“Theoretical” Introduction

David Shih
NHETC, Rutgers University

BOOST 2019

July 22, 2019
My first BOOST conference

Why am I giving this talk?
My first BOOST conference

Why am I giving this talk?

Give ’em enough rope…
My first BOOST conference

“We want an outsider’s perspective…” — P. Harris
The first BOOST conference: 2009

**BOOST 2009**

**Giving New Physics a Boost**

Thursday and Friday, July 9-10, 2009 from 8:00 am to 5:00 pm.
Kavli Auditorium
SLAC National Accelerator Laboratory
Menlo Park, California

Many signatures of new physics that could be discovered at the Tevatron or LHC involve highly boosted objects, which can confuse standard event reconstruction techniques due to the overlapping nature of their decay products in the detector. SLAC is hosting a two-day workshop to bring together the leading theorists and experimentalists in order to better understand the physics behind these novel signatures and how to detect them. The list of topics includes "lepton jets", boosted top jets ("t-tagging"), boosted Higgs ("fat-bottom") jets, di-tau jets, displayed-vertex jets, and light-gluino jets. We look forward to presentations and lively discussions.
The first BOOST conference: 2009

**BOOST 2009**

**Giving New Physics a Boost**

Thursday and Friday, July 9-10, 2009 from 8:00 am to 5:00 pm.
Kavli Auditorium
SLAC National Accelerator Laboratory
Menlo Park, California

Many signatures of new physics that could be discovered at the Tevatron or LHC involve highly boosted objects, which can confuse standard event reconstruction techniques due to the overlapping nature of their decay products in the detector. SLAC is hosting a two-day workshop to bring together the leading theorists and experimentalists in order to better understand the physics behind these novel signatures and how to detect them. The list of topics includes "lepton jets", boosted top jets ("t-tagging"), boosted Higgs ("fat-bottom") jets, di-tau jets, displayed-vertex jets, and light-gluino jets. We look forward to presentations and lively discussions.
The first BOOST conference: 2009

2 days, 35 participants, ~16 talks

Boosted tops: 10 talks

Thursday and Friday, July 9-10, 2009 from 8:00 am to 5:00 pm.
Kavli Auditorium
SLAC National Accelerator Laboratory
Menlo Park, California

Many signatures of new physics that could be discovered at the Tevatron or LHC involve highly boosted objects, which can confuse standard event reconstruction techniques due to the overlapping nature of their decay products in the detector. SLAC is hosting a two-day workshop to bring together the leading theorists and experimentalists in order to better understand the physics behind these novel signatures and how to detect them. The list of topics includes "lepton jets", boosted top jets ("t-tagging"), boosted Higgs ("fat-bottom") jets, di-tau jets, displayed-vertex jets, and light-gluino jets. We look forward to presentations and lively discussions.
The first BOOST conference: 2009

2 days, 35 participants, ~16 talks

Boosted tops: 10 talks

Lepton jets: 4 talks

Many signatures of new physics that could be discovered at the Tevatron or LHC involve highly boosted objects, which can confuse standard event reconstruction techniques due to the overlapping nature of their decay products in the detector. SLAC is hosting a two-day workshop to bring together the leading theorists and experimentalists in order to better understand the physics behind these novel signatures and how to detect them. The list of topics includes "lepton jets", boosted top jets ("t-tagging"), boosted Higgs ("fat-bottom") jets, di-tau jets, displayed-vertex jets, and light-gluino jets. We look forward to presentations and lively discussions.
The first BOOST conference: 2009

2 days, 35 participants, ~16 talks

Boosted tops: 10 talks

Lepton jets: 4 talks

BOOST = boosted tops and lepton jets !!!?
This BOOST conference: 2019

5 days, ~110 participants, ~60 talks

A rich and vibrant field at the interface of theory and experiment
Purpose of this talk

Motivation

Overview / Setting the Stage

Inspiration
Purpose of this talk

Motivation

Overview / Setting the Stage

Inspiration

**Disclaimer:** a highly personal take, focused primarily on *BSM* and *deep learning*.

Apologies in advance that I don’t cover every topic. Stay tuned for many interesting talks at this conference!
Purpose of this talk

Motivation

Overview / Setting the Stage

Inspiration
Where is the new physics????

dark matter

neutrino masses

matter/anti-matter asymmetry
Where is the new physics????

dark matter

We know physics beyond the SM must exist...

neutrino masses

matter/anti-matter asymmetry
Where is the new physics????

hierarchy problem

grand unification

flavor puzzle

\[ \mathcal{L} \supset \theta \frac{\alpha_s}{8\pi} G_{\mu\nu} \tilde{G}^{\mu\nu} \]

\[ \theta \lesssim 10^{-10} \]

strong CP problem
Where is the new physics???

