

Image Processing, Computer Vision, and Deep Learning: new approaches to the analysis and physics interpretation of LHC events

A. Schwartzman¹, M. Kagan¹, L. Mackey², B. Nachman¹ and L. De Oliveira³

¹ SLAC National Accelerator Laboratory, Stanford University, 2575 Sand Hill Road, Menlo Park, CA 94025, USA

² Department of Statistics, Stanford University, Stanford, CA 94305, USA

³ Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305, USA

E-mail: sch@slac.stanford.edu

Abstract. This review introduces recent developments in the application of image processing, computer vision, and deep neural networks to the analysis and interpretation of particle collision events at the Large Hadron Collider (LHC). The link between LHC data analysis and computer vision techniques relies on the concept of *jet-images*, building on the notion of a particle physics detector as a digital camera and the particles it measures as images. We show that state-of-the-art image classification techniques based on deep neural network architectures significantly improve the identification of highly boosted electroweak particles with respect to existing methods. Furthermore, we introduce new methods to visualize and interpret the high level features learned by deep neural networks that provide discrimination beyond physics-derived variables, adding a new capability to understand physics and to design more powerful classification methods at the LHC.

1. Introduction

The Large Hadron Collider is exploring physics at the energy frontier, probing some of the most fundamental questions about the nature of our universe. The datasets of the LHC experiments are among the largest in all science. Each particle collision event at the LHC is rich in information, particularly in the detail and complexity of each event picture, making it ideal for the application of machine learning techniques to extract the maximum amount of physics information. Machine learning has been widely used for the analysis of particle collision data. Examples include unsupervised learning algorithms, such as jet clustering [1], and supervised algorithms for different particle identification and classification tasks, including the identification of different kind of quark flavors, such as b - and c -jets [2], tau leptons [9], and photons [4], among many others. Particle classification algorithms in particle colliders have traditionally been designed using physics intuition to extract a small set of observables or features that were then input into a multivariate classifier. For example, in the case of b flavor tagging in the ATLAS experiment, a Boosted Decision Tree consisting of 24 input variables is used [5]. The 24 input variables provide information about the physics of b -quark decays that is distinct from

the light-quark and gluon jet backgrounds. Much of the work in developing such techniques relies on finding the physics variables that best capture the differences between different classes of particles.

During the last several years, spectacular advances in the fields of artificial intelligence, computer vision, and deep learning have resulted in remarkable performance in image classification and vision tasks [6], in particular through the use of deep convolutional neural networks (CNN) [7]. One of the key innovations in these methods is that they operate directly from raw images instead of requiring the extraction of features from the images. Motivated by the success of these techniques, this paper investigates the applicability of deep neural networks for image classification to the analysis of LHC data. LHC events are in fact ideal for the use of computer vision methods, since each LHC event picture can be interpreted as an image from a digital (100M pixel) camera.

2. Jets at the LHC

One of the most important signatures for the analysis of data from the Large Hadron Collider (LHC) are collimated streams of particles known as jets. 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. For example, W bosons are reconstructed by the two jets from the decay process $W \rightarrow qq$, where q represents a quark. These heavy particles, appear in many models of new physics that predict the existence of new exotic sub-atomic particles. Almost every analysis of LHC data rests on the efficient and accurate reconstruction of jets and their properties. An example of an event display from the ATLAS experiment¹ containing two high p_T ² jets displayed in different colors is shown in Figure 1.

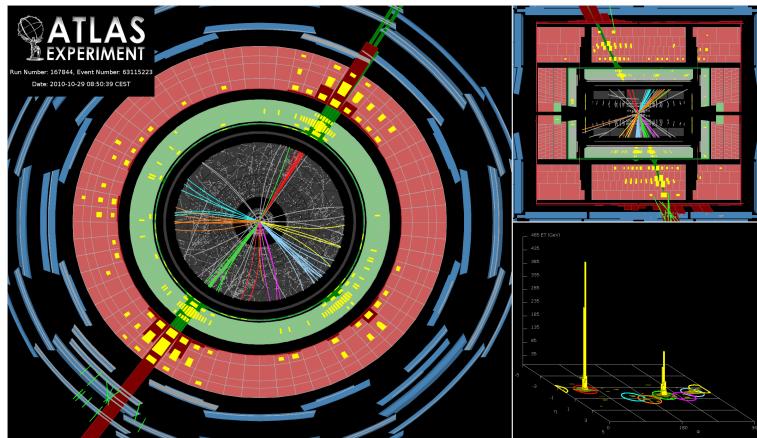


Figure 1. ATLAS event display containing two high p_T jets.

