Jets in the 21st Century

Tilman Plehn

Universität Heidelberg

Annecy 11/2018

Rise of the Machines

All starting with a gedankenexperiment

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

1994 jet-algo top tagger for fun [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]

. . .

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 adversarial autoencoder [Heinel, Kasieczka, Plehn, Thompson; Shi etal]

Jet-level analyses (1990s)

Jets as analysis objects

- jet–parton duality $\Leftrightarrow$ what are partons in detector?
- crucial for any LHC analysis [really???]
- infrared safety the main issue

$\Rightarrow$ any of this true for LHC?

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

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

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

- (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)

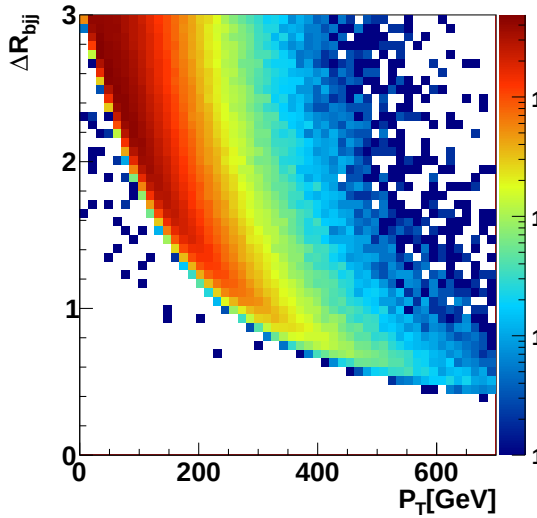
$\Rightarrow$ fat jets/substructure: use clustering history [k_T, C/A only]

Heavy resonance taggers (2000s)

Tagging boosted tops

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

$\Rightarrow$ substructure playground

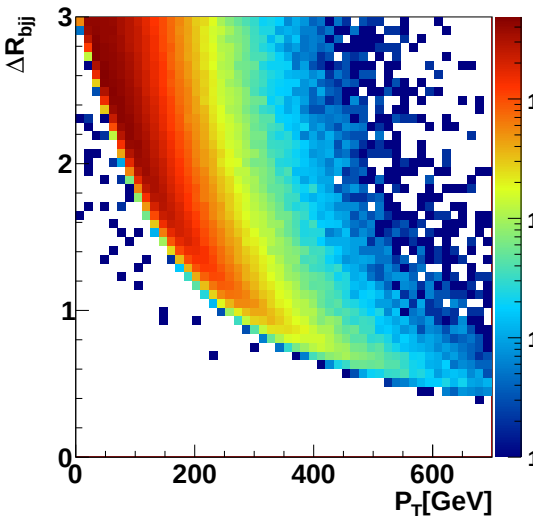


Heavy resonance taggers (2000s)

Tagging boosted tops

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

$\Rightarrow$ substructure playground



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

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

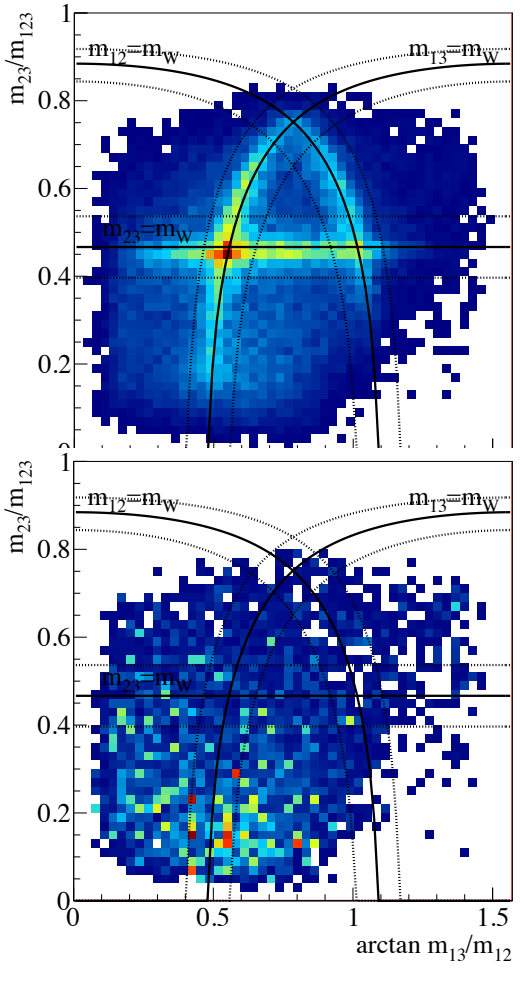
2– mass drop, cutoff $m_{\text{sub}} > 30$ GeV

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

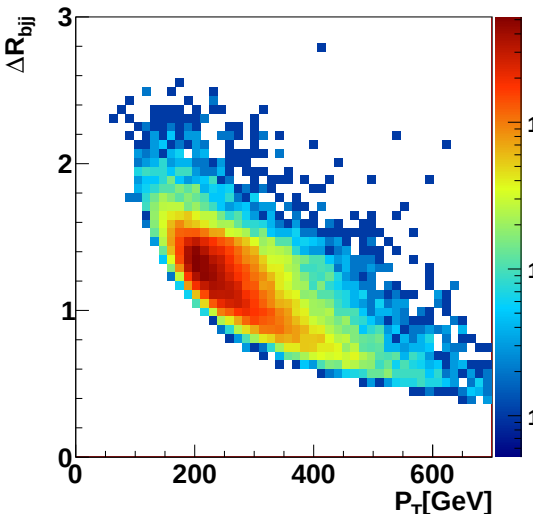


Heavy resonance taggers (2000s)

