Machine Learning just because it is Great Fun

Tilman Plehn

Universität Heidelberg

Hamburg 2/2019

Change of title

SciPost Physics — Submission

The Machine Learning Landscape of Top Taggers

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

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

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

February 26, 2019

Abstract

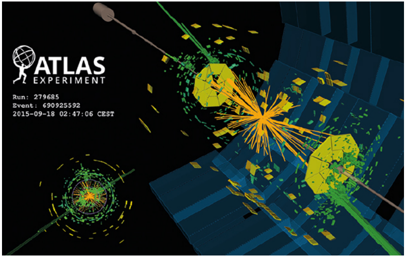
Based on the established task of identifying boosted, hadronically decaying top quarks, we compare a wide range of modern machine learning approaches. We find that they are extremely powerful and great fun.

Why LHC, why jets

Big jet data by 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 big data



Why LHC, why jets

Big jet data by 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 big data

Interesting physics in jets

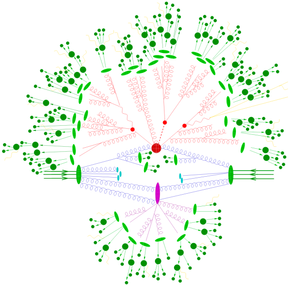
– re-summed perturbative QFT prediction for QCD splittings

– 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' [Jennifer Thompson's talk]

$\Rightarrow$ It's interesting



Why LHC, why jets

Big jet data by 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 big data

Interesting physics in jets

– re-summed perturbative QFT prediction for QCD splittings

– 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' [Jennifer Thompson's talk]

$\Rightarrow$ It's interesting

LHC simulations

– QCD simulation: Pythia, Sherpa, Herwig

– fast detector simulation: Delphes

– excellent agreement with data

$\Rightarrow$ We can simulate it

Inside jets

Jets and machine learning from 1990s to 2020s

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 ~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.

Inside jets

Jets and machine learning from 1990s to 2020s

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

1994 jet-algo W/top-tagger for heavy Higgs [Seymour]

2008 jet-algo Higgs tagger [Butterworth, Davison, Rubin, Salam; Kribs, Martin, Spannowsky]

2008 jet-algo top tagger [Kaplan, Rehermann, Schwartz, Tweedie]

2009 jet-algo HEPTopTagger [TP, Salam, Spannowsky; 1st user Gregor Kasieczka]

. . .

2009 template top tagger [Almeida, Lee, Perez, Sterman, Sung, Virzi]

2011 Shower Deconstruction [Soper, Spannowsky]

2015 Multi-variate HEPTopTagger [Kasieczka, TP, Schell, Strebler, Salam]

. . .

2014 image recognition W-tagger [Cogan, Kagan, Strass, Schwartzman]

2015 jet images [de Oliveira, Kagan), Mackey, Nachman, Schwartzman]

2017 image recognition top tagger [Kasieczka, Plehn, Russell, Schell]

2017 language recognition W-tagger [Louppe, Cho, Becot, Cranmer]

2017 4-vector-based top tagger [Butter, Kasieczka, Plehn, Russel]

. . .

2018 jet autoencoder [Heinel, Kasieczka, Plehn, Thompson; Shi etal]

Jet-level analyses (1990s)

Jets as analysis objects

– partonic predictions from QCD $\Leftrightarrow$ jets describing partons in reality
– infrared safety crucial to compare with perturbative QCD rates
– data-to-data analyses more flexible
– data-to-simulation analyses similarly free?

QCD recombination algorithms [FASTJET]

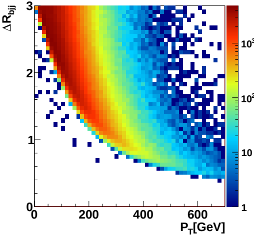
– define jet–jet and jet–beam distances [exclusive with resolution y_{cut}]

$$\begin{array}{llll} k_T & y_{ij} = \dfrac{\Delta R_{ij}}{R} \; \min\left(p_{T,i}, p_{T,j}\right) & & y_{iB} = p_{T,i} \\ \text{C/A} & y_{ij} = \dfrac{\Delta R_{ij}}{R} & & y_{iB} = 1 \\ \text{anti-}k_T & y_{ij} = \dfrac{\Delta R_{ij}}{R} \min\left(p_{T,i}^{-1}, p_{T,j}^{-1}\right) & & y_{iB} = p_{T,i}^{-1} \; . \end{array}$$

– (1) find minimum $y^{\text{min}} = \min_{ij}(y_{ij}, y_{iB})$
(2a) if $y^{\text{min}} = y_{ij}$ merge subjets i and j, back to (1)
(2b) if $y^{\text{min}} = y_{iB}$ remove i from subjets, go to (1)

$\Rightarrow$ clustering history usable?

