From black box to tool box: classification

Tilman Plehn

Universität Heidelberg

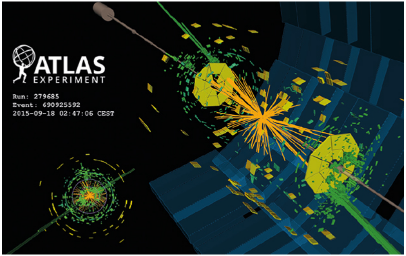
GridKa — The Art of Data 8/2019

Why LHC? Why jets?

Data from ATLAS & CMS

- colliding protons on protons at $E \approx 13000 \times m_p$
- most interactions $q\bar{q}, gg \to q\bar{q}, gg$
- quarks/gluon visible as jets $\sigma_{pp\to jj} \times \mathcal{L} \approx 10^8\text{fb} \times 80/\text{fb} \approx 10^{10}$ events

$\Rightarrow$ It's proper big data



Why LHC? Why jets?

Data from ATLAS & CMS

– colliding protons on protons at $E \approx 13000 \times m_p$

– most 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 proper big data

Physics in jets

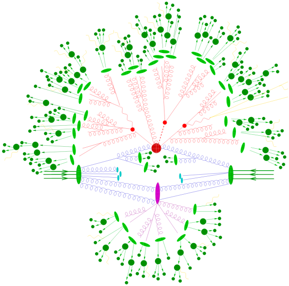
– re-summed perturbative QFT prediction from QCD

– 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 interesting



Why LHC? Why jets?

Data from ATLAS & CMS

– colliding protons on protons at $E \approx 13000 \times m_p$

– most 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 proper 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'

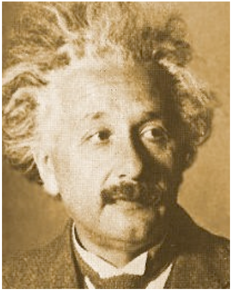
$\Rightarrow$ It's interesting

Monte Carlo data

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

– data-to-data comparison: MC vs LHC

$\Rightarrow$ We can simulate it



Why not LHC?

ATLAS & CMS

– 3000 know-it-alls per experiment

– strong top-down structures

– strongly organized analysis groups

⇒ Small groups driving innovation

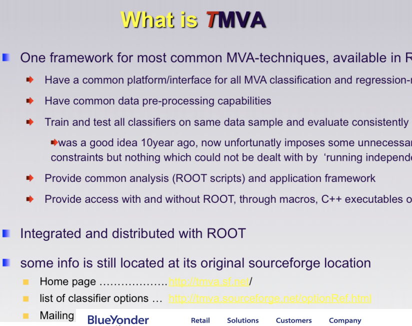
Expertize

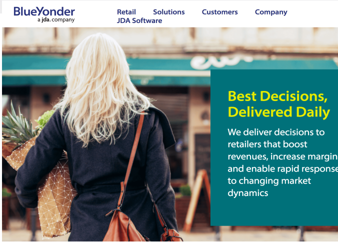
– LHC data format: ROOT

– multi-variate analyses tool: TMVA

– Tensorflow from TMVA/ROOT

⇒ Limited sense of ML-urgency





Experiment vs theory

– theorists linked to lack of team compatibility

– simulated data as good as actual data

– excellent personal ex-th connections

⇒ Theory driving developments we should not...

Jets classification: Nothing is ever new

LHC visionaries

– 1991: NN-based quark-gluon tagger [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.

Jets classification: Nothing is ever new

LHC visionaries

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

– 1994: jet algorithm for W, top... [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 fo
is to find an efficient mapping be
quark-gluon identity. This is don
functions using a gradient descen
network. With this method we ar
Carlo generated e^+e^- events w
model used. This approach for i
effect.

