

Tagging Machines

Tilman Plehn

Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

DeepTop

DeepTopLoLa

The Rise of the Tagging Machines

Tilman Plehn

Universität Heidelberg

Madrid 3/2018

Tagging Machines

Tilman Plehn

Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

DeepTop

DeepTopLoLa

Generation One to Three

Evolving from deterministic taggers to deep networks

1994 QCD-algorithm *W*-tagger for heavy Higgs [Seymour]



1994 QCD-algorithm top tagger for fun [Seymour]

2008 QCD-algorithm BDRS Higgs tagger [Butterworth, Davison, Rubin, Salam]

2008 QCD-algorithm JH/CMS top tagger [Kaplan, Rehermann, Schwartz, Tweedie]

2009 **QCD-algorithm HEPTopTagger** [TP, Salam, Spannowsky]

...

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]



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]

Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

DeepTop

DeepTopLoLa

Jet algorithms

Jets

- jet–parton duality \Leftrightarrow what are partons in detector?
- crucial for any LHC analysis [really???]
- quarks vs gluons tricky in perturbative QCD
- \Rightarrow **what about hadronic decays?**

QCD recombination algorithms [FASTJET: Cacciari, Salam, Soyez]

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

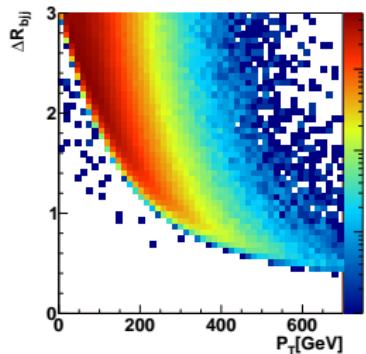
$$\begin{array}{lll} k_T & y_{ij} = \frac{\Delta R_{ij}}{R} \min(p_{T,i}, p_{T,j}) & y_{iB} = p_{T,i} \\ \text{C/A} & y_{ij} = \frac{\Delta R_{ij}}{R} & y_{iB} = 1 \\ \text{anti-}k_T & y_{ij} = \frac{\Delta R_{ij}}{R} \min(p_{T,i}^{-1}, p_{T,j}^{-1}) & y_{iB} = p_{T,i}^{-1} \end{array}$$

- (1) find minimum $y^{\min} = \min_{ij}(y_{ij}, y_{iB})$
- (2a) if $y^{\min} = y_{ij}$ merge subjets i and j , back to (1)
- (2b) if $y^{\min} = y_{iB}$ remove i from subjets, go to (1)
- \Rightarrow **fat jets/substructure: use clustering history [k_T , C/A only]**

Hadronic $t\bar{t}$ resonances

Tagging boosted tops

- hadronic decays vs QCD splittings
 - SM sample: semileptonic $t\bar{t}$ events
- ⇒ **substructure playground**



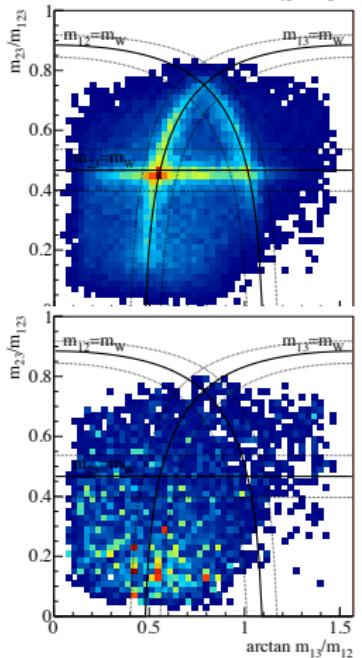
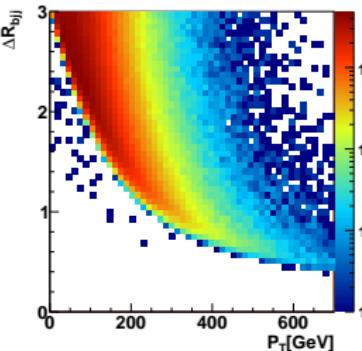
Hadronic $t\bar{t}$ resonances

Tagging boosted tops

- hadronic decays vs QCD splittings
- SM sample: semileptonic $t\bar{t}$ events
- ⇒ **substructure playground**

HEPTopTagger [BDRS; TP, Salam, Spannowsky, Takeuchi]

