QCD Jets: Rise of the Machines

Tilman Plehn

Universität Heidelberg

Budapest 7/2017

Machine Jets

Tilman Plehn

Top tagging

HEPTopTagger

Jet images

DeepTop

Brand new

Fat jets

Boosted particles at the LHC

1994 boosted $W \to 2$ jets from heavy Higgs [Seymour]

1994 boosted $t \to 3$ jets [Seymour]

2006 boosted $t \to 3$ jets from resonances [Agashe, Belyaev, Krupovnickas, Perez, Virzi]

2008 boosted $H \to b\bar{b}$ [BDRS Higgs tagger: Butterworth, Davison, Rubin, Salam]

2008 boosted $t \to 3$ jets from resonances [JH/CMS tagger: Kaplan, Rehermann, Schwartz, Tweedie]

2009 boosted $t \to 3$ jets in Higgs production [HEPTopTagger: TP, Salam, Spannowsky]

...

2009 boosted $t \to 3$ jets from resonances [Template Tagger: Almeida, Lee, Perez, Sterman, Sung, Virzi]

2011 N-Subjettiness [Thaler, van Tilburg]

2011 Shower Deconstruction [Soper, Spannowsky]

...

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]

20?? subjet recognition top tagger [Butter, Kasieczka, Plehn, Russel]

Motivation: $Z' \to t\bar{t}$

Exmple LHC signature $Z' \to t\bar{t}$

– purely leptonic decays rate limited

– semi-leptonic decays with approximate reconstruction

– purely hadronic decays deemed not useful

$\Rightarrow$ challenge to analysis objects

Motivation: $Z' \to t\bar{t}$

Exmple LHC signature $Z' \to t\bar{t}$

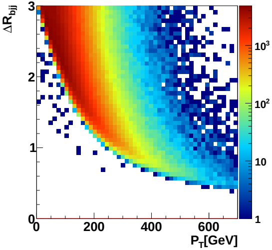
- purely leptonic decays rate limited
- semi-leptonic decays with approximate reconstruction
- purely hadronic decays deemed not useful

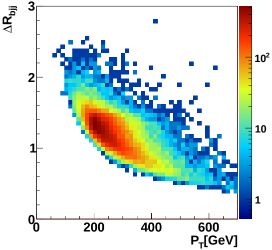
$\Rightarrow$ challenge to analysis objects

Biggest success in sub-jet tools

- top tagger: hadronic identification and reconstruction
- hadronic decays vs QCD splittings
- tagging easy for higher boost, $p_{T,t} > 600$ GeV
- SM sample: semileptonic $t\bar{t}$ events

$\Rightarrow$ t-tagging the new b-tagging?





Deterministic HEPTopTagger

Mass drop algorithm [TP, Salam, Spannowsky, Takeuchi]

1– C/A fat jet, $R = 1.5$ and $p_T > 200$ GeV [FastJet limitation]

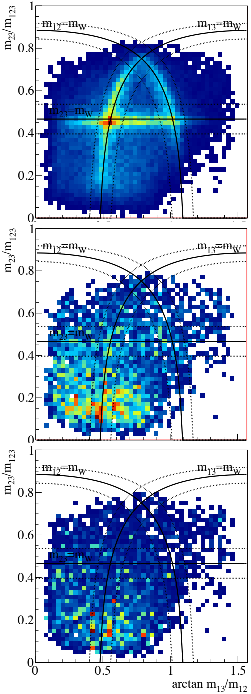
2– mass drop, cutoff $m_{\text{sub}} > 30$ GeV

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



Deterministic HEPTopTagger

Mass drop algorithm [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

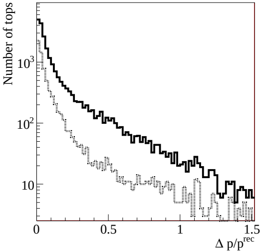
6– consistency condition $p_T^{(\text{tag})} > 200$ GeV