The event display shows three different views: a transverse (relative to the beam axis) $x - y$ view on the left panel, a parallel $x - z$ view on the top right panel, and a *lego* view on the bottom right panel. The lego view shows the projection of the energy deposits on the calorimeter into a two-dimensional (η, ϕ) plane that is obtained by unfolding the detector along the ϕ direction. This view is particularly relevant for image processing and computer vision approaches as it can be related to a fixed size *detector image*. The variable η is called *pseudo-rapidity* and is related to the polar angle θ .

¹ https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2011-047/fig_28.png

² Analyses of hadron collider data use the jet transverse momentum, p_T , rather than total energy. The jet p_T is defined as the magnitude of jet momentum transverse to the beam axis: $p_T = \sqrt{p_x^2 + P_y^2}$.

Jets are reconstructed using dedicated algorithms that group calorimeter energy deposits into clusters of fixed two-dimensional spatial (η, ϕ) size, corresponding to individual jets. Circles in the bottom right panel of Figure 1 represent the area of jets found with the *anti- k_t* clustering algorithm with radius parameter $R = 0.4$, whereas in yellow is shown the magnitude of the transverse energy of the calorimeter deposits within jets. For a review of jet algorithms see [8].

At the LHC, electroweak heavy particles (W, Z, and Higgs bosons, and top quarks) can be produced with large Lorentz boosts. The boost of the electroweak particle determines the distance in the (η, ϕ) plane between their decay products. Highly boosted heavy electroweak particles can hence lead to signatures of *jet substructure*, consisting of single (large radius) jets containing multiple nearby *sub-jets*. For example, experimental signatures boosted W bosons will contain two close-by sub-jets. Figure 2 shows an event display containing two boosted top quarks³. The lego view in the top right panel shows top-jet candidates as green circles with red circles indicating individual sub-jets. The size of the yellow squares is proportional to the transverse component of the energy deposits. A three-prong structure is clearly visible for the jet labeled as *hadronic top candidate*. In this case, the boosted jet corresponds to the process $t \rightarrow Wb \rightarrow qqb$.

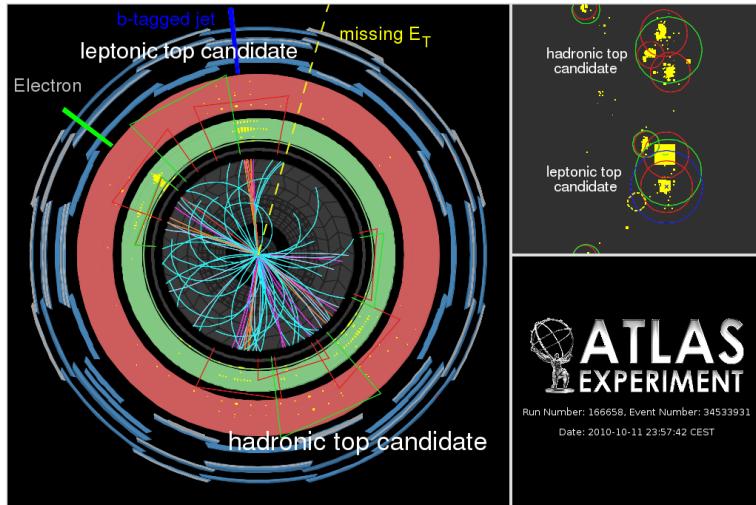


Figure 2. ATLAS event display containing two boosted top-quark jets.

The most difficult challenge to the identification of boosted jets from heavy electroweak particles at the LHC is the rejection of the very large background from multi-jet production in QCD. The key idea to classify boosted top or boson jets is to identify the internal structure of individual jets. The main observables used in jet substructure analyses that best separate signal (top and boson jets) from background (quark and gluon jets, generically referred as *QCD jets*) are the jet mass, the n-prong structure of the jets, for example measured with the variable *n-subjettiness* [9], and the radiation pattern within the jet. For a review of jet substructure techniques see [10][11][12].