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

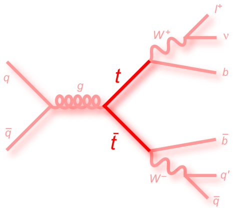
Abstract. We discuss the reconstruction of the hadronic decays of heavy particles using jet algorithms. The ability to reconstruct the mass of the decaying particle is compared between a traditional cone-type algorithm and a recently proposed cluster-type algorithm. The specific examples considered are 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-hadron collider experiments. In both cases, which decay

Benchmark: fat jets (2000s)



Look inside jets [Pre-LHC, jet 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

Simple top tagging [BDRS; TP, Salam, Spannowsky, Takeuchi]

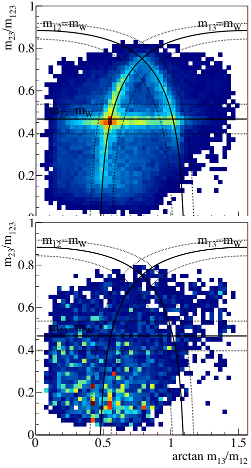
1– fat jet with $p_T > 200$ GeV

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 (2010s)

Developing the benchmark

- multivariate analysis generally old news
- optimal fat jet size R_opt [large to decay jets, small to avoid combinatorics, compute from kinematics]

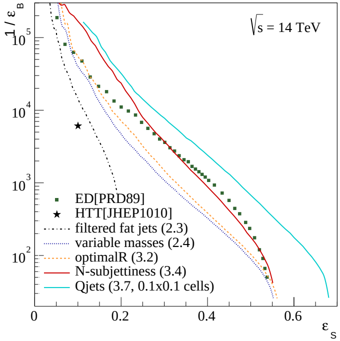
$$|m_{123} - m_{123}^{(R\text{max})}| < 0.2\, m_{123}^{(R\text{max})} \quad \Rightarrow \quad R_\text{opt}$$

- add N-subjettiness [Thaler, van Tilburg]
- $\{m_{123}, f_W, R_\text{opt} - R_\text{opt}^\text{(calc)}, \tau_j, \tau_j^\text{(filt)}\}$

Fat jet and top kinematics

- jet radiation major problem for Z' search
- combine top and fat jet information
- $\{\dots, m_{tt}, p_{T,t}, m_{jj}^\text{(filt)}, p_{T,j}^\text{(filt)}\}$

$\Rightarrow$ Performance increase, as expected



Jet images (2020s)

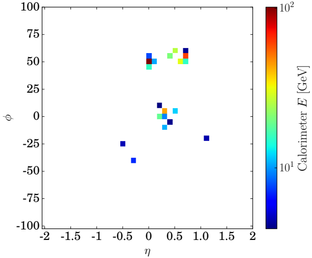
Natural next step [Cogan etal, Oliveira, Nachman etal, Baldi, Whiteson etal (2014/15)]

– why intermediate high-level variables?

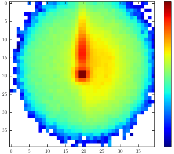
– calorimeter output as image

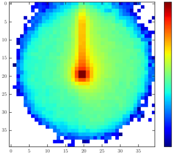
– learn theory through more NN layers

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



Jet images (2020s)





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

– why intermediate high-level variables?

– calorimeter output as image

– learn theory through more NN layers

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

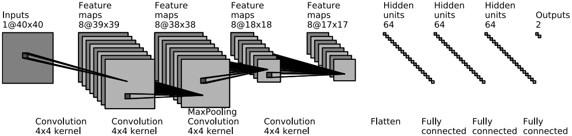
Convolutional network [Kasieczka, TP, Russell, Schell; Macaluso, Shih]

– image recognition standard ML task

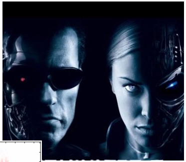
– rapidity vs azimuthal angle, colored by energy deposition

