau_j^\text{(filt)}\}$

$\Rightarrow$ Theory all but precision

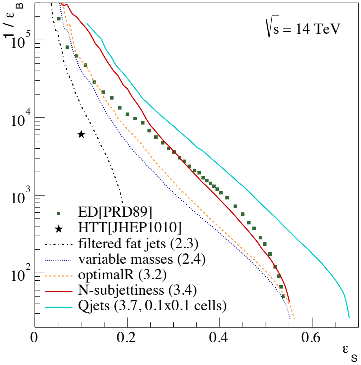
Fat jet and top kinematics

– jet radiation major problem for Z' search

– tag and reconstruction in each other's way

– $\{..., m_{tt}, p_{T,t}, m_{jj}^\text{(filt)}, p_{T,j}^\text{(filt)}\}$

$\Rightarrow$ Best we can do?



Jet image machines

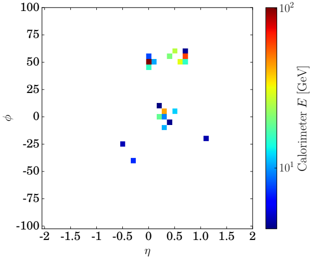
Natural next step [Cogan etal, Oliveira, Nachman etal, Baldi, Whiteson etal (2014/15)]

– why intermediate high-level variables?

– as much data as possible

– calorimeter output as image

⇒ Deep learning = modern networks on low-level observables



Jet image machines

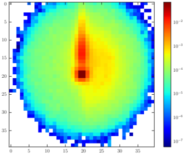
Natural next step [Cogan etal, Oliveira, Nachman etal, Baldi, Whiteson etal (2014/15)]

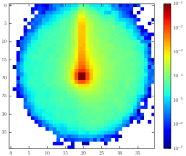
– why intermediate high-level variables?

– as much data as possible

– calorimeter output as image

$\Rightarrow$ Deep learning = modern networks on low-level observables





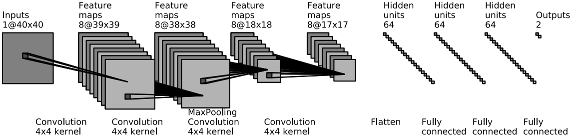
Convolutional network [Kasieczka, TP, Russell, Schell; Macaluso, Shih]

– image recognition standard ML task

– rapidity vs azimuthal angle, colored by energy deposition

– top tagging on 2D jet images

– 40×40 bins through calorimeter resolution

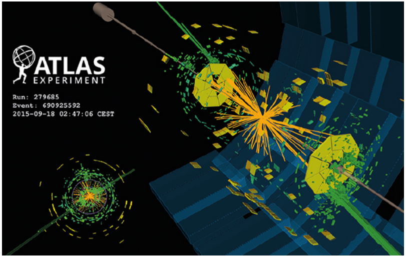


Why LHC? Why jets?

Data from ATLAS & CMS

- most LHC interactions $q\bar{q}, gg \to q\bar{q}, gg$
- quarks/gluon visible as jets $\quad \sigma_{pp\to jj} \times \mathcal{L} \approx 10^8\text{fb} \times 80/\text{fb} \approx 10^{10}$ events

$\Rightarrow$ It's big data



Why LHC? Why jets?

Data from ATLAS & CMS

- most LHC interactions $q\bar{q}, gg \to q\bar{q}, gg$
- quarks/gluon visible as jets $\quad \sigma_{pp\to jj} \times \mathcal{L} \approx 10^8 \text{fb} \times 80/\text{fb} \approx 10^{10}$ events

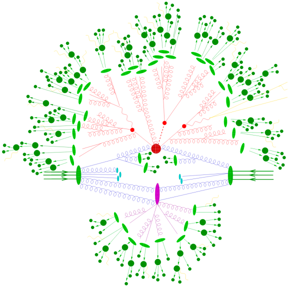
$\Rightarrow$ It's big data

Physics in jets

- re-summed perturbative QFT prediction from QCD
- jets as decay products

  67% $W \to jj$ $\quad$ 70% $Z \to jj$ $\quad$ 60% $H \to jj$ $\quad$ 67% $t \to jjj$ $\quad$ 60% $\tau \to j$ ...
- new physics in ‘dark showers’

$\Rightarrow$ It's fundamentally interesting



Why LHC? Why jets?