- 1– C/A fat jet, $R = 1.5$ and $p_T > 200$ GeV [FastJet limitation]
- 2– mass drop, cutoff $m_{\text{sub}} > 30$ GeV
- 3– filtering leading to hard substructure triple
- 4– top mass window $m_{123} = [150, 200]$ GeV
- 5– A-shaped mass plane cuts as function of m_W/m_t
- 6– consistency condition $p_T^{(\text{tag})} > 200$ GeV



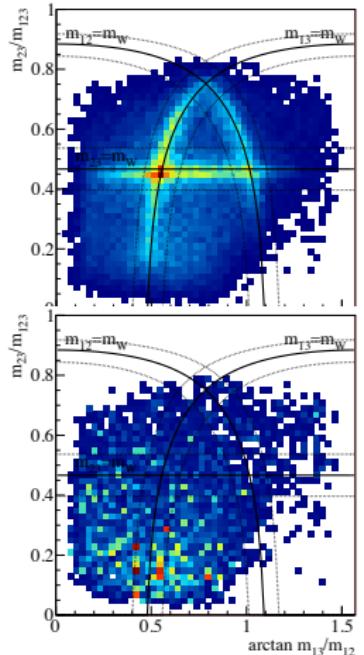
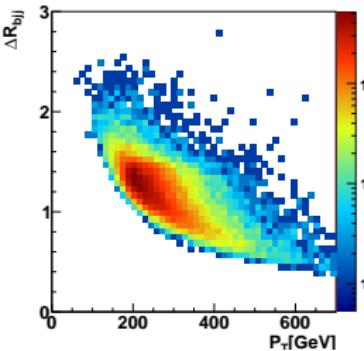
Hadronic $t\bar{t}$ resonances

Tagging boosted tops

- hadronic decays vs QCD splittings
 - SM sample: semileptonic $t\bar{t}$ events
- ⇒ **substructure playground**

HEPTopTagger [BDRS; TP, Salam, Spannowsky, Takeuchi]

- 1– C/A fat jet, $R = 1.5$ and $p_T > 200$ GeV [FastJet limitation]
 - 2– mass drop, cutoff $m_{\text{sub}} > 30$ GeV
 - 3– filtering leading to hard substructure triple
 - 4– top mass window $m_{123} = [150, 200]$ GeV
 - 5– A-shaped mass plane cuts as function of m_W/m_t
 - 6– consistency condition $p_T^{(\text{tag})} > 200$ GeV
- ⇒ **G1: LHC break-through**



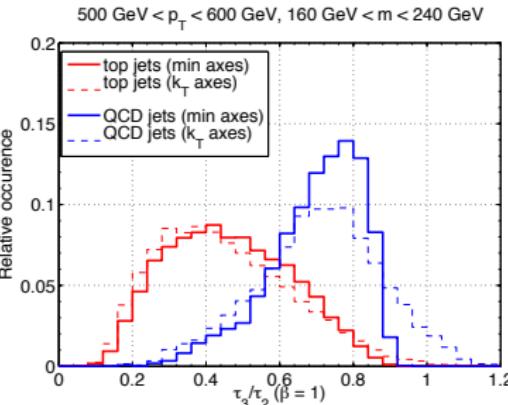
N-Subjettiness

N-Jettiness to count subjets inside fat jet [Thaler, van Tilburg]

- how many subjets do the calo entries correspond to?

$$\tau_N = \frac{1}{\sum_{\alpha \in \text{jet}} p_{T,\alpha} R_0^\beta} \sum_{\alpha \in \text{jet}} p_{T,\alpha} \min_{k=1, \dots, N} (\Delta R_{k,\alpha})^\beta$$

- choice of reference axes
 - 1- from subjet algorithm
 - 2- from minimization of τ_N
 - $\tau_N \rightarrow 1$ means many calo entries away from N axes
 - $\tau_N \rightarrow 0$ means perfect matching
 - systematics cancelled in ratios
 - $\tau_{N+1}/\tau_N \rightarrow 0$ for $N + 1$ subjets
- ⇒ easily added to any other tagger



Multi-variate top taggers

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]
 $|m_{123} - m_{123}^{(R_{\text{max}})}| < 0.2 m_{123}^{(R_{\text{max}})} \Rightarrow R_{\text{opt}}$
- add N-subjettiness [Thaler, van Tilburg]
- $\{m_{123}, f_W, R_{\text{opt}} - R_{\text{opt}}^{(\text{calc})}, \tau_j, \tau_j^{(\text{filt})}\}$

Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

DeepTop

DeepTopLoLa

Multi-variate top taggers

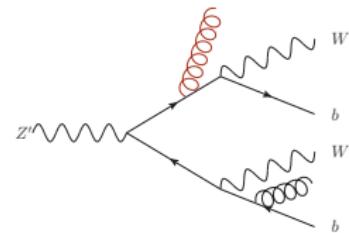
OptimalR and N-Subjettiness [Kasieczka, TP, Salam, Schell, Strebler]

- 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]

$$|m_{123} - m_{123}^{(R_{\text{max}})}| < 0.2 m_{123}^{(R_{\text{max}})} \Rightarrow R_{\text{opt}}$$
- 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
- $\Rightarrow \{..., m_{tt}, p_{T,t}, m_{jj}^{(\text{filt})}, p_{T,j}^{(\text{filt})}\}$



Multi-variate top taggers

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

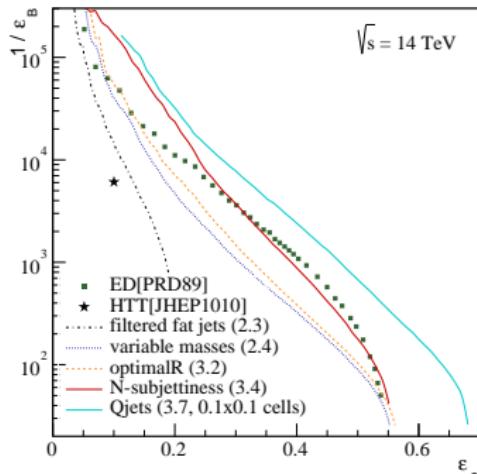
- multivariate analysis old idea [Lonnblad, Peterson, Rognvaldsson]
HEPTopTaggerv2 to keep up with shower deconstruction [Soper, Spannowsky]
- optimal fat jet size R_{opt} [large to decay jets, small to avoid combinatorics, compute from kinematics]

$$|m_{123} - m_{123}^{(R_{\text{max}})}| < 0.2 m_{123}^{(R_{\text{max}})} \Rightarrow R_{\text{opt}}$$
- 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 G2: deterministic taggers terminated!



Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

DeepTop

DeepTopLoLa

Jet images

Image recognition replacing high-level observables

- wavelet transformation [Rentala, Shepherd, Tait; Monk]
 - W -tagging with image recognition [Cogan et al, Oliveira et al, Baldi et al]
 - top-tagging attempt [Almeida, Backovic, Cliche, Lee, Perelstein]
 - QCD and shower study [Barnard et al]
 - quark-gluon discrimination including tracks [Komiske et al]
- ⇒ G3: new hammer for jet nails



Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

DeepTop

DeepTopLoLa

Jet images

Image recognition replacing high-level observables

- wavelet transformation [Rentala, Shepherd, Tait; Monk]
 - W -tagging with image recognition [Cogan et al, Oliveira et al, Baldi et al]
 - top-tagging attempt [Almeida, Backovic, Cliche, Lee, Perelstein]
 - QCD and shower study [Barnard et al]
 - quark-gluon discrimination including tracks [Komiske et al]
- ⇒ G3: new hammer for jet nails



Experimental questions

- does it work?
- what is the training sample? [Metodiev, Nachman, Thaler]
- how do we get it past the jets experts?

Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

DeepTop

DeepTopLoLa

Jet images

Image recognition replacing high-level observables

- wavelet transformation [Rentala, Shepherd, Tait; Monk]
 - W -tagging with image recognition [Cogan et al, Oliveira et al, Baldi et al]
 - top-tagging attempt [Almeida, Backovic, Cliche, Lee, Perelstein]
 - QCD and shower study [Barnard et al]
 - quark-gluon discrimination including tracks [Komiske et al]
- ⇒ G3: new hammer for jet nails



Experimental questions

- does it work?
- what is the training sample? [Metodiev, Nachman, Thaler]
- how do we get it past the jets experts?

Theoretical questions

- how much of all this is QCD?
 - how can be improve the setup? [the future has not been written]
- ⇒ what does the network learn?

