DIY Uala al LITI

2000s Taggers

2010s Multi-variat

2020s Jet image

Autoencode

# Machine Learning in Particle Physics

Tilman Plehn

Universität Heidelberg

Durham 3/2019



Big data at LHC

2000s Taggers

2020s Jet images

DeepTop

tutocrio

# Why LHC? Why jets?

#### Data from ATLAS & CMS

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





Big data at LHC

2000s Taggers

2020s Jet image

Autoence

# Why LHC? Why jets?

#### Data from ATLAS & CMS

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

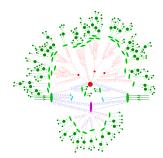
### Physics in jets

- re-summed perturbative QFT prediction from QCD
- jets as decay products

67% 
$$W \rightarrow jj$$
 70%  $Z \rightarrow jj$  60%  $H \rightarrow jj$  67%  $t \rightarrow jjj$  60%  $\tau \rightarrow j ...$ 

- new physics in 'dark showers'
- ⇒ It's interesting





Big data at LHC

2000s Taggers

2020s Jet images

Autoend

# Why LHC? Why jets?

#### Data from ATLAS & CMS

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

### Physics in jets

- re-summed perturbative QFT prediction from QCD
- jets as decay products

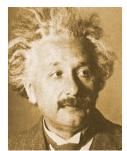
67% 
$$W \rightarrow jj$$
 70%  $Z \rightarrow jj$  60%  $H \rightarrow jj$  67%  $t \rightarrow jjj$  60%  $\tau \rightarrow j \dots$ 

- new physics in 'dark showers'
- ⇒ It's interesting

#### Monte Carlo data

- QCD simulation: Sherpa, Herwig [Pythia, Madgraph]
- data-to-data comparison: MC vs LHC
- ⇒ We can simulate it





Big data at LHC

# Why LHC? Why jets?

#### Data from ATLAS & CMS

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

### Physics in jets

- re-summed perturbative QFT prediction from QCD
- jets as decay products

67% 
$$W \rightarrow jj$$
 70%  $Z \rightarrow jj$  60%  $H \rightarrow jj$  67%  $t \rightarrow jjj$  60%  $\tau \rightarrow j \dots$ 

- new physics in 'dark showers'
- ⇒ It's interesting

#### Monte Carlo data

- QCD simulation: Sherpa, Herwig [Pythia, Madgraph]
- data-to-data comparison: MC vs LHC
- ⇒ We can simulate it





Big data at LHC

2000s Taggers

2020s Jet images

Autoend

# Why LHC? Why jets?

#### Data from ATLAS & CMS

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

### Physics in jets

- re-summed perturbative QFT prediction from QCD
- jets as decay products

67% 
$$W \rightarrow jj$$
 70%  $Z \rightarrow jj$  60%  $H \rightarrow jj$  67%  $t \rightarrow jjj$  60%  $\tau \rightarrow j \dots$ 

- new physics in 'dark showers'
- ⇒ It's interesting

#### Monte Carlo data

- QCD simulation: Sherpa, Herwig [Pythia, Madgraph]
- data-to-data comparison: MC vs LHC
- ⇒ We can simulate it





Big data at LHC

2000s Taggers

2020s Jet image

Autoent

### Why not LHC?

#### ATLAS & CMS

- 3000 know-it-alls per experiment
- strong top-down structures
- strongly organized analysis groups
- ⇒ Small groups driving innovation

#### Expertize

- LHC data format: ROOT
- multi-variate analyses tool: TMVA
- Tensorflow from TMVA/ROOT
- ML challenges running
- ⇒ Little sense of ML-urgency

### Experiment vs theory

- theorists linked to lack of team compatibility
- simulated data as good as actual data
- excellent personal ex-th connections
- ⇒ Theory driving non-theory developments

### What is 7MVA

- One framework for most common MVA-techniques, available in F
  - Have a common platform/interface for all MVA classification and regression-
  - Have common data pre-processing capabilities
  - Train and test all classifiers on same data sample and evaluate consistently
     was a good idea 10year ago, now unfortunatly imposes some unnecessa constraints but nothing which could not be dealt with by 'running independ
  - Provide common analysis (ROOT scripts) and application framework
  - Provide access with and without ROOT, through macros, C++ executables of
- Integrated and distributed with ROOT
- some info is still located at its original sourceforge location
  - Home page ......http://tmva.sf.net/
  - list of classifier options ... <a href="http://tmva.sourceforge.net/optionRef.ht">http://tmva.sourceforge.net/optionRef.ht</a>
  - Mailing BlueYonder Retail Solutions Customers Company