Top reconstruction

– direction less critical

– energy requiring calibration

$\Rightarrow$ standard experimental tool [unfortunately, I'm an obsolete design]



Multivariate HEPTopTagger2

OptimalR [Kasieczka, TP, Salam, Schell, Strebler]

- optimal fat jet size R_opt [large to include decay jets, small to avoid combinatorics]
- reduce R until we leave jet mass plateau

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

- estimate $R_\text{opt}^\text{(calc)}$ from kinematics
- $\{m_{123}, f_W, R_\text{opt} - R_\text{opt}^\text{(calc)}\}$ [$f_W = \text{opt}|m_{ij}/m_{123} - m_W/m_t|$]

Multivariate HEPTopTagger2

OptimalR [Kasieczka, TP, Salam, Schell, Strebler]

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

– reduce R until we leave jet mass plateau

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

– estimate $R_\text{opt}^\text{(calc)}$ from kinematics

– $\{m_{123}, f_W, R_\text{opt} - R_\text{opt}^\text{(calc)}\}$ [$f_W = \text{opt}|m_{ij}/m_{123} - m_W/m_t|$]

N-Subjettiness [Thaler, van Tilburg]

– also include rejected events [$R_\text{filt} = 0.2, N_\text{filt} = 5$]

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

Multivariate HEPTopTagger2

OptimalR [Kasieczka, TP, Salam, Schell, Strebler]

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

– reduce R until we leave jet mass plateau

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

– estimate $R_\text{opt}^\text{(calc)}$ from kinematics

– $\{m_{123}, f_W, R_\text{opt} - R_\text{opt}^\text{(calc)}\}$ [$f_W = \text{opt}|m_{ij}/m_{123} - m_W/m_t|$]

N-Subjettiness [Thaler, van Tilburg]

– also include rejected events [$R_\text{filt} = 0.2, N_\text{filt} = 5$]

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

Qjets [Ellis, Hornig, Roy, Krohn, Schwartz]

– more than one clustering history, weighted by

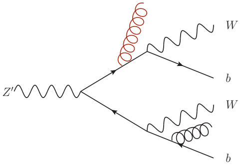
$$\omega_{ij} = \exp\left[-\alpha \frac{y_{ij} - y_{ij}^\text{min}}{y_{ij}^\text{min}}\right]$$

then using distributions like $\langle m^2 \rangle - \langle m \rangle^2$

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

Multivariate HEPTopTagger2

Top resonance search [Kasieczka, TP, Salam, Schell, Strebler]



– Z' search without ISR/FSR
– FSR major problem

Multivariate HEPTopTagger2

Top resonance search [Kasieczka, TP, Salam, Schell, Strebler]

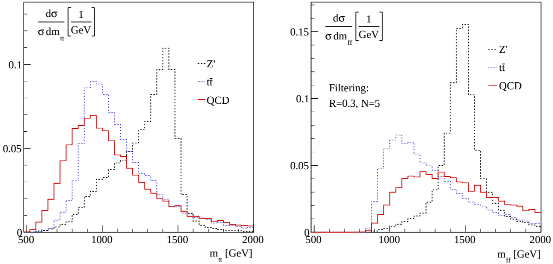
– Z' search without ISR/FSR

– FSR major problem

– using reconstructed 4-momentum of tagged top

– tail in $m_{tt}^{(\text{rec})}$ from tagged momentum and FSR

$\Rightarrow$ add 4-momentum of fat jet: $\{m_{tt}, p_{T,t}, m_{jj}^{(\text{filt})}, p_{T,j}^{(\text{filt})}\}$



Multivariate HEPTopTagger2

Top resonance search [Kasieczka, TP, Salam, Schell, Strebler]

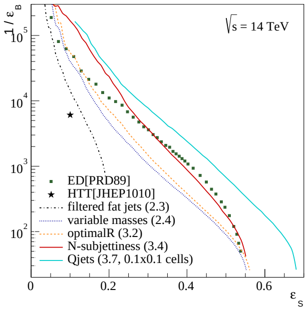
- Z' search without ISR/FSR
- FSR major problem
- using reconstructed 4-momentum of tagged top
- tail in $m_{tt}^{(\text{rec})}$ from tagged momentum and FSR