Tagging boosted tops

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

$\Rightarrow$ 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

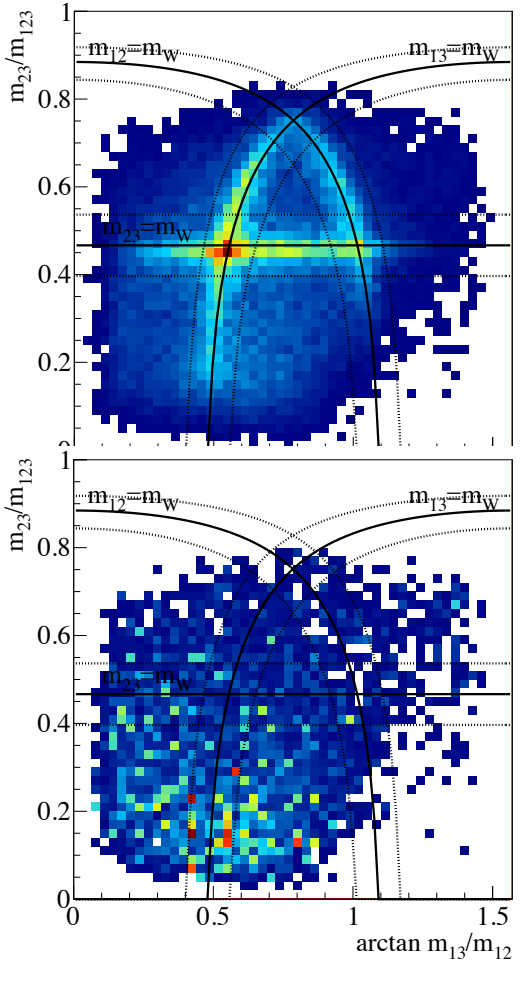
3– filtering leading to hard substructure triple

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

$\Rightarrow$ LHC break-through



Multi-variate top taggers (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 top taggers (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]

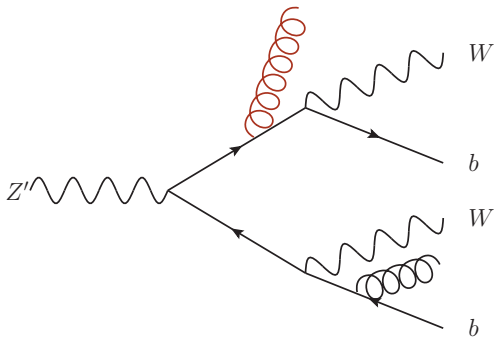
  $$|m_{123} - m_{123}^{(R_{\text{max}})}| < 0.2\, m_{123}^{(R_{\text{max}})} \quad \Rightarrow \quad R_{\text{opt}}$$

- 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

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



Multi-variate top taggers (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]

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

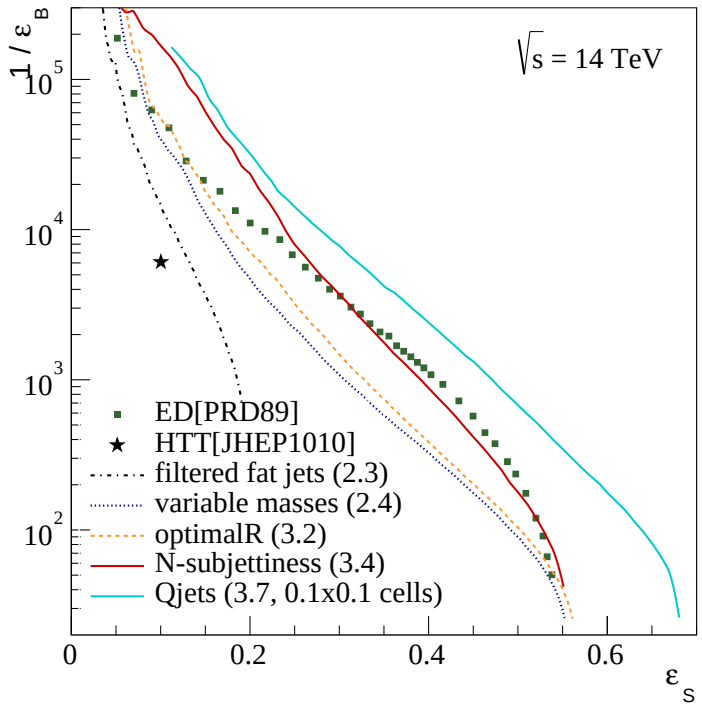
- add N-subjettiness [Thaler, van Tilburg]
- $\{m_{123}, f_W, R_\text{opt} - R_\text{opt}^{(\text{calc})}, \tau_j, \tau_j^{(\text{filt})}\}$

Fat jet and top kinematics

- FSR major problem for Z' search
- tag and reconstruction in each other's way

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

$\Rightarrow$ outperforming deterministic taggers

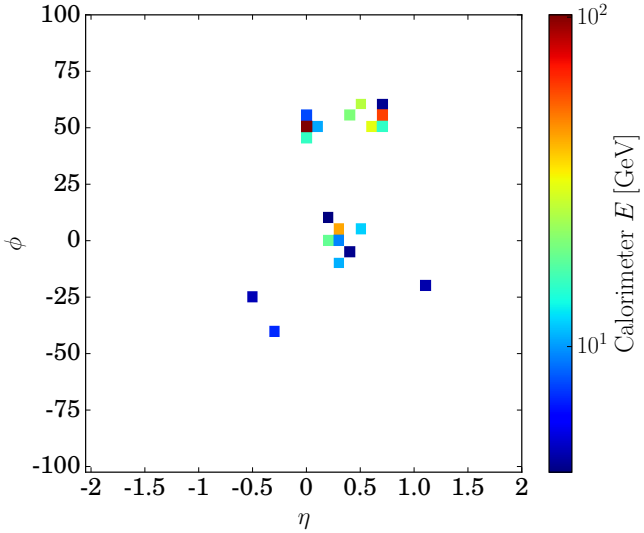


Jet images (2020s)