- hierarchy problem
- grand unification
- flavor puzzle

And many other puzzles hint at new physics...

\[ \mathcal{L} \supset \theta \frac{\alpha_S}{8\pi} G_{\mu\nu} \tilde{G}^{\mu\nu} \]

\[ \theta \lesssim 10^{-10} \]

strong CP problem
Hierarchy problem

In particular, naturalness strongly motivates “top partners” at the TeV-scale.

Example: SUSY scalar top partners (stops)

\[ \delta m_h^2 \sim \frac{y_t^2}{16\pi^2} \Lambda^2 - \frac{\lambda_{t\tilde{t}}}{16\pi^2} \Lambda^2 \]
In particular, naturalness strongly motivates “top partners” at the TeV-scale.

Example: composite Higgs
fermonic top partners

fig. from 1205.0013
Boosted jets from top partners

Decays of heavy top partners

⇒ highly boosted tops, W/Z’s and Higgs
Boosted new physics

NP itself could be highly boosted as well, e.g.

$\tilde{t}^* \rightarrow \tilde{t} \tilde{\chi}_3^{12} \rightarrow t q \bar{q}'$

$\tilde{t} \rightarrow \tilde{\chi}_3^{12} \rightarrow q q' \bar{q}$

+$\text{countless other scenarios...}$
Boosted new physics

NP itself could be highly boosted as well, e.g.

(Can we find it if we don’t know what we’re looking for??)
Boosted new physics

NP itself could be highly boosted as well, e.g.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{diagram.png}
\caption{Diagram for the benchmark models used in this analysis: pair production of top squarks ($t\tilde{t}$), two of which are down-type quarks. Two possible choices of hadronic RPV coupling terms in Eq. (1) are considered: $\lambda_{312}^e l_i q_j q_k$ and $\lambda_{312}^p p_i p_j p_k$.}
\end{figure}

+ countless other scenarios…

(Can we find it if we don’t know what we’re looking for??)

Many opportunities for boosted jet substructure at the LHC!
Purpose of this talk

Motivation

Overview / Setting the Stage

Inspiration
Where is the new physics???
Where is the new physics????

Countless searches for new physics beyond the SM.
Where is the new physics????

Countless searches for new physics beyond the SM.

So far no concrete evidence, only lower limits on the NP scale.
Where is the new physics????

Countless searches for new physics beyond the SM.

So far no concrete evidence, only lower limits on the NP scale.

With stronger limits, boosted jet substructure becomes increasingly crucial.
Example: stop searches

**Figure 2:** Efficiency of the top quark tagger as a function of generator-level top quark $p_T$ for the monojet (red boxes), dijet (magenta triangles), and trijet (green upside-down triangles) categories and for their combination (blue circles), as determined using T2tt signal events with a top squark mass of 850 GeV and an LSP mass of 100 GeV. The vertical bars indicate the statistical uncertainties.

For the monojet and dijet categories, rather than strictly AK4 jets, and through implementation of the random forest tree for the trijet category. These improvements provide a factor of two reduction in the top quark misidentification rate while maintaining a similar efficiency.

### 8 Event selection and search regions

Our study is an inclusive search for events containing $p_{T}^{\text{miss}}$ and reconstructed top quarks. The selection criteria are intended, in general, to be nonrestrictive, while still providing high trigger efficiency and sensitivity to a wide variety of new-physics scenarios. All events must satisfy filters designed to remove detector- and beam-related noise. The events are subjected to the lepton, isolated-track, and charged-hadron vetoes of Section 5. To improve the rejection of background, the two tight AK4 jets with highest $p_T$ must have $p_T > 50$ GeV. Events are required to have $N_j = 4$, $N_b = 1$, $N_t = 1$, $p_{T}^{\text{miss}} > 250$ GeV, and $H_T > 300$ GeV. The QCD multijet background mostly arises when the $p_T$ of one of the highest $p_T$ jets is undermeasured, causing $\sim p_{T}^{\text{miss}}$ to be aligned with that jet. This undermeasurement can occur because of jet misreconstruction or, in the case of semileptonic $b$ or $c$ quark decays, an undetected neutrino. To reduce this background, requirements are placed on the azimuthal angle between $\sim p_{T}^{\text{miss}}$ and the three loose AK4 jets with highest $p_T$, denoted $j_1$, $j_2$, and $j_3$ in order of decreasing $p_T$. Specifically, we require $D_f(\sim p_{T}^{\text{miss}}, j_1) > 0.5$, $D_f(\sim p_{T}^{\text{miss}}, j_2) > 0.5$, and $D_f(\sim p_{T}^{\text{miss}}, j_3) > 0.3$.