Data from ATLAS & CMS

– most LHC interactions $q\bar{q}, gg \to q\bar{q}, gg$

– quarks/gluon visible as jets $\quad \sigma_{pp \to jj} \times \mathcal{L} \approx 10^8 \text{fb} \times 80/\text{fb} \approx 10^{10}$ events

$\Rightarrow$ It's big data

Physics in jets

– re-summed perturbative QFT prediction from QCD

– jets as decay products

67% $W \to jj$ 70% $Z \to jj$ 60% $H \to jj$ 67% $t \to jjj$ 60% $\tau \to j$ ...

– new physics in 'dark showers'

$\Rightarrow$ It's fundamentally interesting

Monte Carlo data

– QCD simulation: Sherpa, Pythia, Herwig [Madgraph]

– fast detector simulation: Delphes

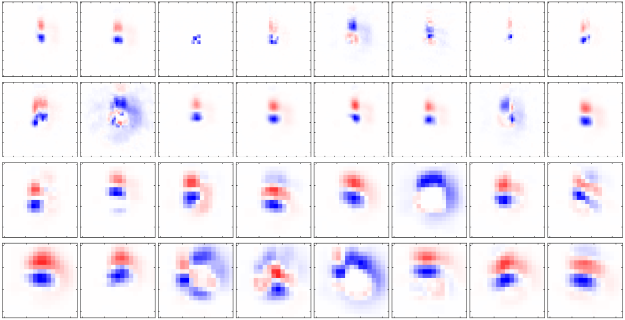
– data-to-data comparison: MC vs LHC

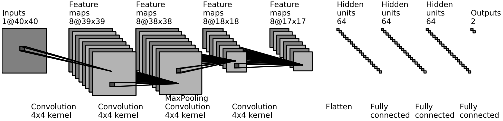
$\Rightarrow$ We can simulate it

Inside DeepTop

Particle physicists as 'users' [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

– 2+2 convolutional layers

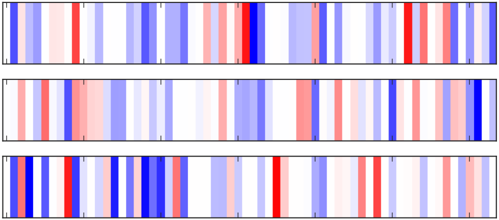


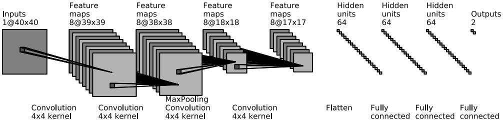


Inside DeepTop

Particle physicists as ‘users’ [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

- 2+2 convolutional layers
- 3 fully connected layers





Inside DeepTop

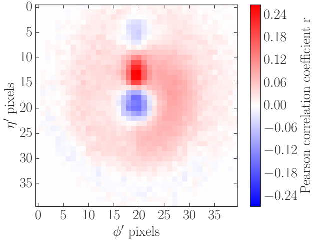
Particle physicists as 'users' [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

– 2+2 convolutional layers

– 3 fully connected layers

– Pearson input-output correlation [pixel x vs label y]

$$r_{ij} \approx \sum_{\text{images}} \left(x_{ij} - \bar{x}_{ij} \right) \left(y - \bar{y} \right)$$



Inside DeepTop

Particle physicists as ‘users’ [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

