New approaches to the analysis and physics interpretation of LHC events

Ariel Schwartzman
SLAC National Accelerator Laboratory, Stanford University

ACAT 2016
Valparaiso, Chile, January 2016
Machine Learning in HEP

• Machine learning is widely used on HEP
  o Unsupervised learning: clustering (jet finding)
  o Supervised learning for classification (b-tagging)

• Use physics intuition to reduce information down to a small set of input variables into a multivariate discriminant
  o Do not make use of all information available
  o Require the invention of variables that capture the main features of signals and backgrounds

• Outstanding advances and new methods in artificial intelligence and image classification
  o Potential to enhance HEP analyses using new powerful tools
Machine Learning in HEP

Example: b-quark jet tagging

24-variable Boosted Decision Tree

Broadly used in particle identification and physics signature classification: feature extraction + multivariate discriminant
IMAGENET: Large Scale Visual Recognition Challenge


ILSVRC top-5 error on ImageNet

Convolutional Deep Neural Networks
Image Classification

Before 2012

Feature extraction → Classification

After 2012

Convolutional Neural Networks → Classification

http://cs231n.github.io/assets/cnn/convnet.jpeg
Convolutional Deep Neural Networks for Image Classification

http://cs231n.github.io/assets/cnn/convnet.jpeg
Convolutional Deep Neural Networks for Image Classification

http://cs231n.github.io/assets/cnn/convnet.jpeg
A new approach

• Connect the fields of experimental particle physics with computer vision, and image processing

• Apply state-of-the art image classification and machine learning methods to analyze LHC data in new ways

• Use these techniques to understand physics and to help design more powerful tagging methods
  o Are there new discriminant features that we have not considered?
  o How do we visualize new, high level, representations of LHC events?
Learning representations

- **Two goals:** Improve classification (jet tagging) and physics understanding with learned jet representations

- Use deep convolutional neural networks to learn new discriminative, interpretable representations for jet tagging:
  - Can we learn more discriminative representations of jets?
  - What do these representations teach us about physics?
  - Can we improve classification and physical understanding with learned jet representations?

- **Physics learning from machine learning:**
  - Extract physics information about what machine learning algorithms are learning

- **Broad applicability**
  - Jet tagging at the LHC used as example
The ATLAS detector is one of the two general-purpose experiments at the LHC. The 100 million hypothetical new physics event, done in Computer Vision, to account for non-discriminative difference in pixel intensities. We transform each image in enabling the connection between LHC physics event reconstruction and computer vision. We take these energy levels, and use them as the pixel intensities in a greyscale analogue. We transform the ATLAS coordinate system (φ, W') to a rectangular grid that allows for an image-image between signal difference-visualization technique helps understand what the network learns. Below, we have the learned convolutional filters (left) and the difference in between the average accuracy. The learned filters from the convolutional layers exhibit a two prong and location based accuracy.
Jets at the LHC
Jets at the LHC

- Jets are key observables in any event at the LHC
  - Individual jets identify quarks and gluons emitted in high-energy interactions
  - Combinations of jets are used to identify unstable massive particles such as the top quark, the W, Z, and Higgs bosons
  - Almost every analysis of LHC data rests on the efficient and accurate reconstruction of jet properties

- As the center-of-mass energy of the LHC increases, EW heavy states in the SM (V, H, top quark) can be produced with large Lorentz boosts and reconstructed as single (merged) jets
Jets at the LHC

Main challenge: very large QCD background → jet tagging

Boosted top pair candidate event
Jet tagging using jet substructure

- **Key idea:** Identify the internal structure of individual jets
- Jet tagging methods based on three main observables:
  - **Jet mass**
  - **N-prong structure:**
    - 1-prong (QCD)
    - 2-prong (W,Z,H)
    - 3-prong (top)
  - **Radiation pattern:**
    - QCD gluon emission is soft
    - Color flow

[Diagram showing a single large-R jet, 2-prong structure, and soft emission]

**ATLAS Simulation**
- $\sqrt{s}=8$ TeV
- $|\eta|^{\text{truth}}<1.2$
- $200 < p_{T,\text{truth}} < 350$ GeV
- anti-kt, R=1.0 jets
- Trimmed ($f_{\text{cut}}=5\%, R_{\text{sub}}=0.2$)