The $m_{T2}$ variable [20–22] is used to reduce background from $t\bar{t}$ events. This variable is designed to provide an estimate of the transverse mass of pair-produced heavy objects that decay to both visible and undetected particles. It has a kinematic upper limit at the mass of the heavy object undergoing decay. Thus the upper limit for SM $t\bar{t}$ events is $m_t$, while the upper limit for TeV-scale squarks and gluinos is much larger. If there are two tagged top quarks in an event, $m_{T2}$ is calculated using the pair of tagged top quarks and $\sim p_{T}^{\text{miss}}$. If there are more than two tagged top quarks, we compute $m_{T2}$ for all combinations and choose the combination with the smallest $m_{T2}$. If there is only one tagged top quark, we construct a proxy for the other top quark using

(CMS 1710.11188)
Search relies heavily on a (not very sophisticated) boosted top tagger.
Search relies heavily on a (not very sophisticated) boosted top tagger.

How would reach improve with state-of-the-art tagger?
Example: squark RPV

\[ \sigma(pp \rightarrow \tilde{t}\tilde{t}) \text{[pb]} \]

Boosted search

Resolved search

CMS

1808.03124

35.9 fb\(^{-1}\) (13 TeV)

Selection

Boosted search

60 < \bar{m} < 450 GeV

(80 \leq m_{\tilde{t}} < 400 GeV)

Inclusive

AK8 jets

and

jet \( p_T > 150 \text{ GeV} \)

b-tagged

jet \( |\eta| < 2.5 \)

Number of jets \( \geq 2 \)

\( H^\text{AK8}_T > 900 \text{ GeV} \)

\( m_{\text{asym}} < 0.1 \)

\( \tau_{21} < 0.45 \)

\( \tau_{32} > 0.57 \)

\( \Delta \eta < 1.5 \)

b-tagged

two loose b-tagged jets

Inclusive selection

\( \tilde{t} \rightarrow qq' (\lambda'^{-}_{312}) \)

95% CL upper limits

- Observed
- Median expected
- 68% expected
- 95% expected

Top squark pair production
Example: squark RPV

**Figure 11:** Observed and expected 95% CL upper limits on the signal cross section as a function of the top squark mass. We exclude top squark masses with the limits set at 95% confidence level on the pair production cross section of top squarks as a function of the top squark mass. We exclude top squark masses with the NLO+NLL theoretical prediction for top squark pair production [40, 41].

**Table:**

<table>
<thead>
<tr>
<th>Selection</th>
<th>Boosted search</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>60 &lt; ( \bar{m} ) &lt; 450 GeV</td>
</tr>
<tr>
<td></td>
<td>(80 ≤ ( m_{t\tilde{t}} ) ≤ 400 GeV)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inclusive</th>
<th>AK8 jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>and</td>
<td>jet ( p_T &gt; 150 ) GeV</td>
</tr>
<tr>
<td>b-tagged</td>
<td>jet (</td>
</tr>
<tr>
<td></td>
<td>Number of jets ≥ 2</td>
</tr>
<tr>
<td></td>
<td>( H_T^{AK8} &gt; 900 ) GeV</td>
</tr>
<tr>
<td></td>
<td>( m_{asym} &lt; 0.1 )</td>
</tr>
<tr>
<td></td>
<td>( \tau_{21} &lt; 0.45 )</td>
</tr>
<tr>
<td></td>
<td>( \tau_{32} &gt; 0.57 )</td>
</tr>
<tr>
<td></td>
<td>( \Delta \eta &lt; 1.5 )</td>
</tr>
<tr>
<td>b-tagged</td>
<td>two loose b-tagged jets</td>
</tr>
</tbody>
</table>

**Relatively simple 2-prong substructure tagger…**
Example: squark RPV

Relatively simple 2-prong substructure tagger...