– 2+2 convolutional layers

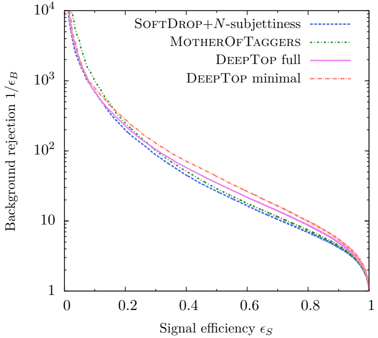
– 3 fully connected layers

– Pearson input-output correlation [pixel x vs label y]

$$r_{ij} \approx \sum_{\text{images}} \left(x_{ij} - \bar{x}_{ij}\right) \left(y - \bar{y}\right)$$

– comparison to MotherOfTaggers BDT

⇒ Understandable performance gain



Inside DeepTop

Particle physicists as ‘users’ [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

– 2+2 convolutional layers

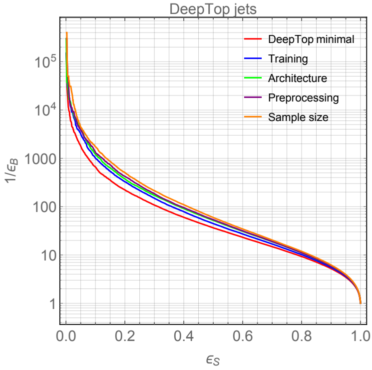
– 3 fully connected layers

– Pearson input-output correlation [pixel x vs label y]

$$r_{ij} \approx \sum_{\text{images}} \left(x_{ij} - \bar{x}_{ij}\right)\left(y - \bar{y}\right)$$

– comparison to MotherOfTaggers BDT

$\Rightarrow$ Understandable performance gain



Inside DeepTop

Particle physicists as ‘users’ [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

– 2+2 convolutional layers

– 3 fully connected layers

– Pearson input-output correlation [pixel x vs label y]

$$r_{ij} \approx \sum_{\text{images}} \left(x_{ij} - \bar{x}_{ij} \right) \left(y - \bar{y} \right)$$

⇒ Understandable performance gain

Typical reaction: ‘F*** you, you f***ing machine’

– full control for supervised learning
easy checks for correctly identified signal/background

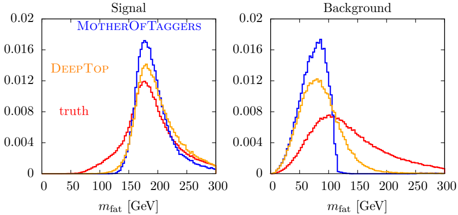
– MC truth vs MotherOfTaggers vs DeepTop

fat jet mass

N-subjettiness

transverse momenta

⇒ The box is not black



Theory inspiration

4-vector input — graph CNN [Butter, Kasieczka, TP, Russell; much better versions by now]

– physics objects from calorimeter and tracker

– distance measure known from e&m [alternatively: Erdmann, Rath, Rieger]

Inspired by QFT

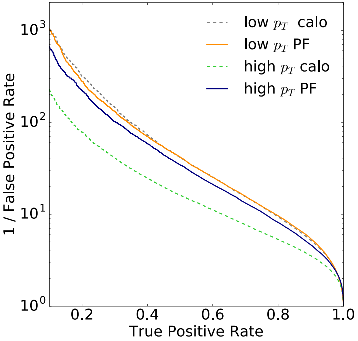
– input 4-vectors $(k_{\mu,i})$

– jet algorithm $\longrightarrow$ combination layer

$$k_{\mu,i} \xrightarrow{\text{CoLa}} \widetilde{k}_{\mu,j} = k_{\mu,i}\, C_{ij}$$

– observables $\longrightarrow$ Lorentz layer

$$\tilde{k}_j \xrightarrow{\text{LoLa}} \hat{k}_j = \begin{pmatrix} m^2(\tilde{k}_j) \\ p_T(\tilde{k}_j) \\ \vdots \end{pmatrix}$$



$\Rightarrow$ **Learn Minkowski metric**

g =diag(0.99$\pm$0.02, -1.01 ± 0.01, -1.01 ± 0.02, -0.99 ± 0.02)