Fat jet taggers (2000s)



For instance: boosted tops

- hadronic decays vs QCD splittings
- perfectly described by perturbative QCD
- labelled sample: semileptonic $t\bar{t}$ events

$\Rightarrow$ substructure playground

Fat jet taggers (2000s)

For instance: boosted tops

– hadronic decays vs QCD splittings

– perfectly described by perturbative QCD

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

$\Rightarrow$ substructure playground

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

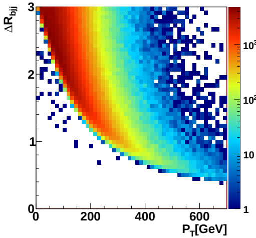
1– C/A fat jet with $p_T > 200$ GeV

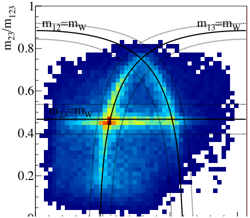
2– filtering defining 3-5 decay jets

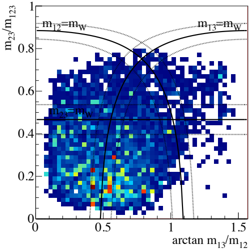
3– top mass window $m_{123} = [150, 200]$ GeV

4– A-shaped mass plane cuts probing m_W

$\Rightarrow$ not rocket science, but experimental break-through







Multi-variate subjet physics (2010s)

OptimalR and N-Subjettiness [Kasieczka, TP, Salam, Schell, Strebler]

– multivariate analysis old idea [Lonnblad, Peterson, Rognvaldsson]
HEPTopTaggerv2 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]

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

Multi-variate subjet physics (2010s)

OptimalR and N-Subjettiness [Kasieczka, TP, Salam, Schell, Strebler]

– multivariate analysis old idea [Lonnblad, Peterson, Rognvaldsson]
  HEPTopTaggerv2 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]

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

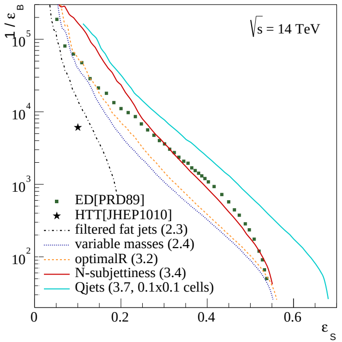
Fat jet and top kinematics

– FSR major problem for Z' search

– tag and reconstruction in each other's way

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

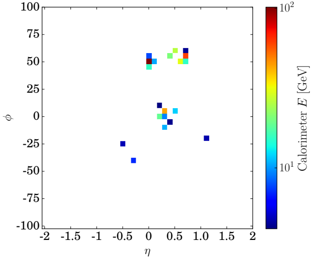
$\Rightarrow$ expected performance increase



Jet images (2020s)

'Deep learning' = modern architectures on low-level observables

- wavelet transformation [Rentala, Shepherd, Tait; Monk]
- *W*-tagging with image recognition [Cogan etal, Oliveira etal, Baldi etal]
- impact of shower? [Barnard etal]
- combining calorimeter and tracking? [Komiske etal]
- understanding additional information? [Datta & Larkosky]
- link to infrared safety? [Choi, Lee, Perelstein; Friday speakers]



Jet images (2020s)

‘Deep learning’ = modern architectures on low-level observables

– wavelet transformation [Rentala, Shepherd, Tait; Monk]

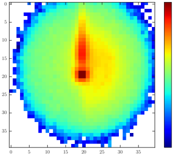
– W-tagging with image recognition [Cogan etal, Oliveira etal, Baldi etal]

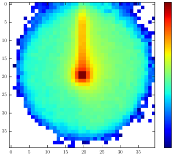
– impact of shower? [Barnard etal]

– combining calorimeter and tracking? [Komiske etal]

– understanding additional information? [Datta & Larkosky]

– link to infrared safety? [Choi, Lee, Perelstein; Friday speakers]



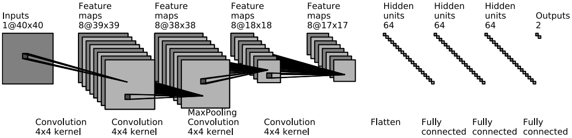


Convolutional network [Kasieczka, TP, Russell, Schell; Macaluso, Shih]

– run on 2-D jet images [$p_T = 350, \dots, 450$ GeV]