Tilman Plehn

Big data at LHC

2000s Tagger

2010s Multi-vai

2020s Jet

DeepTop

### Jets story's starting point: Nothing is ever new

#### LHC visionaries

- 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 elecent 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<sup>+</sup>e<sup>-</sup> 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.





Tilman

Big data at LHC

2000s Tagger

2020s Jet image

DoopTop

Autoencod

### Jets story's starting point: Nothing is ever new

#### LHC visionaries

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

1994: jet-algorithm W/top-tagger [Seymour]

#### USING NEURAL NETWORKS TO IDENTIFY JETS

Leif LÖNNBLAD\*, Carsten PETERSON\*\* and Thorsteinn RÖGNVALDSSON\*\*\*

Department of Theoretical Physics, University of Lund, Sölvegatan 14A, S-22362 Lund, Sweden



A neural network method fo is to find an efficient mapping by quark-gluon identity. This is don functions using a gradient descer network. With this method we ar Carlo generated e\*c\* events v model used. This approach for i effect.

# Searches for new particles using cone and cluster jet algorithms: a comparative study

Michael H. Seymour

Department of Theoretical Physics, University of Lund, Sölvegatan 14A, S-22362 Lund, Sweden

Received 18 June 1993; in revised form 16 September 1993

Abstract. We discuss the reconstruction of the hadronic decays of heavy particles using jet algorithms. The ability to reconstruct the mass of the decaying particle is compared between a traditional cone-type algorithm. The specific camples considered are the semileptonic decays of a heavy Higgs boson at  $\sqrt{s} = 16 \, \text{TeV}$ , and of top quark-antiquark pairs at  $\sqrt{s} = 18 \, \text{TeV}$ . We find that the cluster algorithm offers considerable advantages in the former case, and a sight advantage in the latter. We are the special considerable advantages in the special considerable and the special considerable and show that a typical resolution dilutes these advantages, but does not remove them entire. except that the invariant mass of a pair is replaced by the transverse momentum of the softer particle relative to the other.

More recently, this algorithm was extended to collisions with incoming hadrons [5], and a longitudinallyinvariant k<sub>2</sub>-clustering algorithm for hadron-hadron collisions was proposed [6]. This algorithm has been compared with the more commonly used cone algorithm from the viewonits of a parton-shower. Monte Carlo

program [6, 7], and a fixed-order matris lation [8], and advantages of the cluster reported in both cases. This paper is a comparison between the algorithms reconstructing the hadronic decays of which was also studied in a preliminary The only as-yet unobserved particles

Standard Model are the top quark and H search for, and study of, these particles most important goals of current and p hadron collider experiments. In both cas





~ 1970: People with visions should see a doctor [Helmut Schmidt, wrong for once]

Big data at LHC

2010s Multi-vari

DeepTop

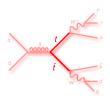
Fat jet taggers (2000s)

Look what makes jets [Pre-LHC, jet were just annoying]

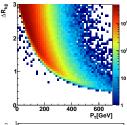
- top jets from  $t \rightarrow bq\bar{q}'$  vs QCD jets
- top decays well-defined in theory
- labelled sample: semileptonic  $t\bar{t}$  events
- ⇒ LHC physics playground

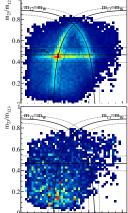
Simple top tagging [BDRS; TP, Salam, Spannowsky, Takeuchi]

- 1– fat jet with  $p_T > 200 \text{ GeV}$
- 2- filtering defining 3-5 decay jets
- 3- top mass window  $m_{123} = [150, 200]$  GeV
- 4- mass plane cuts extracting  $m_{ii} \approx m_W$
- ⇒ Not rocket science, but crucial to build trust









arctan m13/m12

Big data at LHC

2010s Multi-variate

DeepTop

Autoencoder

### Multi-variate taggers (2010s)

### Developing the benchmark

- multivariate analysis generally old news multivariate tagger to keep up with shower deconstruction [Soper, Spannowsky]
- optimal fat jet size R<sub>opt</sub> [large to decay jets, small to avoid combinatorics, compute from kinematics]