Jet classification done

SciPost Physics Submission

The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], B. M. Dillon[5], M. Fairbairn[6], D. A. Faroughy[5], W. Fedorko[7], C. Gay[7], L. Gouskos[8], J. F. Kamenik[5,9], P. T. Komiske[10], S. Leiss[1], A. Lister[7], S. Macaluso[3,4], E. M. Metodiev[10], L. Moore[11], B. Nachman,[12,13], K. Nordström[14,15], J. Pearkes[7], H. Qu[8], Y. Rath[16], M. Rieger[16], D. Shih[4], J. M. Thompson[2], and S. Varma[6]

1 Institut für Experimentalphysik, Universität Hamburg, Germany
2 Institut für Theoretische Physik, Universität Heidelberg, Germany
3 Center for Cosmology and Particle Physics and Center for Data Science, NYU, USA
4 NHECT, Dept. of Physics and Astronomy, Rutgers, The State University of NJ, USA
5 Jozef Stefan Institute, Ljubljana, Slovenia
6 Theoretical Particle Physics and Cosmology, King's College London, United Kingdom
7 Department of Physics and Astronomy, The University of British Columbia, Canada
8 Department of Physics, University of California, Santa Barbara, USA
9 Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia
10 Center for Theoretical Physics, MIT, Cambridge, USA
11 CP3, Universitéxx Catholique de Louvain, Louvain-la-Neuve, Belgium
12 Physics Division, Lawrence Berkeley National Laboratory, Berkeley, USA
13 Simons Inst. for the Theory of Computing, University of California, Berkeley, USA
14 National Institute for Subatomic Physics (NIKHEF), Amsterdam, Netherlands
15 LPTHE, CNRS & Sorbonne Université, Paris, France
16 III. Physics Institute A, RWTH Aachen University, Germany

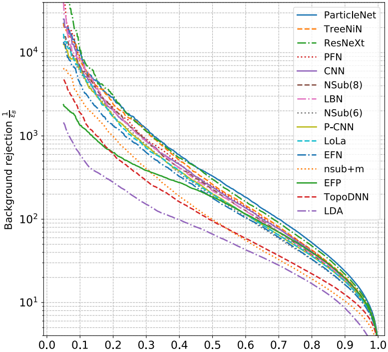
gregor.kasieczka@uni-hamburg.de
plehn@uni-heidelberg.de

July 24, 2019

Abstract

Based on the established task of identifying boosted, hadronically decaying top quarks, we compare a wide range of modern machine learning approaches. Unlike most established methods they rely on low-level input, for instance calorimeter output. While their network architectures are vastly different, their performance is comparatively similar. In general, we find that these new approaches are extremely powerful and great fun.

Content



– many networks successful

⇒ Which one to pick?

When reality hits

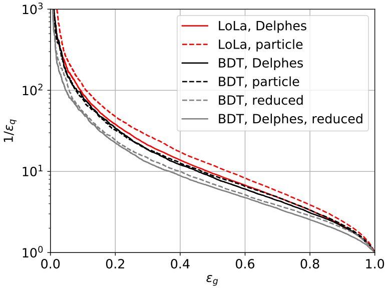
ML-Life is not always nice to us [Kasieczka, Kiefer, TP, Thompson]

– quark-gluon tagging a problem since 1991

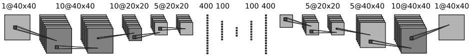
– quark jets typical for resonance searches
gluon jets typical as dark matter recoil
...

– BDT/NN on high-level variables established

$\Rightarrow$ deep-learning advantage gone after detector simulation, REALLY???