QCD jet

Jet mass

W jet

[Graph showing normalised entries vs. M (GeV)]

arXiv:1510.05821
W jet tagging

- n-subjettiness:
  - Measures the n-prong structure of jets

\[ \tau_N = \frac{1}{d_0} \sum p_{T,k} \min\{\Delta R_{k,\text{axis}-1}, \ldots, \Delta R_{k,\text{axis}-n}\} \]
The Jet-Image

- New data representation for jet tagging: the jet-image
  - Calorimeter towers → pixels in a camera
- Use all available information for jet classification
- Enable the use of computer vision image classification algorithms
The Jet-Image

- New data representation for jet tagging: the jet-image
  - Calorimeter towers $\rightarrow$ pixels in a camera
- Use all available information for jet classification
- Enable the use of computer vision image classification algorithms
Image pre-processing

Pythia 8, $W' \rightarrow WZ$, $\sqrt{s} = 13$ TeV

$240 < p_T/\text{GeV} < 260 \text{ GeV, } 65 < \text{mass}/\text{GeV} < 95$

Pixelate $\rightarrow$ Translate $\rightarrow$ Rotate $\rightarrow$ Re-grid $\rightarrow$ Flip

Use subjects to align images. Make use of symmetries: center, rotate, translate
Image pre-processing and the symmetries of space-time (I)

Information about the jet mass can be washed out during pre-processing

- Translations in eta are Lorentz boosts along the z axis, which do no preserve the pixel energies → Use $p_T$ rather than $E$ as pixel intensity

- Jet mass is not invariant under Image normalization

Pythia 8, $\sqrt{s} = 13$ TeV

$240 < p_T/\text{GeV} < 260 \text{ GeV}, 65 < \text{mass}/\text{GeV} < 95$

$$m^2_J = \sum_{i<j} E_i E_j (1 - \cos(\theta_{ij}))$$

- No pixelation
- Only pixelation
- Pix+Translate (naive) ($\times 0.75$)
- Pix+Translate
- Pix+Translate+Flip
- Pix+Translate+$\pi/2$ Rotation
- Pix+Translate+$p_T^2$ norm ($\times 170$)

Naïve Translation

Image normalization
In both pictures the total intensity of Einstein’s face is about the same. However, the image mass is different!
Image pre-processing and the symmetries of space-time (II)

In both pictures the total intensity of Einstein’s face is about the same. However, the image mass is different!

Standard computer vision methods might not be directly applicable to physics analyses and require physics-domain knowledge input!

http://mentalfloss.com/article/49222/11-unserious-photos-albert-einstein
Computer vision
W tagging: Fisher jets

Analog to facial recognition with Fisher Faces
Construct a Fisher Linear Discriminant to find the direction in the image space that maximizes the variance between jet classes (W vs. QCD images) over the variance within each jet class.

Average W jet

Average Light jet from QCD

Fisher Face

Example: Fisher Face to determine if a person is wearing glasses

This fisher image can be interpreted as a template to find glasses in faces, ignoring other characteristics of the face
Fisher Discriminant

**Average W jet**

**Average QCD jet**

Radiation around 1\(^{st}\) subjet in QCD jets

No info in presences of 1\(^{st}\) subjet in W-jets

Hard 2\(^{nd}\) subjet in W-jets

Wide 2\(^{nd}\) subjet in QCD jets

0.6 < Subjet $\Delta R$ < 0.8

QCD-like

W-like

**Fisher Jet**

How can we extract the important features?