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

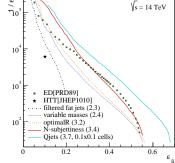
- add N-subjettiness [Thaler, van Tilburg]
- $-\{m_{123}, f_W, R_{\text{opt}} R_{\text{opt}}^{(\text{calc})}, \tau_j, \tau_j^{(\text{filt})}\}$

### Fat jet and top kinematics

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

$$-\{..., m_{tt}, p_{T,t}, m_{ii}^{(filt)}, p_{T,i}^{(filt)}\}$$

⇒ Performance increase, as expected





Big data at LH

2000s Taggers

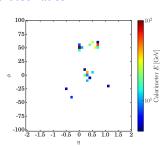
2020s Jet images

Autooncodor

### Jet image machines (2020s)

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

- why intermediate high-level variables?
- learn theory through more NN layers
- calorimeter output as image
- ⇒ Deep learning = modern networks on low-level observables





Big data at LHC

2000s Taggers

2020s Jet images

Autoend

Jet image machines (2020s)

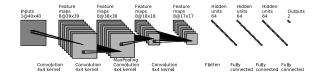
The natural next step [Cogan etal, Oliveira, Nachman etal, Baldi, Whiteson etal (2014/1!

- why intermediate high-level variables?
- learn theory through more NN layers
- calorimeter output as image
- ⇒ Deep learning = modern networks on low-level observables



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

- image recognition standard ML task
- rapidity vs azimuthal angle, colored by energy deposition
- top tagging on 2D jet images
- 40 × 40 bins through calorimeter resolution





Big data at LHC

2000s Taggers

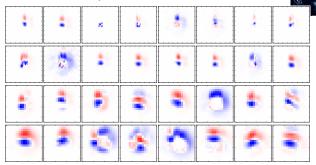
2020s Jet

Autoencod

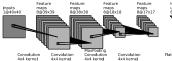
# Inside DeepTop

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

- 2+2 convolutional layers



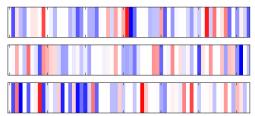




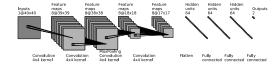


# Inside DeepTop

- 2+2 convolutional layers
- 3 fully connected layers









Big data at LH

2000s Taggers

2020s Jet imag

DeepTop

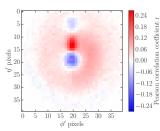
Autoencode

# Inside DeepTop

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

$$r_{ij}pprox\sum_{ ext{images}}\left(x_{ij}-ar{x}_{ij}
ight)\left(y-ar{y}
ight)$$







Big data at LHC

2010s Multi-varia

DeepTop

Autoencode

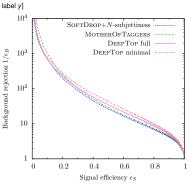
# Inside DeepTop

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

$$r_{ij}pprox\sum_{ ext{images}}\left(x_{ij}-ar{x}_{ij}
ight)\left(y-ar{y}
ight)$$

- comparison to MotherOfTaggers BDT
- ⇒ Understandable performance gain







Big data at LH

2010s Multi-varia

DeepTop

Autoencod

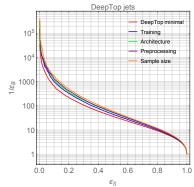
# Inside DeepTop

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

$$r_{ij}pprox\sum_{ ext{images}}\left(x_{ij}-ar{x}_{ij}
ight)\left(y-ar{y}
ight)$$

- comparison to MotherOfTaggers BDT
- ⇒ Understandable performance gain







### Inside DeepTop

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

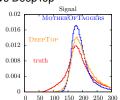
- 2+2 convolutional layers
- 3 fully connected layers
- Pearson input-output correlation [pixel x vs label y]

$$r_{ij}pprox\sum_{ ext{images}}\left(x_{ij}-ar{x}_{ij}
ight)\left(y-ar{y}
ight)$$