Learning background only



Fully supervised classification boring

[Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih; David's talk]

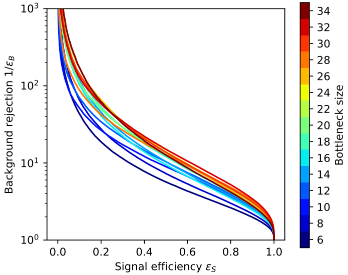
– anomaly searches, only training on 'background'

– established ML concept: autoencoder

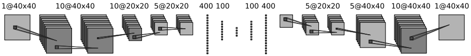
– reconstruct typical QCD jet image from many QCD jets
reduce weights in central layer, compress information to 'typical'

– outliers hard to describe, (hopefully) non-QCD less compressible

⇒ Making an okay tagger



Learning background only



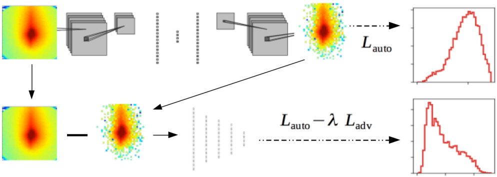
Fully supervised classification boring [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih; David's talk]

- anomaly searches, only training on 'background'
- established ML concept: autoencoder
- reconstruct typical QCD jet image from many QCD jets
  reduce weights in central layer, compress information to 'typical'
- outliers hard to describe, (hopefully) non-QCD less compressible

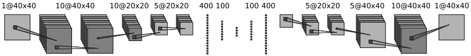
⇒ Making an okay tagger

De-correlate background shaping, define side bands

- established concept: adversary [Shimmin,...]



Learning background only



Fully supervised classification boring [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih; David's talk]

– anomaly searches, only training on 'background'

– established ML concept: autoencoder

– reconstruct typical QCD jet image from many QCD jets
reduce weights in central layer, compress information to 'typical'

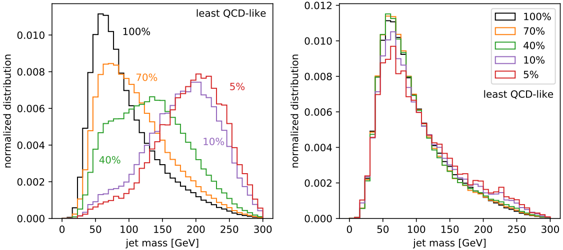
– outliers hard to describe, (hopefully) non-QCD less compressible

⇒ Making an okay tagger

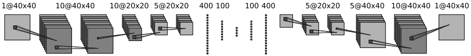
De-correlate background shaping, define side bands

– established concept: adversary [Shimmin,...]

– atypical QCD jets typially with large jet mass
remove jet mass from network training



Learning background only



Fully supervised classification boring [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih; David's talk]

– anomaly searches, only training on ‘background’

– established ML concept: autoencoder

– reconstruct typical QCD jet image from many QCD jets
reduce weights in central layer, compress information to ‘typical’

– outliers hard to describe, (hopefully) non-QCD less compressible

$\Rightarrow$ Making an okay tagger

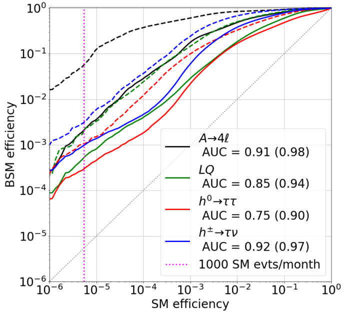
The whole thing on anomalous LHC events [Cerri, Nguyen, Pierini, Spiropulu, Vlimant]

– same thing on full events

– training data a problem

– variational autoencoder more powerful

$\Rightarrow$ Proof of concept...