How can we convert this into discrimination power?
The performance of the classifier does not change dramatically whether it is applied to a
approximately -0.25 for both W and QCD jets indicating a small degree of anti-correlation,
Fisher-jet output is only slightly correlated with mass, with a correlation coe-

The Fisher-jet approach is able to reproduce, as shown in Figures
of jet output when applied to W-jets and Light-jets (right), when the FLD is

Figure 2:

It should be noted that the output of FLD and N-subjettiness are correlated, as shown

The background rejection vs. signal e

Figure 2:

A Fisher's linear discriminant presented as an image (left) and the distributions

Extracting a Value

\[ D = \sum_{\text{pixels}} p_T^i F_i \]

Dot product
of two images
Computer vision W tagging

- Comparable performance with respect to state-of-the-art methods (n-subjettiness)
- Visualization of the discriminant adds a new capability to understand the physics within jets and to design more powerful jet tagging methods
Our work focuses on the idea of the ATLAS detector as a camera, with events captured as Data. Scientists use machine learning for rare-event detection, and hope to catch glimpses of new physics at the Large Hadron Collider (LHC) at CERN, the largest and most powerful particle accelerator in the world.

We focus our attention on the Calorimeter, which we treat as a digital camera in cylindrical space. The channel detector captures snapshots of particle collisions occurring 40 million times per second. In our experiments, we build discriminants on top of Jet Images to distinguish between a hypothetical new physics event and standard backgrounds.

To account for non-discriminative differences in pixel intensities, we transform each image in a based grid arrangement. During a collision, energy from particles are deposited in pixels in \( \eta, \phi \) space, rotate around the jet-axis, and normalize each image, as is often done in Computer Vision, to account for non-discriminative difference in pixel intensities.

We transform the ATLAS coordinate system \((\eta, \phi)\) to a rectangular grid that allows for an image-image between signal and background. The deep convolutional network is used to enhance discrimination power. This indicates that the deep convolutional network retains some discriminative power. This suggests that the deep convolutional network is capturing features and representations beyond what is captured by standard methods.

Notice that removing out the individual effects of discriminative power. This indicates that the deep convolutional network is capturing features and representations beyond what is captured by standard methods.

Our analysis shows that Deep Convolutional Networks significantly improve the classification of W vs. QCD jets. We hope to both inspire future research and improve our understanding of how the LHC works.

Deep Learning — convolutional networks in particular — currently represent the state of the art in image classification and object detection.
Convolutional Neural Networks

ConvNet architecture:

1. Convolutional layer
2. Fully connected layer
During the learning process the values of the filters will filter however the filters learned will be much different. Convolutions. This design is important because spatial value. A picture is formed from the results of these value and summed together which results in a single this is the interface between the input image and the process of a convolutional...
The Large Hadron Collider (LHC) at CERN is the largest and most powerful particle accelerator in the world. Below, we see a snapshot of a 13 TeV proton-proton collision. The ATLAS detector is one of the two general-purpose experiments at the LHC. The 100 million pixel images from these collisions are captured and processed to extract meaningful information.

We transform each image in the ATLAS coordinate system ($\eta$, $\phi$), rotate around the jet-axis, and normalize each image, as is often done. We apply deep learning techniques on jet images, enabling the connection between LHC physics event reconstruction and computer vision. We found that architectures with large filters captured the physics response with a higher level of accuracy. The learned filters from the convolutional layers exhibit a two prong and location based structure that sheds light on phenomenological structures within jets.

We show that modern Deep Convolutional Architectures can significantly enhance the discovery potential of the LHC. More importantly, the improved performance towards increased discovery potential for new physics processes compared to state-of-the-art methods based on physics features, meaning that physical variables have no discrimination power. Then, we apply our learned sample weights to test this — we derive sample weights to apply such that we ensure that the representations we learn are more than simple recombinations of basic physical variables.

We hypothesize these may have to do with discriminative power. This indicates that the deep learning models are learning representations that capture the physics response in a way that is not simply a recombination of basic physical variables. We hope to both inspire future research towards increased discovery potential for new physics and deepen our understanding of the physics-motivated variables.

We introduce a new method to test this — we derive sample weights to apply such that we ensure that the representations we learn are more than simple recombinations of basic physical variables.

We show that modern Deep Convolutional Architectures can significantly enhance the discovery potential of the LHC. More importantly, the improved performance towards increased discovery potential for new physics processes compared to state-of-the-art methods based on physics features, meaning that physical variables have no discrimination power. Then, we apply our learned sample weights to test this — we derive sample weights to apply such that we ensure that the representations we learn are more than simple recombinations of basic physical variables.