How would reach improve with state-of-the-art tagger?
Example: vectorlike quark searches

1808.01771

Uses variable-cone jets and DNN to identify boosted W/Z, Higgs and tops!
Future of the LHC
Future of the LHC

Graph showing the projected luminosity and integrated luminosity over time. The graph includes different labeling periods (LS1, LS2, LS3, LS4, LS5) and peak luminosity points marked with red dots. The x-axis represents the year, ranging from 2010 to 2037, and the y-axis shows luminosity in cm$^{-2}$s$^{-1}$, with integrated luminosity in fb$^{-1}$ on the right side.
Future of the LHC

![Graph showing LHC phases and HL-LHC luminosities](image)
Future of the LHC

The HL-LHC now aims for a 5-10 x instantaneous luminosity.
Future of the LHC

- HL-LHC: now
- 5-10 x instantaneous luminosity
- 20x current dataset
Future of the LHC

- **HL-LHC** now
- 5-10 x instantaneous luminosity
- ~20x current dataset
- ~20 more years

Graph showing the evolution of luminosity and integrated luminosity over the years.
Future of the LHC

How do we maximize the discovery potential of this enormous dataset??
Naive projections of future search sensitivity, assuming analyses maintain the status quo

Salam & Weiler [http://collider-reach.web.cern.ch/collider-reach/]
Naive projections of future search sensitivity, assuming analyses maintain the status quo

Salam & Weiler
http://collider-reach.web.cern.ch/collider-reach/
Stop Search Projections

Naive projections of future search sensitivity, assuming analyses maintain the status quo

Salam & Weiler http://collider-reach.web.cern.ch/collider-reach/
Naive projections of future search sensitivity, assuming analyses maintain the status quo

Salam & Weiler [http://collider-reach.web.cern.ch/collider-reach/]
Naive projections of future search sensitivity, assuming analyses maintain the status quo

Salam & Weiler [http://collider-reach.web.cern.ch/collider-reach/]
In the early days of LHC, progress was relatively easy.

Energy increase (8 TeV→13 TeV) and rapid luminosity gains led to huge gains in sensitivity.

Analyses did not need to be very sophisticated.

Could go after many low-hanging fruits.
In the future, the status quo will mean much slower progress. The data-taking rate will plateau, and no increases in energy are foreseen.

To maintain the rapid growth in sensitivity, we need new, more sophisticated analysis techniques.

Also, with this enormous dataset, we need to make sure we haven’t overlooked any subtle and unexpected signals of new physics.

We need new ideas for how to look for new physics in the data!
In the future, the status quo will mean much slower progress. The data-taking rate will plateau, and no increases in energy are foreseen. To maintain the rapid growth in sensitivity, we need new, more sophisticated analysis techniques.

Also, with this enormous dataset, we need to make sure we haven’t overlooked any subtle and unexpected signals of new physics.

We need new ideas for how to look for new physics in the data!
In the future, the status quo will mean much slower progress. The data-taking rate will plateau, and no increases in energy are foreseen. To maintain the rapid growth in sensitivity, we need new, more sophisticated analysis techniques.

Also, with this enormous dataset, we need to make sure we haven’t overlooked any subtle and unexpected signals of new physics. We need new ideas for how to look for new physics in the data!
Purpose of this talk

Motivation

Overview / Setting the Stage

Inspiration
Potential of Deep Learning

- High-level concepts from low-level inputs
- Automated feature engineering
- Robust against overfitting
Big Data and Deep Learning

https://www.wired.com/2013/04/bigdata/
Pasquale Musella, ETH-Zurich seminar
Big Data and Deep Learning

Key prerequisite to successful deep learning: large, complex, well-understood dataset.

https://www.wired.com/2013/04/bigdata/
Pasquale Musella, ETH-Zurich seminar
Big Data and Deep Learning

Key prerequisite to successful deep learning: large, complex, well-understood dataset.

Ability to cheaply generate realistic simulations also very beneficial for supervised ML.

https://www.wired.com/2013/04/bigdata/
Pasquale Musella, ETH-Zurich seminar
Big Data and Deep Learning

Key prerequisite to successful deep learning: large, complex, well-understood dataset.