Classification with error bars

Propagating uncertainties

- $(60\pm??)\%$ top, uncertainty from training
- probability for test event $p(c^*|C)$ [classifier output C, network ω]

$$p(c^*|C) = \int d\omega \; p(c^*|\omega, C) \; p(\omega|C) = \int d\omega \; p(c^*|\omega, C) \; q(\omega)$$

- for instance minimize Kullbeck-Leibler divergence [Bayes' theorem]

$$\begin{aligned} \mathrm{KL}[q(\omega), p(\omega|C)] &= \int d\omega \; q(\omega) \; \log \frac{q(\omega)}{p(\omega|C)} \\ &= \int d\omega \; q(\omega) \; \log \frac{q(\omega) p(C)}{p(C|\omega) p(\omega)} \\ &= \underbrace{\mathrm{KL}[q(\omega), p(\omega)]}_{\text{L2-regularization}} + \underbrace{\log p(C) \int d\omega \; q(\omega)}_{\text{normalization of } q\text{, irrelevant}} - \underbrace{\int d\omega \; q(\omega) \log p(C|\omega)}_{\text{likelihood, maximized}} \end{aligned}$$

- minimum condition [Gaussian $\omega = \{\mu, \sigma\}$]

$$\frac{\partial}{\partial \omega} \int d\omega \; q(\omega) \log p(C|\omega) = 0$$

- sample in ω to extract $(\mu_{\text{pred}}, \sigma_{\text{pred}})$ jet by jet...

  ...and check prior dependence [Gaussian, 5 orders in width]

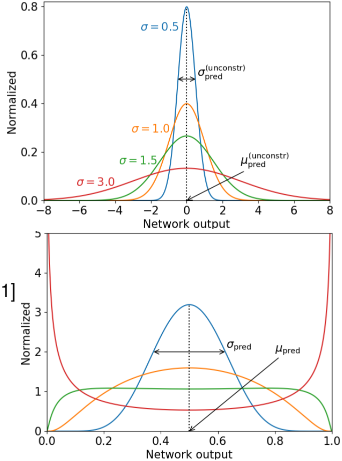
Classification with error bars

Propagating uncertainties

- $(60\pm??)$% top, uncertainty from training
- probability for test event $p(c^*|C)$ [classifier output C, network ω]

$$p(c^*|C) = \int d\omega\; p(c^*|\omega, C)\; p(\omega|C) = \int d\omega\; p(c^*|\omega, C)\; q(\omega)$$

- sample in ω to extract $(\mu_\text{pred}, \sigma_\text{pred})$ jet by jet...



Complication with classification

- sigmoid to map on closed interval [0, 1]

$$\text{Sigmoid}(x) = \frac{e^x}{1+e^x}$$

- predictive mean

$$\mu_\text{pred} = \int_{-\infty}^{\infty} d\omega\; \text{Sigmoid}(\omega)\; G_{\mu,\sigma}(\omega)$$
$$= \int_0^1 dx\; \frac{x}{x(1-x)}\; G_{\mu,\sigma}\left(\log\frac{x}{1-x}\right) \in [0,1]$$

- predictive standard deviation

$$\sigma_\text{pred} \approx \mu_\text{pred}\left(1-\mu_\text{pred}\right)\; \sigma_\text{pred}^\text{(unconstr)}$$

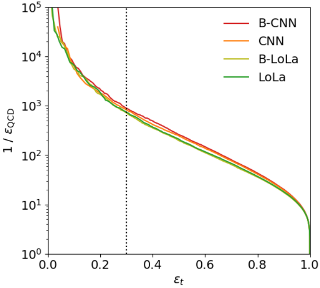
$\Rightarrow$ Additional complication...

Statistics & systematics

Training statistics [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson; Ben's talk]

– Bayesian version of DeepTop and LoLa

– similar performance as deterministic network
training time somewhat increased



Statistics & systematics

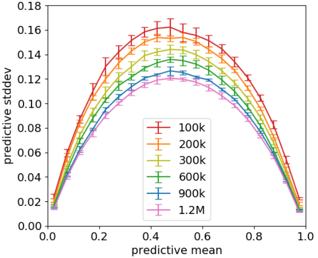
Training statistics [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson; Ben's talk]