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

- why high-level observables as input?
- 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 tagging with tracks [Komiske etal]

$\Rightarrow$ new hammer for jet nails



Jet images (2020s)

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

- why high-level observables as input?
- 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 tagging with tracks [Komiske etal]

$\Rightarrow$ 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]

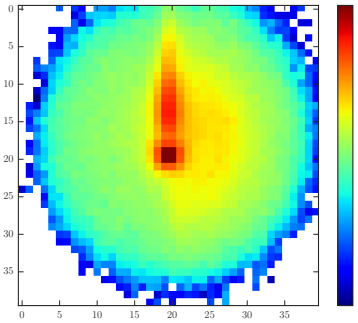
$\Rightarrow$ what does the network learn?

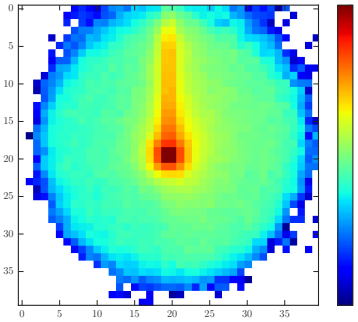
Jet images (2020s)

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

- why high-level observables as input?
- 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 tagging with tracks [Komiske etal]

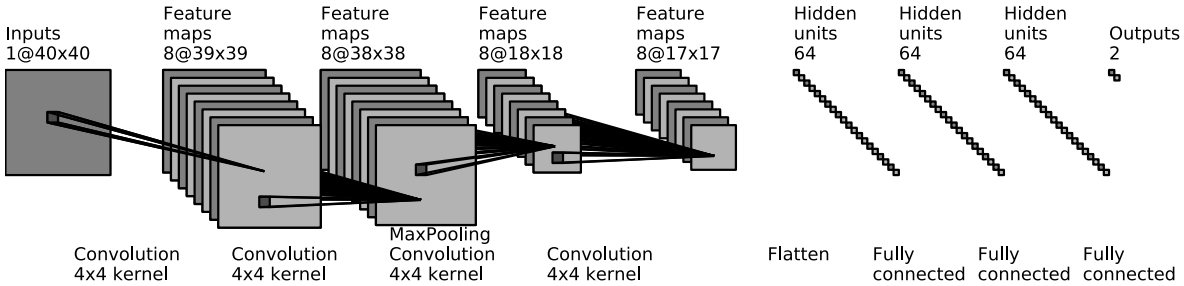
$\Rightarrow$ new hammer for jet nails

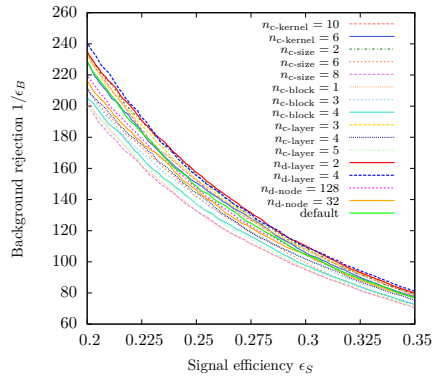




Neural network [Kasieczka, TP, Russell, Schell; Macaluso, Shih]

- run on 2-D jet images [$p_T = 350, \ldots, 450$ GeV]
- colored image as input
- binning through calorimeter resolution [$\Delta\eta = 0.1$ vs $\Delta\phi = 5^\circ$]
- analyze geometric patterns [Facebook: convolutional network]