Ability to cheaply generate realistic simulations also very beneficial for supervised ML.

The LHC is the perfect setting for deep learning!

https://www.wired.com/2013/04/bigdata/
Pasquale Musella, ETH-Zurich seminar
Deep Learning Papers

INSPIRE search: ("machine learning" or "deep learning" or neural) and (hep-ex or hep-ph)

An explosion of interest in machine learning!
Deep Learning at BOOST

Jet substructure is a natural arena for deep learning!
The Landscape of DL @ LHC

Unsupervised Learning
- Generation
- Clustering
- Anomaly Detection
- Autoencoders
  - CaloGAN
  - LaGAN
  - JUNIPR
- Dimensionality Reduction
- (unlabeled data)

Machine Learning
- Autoencoders
  - PCA
- (unlabeled data)

Supervised Learning
- Regression
- Classification
  - top tagging
  - b tagging
  - W/Z tagging
  - q/g tagging
  - strange tagging
  - full event tagging
- pile-up reduction

Triggering
- CWoLa
- Triggering

Jet finding algorithms
- (labeled data)
Deep Learning @ BOOST ‘18

- Deep Learning for Jet Tagging at CMS and ATLAS
- Recursive NNs for Jet Tagging
- Autoencoders
- Classification Without Labels
- Jet Topics
- JUNIPR
- Energy Flow Networks
- New observables from DL

New ideas for jet tagging
Unsupervised deep learning
Learning from deep learning
Shower Deconstruction (SD)

- Encoding physics knowledge (~ME calc.) in observable
- Between cut-based and ML taggers
- Multivariate (MVA) combination of inputs possible with deep neural networks (DNN) and boosted decision trees (BDT)

- TopoDNN
- Top tagging using jet LC topo cluster constituents directly
- Implemented and evaluated in data

Heavy flavor identification

- CSVv2 (track, SV)
- DeepCSV (track, SV + more charged tracks)
- DeepFlavour (charged and neutral PF + SV)

Future work

- Beyond double-b, deepDoubleB
Huge improvements in tagging from deep learning.

For 2019 updates, see talks by H. Qu and S. Macaluso tomorrow.
Huge improvements in tagging from deep learning.

For 2019 updates, see talks by H. Qu and S. Macaluso tomorrow

Realized in data???
See ATLAS and CMS talks by Schramm and Narain later today
Beyond tagging: mass decorrelation

Raw tagger performance not the only consideration.

For robust background estimation, often need to ensure tagger does not bias the background mass distribution.

State of the art in mass decorrelation methods was presented by ATLAS for W-tagging at BOOST ’18 (ATL-PHYS-PUB-2018-014)
ATLAS Simulation Preliminary

$\sqrt{s} = 13$ TeV, $W$ jet tagging

Cuts at $\varepsilon_{\text{rel}}^{\text{sig}} = 50\%$

Inclusive selection:
- Multijets
- $W$ jets

$D_2$ vs. Large-$R$ jet mass [GeV]
7.3 Combined metric

A combined metric, reflecting both classification performance and mass-decorrelation, is necessary to assess the trade-offs balanced by each of the mass-decorrelation procedures. A more complete picture of the performance is found by plotting the two metrics together. Figure 11 shows the mass-decorrelation (1/JSD) versus the background rejection (1/\(\varepsilon_{rel}\)) for tagger cuts at \(\varepsilon_{rel}^{rel} = 50\%\), in two \(p_T\) bins. The x-axis measures classification power and the y-axis measures mass-decorrelation, with larger values along each indicating better performance. For any given task, a specific direction in the plane of Figure 11 will correspond to the best trade-offs.

For each of the mass-decorrelated MV A taggers, several working points are evaluated, by scanning for the ANN tagger and for uBoost. For high values of \(\alpha\) (\(\leq 10\)), the ANN method starts to saturate given the chosen network configurations, training procedures, and datasets.
7.3 Combined metric

A combined metric, reflecting both classification performance and mass-decorrelation, is necessary to assess the trade-offs balanced by each of the mass-decorrelation procedures. A more complete picture of the performance is found by plotting the two metrics together. Figure 11 shows the mass-decorrelation \( \frac{1}{JSD} \) versus the background rejection \( \frac{1}{\varepsilon_{rel}} \) for tagger cuts at \( \varepsilon_{rel}^{sig} = 50\% \), in two \( p_T \) bins. The \( x \)-axis measures classification power and the \( y \)-axis measures mass-decorrelation, with larger values along each indicating better performance. For any given task, a specific direction in the plane of Figure 11 will correspond to the best trade-off.