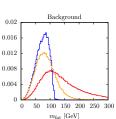
- comparison to MotherOfTaggers BDT
- ⇒ Understandable performance gain

### Typical reaction: 'F\*\*\* you, you f\*\*\*ing machine'

- full control for supervised learning easy checks for correctly identified signal/background
- MC truth vs MotherOfTaggers vs DeepTop
  - fat jet mass N-subjettiness transverse momenta
- ⇒ It works and we know why



 $m_{\text{fat}}$  [GeV]







DeepTop

### Grand theory ideas

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

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

### Inspired by jet algorithm — combination layer

input 4-vectors

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

- on-shell conditions for top tag

$$\begin{array}{c} - \text{ combined 4-vectors} \\ k_{\mu,i} \xrightarrow{\text{Col.a}} \widetilde{k}_{\mu,j} = k_{\mu,i} \ \textit{C}_{ij} \end{array} \qquad C = \begin{pmatrix} 1 & 0 & \cdots & 0 & \textit{C}_{1,N+2} & \cdots & \textit{C}_{1,M} \\ 0 & 1 & & \vdots & \textit{C}_{2,N+2} & \cdots & \textit{C}_{2,M} \\ \vdots & \vdots & \ddots & 0 & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & \textit{C}_{N,N+2} & \cdots & \textit{C}_{N,M} \end{pmatrix}$$

⇒ Physics step, easy to interpret



g data at LHC

2000s Taggers

2020s Jet image:

DoopTop

Autoencode

# Grand theory ideas

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

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

### Inspired by jet algorithm — combination layer

- input 4-vectors  $(k_{\mu,i})$
- on-shell conditions for top tag
- combined 4-vectors  $k_{\mu,i} \stackrel{\text{CoLa}}{\longrightarrow} \widetilde{k}_{\mu,j} = k_{\mu,i}$  10<sup>3</sup>
- ⇒ Physics step, easy to interpret

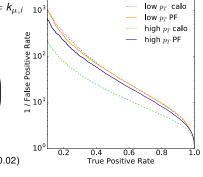
### Inspired by Jackson — Lorentz layer

- DNN on Lorentz scalars

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

⇒ Learn Minkowski metric

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





# Meet the professionals

#### A brief history of hurry

- 2014/15: first jet image papers
- 2017: first (working) ML top tagger
- MI 4 lets 2017: what architecture best
- ML4Jets 2018: Lots of architectures work [1902.09914]

SciPost Physics

Submission

#### The Machine Learning Landscape of Top Taggers

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

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.

A CONTRACTOR OF THE PARTY OF TH		
STEP STEP	Jü.	

nt	er	ıt	

References

C	onter	nt	
1	3		
2	Dat	a set	4
3	Tag	gers	4
	3.1	Imaged-based taggers	4
		3.1.1 CNN	5
		3.1.2 ResNeXt	5
	3.2	4-Vector-based taggers	6
		3.2.1 TopoDNN	6
		3.2.2 Multi-Body N-Subjettiness	7 7
		3.2.3 Recurrent Networks	
		3.2.4 P-CNN	8
		3.2.5 ParticleNet	8
	3.3	Theory-inspired taggers	9
		3.3.1 Lorentz Boost Network	9
		3.3.2 Lorentz Layer	10
		3.3.3 Energy Flow Polynomials	11
		3.3.4 Energy Flow Networks	11
		3.3.5 Particle Flow Networks	12
4 Comparison			13
5 Conclusion			16

#### Tilman

Big data at LHC

2000s Taggers

2020s lot image

DeepTo

Autoencode

### Meet the professionals

#### A brief history of hurry

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

SciPost Physics

Submission

#### The Machine Learning Landscape of Top Taggers

G. Kasicczka (ed)<sup>1</sup>, T. Plehn (ed)<sup>2</sup>, A. Butter<sup>2</sup>, D. Debnath<sup>3</sup>, M. Fairbairn<sup>4</sup>, W. Fedorko<sup>5</sup>, C. Gay<sup>5</sup>, L. Gouskos<sup>6</sup>, P. T. Komiska<sup>5</sup>, S. Leiss<sup>1</sup>, A. Lister<sup>5</sup>, S. Macaluso<sup>3</sup>, E. M. Hetodie<sup>7</sup>, L. Moore<sup>8</sup>, B. Nachman<sup>3</sup>, O.S. K. Nordströml<sup>1,1,2</sup>, J. Pearkes<sup>6</sup>, H. Qu<sup>6</sup>, Y. Rath<sup>13</sup>, M. Riegler<sup>13</sup>, D. Shih<sup>3</sup>, J. M. Thompson<sup>2</sup>, and S. Varme