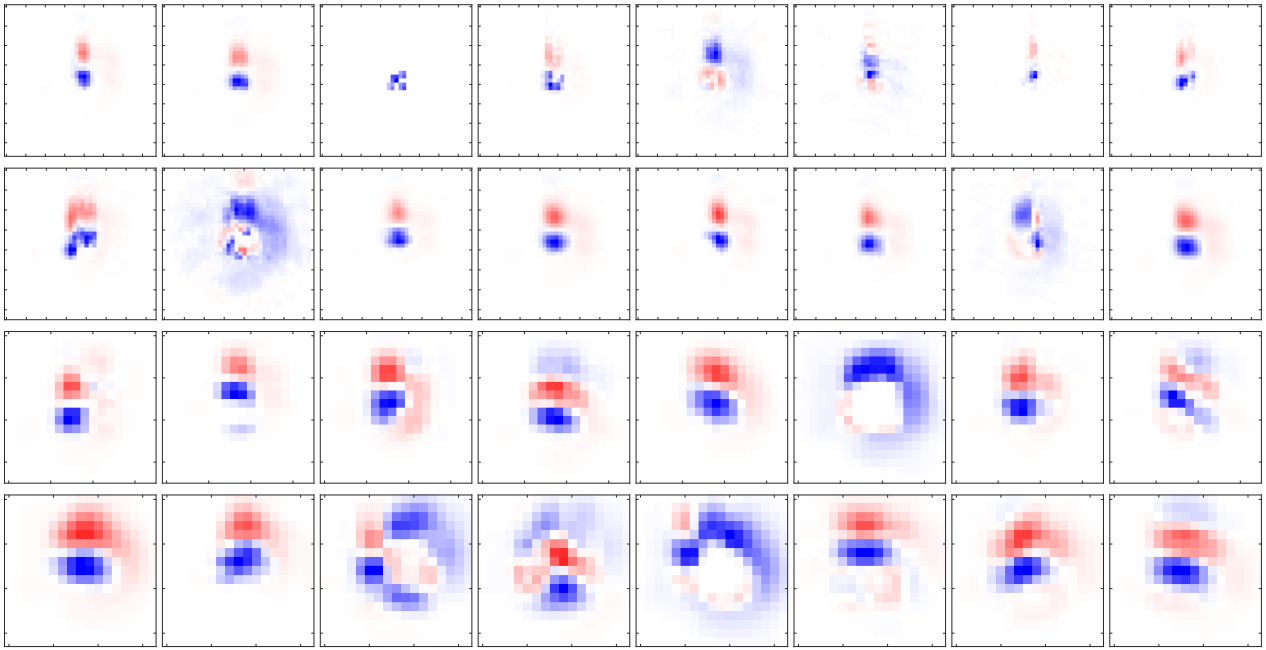


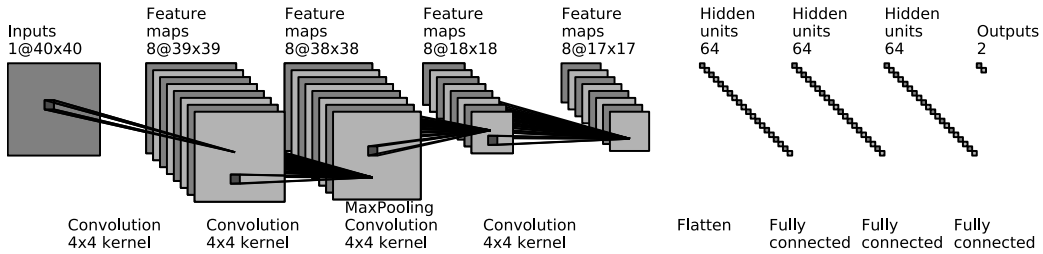


Inside DeepTop

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

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

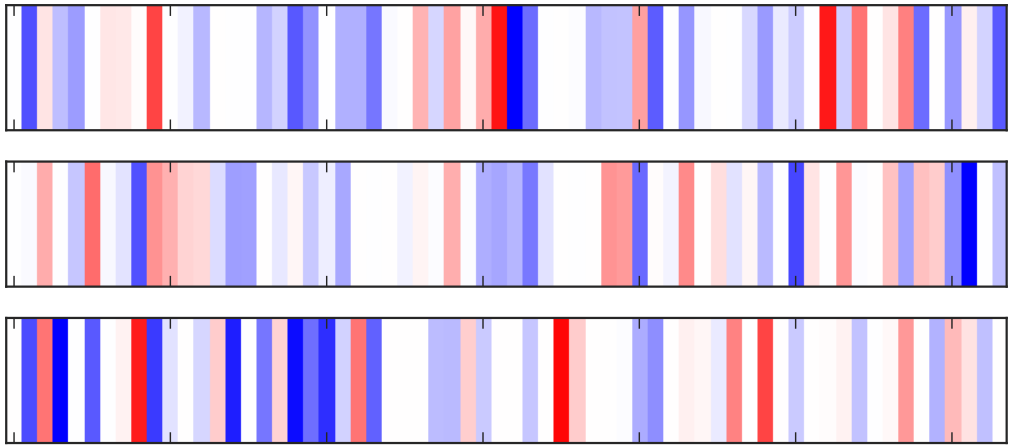


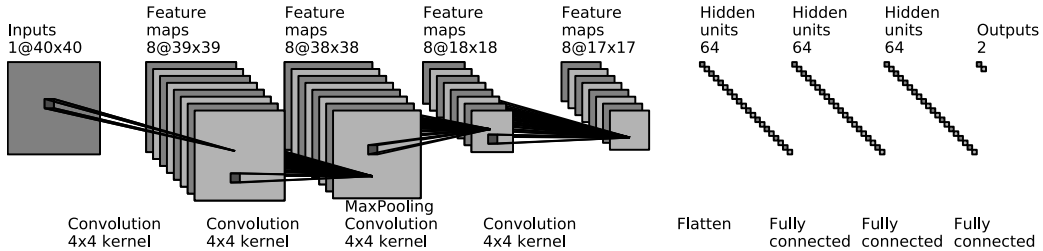


Inside DeepTop

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

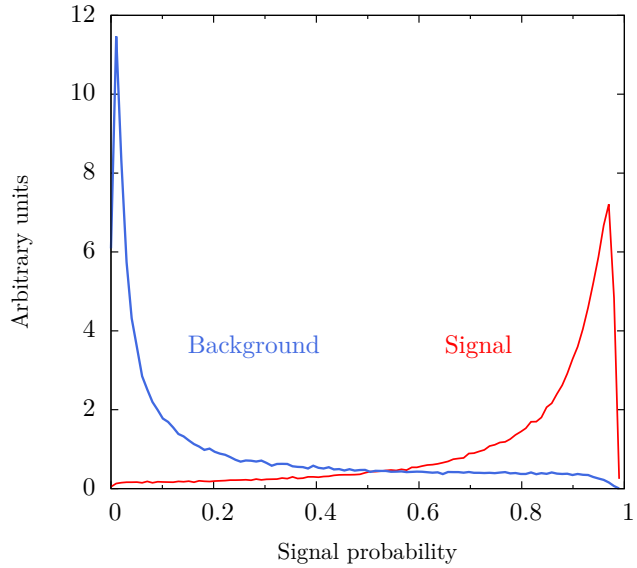


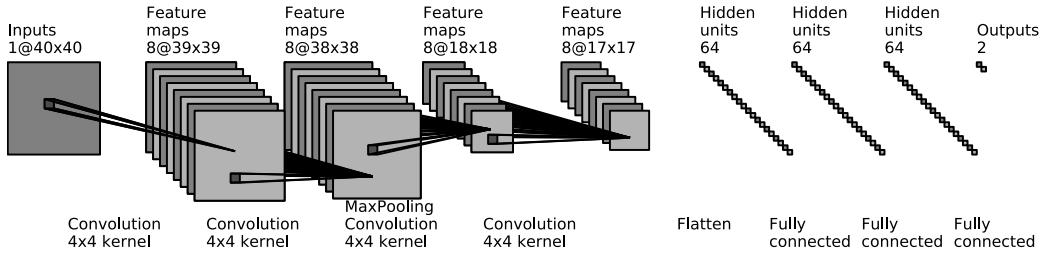


Inside DeepTop

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?]