– colored image as input

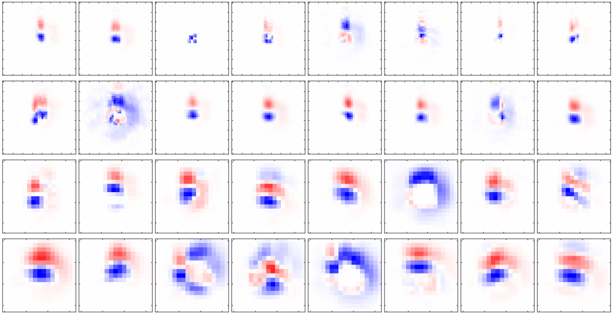
– binning through calorimeter resolution [$\Delta\eta = 0.1$ vs $\Delta\phi = 5^\circ$]

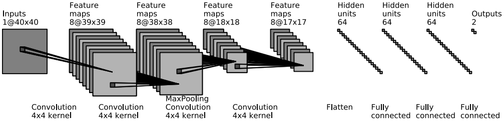


Inside DeepTop

Benchmarking image-based top tagger [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

– 2+2 convolutional layers



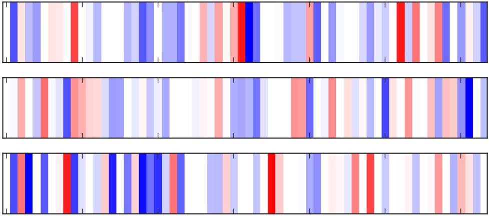


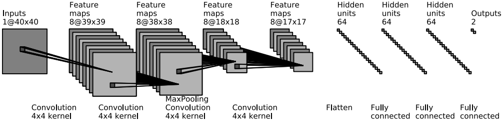
Inside DeepTop

Benchmarking image-based top tagger [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

– 2+2 convolutional layers

– 3 fully connected layers





Inside DeepTop

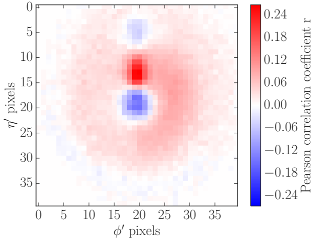
Benchmarking image-based top tagger [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

– 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

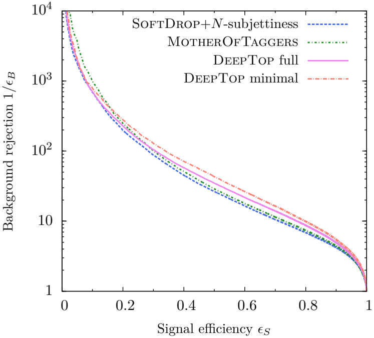
Benchmarking image-based top tagger [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

– 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

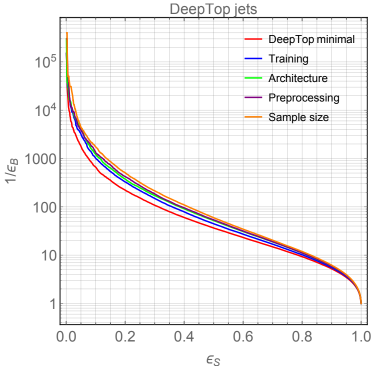
Benchmarking image-based top tagger [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

- 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

Benchmarking image-based top tagger [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

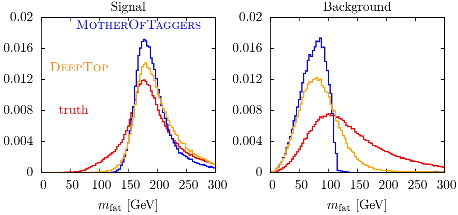
⇒ understandable performance gain

Typical reaction: ‘Fuck you, you fucking machine’

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

– MC truth vs MotherOfTaggers vs DeepTop

fat jet mass



Inside DeepTop

Benchmarking image-based top tagger [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

– 2+2 convolutional layers

– 3 fully connected layers

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

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

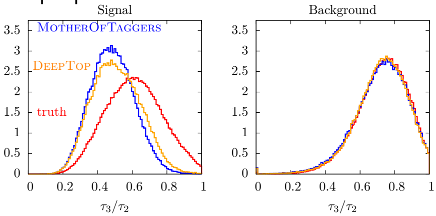
– comparison to MotherOfTaggers BDT

⇒ understandable performance gain

Typical reaction: ‘Fuck you, you fucking machine’

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

– MC truth vs MotherOfTaggers vs DeepTop
fat jet mass
N-subjettiness



Inside DeepTop

Benchmarking image-based top tagger [Kasieczka, TP, Russell, Schell; Macaluso & Shih]