$\Rightarrow$ add 4-momentum of fat jet: $\{m_{tt}, p_{T,t}, m_{jj}^{(\text{filt})}, p_{T,j}^{(\text{filt})}\}$

Fully multi-variate HEPTopTagger2

- accepting Spanno's SD challenge
- Z' to two tagged tops
- tagging highly complex

$\Rightarrow$ deterministic QCD-based taggers terminated



Jet images

Applied image regognition

- wavelet transformation [Rentala, Shepherd, Tait; Monk]
- W-tagging with image recognition [Cogan etal, Oliveira etal, Baldi etal]
- top-tagging attempt [Almeida, Backovic, Cliche, Lee, Perelstein]
- QCD and shower study [Barnard etal]
- quark-gluon discrimination including tracks [Komiske etal]

⇒ new avenue in jet physics

Jet images

Applied image regognition

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

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

– top-tagging attempt [Almeida, Backovic, Cliche, Lee, Perelstein]

– QCD and shower study [Barnard etal]

– quark-gluon discrimination including tracks [Komiske etal]

$\Rightarrow$ new avenue in jet physics

Questions waiting to be answered

– what is the best input format?

– how can be use the tracker resolution?

– what is the training data?

– what does the neutal net learn?

– how much of this is soft activity?

– how does it compare to QCD tools? [judgement day is inevitable]

– are there better/alternative approaches? [the future has not been written]

Jet images

Start with anti-k_T fat jet $[p_T = 350 \dots 450$ GeV, $R = 1.5]$

– shift move image to center the global maximum

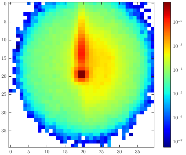
– rotation rotate the second maximum to 12 o'clock

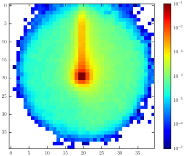
– flip ensure third maximum is in the right half-plane

– crop crop the image to 40×40 pixels

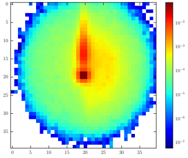
– decide on E vs E_T for rapidities $\eta \gtrsim 2$

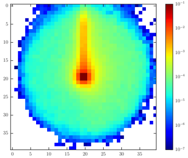
$\Rightarrow$ pre-processing strictly necessary?





Jet images





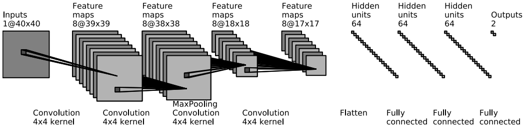
Start with anti-k_T fat jet [$p_T = 350 \dots 450$ GeV, $R = 1.5$]

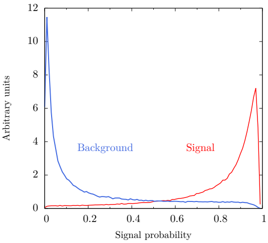
- shift move image to center the global maximum
- rotation rotate the second maximum to 12 o'clock
- flip ensure third maximum is in the right half-plane
- crop crop the image to 40×40 pixels
- decide on E vs E_T for rapidities $\eta \gtrsim 2$

$\Rightarrow$ pre-processing strictly necessary?