– top tagging on 2D jet images

– 40×40 bins through calorimeter resolution

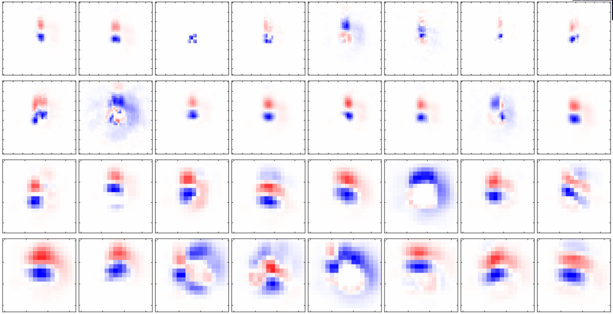


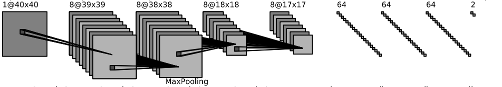
Inside DeepTop



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

– 2+2 convolutional layers



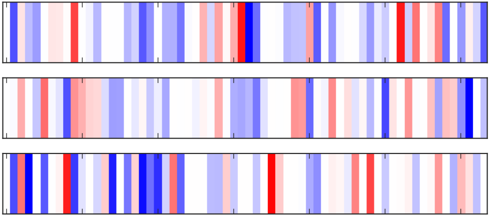


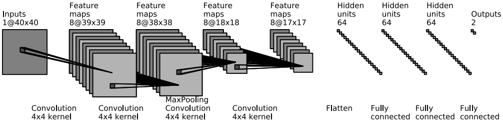
Inside DeepTop

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

– 2+2 convolutional layers

– 3 fully connected layers





Inside DeepTop

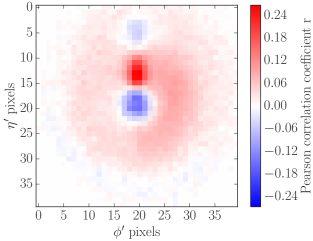
Particle physicists as users [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

– 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

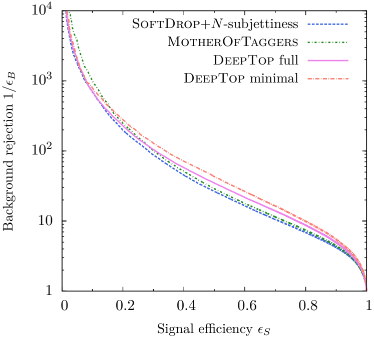
Particle physicists as users [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

- 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

$\Rightarrow$ Understandable performance gain



Inside DeepTop

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

– 2+2 convolutional layers

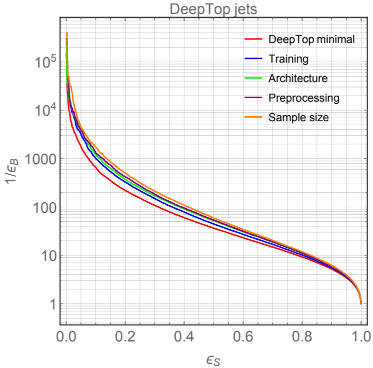
– 3 fully connected layers

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

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

– 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]

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

– comparison to MotherOfTaggers BDT

$\Rightarrow$ Understandable performance gain

Colleagues my age: ‘F*** you, you f***ing machine’

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

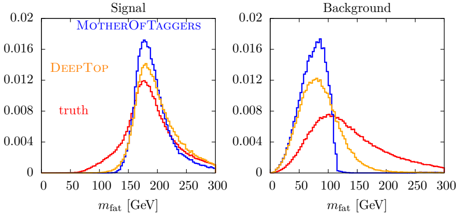
– MC truth vs MotherOfTaggers vs DeepTop

fat jet mass

N-subjettiness

transverse momenta

$\Rightarrow$ It works and we know why



Grand theory ideas

Networks with 4-vector input [Butter, Kasieczka, TP, Russell; many more by now]

– sparsely filled picture: graph CNN

– physics objects from calorimeter and tracker

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

Inspired by jet algorithm — combination layer

– input 4-vectors

$$(k_{\mu,i}) = \begin{pmatrix} k_{0,1} & k_{0,2} & \cdots & k_{0,N} \\ k_{1,1} & k_{1,2} & \cdots & k_{1,N} \\ k_{2,1} & k_{2,2} & \cdots & k_{2,N} \\ k_{3,1} & k_{3,2} & \cdots & k_{3,N} \end{pmatrix}$$

– on-shell conditions for top tag

– combined 4-vectors

$$k_{\mu,i} \xrightarrow{\text{CoLa}} \widetilde{k}_{\mu,j} = k_{\mu,i}\, C_{ij} \qquad C = \begin{pmatrix} 1 & 0 & \cdots & 0 & C_{1,N+2} & \cdots & C_{1,M} \\ 0 & 1 & & \vdots & C_{2,N+2} & \cdots & C_{2,M} \\ \vdots & \vdots & \ddots & 0 & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & C_{N,N+2} & \cdots & C_{N,M} \end{pmatrix}$$

$\Rightarrow$ Physics step, easy to interpret

Grand theory ideas

Networks with 4-vector input [Butter, Kasieczka, TP, Russell; many more by now]

- sparsely filled picture: graph CNN
- physics objects from calorimeter and tracker
- distance measure known from e&m [alternatively: Erdmann, Rath, Rieger]