- 2+2 convolutional layers
- 3 fully connected layers
- Pearson input-output correlation [pixel x vs label y]

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

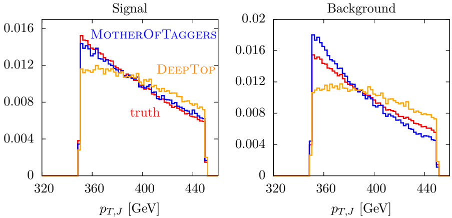
- comparison to MotherOfTaggers BDT

⇒ understandable performance gain

Typical reaction: ‘Fuck you, you fucking machine’

- full control for supervised learning
  easy checks for correctly identified signal/background events
- MC truth vs MotherOfTaggers vs DeepTop
  fat jet mass
  N-subjettiness
  transverse momenta

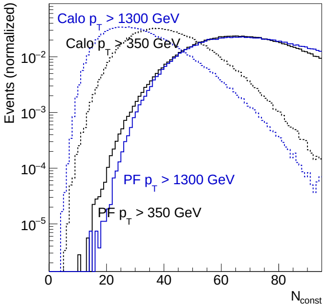
⇒ it works and we know why



DeepTopLoLa

Our version of graph network [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



DeepTopLoLa

Our version of graph network [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

Inspired by jet algorithm — combination layer

– input 4-vectors

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

DeepTopLoLa

Our version of graph network [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

Inspired by jet algorithm — combination layer

- input 4-vectors
- on-shell conditions for top tag

$$\tilde{k}^2_{\mu,1} = (k_{\mu,1} + k_{\mu,2} + k_{\mu,3})^2 \stackrel{!}{=} m_t^2$$

$$\tilde{k}^2_{\mu,2} = (k_{\mu,1} + k_{\mu,2})^2 \stackrel{!}{=} m_W^2$$

DeepTopLoLa

Our version of graph network [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

Inspired by jet algorithm — combination layer

– input 4-vectors
– on-shell conditions for top tag
– combined 4-vectors

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

– after combination of input 4-vectors

original momenta k_i
$M - N$ trainable linear combinations [M-N=15]

$\Rightarrow$ physics step, easy to interpret

DeepTopLoLa

Our version of graph network [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

Inspired by jet algorithm — combination layer

- combined 4-vectors $k_{\mu,i} \xrightarrow{\text{CoLa}} \widetilde{k}_{\mu,j} = k_{\mu,i}\, C_{ij}$

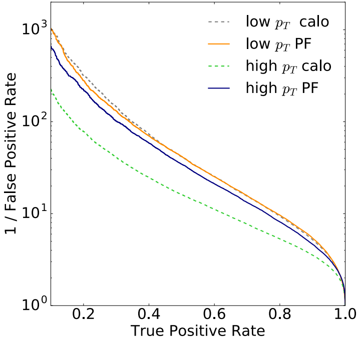
Inspired by Jackson — Lorentz layer

- DNN on Lorentz scalars

$$\tilde{k}_j \xrightarrow{\text{LoLa}} \hat{k}_j = \begin{pmatrix} m^2(\tilde{k}_j) \\ p_T(\tilde{k}_j) \\ w_{jm}^{(E)}\, E(\tilde{k}_m) \\ w_{jm}^{(d)}\, d_{jm}^2 \end{pmatrix}$$

- learn Minkowski metric

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



Moving on

Simple questions

– ML4Jets 2017: what architecture?

– ML4Jets 2018: top tagging study

SciPost Physics Submission

The Machine Learning Landscape of Top Taggers

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

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

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

February 26, 2019

Abstract

Based on the established task of identifying boosted, hadronically decaying top quarks, we compare a wide range of modern machine learning approaches. We find that they are extremely powerful and great fun.

Content

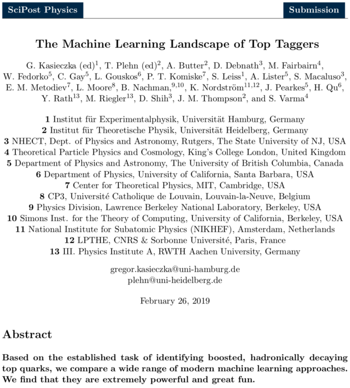
Moving on

Simple questions

– ML4Jets 2017: what architecture?