Motivation

G1 Taggers

G2 Multi-variate

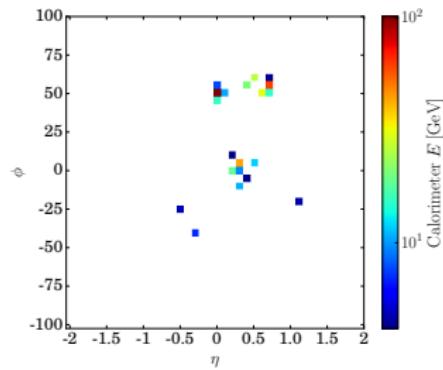
G3 Jet images

DeepTop

DeepTopLoLa

Jet images

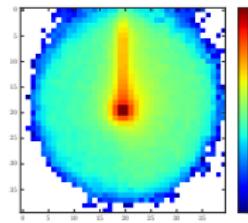
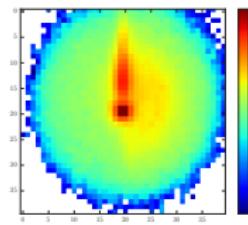
Fat jet defining the geometric area [$p_T = 350 \dots 450 \text{ GeV}$, $R = 1.5$]



Jet images

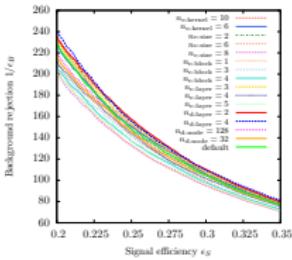
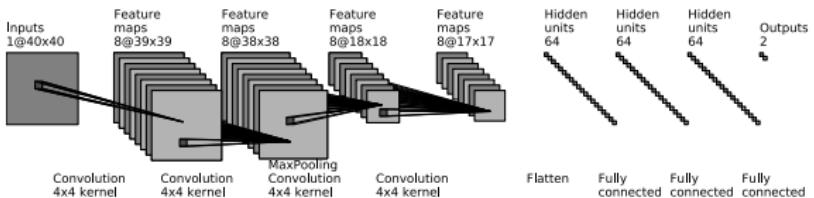
Fat jet defining the geometric area $[p_T = 350 \dots 450 \text{ GeV}, R = 1.5]$

- shift move image to center the global maximum
 - rotation rotate the second maximum to 12 o'clock
 - flip ensure third maximum is in the right half-plane
 - crop crop the image to 40×40 pixels
- ⇒ pre-processing for illustration



Neural network [Kasieczka, TP, Russell, Schell]

- colored image as input
- binning through calorimeter resolution $[\Delta\eta = 0.1 \text{ vs } \Delta\phi = 5^\circ]$
- analyze geometric patterns [convolutional network]



Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

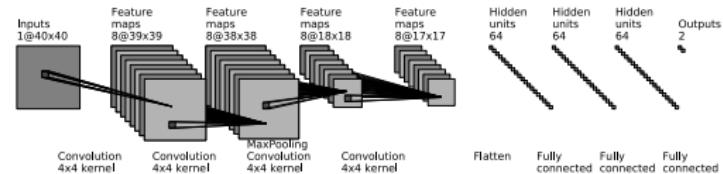
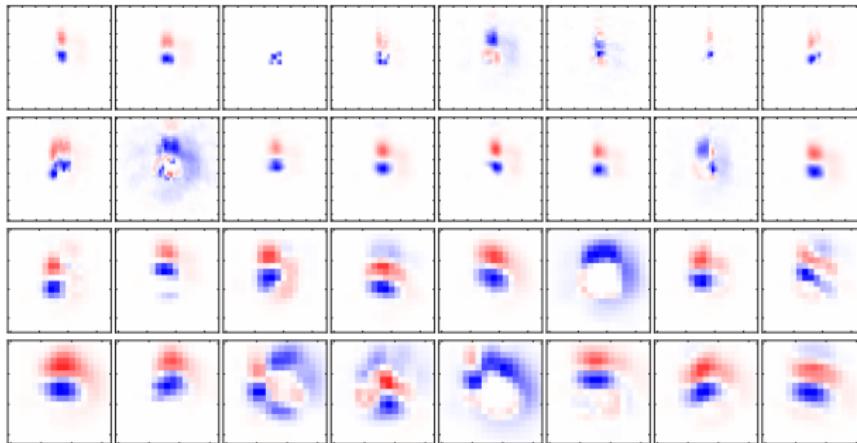
DeepTop