Inspired by jet algorithm — combination layer

- input 4-vectors $(k_{\mu,i})$
- on-shell conditions for top tag
- combined 4-vectors $k_{\mu,i} \xrightarrow{\text{CoLa}} \widetilde{k}_{\mu,j} = k_{\mu,i}$

$\Rightarrow$ Physics step, easy to interpret

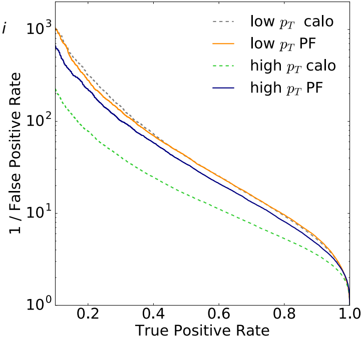
Inspired by Jackson — Lorentz layer

- DNN on Lorentz scalars

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

$\Rightarrow$ Learn Minkowski metric

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



Meet the professionals

Brief history of jet classification

– 2014/15: first jet image papers
– 2017: first (working) ML top tagger
– ML4Jets 2017: what architecture best
– ML4Jets 2018: Lots of architectures work [1902.09914, point clouds win]

SciPost SciPost Phys. 7, 014 (2019)

The Machine Learning landscape of top taggers

Gregor Kasieczka[1*], Tilman Plehn[2†], Anja Butter[2], Kyle Cranmer[3], Dipsikha Debnath[4], Barry M. Dillon[5], Malcolm Fairbairn[6], Darius A. Faroughy[5], Wojtek Fedorko[7], Christophe Gay[7], Loukas Gouskos[8], Jernej F. Kamenik[5,9], Patrick T. Komiske[10], Simon Leiss[1], Alison Lister[7], Sebastian Macaluso[3,4], Eric M. Metodiev[10], Liam Moore[11], Ben Nachman[12,13], Karl Nordström[14,15], Jannicke Pearkes[7], Huilin Qu[8], Yannik Rath[16], Marcel Rieger[16], David Shih[4] Jennifer M. Thompson[2], and Sreedevi 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

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

Meet the professionals

Brief history of jet classification

– 2014/15: first jet image papers
– 2017: first (working) ML top tagger
– ML4Jets 2017: what architecture best
– ML4Jets 2018: Lots of architectures work [1902.09914, point clouds win]


SciPost SciPost Phys. 7, 014 (2019)

The Machine Learning landscape of top taggers

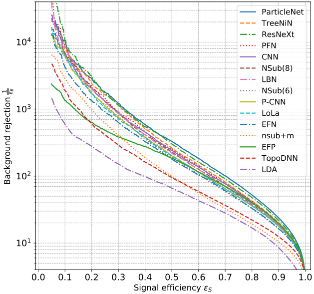
Gregor Kasieczka[1*], Tilman Plehn[2†], Anja Butter[2], Kyle Cranmer[3], Dipsikha Debnath[4], Barry M. Dillon[5], Malcolm Fairbairn[6], Darius A. Faroughy[5], Wojtek Fedorko[7], Christophe Gay[7], Loukas Gouskos[8], Jernej F. Kamenik[5,9], Patrick T. Komiske[10], Simon Leiss[1], Alison Lister[7], Sebastian Macaluso[3,4], Eric M. Metodiev[10], Liam Moore[11], Ben Nachman[12,13], Karl Nordström[14,15], Jannicke Pearkes[7], Huilin Qu[8], Yannik Rath[16], Marcel Rieger[16], David Shih[4] Jennifer M. Thompson[2], and Sreedevi 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

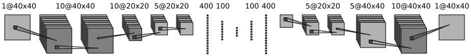
⋆ gregor.kasieczka@uni-hamburg.de, † plehn@uni-heidelberg.de

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.