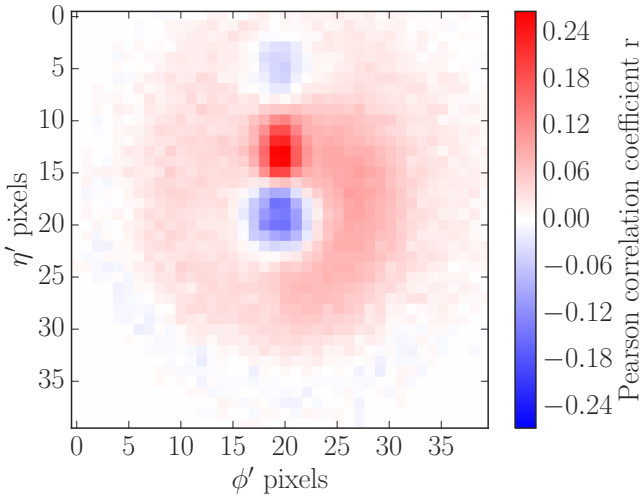


Inside DeepTop

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]

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



Inside DeepTop

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

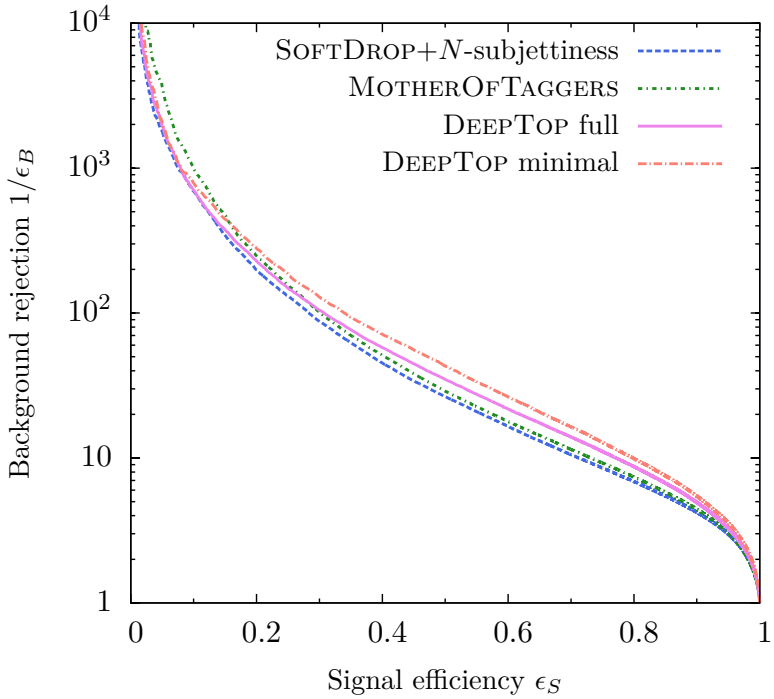
- 2+2 convolutional layers probing 2D structure with kernel matrix
- 3 fully connected layers weight function linking input and output
- signal-ness vs background-ness output [probability?]
- Pearson input-output correlation [pixel x vs label y]