DeepTopLoLa

DeepTop tagger

Benchmarking image-based top tagger [Kasieczka, TP, Russell, Schell]

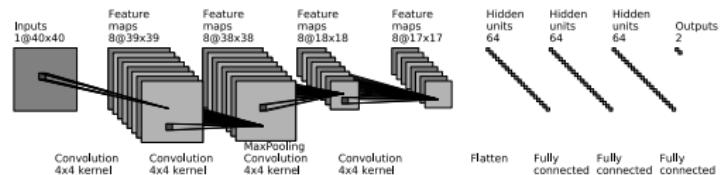
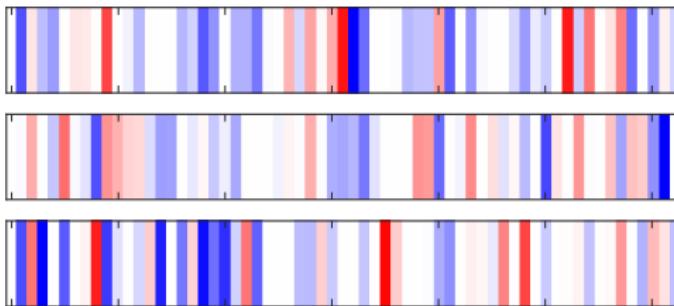
– 2+2 convolutional layers probing 2D structure with kernel matrix



DeepTop tagger

Benchmarking image-based top tagger [Kasieczka, TP, Russell, Schell]

- 2+2 convolutional layers probing 2D structure with kernel matrix
- 3 fully connected layers weight function linking input and output



Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

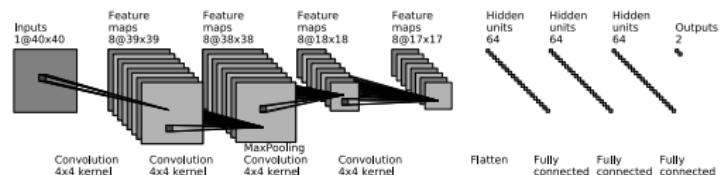
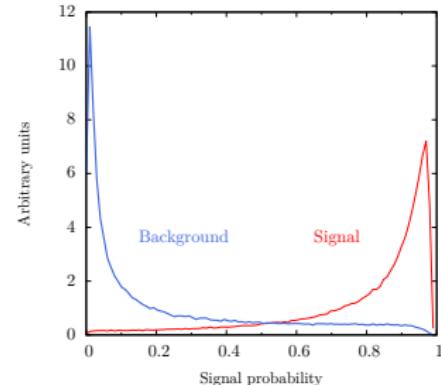
DeepTop

DeepTopLoLa

DeepTop tagger

Benchmarking image-based top tagger [Kasieczka, TP, Russell, Schell]

- 2+2 convolutional layers probing 2D structure with kernel matrix
- 3 fully connected layers weight function linking input and output
- signal-ness vs background-ness output [probability?]

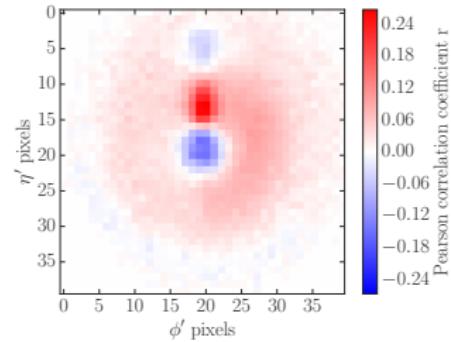


DeepTop tagger

Benchmarking image-based top tagger [Kasieczka, TP, Russell, Schell]

- 2+2 convolutional layers probing 2D structure with kernel matrix
- 3 fully connected layers weight function linking input and output
- signal-ness vs background-ness output [probability?]
- Pearson input-output correlation [pixel x vs label y]

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



DeepTop tagger

Benchmarking image-based top tagger [Kasieczka, TP, Russell, Schell]