– ML4Jets 2018: top tagging study

⇒ lots of architectures work



SciPost Physics Submission

The Machine Learning Landscape of Top Taggers

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

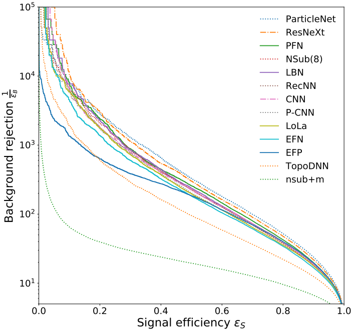
1 Institut für Experimentalphysik, Universität Hamburg, Germany
2 Institut für Theoretische Physik, Universität Heidelberg, Germany
3 NHECT, Dept. of Physics and Astronomy, Rutgers, The State University of NJ, USA
4 Theoretical Particle Physics and Cosmology, King's College London, United Kingdom
5 Department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

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

February 26, 2019

Abstract

Based on the established task of identifying boosted, hadronically decaying top quarks, we compare a wide range of modern machine learning approaches. We find that they are extremely powerful and great fun.



Moving on

Simple questions

– ML4Jets 2017: what architecture?

– ML4Jets 2018: top tagging study

$\Rightarrow$ lots of architectures work

More questions

– what about uncertainties?

– how stable are taggers in experimental reality?

– can we go beyond fully supervised learning?

– how do we go beyond jets?

– is classification all we can use ML for?

– are there analyses only ML will allow us to do?

– what is the particle nature of dark matter?

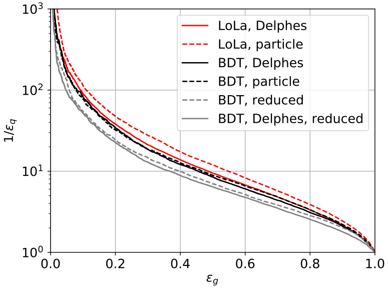
etc

When reality hits

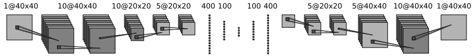
ML-Life is not always nice to us [Kasieczka, Kiefer, TP, Thompson]

- 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

⇒ deep-learning advantage gone after detector simulation, REALLY???



Getting seriously inspired



Anomaly search only trained on background [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih]

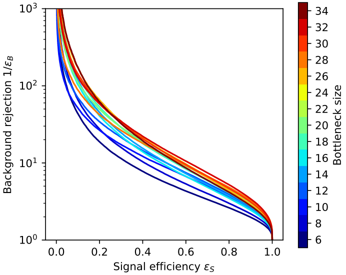
– established ML concept: autoencoder

– reconstruct typical QCD jet image from many QCD jets

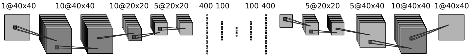
– reduce weights in central layer
  compress information on 'typical'

– search for outliers hard to describe

– benchmark on top jets, search for Higgs or dark showers



Getting seriously inspired



Anomaly search only trained on background

[Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih]

– established ML concept: autoencoder

– reconstruct typical QCD jet image from many QCD jets

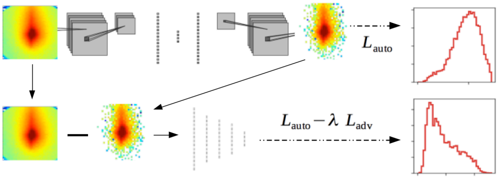
– reduce weights in central layer
  compress information on ‘typical’

– search for outliers hard to describe

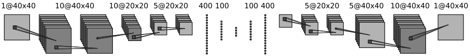
– benchmark on top jets, search for Higgs or dark showers

De-correlate background shaping

– established concept: adversary



Getting seriously inspired

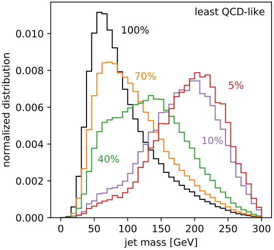


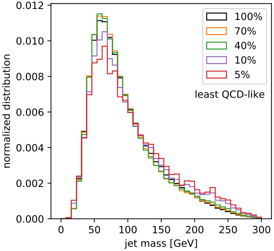
Anomaly search only trained on background [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih]

- established ML concept: autoencoder
- reconstruct typical QCD jet image from many QCD jets
- reduce weights in central layer
  compress information on ‘typical’
- search for outliers hard to describe
- benchmark on top jets, search for Higgs or dark showers

De-correlate background shaping

- established concept: adversary
- atypical QCD jets typially with large jet mass
  remove jet mass from network training





The future

Times are moving fast...

...jets are containers for subjet physics [was 1990s]

...deterministic taggers are established/old/boring [was 2000s]

...multi-variate taggers are an intermediate step [dying with the 2010s]

...imagine recognition is a starting point [will be 2020s]

...deep learning is not just classification

Join the fun!