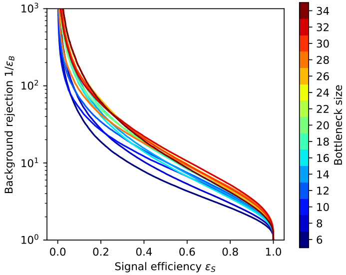
New: autoencoder



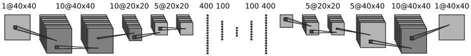
Fully supervised classification boring [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih]

– 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'
– search for outliers hard to describe

⇒ Making an okay tagger



New: autoencoder



Fully supervised classification boring [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih]

– anomaly searches, only training on ‘background’

– established ML concept: autoencoder

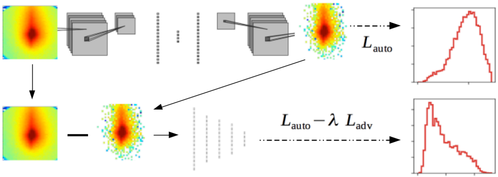
– reconstruct typical QCD jet image from many QCD jets
reduce weights in central layer, compress information to ‘typical’

– search for outliers hard to describe

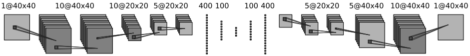
⇒ Making an okay tagger

De-correlate background shaping

– established concept: adversary



New: autoencoder



Fully supervised classification boring [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih]

– 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’

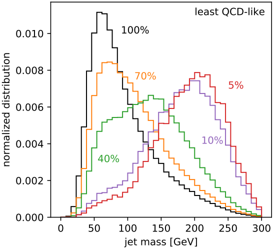
– search for outliers hard to describe

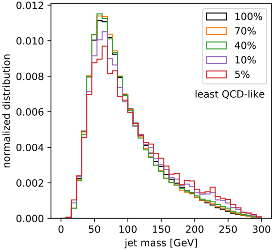
$\Rightarrow$ Making an okay tagger

De-correlate background shaping

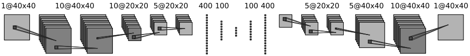
– established concept: adversary

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





New: autoencoder



Fully supervised classification boring [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih]

– 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’

– search for outliers hard to describe

⇒ Making an okay tagger

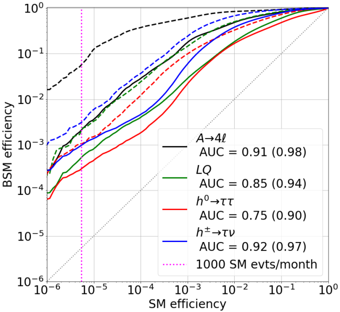
The whole thing on anomalous LHC events [Cerri, Nguyen, Pierini, Spiropulu, Vlimant]

– same thing on full events

– training data a problem

– variational autoencoder more powerful

⇒ Proof of concept...

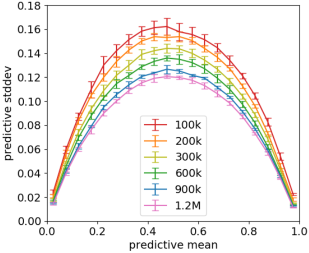


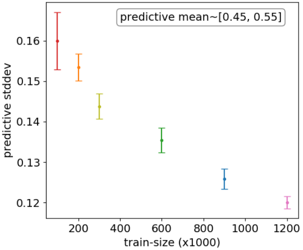
New: B****ian networks

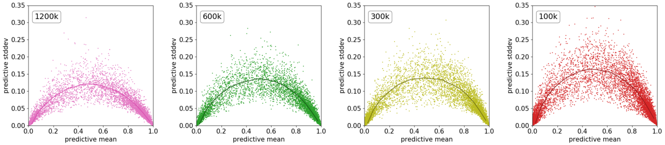
Simply better networks [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson]

– learn classification output and uncertainty [(60 ± 0)% top different from (60 ± 1)% top]

– error bars: limited training statistics



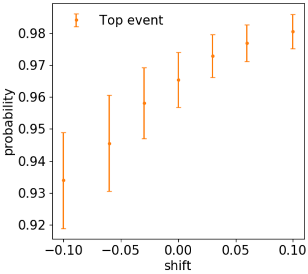




New: B****ian networks

Simply better networks [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson]

- learn classification output and uncertainty [(60 ± 0)% top different from (60 ± 1)% top]
- error bars: limited training statistics
- error bar: jet energy scale (correlated)



New: B****ian networks

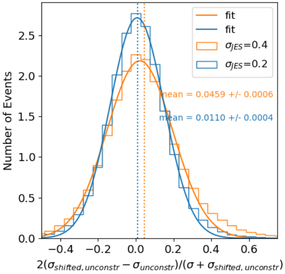
Simply better networks [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson]

– learn classification output and uncertainty [(60 ± 0)% top different from (60 ± 1)% top]

– error bars: limited training statistics

– error bar: jet energy scale (correlated)