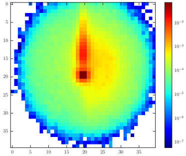
Set up network [Kasieczka, TP, Russell, Schell]

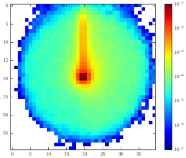
- run on 2-D jet images
- binning through calorimeter resolution [$\Delta\eta = 0.1$ vs $\Delta\phi = 5^\circ$]
- 150k events for training
- analyze geometric patterns [CNN]
- output: signal or background





Jet images





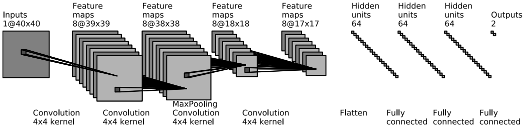
Start with anti-k_T fat jet [$p_T = 350...450$ GeV, $R = 1.5$]

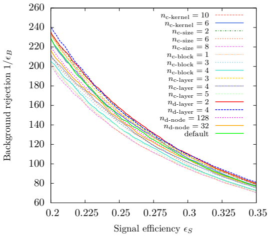
- shift move image to center the global maximum
- rotation rotate the second maximum to 12 o'clock
- flip ensure third maximum is in the right half-plane
- crop crop the image to 40×40 pixels
- decide on E vs E_T for rapidities $\eta \gtrsim 2$

$\Rightarrow$ pre-processing strictly necessary?

Set up network [Kasieczka, TP, Russell, Schell]

- run on 2-D jet images
- binning through calorimeter resolution [$\Delta\eta = 0.1$ vs $\Delta\phi = 5^\circ$]
- 150k events for training
- analyze geometric patterns [CNN]
- output: signal or background





Convolutional network

Getting binned calorimeter images ready [$n_\eta \times n_\phi$ energy bins]

- convolutional layers to transform into optimal basis with kernel matrix $\widetilde{W}$
- fully connected DNN weight function linking input and output

$$y_i = \max\left(0, \sum_{j=1}^{n^2} W_{ij} x_j + b_i\right)$$

Convolutional network

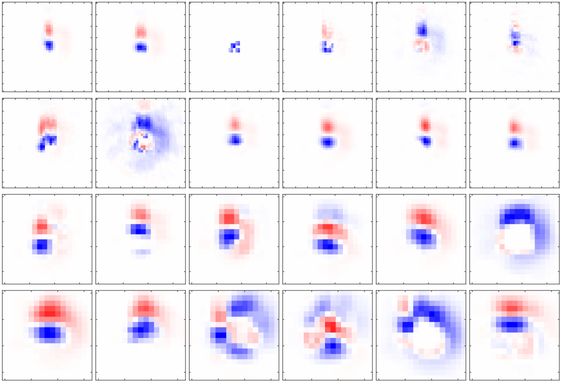
Getting binned calorimeter images ready [$n_\eta \times n_\phi$ energy bins]

– convolutional layers to transform into optimal basis with kernel matrix $\widetilde{W}$

– fully connected DNN weight function linking input and output

$$y_i = \max\left(0, \sum_{j=1}^{n^2} W_{ij} x_j + b_i\right)$$

– 4 convolutional layers with 8 feature maps [averaged S-B]



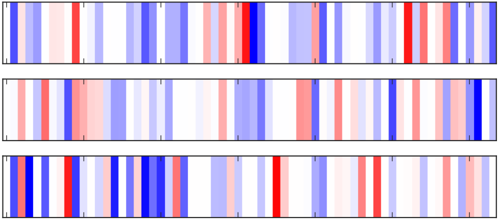
Convolutional network

Getting binned calorimeter images ready [$n_\eta \times n_\phi$ energy bins]

- convolutional layers   to transform into optimal basis with kernel matrix $\widetilde{W}$
- fully connected DNN   weight function linking input and output

$$y_i = \max\left(0, \sum_{j=1}^{n^2} W_{ij} x_j + b_i\right)$$

- 4 convolutional layers with 8 feature maps [averaged S-B]
- 3 linear DNN layers [averaged S-B]



Convolutional network

Getting binned calorimeter images ready [$n_\eta \times n_\phi$ energy bins]