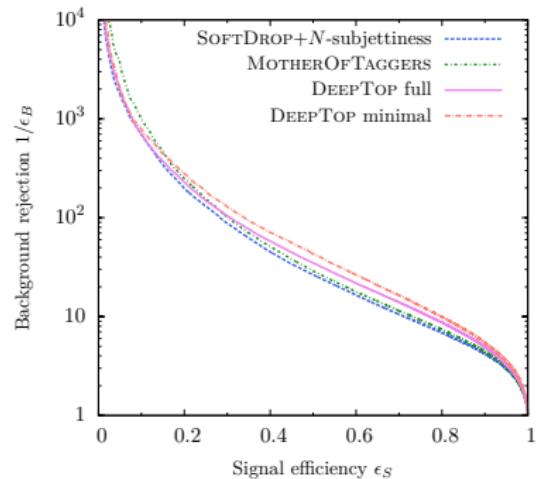
- 2+2 convolutional layers probing 2D structure with kernel matrix
- 3 fully connected layers weight function linking input and output
- signal-ness vs background-ness output [probability?]
- Pearson input-output correlation [pixel x vs label y]

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

- comparison to G2 MotherOfTaggers

$$\{m_{\text{sd}}, m_{\text{fat}}, m_{\text{rec}}, f_{\text{rec}}, \Delta R_{\text{opt}}, \tau_2, \tau_3, \tau_2^{\text{sd}}, \tau_3^{\text{sd}}\}$$

⇒ slight performance gain for G3



Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

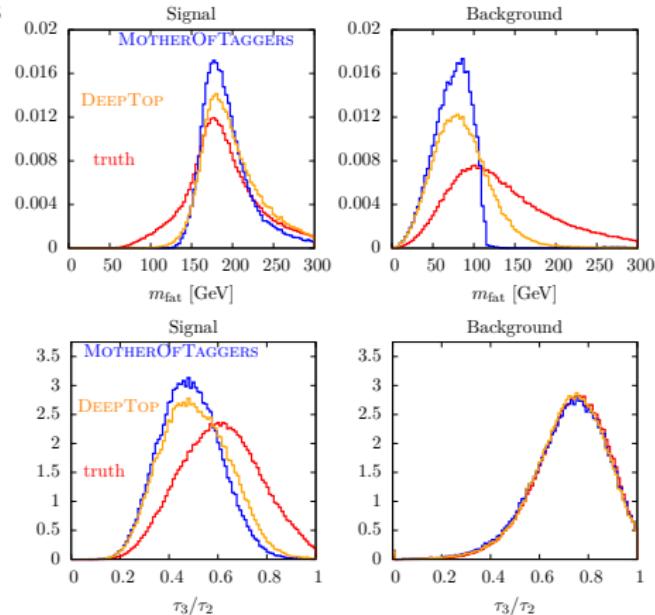
DeepTop

DeepTopLoLa

Checking physics

Typical reaction: 'fuck you, you fucking machine'

- in principle, full control for fully supervised learning
 - lots of events in the grey zone
but checks possible for correctly identified signal/background events
 - compare truth vs MotherOfTaggers vs DeepTop
- 1- fat jet mass and N-subjettiness



Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

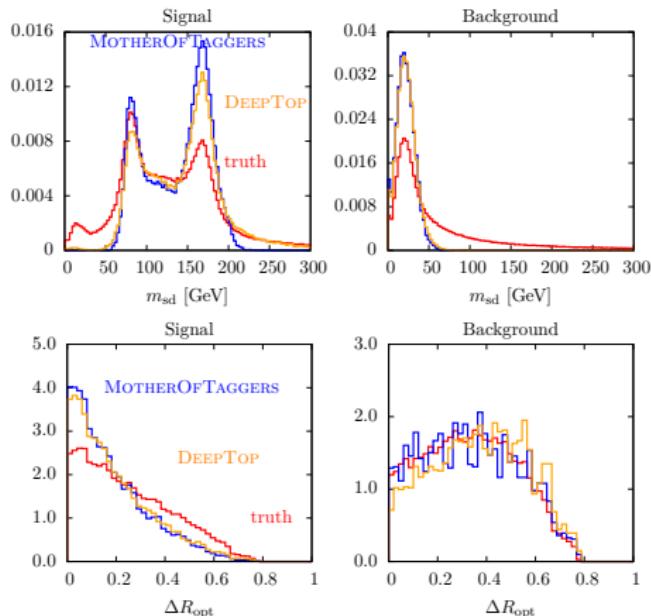
DeepTop

DeepTopLoLa

Checking physics

Typical reaction: 'fuck you, you fucking machine'