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

- comparison to MotherOfTaggers

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

$\Rightarrow$ understandable performance gain

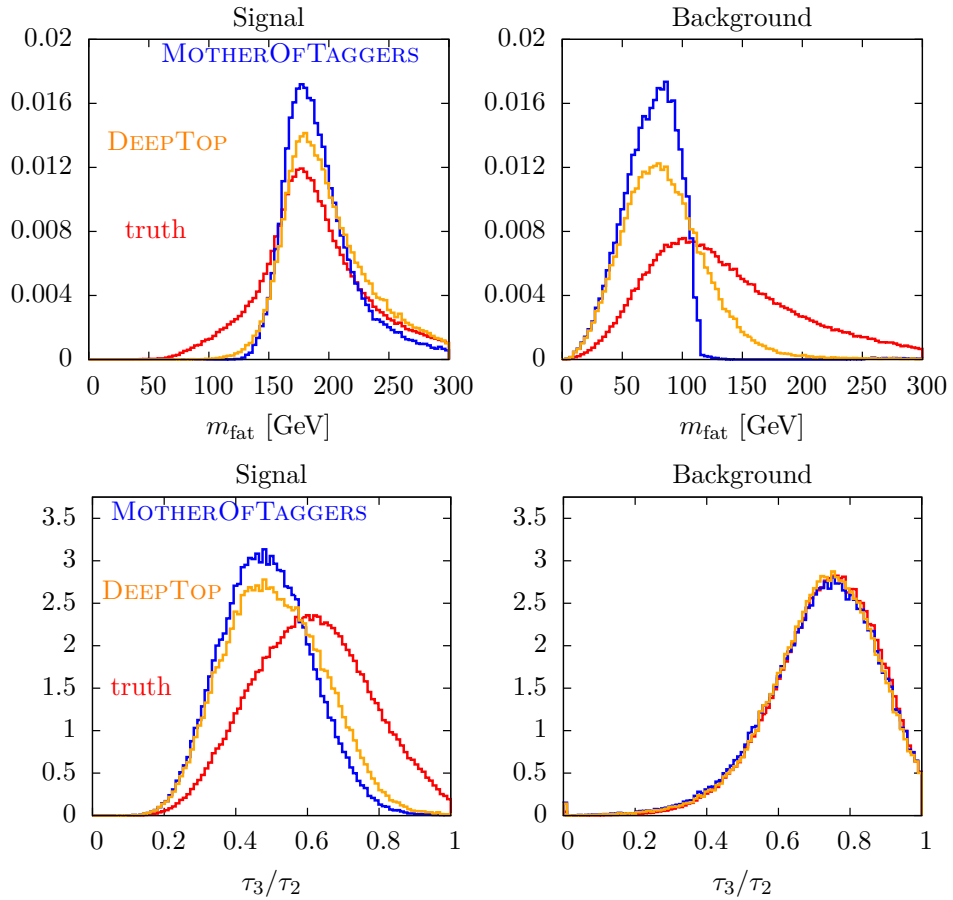


Just a new tool

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



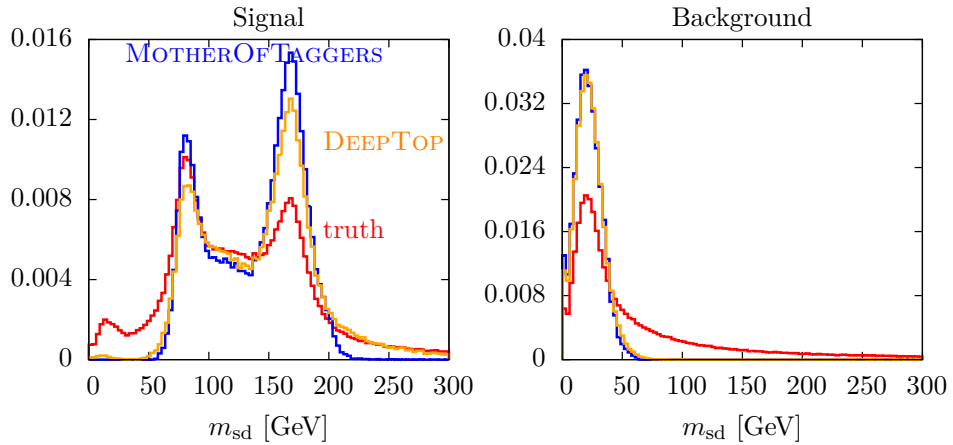
Just a new tool

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



Just a new tool

Typical reaction: 'Fuck you, you fucking machine'

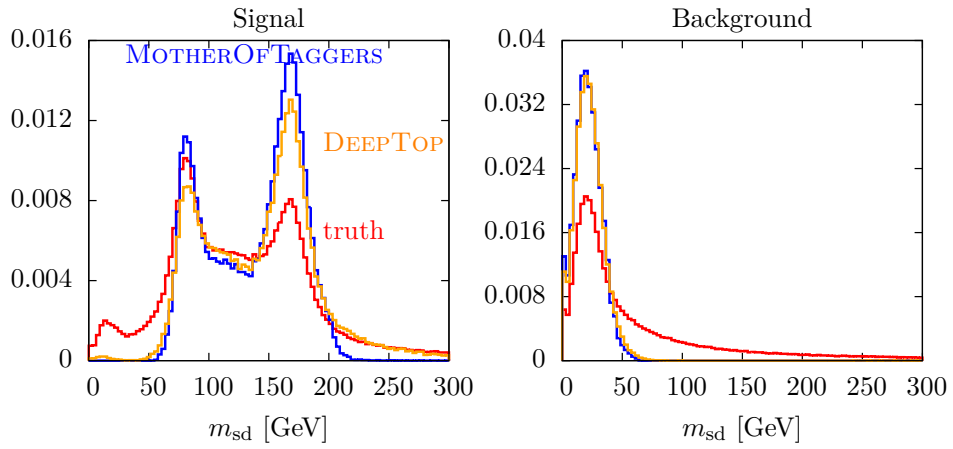
- in principle, full control for fully supervised learning
- lots of events in the grey zone
  but checks possible for correctly identified signal/background events
- compare truth vs MotherOfTaggers vs DeepTop