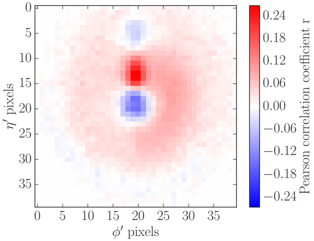
- convolutional layers — to transform into optimal basis with kernel matrix $\widetilde{W}$
- fully connected DNN — weight function linking input and output

$$y_i = \max\left(0, \sum_{j=1}^{n^2} W_{ij} x_j + b_i\right)$$

- 4 convolutional layers with 8 feature maps [averaged S-B]
- 3 linear DNN layers [averaged S-B]
- Pearson input-output correlation [pixel x vs label y]

$$r_{ij} = \frac{\sum_{\text{images}} (x_{ij} - \bar{x}_{ij})\,(y - \bar{y})}{\sqrt{\sum_{\text{images}} (x_{ij} - \bar{x}_{ij})^2}\sqrt{\sum_{\text{images}} (y - \bar{y})^2}}$$

$\Rightarrow$ not a perfect black box



DeepTop tagger

QCD-inspired references

– SoftDrop [separating soft from collinear; CMS default]

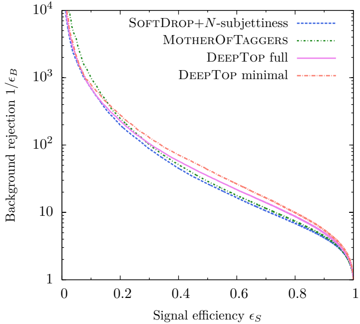
$$\frac{\min(p_{T1}, p_{T2})}{\Delta R_{12}^{\beta}} > z$$

– SoftDrop+N-subjettiness $\{m_{\text{sd}}, m_{\text{fat}}, \tau_2, \tau_3, \tau_2^{\text{sd}}, \tau_3^{\text{sd}}\}$

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

– DeepTop with and without pre-processing

⇒ slight performance gain through CNN



DeepTop tagger

QCD-inspired references

– SoftDrop [separating soft from collinear; CMS default]

$$\frac{\min(p_{T1}, p_{T2})}{\Delta R_{12}^{\beta}} > z$$

– SoftDrop+N-subjettiness $\{m_{\text{sd}}, m_{\text{fat}}, \tau_2, \tau_3, \tau_2^{\text{sd}}, \tau_3^{\text{sd}}\}$

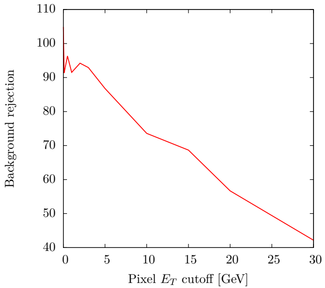
– MotherOfTaggers $\{m_{\text{sd}}, m_{\text{fat}}, m_{\text{rec}}, f_{\text{rec}}, \Delta R_{\text{opt}}, \tau_2, \tau_3, \tau_2^{\text{sd}}, \tau_3^{\text{sd}}\}$

– DeepTop with and without pre-processing

⇒ slight performance gain through CNN

Detector realities [30% signal efficiency]

– soft cutoff per calorimeter pixel



DeepTop tagger

QCD-inspired references

- SoftDrop [separating soft from collinear; CMS default]

$$\frac{\min(p_{T1}, p_{T2})}{\Delta R_{12}^{\beta}} > z$$

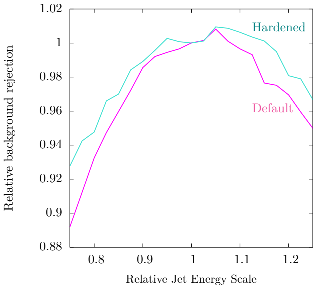
- SoftDrop+N-subjettiness $\{m_{\text{sd}}, m_{\text{fat}}, \tau_2, \tau_3, \tau_2^{\text{sd}}, \tau_3^{\text{sd}}\}$
- MotherOfTaggers $\{m_{\text{sd}}, m_{\text{fat}}, m_{\text{rec}}, f_{\text{rec}}, \Delta R_{\text{opt}}, \tau_2, \tau_3, \tau_2^{\text{sd}}, \tau_3^{\text{sd}}\}$
- DeepTop with and without pre-processing