For each of the mass-decorrelated MV A taggers, several working points are evaluated, by scanning for the ANN tagger and \( \lambda \) for uBoost. For high values of \( \alpha \), the ANN method starts to saturate given the chosen network configurations, training procedures, and datasets.
A new method for mass decorrelation is presented. The method involves adversarial training and achieves the greatest background rejection. The BDT-based MVA taggers have comparable performance to the NN-based taggers. For each of the mass-decorrelated MVA taggers, several working points are evaluated, by scanning the parameters. The performance is found by plotting the two metrics together. A combined metric, reflecting both classification performance and mass-decorrelation, is necessary to assess the trade-off between tagger performance and mass decorrelation.

\[ \sqrt{s} = 13 \text{ TeV} \]
\[ W \text{ jet tagging} \]
\[ p_T \in [200, 500] \text{ GeV} \]

**Trade off between tagger performance and mass decorrelation.**

Best performer was adversarial NN.
Alternatives to adversaries

Work in progress with Gregor Kasieczka

Adversaries are notoriously **tricky to train** — saddle point optimization

\[
\min_{\theta_{\text{clf}}} \max_{\theta_{\text{adv}}} L_{\text{clf}}(y(\theta_{\text{clf}})) - \lambda L_{\text{adv}}(y(\theta_{\text{clf}}), m; \theta_{\text{adv}})
\]

\(y\): NN prediction
\(m\): mass

Would be great if we could achieve the same performance but with a convex regularizer term

\[
\min_{\theta_{\text{clf}}} L_{\text{clf}}(y(\theta_{\text{clf}})) + \lambda C_{\text{reg}}(y(\theta_{\text{clf}}), m)
\]

First idea: can we just use Pearson correlation coefficient?

\[
C_{\text{reg}} = R(y, m) \propto \sum_i y_i m_i
\]

Problem: this only measures linear correlations
Distance (de)correlation

Promising idea: “distance correlation” (Szekely, Rizzo, Bakirov 2007; Szekely & Rizzo 2009)

\[ X_{ij} = |X_i - X_j|, \quad Y_{ij} = |Y_i - Y_j| \]

\[ \hat{X} = CXC, \quad \hat{Y} = CYC \]

\[ \text{dCov}^2(X, Y) \equiv \text{tr} \hat{X}\hat{Y} \]

- Zero iff X,Y are independent; positive otherwise!
- Computationally tractable!
- Doesn’t require binning!
Distance (de)correlation

Work in progress with Gregor Kasieczka

Comparable performance to DNN+adversary.

Much easier to train.
Beyond tagging: unsupervised ML

Jet tagging is a prime example of supervised machine learning. Perfect when you know what you’re looking for.

Increasing interest in applications of unsupervised ML to LHC.

- Learning without labels
- Learning directly from the data
- Anomaly detection
- Triggering
Classification WithOut Labels (CWoLa)

Dery et al 1702.00414, Cohen, Freytsis & Ostdiek 1706.09451, Metodiev, Nachman & Thaler 1708.02949, Komiske et al 1801.10158, Collins, Howe & Nachman 1805.02664, 1902.02634

\[
P_{1}^{\text{signal}} = P_{2}^{\text{signal}}
\]

\[
P_{1}^{\text{background}} = P_{2}^{\text{background}}
\]

Suppose we are given two mixtures of signal and background.

If signal and background are drawn from the same distributions in each sample, then under certain mild assumptions, one can train a classifier to distinguish signal from background directly from the data.
Can look for anomalies in the data without a detailed signal hypothesis in mind.
Unsupervised anomaly detection with deep autoencoders

Heimel et al 1808.08979; Farina, Nakai & DS 1808.08992

An autoencoder maps an input into a “latent representation” and then attempts to reconstruct the original input from it.

The encoding is lossy, so the decoding cannot be perfect.