3. Data samples

This review considers simulated jet data produced from Monte Carlo simulations of LHC collisions. Boosted W bosons were generated as signal using PYTHIA 8.170 [13] at $\sqrt{s} = 14$ TeV via the decay of a hypothetical heavy W' boson forced to decay to a hadronically decaying W boson ($W \rightarrow qq$) and a Z boson which decays invisibly to neutrinos. Multijet production

³ https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2011-073/fig_20.png

of quarks and gluons were generated as background using the same PYTHIA generator and center-of-mass energy as for W bosons.

4. Jet images

In order to use image processing and computer vision approaches for jet tagging at the LHC, a new data representation is introduced: the *jet-image* [14]. Jet images build on the notion of a detector as pixels on a digital camera, and jets as images, enabling the connection between the fields of computer vision and jet substructure and jet physics. Jet images are defined by a 25×25 grid of size (0.1×0.1) in $(\eta \times \phi)$ space centered around the axis of $R=1.0$ anti- k_t jets. The intensity of the pixels given by the transverse momentum p_T of the pixel cell. Prior to the application of computer vision classification techniques, a series of pre-processing steps to account for the space-time symmetries of jets images are applied. Pre-processing for the specific case of the identification of 2-prong jets such as those from the decay of W bosons, is defined by a *translation* that places the leading p_T subjet at the center of jet-image, a *rotation* such that the second p_T leading subjet is placed down below, and a *flip* operation such that the right side of the image has higher total p_T than the left. The goal of these pre-processing steps is to make it easier for image classification algorithms to focus on the important differences in physics between images. Jet-image pre-processing is illustrated in Figure 3 which shows the average jet-image of W (signal) and QCD (background) jets in a narrow p_T and mass bin before and after the the pre-processing steps.

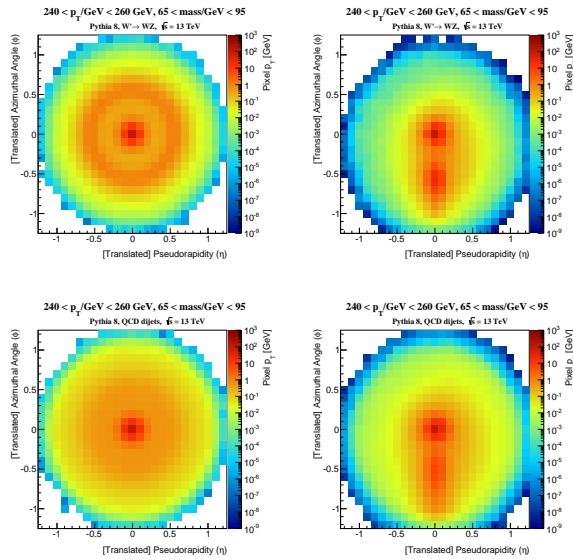


Figure 3. Average jet-image for W jets (top panel) and QCD jets (bottom panel) with $240 < p_T < 260$ GeV and $65 < \text{mass} < 95$ GeV before (left) and after (right) pre-processing.

5. Jet tagging using deep neutral networks

The concept of jet-images has enabled the use image classification methods for the identification (*tagging*) of boosted W boson and top quarks at the LHC. In the first case, through the use of *Fisher jets* [14]: a linear classifier inspired by facial recognition algorithms. In the second case, by the use of neural networks [15]. An earlier use of image-based event reconstruction from the OPAL Collaboration at the LEP collider is described in [16]. This paper focuses on the

application of modern computer vision algorithms based on deep neural networks for jet-image classification.

Using the pre-processed jet-image representation of jets, two different architectures of deep neural networks for image classification were trained to separate W from QCD jets: Convolutional Neural Networks (CNN), and Fully Connected (FC) MaxOut networks. This represents a significant departure from state-of-the-art jet tagging at LHC in that the full data image is used as input to the networks rather than a set of physics-defined jet substructure variables. In analogy to image recognition problems in computer vision, deep neural networks trained on the full jet-image are expected to learn high level representation of the data and capture new, more discriminant information and features between signal and background.

For a detailed description of the deep network architectures see [17]. The MaxOut architecture consists of two FC layers with MaxOut activations units followed by two FC layers with rectified linear units (ReLU), followed by a FC sigmoid layer for classification. The number of units per layer are 256, 128, 64, and 25. The CNN network consists of three convolution layers followed by two fully connected layers. Each convolution layer utilize 64 filters of sizes 11×11 , 3×3 , and 3×3 respectively. The large size of the filters in the first layer is motivated by the typical scale of the most prominent features in W jets, and the fact that images are very sparse, with an average occupancy of about 5%. Each convolutional layer consists of three sequential units: convolution, max-pool, and dropout, with ReLU activation functions. The max-pool down sampling in each convolution layer is (2,2), (3,3), and (3,3). The FC hidden layer consists of 64 units. Sigmoid units are used in the final classification layer. Figure 4 shows a diagram illustrating the CNN architecture.