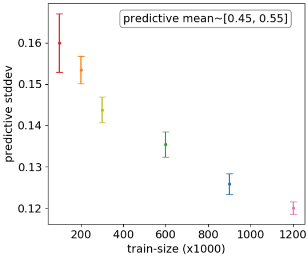
– Bayesian version of DeepTop and LoLa

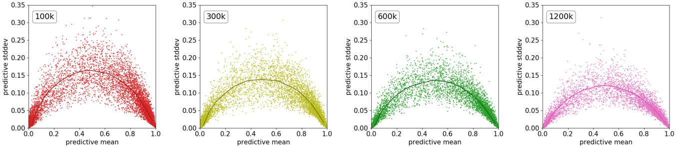
– similar performance as deterministic network
training time somewhat increased

– correlation between μ_{pred} and σ_{pred} [toy network, 10k jets]

– increasing training statistics [parabola from closed interval output]







Statistics & systematics

Training statistics [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson; Ben's talk]

– Bayesian version of DeepTop and LoLa

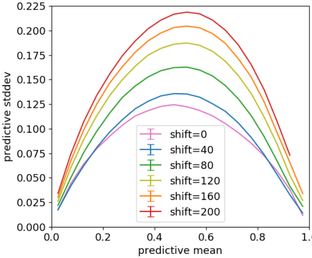
– similar performance as deterministic network
training time somewhat increased

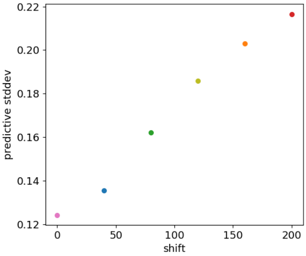
– correlation between μ_{pred} and σ_{pred} [toy network, 10k jets]

– increasing training statistics [parabola from closed interval output]

Noise/pile-up

– increasing pile-up, stable [LoLa, ordered constituents]





Statistics & systematics

Training statistics [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson; Ben's talk]

– Bayesian version of DeepTop and LoLa

– similar performance as deterministic network
  training time somewhat increased

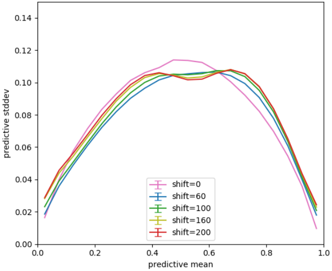
– correlation between μ_pred and σ_pred [toy network, 10k jets]

– increasing training statistics [parabola from closed interval output]

Noise/pile-up

– increasing pile-up, stable [LoLa, ordered constituents]

– increasing pile-up, unstable [DeepTop, jet image]



Statistics & systematics

Training statistics [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson; Ben's talk]

– Bayesian version of DeepTop and LoLa

– similar performance as deterministic network
  training time somewhat increased

– correlation between μ_{pred} and σ_{pred} [toy network, 10k jets]

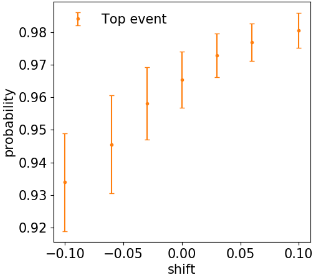
– increasing training statistics [parabola from closed interval output]

Jet energy scale

– systematics effect in test sample

1– shift of hardest constituent

– adversarial example: hierarchical subjets = top



Statistics & systematics

Training statistics [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson; Ben's talk]

– Bayesian version of DeepTop and LoLa

– similar performance as deterministic network
training time somewhat increased

– correlation between μ_{pred} and σ_{pred} [toy network, 10k jets]

– increasing training statistics [parabola from closed interval output]

Jet energy scale

– systematics effect in test sample

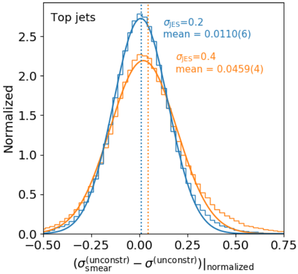
1– shift of hardest constituent

– adversarial example: hierarchical subjets = top

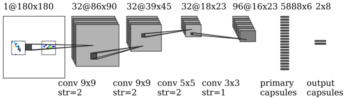
2– uncorrelated shift of all constituents