We hypothesize these may have to do with discriminative power. This indicates that the deep learning models are learning representations that capture the physics response in a way that is not simply a recombination of basic physical variables. We hope to both inspire future research towards increased discovery potential for new physics and deepen our understanding of the physics-motivated variables.

We show that modern Deep Convolutional Architectures can significantly enhance the discovery potential of the LHC. More importantly, the improved performance towards increased discovery potential for new physics processes compared to state-of-the-art methods based on physics features, meaning that physical variables have no discrimination power. Then, we apply our learned sample weights to test this — we derive sample weights to apply such that we ensure that the representations we learn are more than simple recombinations of basic physical variables.

We hypothesize these may have to do with discriminative power. This indicates that the deep learning models are learning representations that capture the physics response in a way that is not simply a recombination of basic physical variables. We hope to both inspire future research towards increased discovery potential for new physics and deepen our understanding of the physics-motivated variables.
Deep network architectures

<table>
<thead>
<tr>
<th>Filter size (1\textsuperscript{st} layer)</th>
<th>(3x3)</th>
<th>(4x4)</th>
<th>(5x5)</th>
<th>(7x7)</th>
<th>(9x9)</th>
<th>(11x11)</th>
<th>(15x15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>14.8</td>
<td>12.5</td>
<td>11.1</td>
<td>13.3</td>
<td>17.3</td>
<td>20.3</td>
<td>18.1</td>
</tr>
</tbody>
</table>

Jet image occupancy

Need larger filters than used for standard image classification
Likely helpful to process sparse images

Pythia 8, W\textrightarrow{} WZ, \(\sqrt{s} = 13\) TeV

240 < \(p_T/\text{GeV} < 260\) GeV, 65 < mass/GeV < 95

\(\sim 3\%\) occupancy
Deep learning W tagging

1. Performance of Individual variables

Convolutional deep neural network:
x2 better rejection than n-subjettiness

Large performance gains beyond mass, n-subjettiness and $\Delta R(j,j)$

Pythia 8, $\sqrt{s} = 13$ TeV
250 < $p_T$/GeV < 300 GeV, 65 < mass/GeV < 95
Deep learning W tagging

1. Performance of Individual variables

**Fully Connected deep neural network:**

- **Higher rejection than ConvNet**
- Views the full (sparse) image from the initial hidden layer

### Large performance gains beyond mass, n-subjettiness and $\Delta R(j,j)$

**Pythia 8, $\sqrt{s} = 13$ TeV**

$250 < p_T/\text{GeV} < 300$ GeV, $65 < \text{mass}/\text{GeV} < 95$
Deep neural networks:
x2 better rejection than n-subjettiness

2. Combination of physics variables

Learned information beyond mass and n-subjettiness
2. Combination of ConvNet with physics variables

**Mass + ConvNet**

**Increased performance**

ConvNet, (ConvNet + $\Delta R$), (ConvNet + $\tau_{21}$)

$\Delta R$ and $\tau_{12}$ fully learned by the neural network

Deep network learns $\Delta R$ and $\tau_{21}$ but it does not fully capture all the discriminating information contained in the jet mass!
2. Combination of ConvNet with physics variables

Mass + ConvNet
Increased performance

ConvNet, (ConvNet + ΔR), (ConvNet + τ_{21})
ΔR and τ_{21} fully learned by the neural network

ΔR and τ_{21} are invariant under normalizations
Mass is not.

Pythia 8, \( \sqrt{s} = 13 \text{ TeV} \)
\( 250 < p_T \text{/GeV} < 300 \text{ GeV}, 65 < \text{mass/GeV} < 95 \)
Classification performance increases when images are normalized

Since it is difficult for these networks to learn the jet mass, the lack of mass information from pre-processing normalization does not lead to worse discrimination, while normalization improves training.
Convolved representations

How the internal structure of the network relates to the properties of $W$ bosons vs. QCD jets

Many features captured by the network:
Differences learned at the first layer to separate $W$ from QCD jets
Physics in deep representations

Deep correlation jet-image: Pearson Correlation Coefficient of the pixels intensity with the network output: how discriminating information is contained within the network
Physics in deep representations