- in principle, full control for fully supervised learning
 - lots of events in the grey zone
but checks possible for correctly identified signal/background events
 - compare truth vs MotherOfTaggers vs DeepTop
- 1- fat jet mass and N-subjettiness
 - 2- soft drop mass and ΔR_{opt}



Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

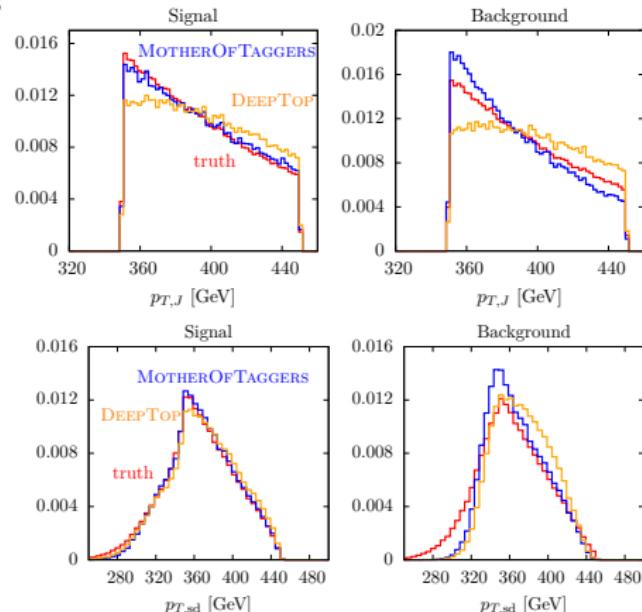
DeepTop

DeepTopLoLa

Checking physics

Typical reaction: 'fuck you, you fucking machine'

- in principle, full control for fully supervised learning
 - lots of events in the grey zone
but checks possible for correctly identified signal/background events
 - compare truth vs MotherOfTaggers vs DeepTop
- 1- fat jet mass and N-subjettiness
 - 2- soft drop mass and ΔR_{opt}
 - 3- transverse momenta



Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

DeepTop

DeepTopLoLa

Checking physics

Typical reaction: 'fuck you, you fucking machine'

- in principle, full control for fully supervised learning
 - lots of events in the grey zone
but checks possible for correctly identified signal/background events
 - compare truth vs MotherOfTaggers vs DeepTop
- 1- fat jet mass and N-subjettiness
 - 2- soft drop mass and ΔR_{opt}
 - 3- transverse momenta
- ⇒ deep learning works and we know why



Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

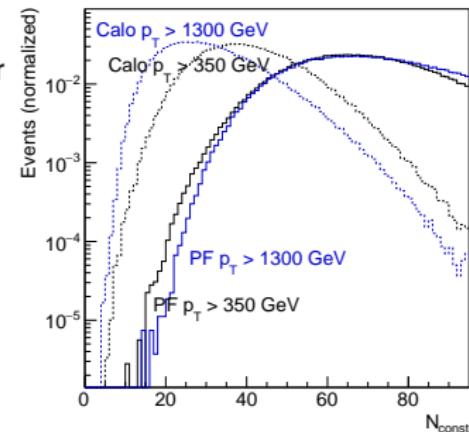
DeepTop

DeepTopLoLa

DeepTop using Lorentz Layer

Our version of graph network [Butter, Kasieczka, TP, Russell; see also Louppe et al, Pearkes et al]

- sparsely filled picture: graph CNN
- physics objects from calorimeter and tracker
- distance measure known from e&m



Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

DeepTop

DeepTopLoLa

DeepTop using Lorentz Layer

Our version of graph network [Butter, Kasieczka, TP, Russell; see also Louppe et al, Pearkes et al]

- 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

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

Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

DeepTop

DeepTopLoLa

DeepTop using Lorentz Layer

Our version of graph network [Butter, Kasieczka, TP, Russell; see also Louppe et al, Pearkes et al]

- 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

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

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

Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

DeepTop

DeepTopLoLa

DeepTop using Lorentz Layer

Our version of graph network [Butter, Kasieczka, TP, Russell; see also Louppe et al, Pearkes et al]

- 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

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

$$C = \begin{pmatrix} 1 & 0 & \cdots & 0 & C_{1,N+2} & \cdots & C_{1,M} \\ 0 & 1 & & & \vdots & C_{2,N+2} & \cdots & C_{2,M} \\ \vdots & \vdots & \ddots & & 0 & \vdots & & \vdots \\ \vdots & \vdots & \ddots & & 0 & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & C_{N,N+2} & \cdots & C_{N,M} \end{pmatrix}$$