– tiny degradation for signal

⇒ More studies needed



Capsules vs CNN



Full events instead of fat jet

– sparse events with sparse objects

– training an open problem

– multi-label for different backgrounds

$\Rightarrow$ Need to go beyond CNN

Capsule networks [Diefenbacher, Frost, Kasieczka, TP, Thompson]

– vector output instead of scalar classification

– agreement by parallel vectors in feature space

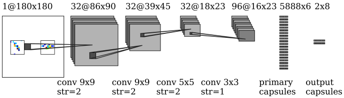
– new squashing prescription

$$v \to \frac{\vec{v}^2}{1+\vec{v}^2}\,\hat{v} \neq \frac{|\vec{v}|}{\sqrt{1+|\vec{v}|^2}}\,\hat{v}$$

– pooling vs stride convolutions?

$\Rightarrow$ properties and geometry in vector entries [eyes, nose, mouth]

Capsules vs CNN



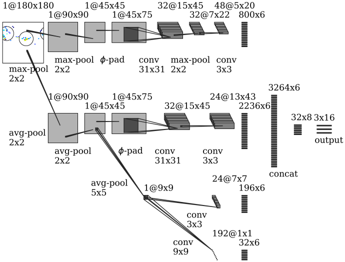
Full events instead of fat jet

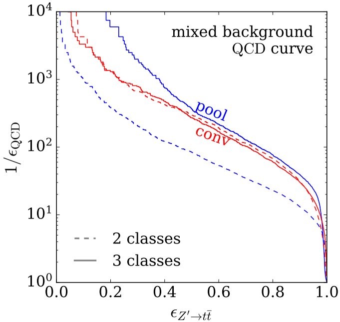
- sparse events with sparse objects
- training an open problem
- multi-label for different backgrounds

$\Rightarrow$ Need to go beyond CNN

$Z' \to t\bar{t}$ resonance

- subjet-level: jj background [conv setup]
- event-level: $t\bar{t}$ continuum [pool w/ 3 classes]
- still not perfect in $t\bar{t}$ continuum rejection
- next step $t\bar{t}H$...

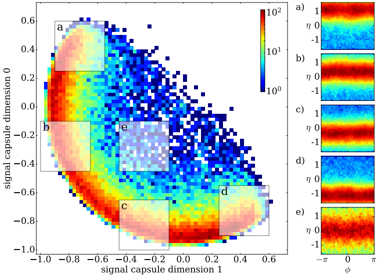




Organizing information

2D toy network for $Z' \to t\bar{t}$

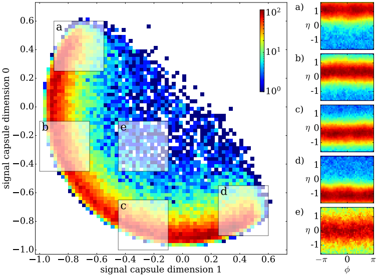
- signal capsule/events
- classification through radius
- azimuthal angle to organize information
- jet rapidity the key



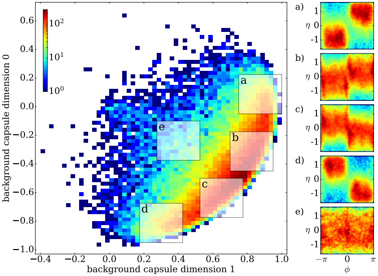
Organizing information

2D toy network for $Z' \to t\bar{t}$

- signal capsule/events
- classification through radius
- azimuthal angle to organize information
- jet rapidity the key



- background capsule/events
- back-to-back topology



The future

Machine learning is an amazing tool box...

...LHC physics really is big data

...imagine recognition is a starting point

...deep learning is not just classification

...jets are not the only interesting objects at LHC

...Bayesian networks are extremely likable

...capsule networks useful for full events

Let's find some really cool applications!