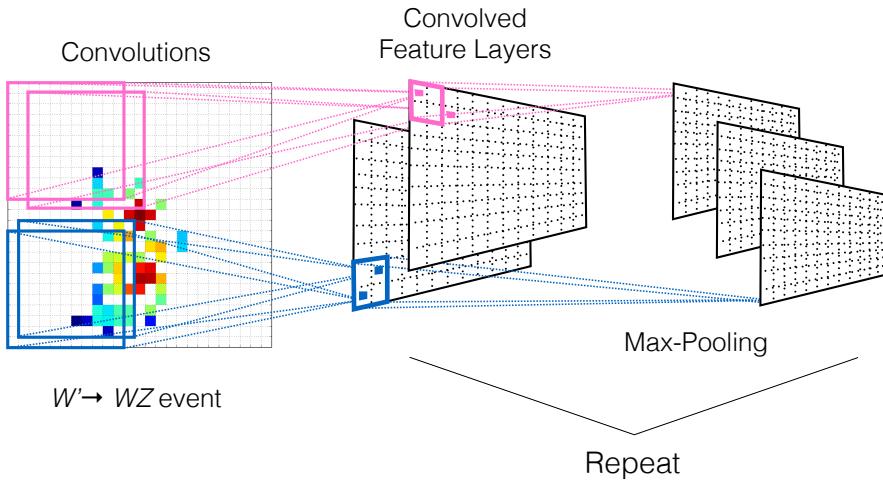


Figure 4. Convolutional neural network architecture applied to a jet-image.

The convolution layer can be intuitively understood as follows. In the first convolution layer, filters of 11×11 pixels are scanned across the jet-image in a sliding window approach. At each point in the jet-image, the convolution between the filter and the patch of the image behind it is computed. This process results in a new image for each filter. The filter values are learned during the CNN training. The following max pool layer then considers patches of size 2×2 in a sliding window approach and only the maximum value within the 2×2 patch is used to form the max-pool resulting image. In the next convolutional layer, new filters are run across all previous convolution images and a similar max-polling is performed. An unique aspect of CNN architectures is that the first convolution layer only have access to local information of size given by the filter size. In the subsequent layers, when the previous convolutions are combined, units

have access to more global information, until reaching the final FC layer where units combine information of the whole image. This architecture enables a hierarchical representation of the data where the first layers focus on local features for discrimination, for example individual subjets or local radiation patterns, whereas deeper layers learn higher level features such as distances between subjets or more global structure features within jets.

6. W boson jet tagging performance

The performance of the deep neural networks to identify W boson and reject QCD jets was evaluated for jets with $250 \text{ GeV} < p_T < 300 \text{ GeV}$ and within a $65 \text{ GeV} < m < 95 \text{ GeV}$ mass window containing the peak of the W boson mass. The primary figure of merit to compare the performance of different jet tagging techniques is the ROC curve, which measures the rejection of background QCD jets as a function of the efficiency to correctly classify signal W jets for different thresholds applied to the output of the classifier. Background rejection is defined as the inverse of the efficiency to incorrectly classify background jets as signal.

Figure 5 shows ROC curves comparing the performance of several single jet substructure variables (mass, n-subjettiness τ_{21} , and the distance ΔR between the two leading p_T subjets) and the two neural networks architectures. An additional *normalized* convolutional deep neural network (Convnet-norm) was also considered. In this setting, the pixel intensities were normalized by the total image intensity before training. Figure 5 shows that the performance of the three DNN classifiers significantly outperform all individual single jet substructure variables. We also observe that the MaxOut architecture performs slightly better than the convolutional networks. The curved denoted *Fisher* corresponds to the Fisher jet-image discriminant described in reference [14].