See also:
Hajer et al “Novelty Detection Meets Collider Physics” 1807.10261
Cerri et al “Variational Autoencoders for New Physics Mining at the Large Hadron Collider” 1811.10276
Figure 2: Distribution of reconstruction error computed with a CNN autoencoder on test samples of QCD background (gray) and two signals: tops (blue) and 400 GeV gluinos (orange). We see that the autoencoder works as advertised: it learns to reconstruct the QCD background that it has been trained on (to be precise, we train on 100k QCD jets and then we evaluate the autoencoder on a separate sample of QCD jets), and it fails to reconstruct the signals that it has never seen before. This is further illustrated in Fig. 3, which shows the average QCD, top and gluino jet image before and after autoencoder reconstruction. We see by eye that the QCD images are reconstructed well on average, while the others contain more errors.

By sliding the reconstruction loss threshold $L > L_S$ around, we can turn the histograms in Fig. 2 into ROC curves. The ROC curves for the different autoencoder architectures are shown in Fig. 4 for the top and gluino signals. For comparison we have also included the ROC curve obtained by cutting on jet mass as an anomaly threshold. While the three architectures have comparable performances it is clear there are some differences. For tops, the CNN outperforms the others, while for gluinos the situation is largely reversed. Surprisingly, for gluinos, the CNN is even outperformed by the humble PCA autoencoder at all but the lowest signal efficiencies! We will explore this in more detail in section 4.2, but a clue as to what's going on is shown in the comparison of the PCA ROC curve with the jet mass ROC curve. For gluinos, they track each other extremely closely, suggesting that the PCA reconstruction error is highly correlated with jet mass. We will confirm this in section 4.2. Evidently, the PCA autoencoder (and to a lesser extent the dense autoencoder) has learned to reconstruct.

The algorithm works when trained on “real data”! (QCD + a small fraction of signal)
Bump hunt with deep autoencoder

Can train directly on data that contains 400 GeV gluinos, use the AE to clean away “boring” SM events, and improve S/N by a lot.

Could potentially discover new physics this way!
Come join us for the LHC Olympics 2020!

Despite an impressive and extensive effort by the LHC collaborations, there is currently no convincing evidence for new particles produced in high-energy collisions. At the same time, there has been a growing interest in machine learning techniques to enhance potential signals using all of the available information.

In the spirit of the first LHC Olympics (circa 2005-2006) [1st, 2nd, 3rd, 4th], we are organizing the 2020 LHC Olympics. Our goal is to ensure that the LHC search program is sufficiently well-rounded to capture "all" rare and complex signals. The final state for this olympics will be focused (generic dijet events) but the observable phase space and potential BSM parameter space(s) are large: all hadrons in the event can be used for learning (be it "cuts", supervised machine learning, or unsupervised machine learning).

For setting up, developing, and validating your methods, we provide background events and a benchmark signal model. You can download these from this page. To help get you started, we have also prepared simple python scripts to read in the data and do some basic processing.

The final test will happen 2 weeks before the ML4Jets2020 workshop. We will release a new dataset where the "background" will be similar to but not identical to the one in the development set (as is true in real data). The goal of the challenge is to see who can "best" identify BSM (yes/no, what mass, what cross-section) in the dataset. There are many ways to quantify "best" and we will use all of the submissions to explore the pros/cons of the various approaches.
Conclusions/Outlook

Boosted jet substructure is a crucial ingredient in the search for NP.

Deep learning is an exciting new tool with enormous potential to enhance the sensitivity to jet substructure and NP in the HL-LHC era.

Boosted heavy resonance tagging is being greatly accelerated by deep learning, along with many other tasks. Important higher-order questions such as mass decorrelation are now being actively investigated.

Besides boosted tops, W/Z’s and Higgses, new physics itself could be highly boosted. Could NP be hiding in jet substructure? Can we find it if we don’t know what we’re looking for?
Thanks for your attention!
HL-LHC projections


Assume future sensitivity set by

\[ N_{\text{sig}}(M_{\text{future}}, L_{\text{future}}) = N_{\text{sig}}(M_{\text{now}}, L_{\text{now}}) \]

\[ N_{\text{sig}}(M, L) \sim \sigma(M) \times L \]

\[ \sigma(M) \sim \frac{1}{M^2} f_{\text{parton}}(x = 2M/E_{CM}) \]

Assume future background negligible with comparable signal efficiency.
Distance correlation vs Pearson correlation

Distance correlation
Distance correlation vs Pearson correlation

Pearson correlation