Additional radiation in QCD jets

Deep correlation jet-image: Pearson Correlation Coefficient of the pixels intensity with the network output: how discriminating information is contained within the network

![Correlation of Deep Network output with pixel activations.](image)

- Soft QCD gluon emission
- Signal-like
- Background-like
Restrict the phase space in very small mass and τ_{21} bins:
Improvement in discrimination from new, unique, information learned by the network
Spatial information indicative of radiation pattern for W and QCD: where in the image the network is looking for discriminating features.
Deep Jets Summary

- Jet images is a powerful paradigm that connects the fields of jet substructure, computer vision, and image processing

- State-of-the-art deep neural networks for image classification provide a new way to analyze and interpret LHC events directly from raw data:
  - Significantly outperform state-of-the-art techniques based on engineered feature extraction
  - Enable new ways to visualize learned high level representations of the data and gain new insights into the underlying physics that provide discrimination power

- Need physics-domain specific knowledge
  - Large convolutional filter sizes for sparse images
  - Fully connected vs. convolutional networks
  - Image pre-processing (normalization)
Future ideas (I)

- **3-dimensional jet images:**
  - Make use of longitudinal calorimeter segmentation (like “RGB color” channels in images)
  - Tracking information
Future ideas (II)

Neutrino detectors

“Whole-event” LHC Convolutional Neural Networks
Summary

• Advanced computer vision and image processing methods add a new capability to understand LHC events, and to design new, more powerful, classification methods.

• Many potential applications within and beyond HEP.

• Next step is to apply these ideas to data!
Acknowledgments

Michael Kagan
SLAC

Lester Mackey
Statistics, Stanford

Ben Nachman
SLAC, Stanford

Luke De Oliveira
ICME, Stanford
Backup
Jet trimming

1. Recluster jet constituents into $k_t$ subjets with small $R$

2. Discard subjets with:

$$p_T < f_{cut} \cdot p_T^{jet}$$

- Jet contamination from pileup, underlying event, and initial state radiation is softer than hard-scatter partons and final state radiation:
  - Remove soft components of the jet
  - Reduce the area of the jet

Jet trimming

- $k_t$ $R = 0.3$
- $f_{cut} = 0.05$
Trimming performance

Signal: $Z \rightarrow q\bar{q}$, Background: QCD jets

- **Improved mass resolution**
  (sharpens the mass peak)

- **Reduced QCD background**
  (improved S/B)
Conditional distributions between the network output and n-subjectness and mass
Reweighted phase space

Apply weights so that mass and $\tau_{21}$ have a uniform distribution for signal and background

- Mass+ $\tau_{21}$ does not provide discrimination

Network learned information beyond mass and $\tau_{21}$

Discrimination performance improved when training is also performed with weights
Deep correlation jet images in reweighted phase space

Unique information learned by the network is related to radiation surrounding the subjets
Top jet tagging

Top tagging using fully connected neural networks

Figure 4. Efficiency vs. Mistag rate curves for the ANN tagger (blue solid lines), for jets in three representative $p_T$ ranges. For comparison, corresponding curves for three existing top taggers are also shown: $d_{12}$ tagger (yellow dashed), top template tagger (green dotted), and N-subjettiness (red dash-dotted).

Almeida, Backovic, Cliche, Lee, & Perelstein (arXiv:1501.05968)
Stanford Data Science Initiative (SDSI)

https://sdsi.stanford.edu

Funded Research Projects in Data Science

OCTOBER 2015

Data scientists and physicists both work with numbers. But they seldom work together. Stanford's Ariel Schwartzman, a particle physicist, and Lester Mackey, a professor of statistics, think that collaborating will give better understanding of the information that is being produced by the Large Hadron Collider (the LHC), the huge particle accelerator in Geneva, Switzerland. It might even help find new subatomic particles.

By applying big data analysis techniques to the petabytes of data that the LHC generates every year, the scientists expect to be able to more accurately identify and differentiate particles like W, Z, and Higgs bosons, as well as top quarks. Bosons and top quarks are important to study because they are predicted by many models of new physics, such as supersymmetry, or models with extra dimensions of space.