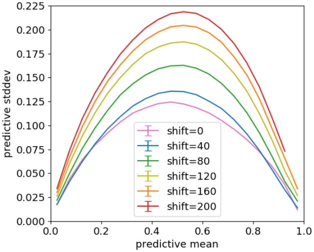
– error bar: jet energy scale (uncorrelated)

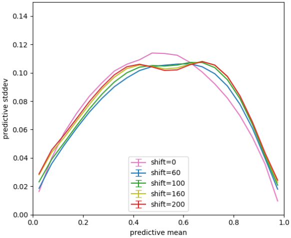


New: B****ian networks

Simply better networks [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson]

- learn classification output and uncertainty
- error bars: limited training statistics
- error bar: jet energy scale (correlated)
- error bar: jet energy scale (uncorrelated)
- error bar and stability: pile-up





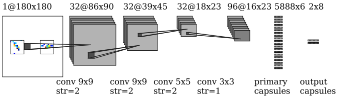
New: B****ian networks

Simply better networks [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson]

- learn classification output and uncertainty [(60 $\pm$ 0)% top different from (60 $\pm$ 1)% top]
- error bars: limited training statistics
- error bar: jet energy scale (correlated)
- error bar: jet energy scale (uncorrelated)
- error bar and stability: pile-up
- tagger calibration part of the training
- systematic approach to regularization and drop-out
- performance just like usual taggers

  ....
- Lots of conceptual and practical advantages at little cost

Capsules vs CNN



Full calorimeter images

– full detector instead of fat jet [forget training for now]

– sparse in objects with sparse objects

– multi-label for different backgrounds

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

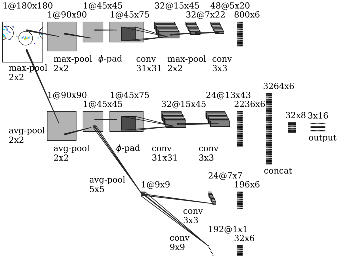
– vector output instead of scalar classification

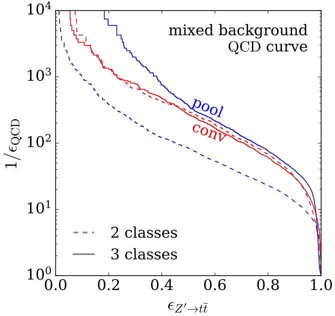
– agreement by parallel vectors in feature space

– vector components for properties and geometry [eyes, nose, mouth]

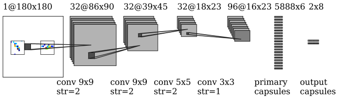
– pooling vs stride convolutions?

⇒ boosted tops from Z' resonance





Capsules vs CNN



Full calorimeter images

– full detector instead of fat jet [forget training for now]

– sparse in objects with sparse objects

– multi-label for different backgrounds

Visualization

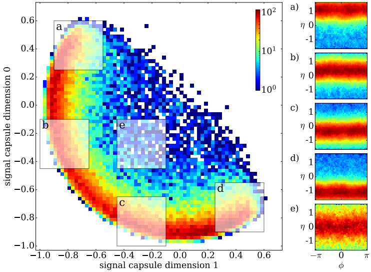
– signal capsule for signal events

– classification through radius

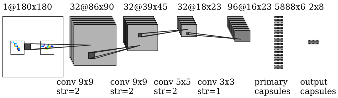
– rotation free to organize information

– average event per region

signal identifying η_j
azimuthal angle insensitive



Capsules vs CNN



Full calorimeter images

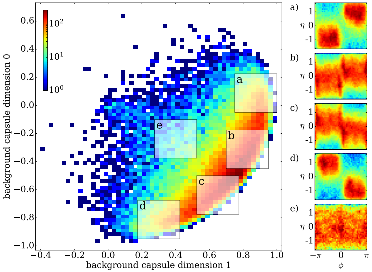
- full detector instead of fat jet [forget training for now]
- sparse in objects with sparse objects
- multi-label for different backgrounds

Visualization

- signal capsule for signal events
- classification through radius
- rotation free to organize information
- average event per region

  signal identifying η_j
  azimuthal angle insensitive

  background identifying back-to-back
- ultimate-pain benchmark: $t\bar{t}H_{bb}$



The future

Machine learning a tool box, not a black 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

physicists like things to play with
visualization/uncertainties becoming the focus
ask me about GANs...