- after combination of input 4-vectors
 - original momenta k_i
 - $M - N$ trainable linear combinations [M-N=15]
- ⇒ physics step, easy to interpret

DeepTop using Lorentz Layer

[Our version of graph network](#) [Butter, Kasieczka, TP, Russell; see also Louppe et al, Pearkes et al]

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

[Remember e&m — Lorentz layer](#)

- DNN on Lorentz scalars

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

- measurement-inspired individual $m_j, p_{T,j}$

trainable correlations: energy, metric

[1+4 copies, sum,max]

DeepTop using Lorentz Layer

Inspired by jet algorithm — combination layer

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

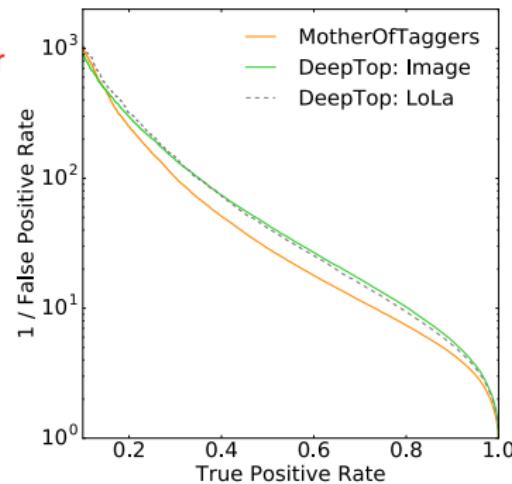
Remember e&m — Lorentz layer

- DNN on Lorentz scalars $\tilde{k}_j \xrightarrow{\text{LoLa}} \hat{k}_j$
- measurement-inspired individual $m_j, p_{T,j}$

trainable correlations: energy, metric

[1+4 copies, sum,max]

- comparison to G2, images for calorimeter



Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

DeepTop

DeepTopLoLa

DeepTop using Lorentz Layer

Inspired by jet algorithm — combination layer

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

Remember e&m — Lorentz layer

- DNN on Lorentz scalars $\tilde{k}_j \xrightarrow{\text{LoLa}} \hat{k}_j$
- measurement-inspired individual $m_j, p_{T,j}$

trainable correlations: energy, metric

[1+4 copies, sum,max]

- comparison to G2, images for calorimeter
- fun: measure Minkowski metric

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

DeepTop using Lorentz Layer

Inspired by jet algorithm — combination layer

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

Remember e&m — Lorentz layer

- DNN on Lorentz scalars $\tilde{k}_j \xrightarrow{\text{LoLa}} \hat{k}_j$
- measurement-inspired individual $m_j, p_{T,j}$

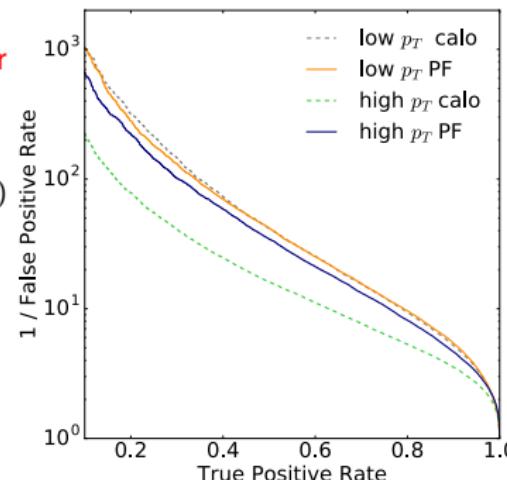
trainable correlations: energy, metric

[1+4 copies, sum,max]

- comparison to G2, images for calorimeter
- fun: measure Minkowski metric

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

$$[p_T = 350, \dots, 450 \text{ GeV and } 1300, \dots, 1400 \text{ GeV}]$$
- 180k training events, 15 GPU minutes
- G3 with little but Lorentz invariance



Motivation

G1 Taggers

G2 Multi-variate

G3 Jet images

DeepTop

DeepTopLoLa

The future

Times are moving fast...

...deterministic taggers established/old/boring

...multi-variate taggers an intermediate step

...imagine recognition a starting point

...DeepTop is not a black box

...back to physics with DeepTopLoLa

ML taggers ready, what's next?