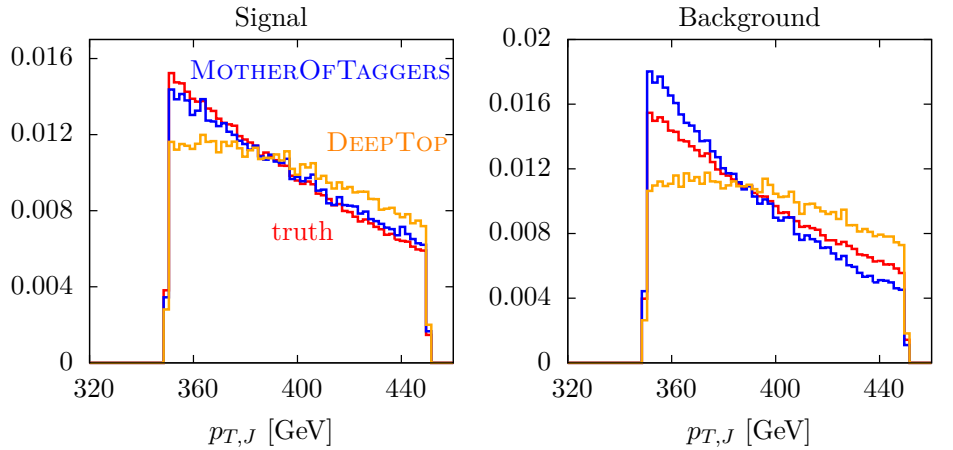
1- fat jet mass and N-subjettiness

2- soft drop mass

3- transverse momenta

$\Rightarrow$ it works and we know why

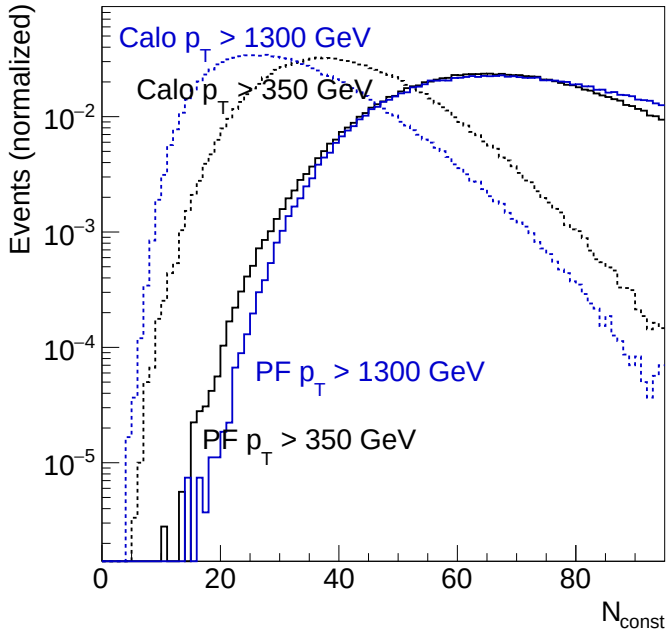




DeepTopLoLa

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

- 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; see also Louppe etal, Pearkes etal]

- 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; see also Louppe etal, Pearkes etal]

- 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; see also Louppe etal, Pearkes etal]

- 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

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

- 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; see also Louppe etal, Pearkes etal]