⇒ slight performance gain through CNN

Detector realities [30% signal efficiency]

- soft cutoff per calorimeter pixel
- (sub-) jet energy scale $\pm 25\%$

⇒ CNN surprisingly stable



Checking on the Machine

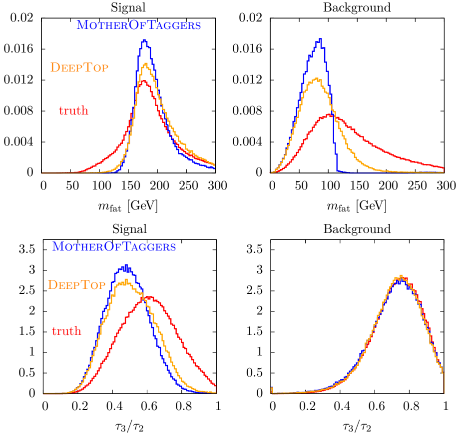
Typical reaction: ‘fuck you, you fucking machine’

– 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



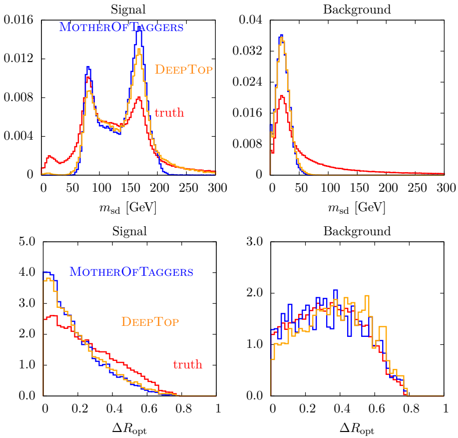
Checking on the Machine

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}

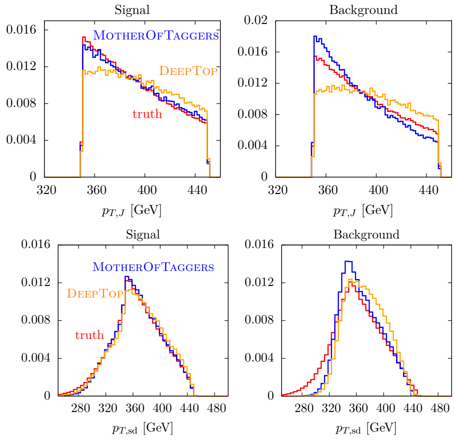


Checking on the Machine

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



Checking on the Machine

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

4- so what do you want checked?

$\Rightarrow$ HEPTop vs DeepTop: just aesthetic question...

Learning Minkowski

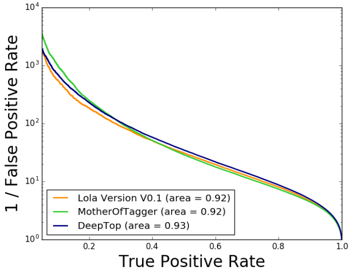
Back to QCD [Butter, Kasieczka, TP, Russell; see also Louppe etal, Pearkes etal]

- physicists are hackers, not users
- start with table of 4-momenta
- DNN layer combining momenta [CoLa]
- DNN layer introducing Lorentz-invariant Minkowski metric [LoLa]
- search for decays vs QCD

Learning Minkowski

Back to QCD [Butter, Kasieczka, TP, Russell; see also Louppe etal, Pearkes etal]

- physicists are hackers, not users
- start with table of 4-momenta
- DNN layer combining momenta [CoLa]
- DNN layer introducing Lorentz-invariant Minkowski metric [LoLa]
- search for decays vs QCD
- first, preliminary results promising



Outlook

Times are moving fast...

...deterministic taggers estalished/old/boring

...information beyond clustering history helps

...imagine recognition for subjets is cool

...DeepTop is not a black box

...maybe a machine can learn constituents and Minkowski metric