1 Institut für Experimentalpkysik, Universität Hamburg, Germany 2 Institut für Horectrische Physik, Universität Heidelberg, Germany 3 NHECT, Dept. of Physics and Astronomy, Rutgers, The State University of NI, USA 4 Heroenteral Particle Physics and Cosmology, King's College London, United Kingdom 5 Department of Physics and Astronomy, The University of British Codumba, Canada 6 Control of Computer Science (Computer Science Computer Computer Science Computer Computer Science Computer Comput

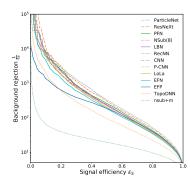
8 CP3, Université Catholique de Louvain, Louvain-la-Neuve, Belgium 9 Physics Division, Lawrence Berkeley National Laboratory, Berkeley, USA 10 Simons Inst. for the Theory of Computing, University of California, Berkeley, USA 11 National Institute for Substantie Physics (NIKHEP), Amsterdam, Netherlands 12 LPTHE, CNRS & Sorbonne Université, Paris, France 13 III, Physics Institute A WIFH Authen 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.





Big data at LHC

2010s Multi-varia

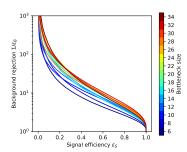
Autoencode

### New analysis ideas



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

- anomaly searches, only training on 'background'
- established ML concept: autoencoder
- reconstruct typical QCD jet image from many QCD jets reduce weights in central layer, compress information to 'typical'
- search for outliers hard to describe
- ⇒ Making an okay tagger





ig data at LH0

2000s Taggers

2020s Jet images

Autoencoder

### New analysis ideas

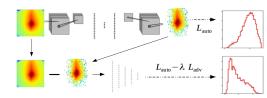


Fully supervised classification boring [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih]

- anomaly searches, only training on 'background'
- established ML concept: autoencoder
- reconstruct typical QCD jet image from many QCD jets reduce weights in central layer, compress information to 'typical'
- search for outliers hard to describe
- ⇒ Making an okay tagger

### De-correlate background shaping

- established concept: adversary [also see Englert, Galler, Harris, Spannowsky]





New analysis ideas



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

- anomaly searches, only training on 'background'

established ML concept: autoencoder

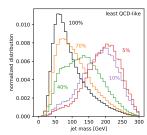
 reconstruct typical QCD jet image from many QCD jets reduce weights in central layer, compress information to 'typical'

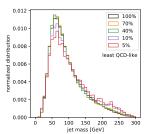
- search for outliers hard to describe

⇒ Making an okay tagger

### De-correlate background shaping

- established concept: adversary [also see Englert, Galler, Harris, Spannowsky]
- atypical QCD jets typially with large jet mass remove jet mass from network training







Autoencoder



Autoencoder

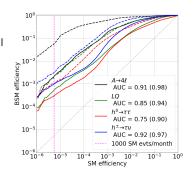
New analysis ideas

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

- anomaly searches, only training on 'background'
- established ML concept: autoencoder
- reconstruct typical QCD jet image from many QCD jets reduce weights in central layer, compress information to 'typical'
- search for outliers hard to describe
- ⇒ Making an okay tagger

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

- same thing on full events
- training data a problem
- variational autoencoder more powerful
- ⇒ Proof of concept...





Big data at LHC

2010s Multi-variati 2020s Jet images

Autoencoder

### The future

### Times are moving fast...

- ...LHC physics really is big data
- ...imagine recognition is a starting point
- ...deep learning is not just classification
- ...jets are not the only interesting objects at LHC
- ...machine learning is an amazing tool box
- ...maybe at some time we can pay back a little

For now, join the fun!