- 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

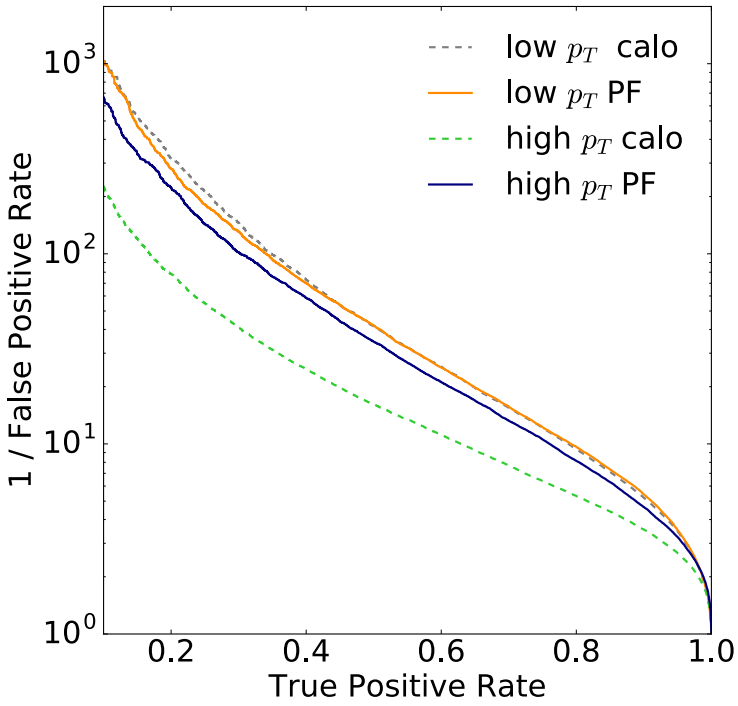
- DNN on Lorentz scalars

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

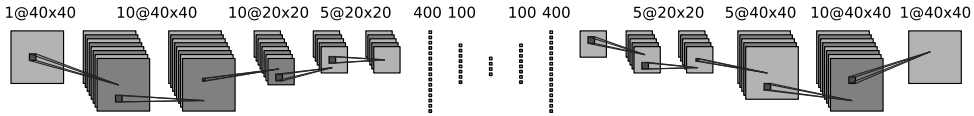
- fun: measure Minkowski metric

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

- extend to particle flow
- network with little but Lorentz invariance

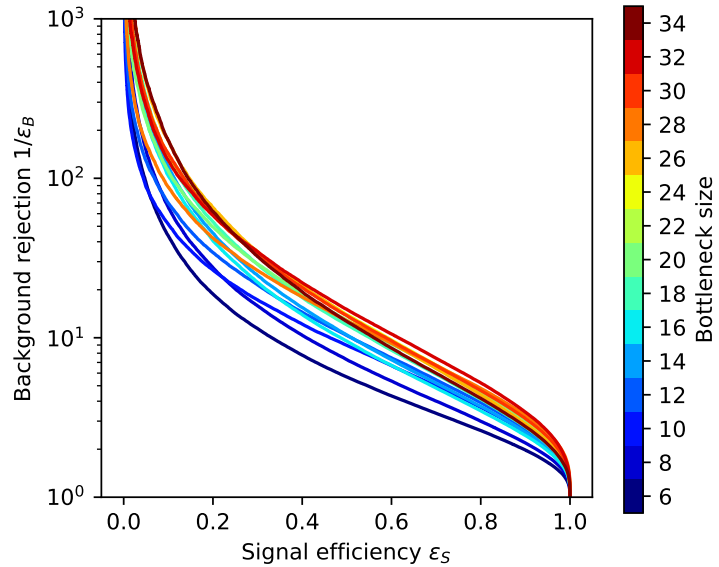


Getting inspired

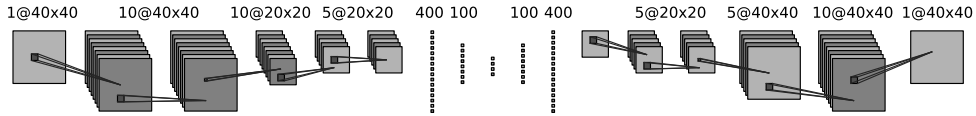


Anomaly search trained on background [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih]

- established 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 inspired

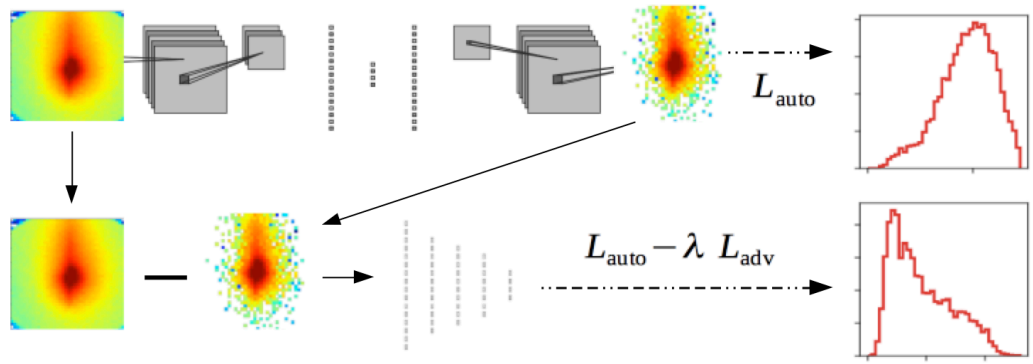


Anomaly search trained on background [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih]

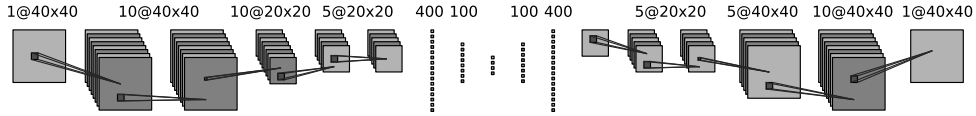
- established concept: autoencoder
- reconstruct typical QCD jet image from many QCD jets
- reduce weights in central layer
  compress information on ‘typical’
- search for outliers hard to describe
- benchmark on top jets, search for Higgs or dark showers

De-correlate background shaping

- established concept: adversary



Getting inspired

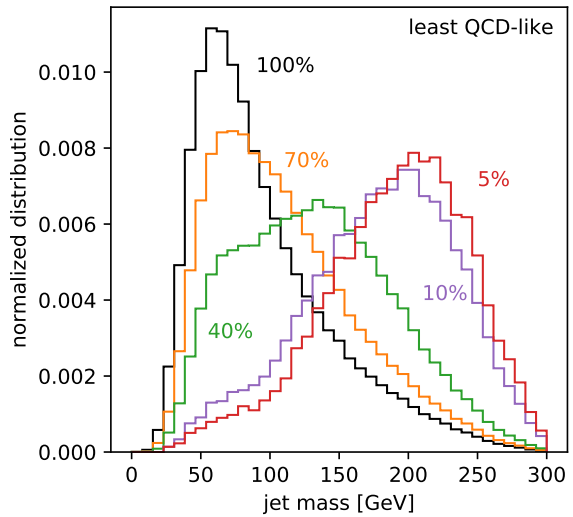


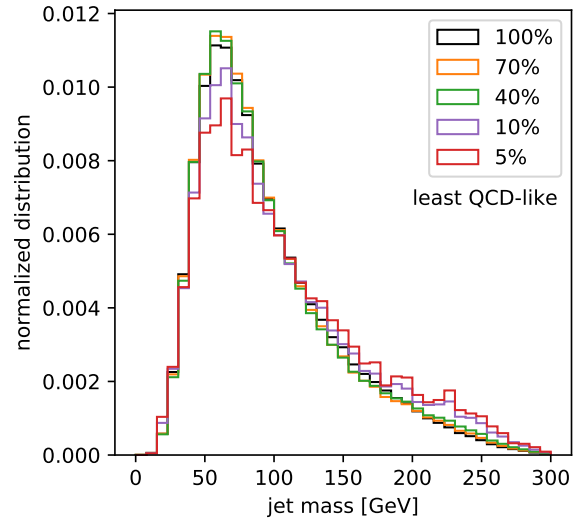
Anomaly search trained on background [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih]

- established 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 established/old/boring [was 2000s]

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

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

...DeepTop is not a black box

...DeepTopLoLa took Jackson

...Autoencoders open avenues

Join the fun!