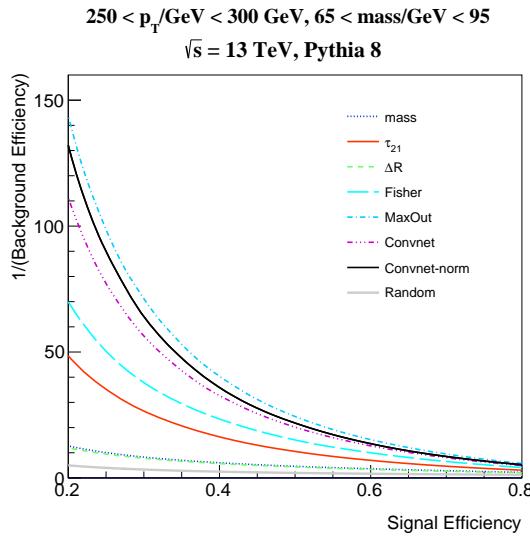


Figure 5. ROC curves for state-of-the-art jet substructure variables and computer vision deep neural network discriminants.

Figure 6(a) compares the performance of the DNN classifiers with two-variable combinations of jet substructure variables. While two-variable combinations achieve better discrimination power than individual variables, DNN classifiers still perform significantly better, indicating that the deep networks are learning new physics information not captured in the jet substructure variables considered.

In order to understand if the DNN have learned the jet substructure variables, Figure 6(b) shows ROC curves for combinations of the CNN classifier and each of the individual jet substructure variables. It can be seen than adding τ_{21} and ΔR to the CNN classifier does not improve the ROC curve, indicating that the network has fully learned these two variables. However, there is clear improvement in classification performance when the CNN is combined with the jet mass. This suggests that, while the CNN is able to learn new physics features to significantly outperform the combination of mass and τ_{21} , it does not fully learn the jet mass.

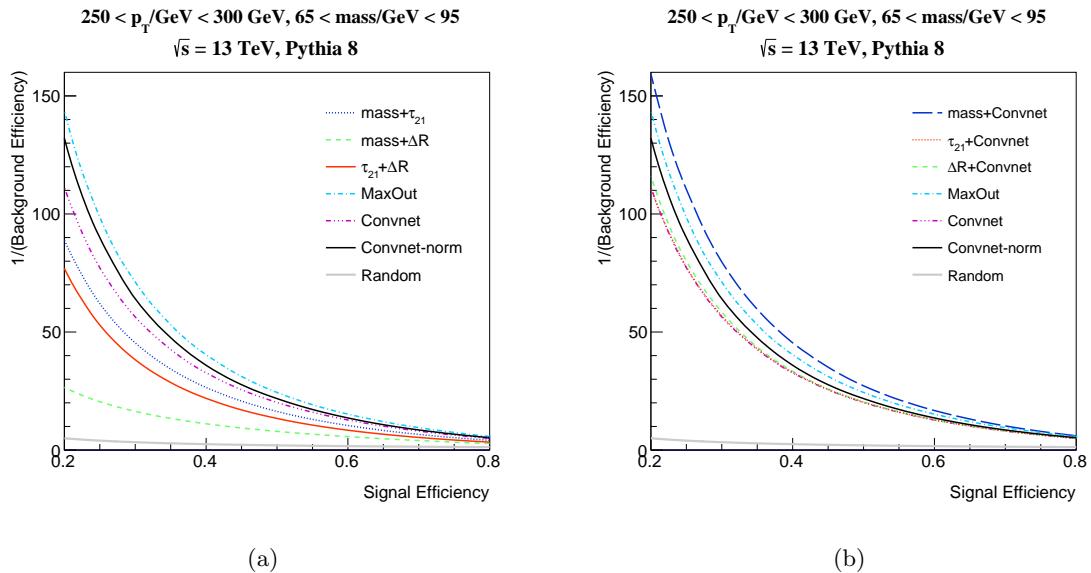


Figure 6. (a) ROC curves for two-variable combination of jet substructure features and computer vision deep neural network discriminants. (b) ROC curves for combinations of CNN output with individual jet substructure features.

7. Visualization of learned features

Image-based approaches to LHC data analysis can provide unique new ways to visualize what physics information is learned by the different classifiers. Here only a few examples are given. For a more detailed discussion see [17]

Figure 7 shows the conditional distribution of n-subjettiness τ_{21} and jet mass given the CNN output. The strong non-linear relationship between CNN and τ_{21} , indicates that the CNN has learned the information contained in n-subjettiness. On the other hand, the correlation between jet mass and CNN is significantly weaker, in agreement with the previous observation that jet mass is not fully learned by the deep neural networks architectures considered.

A very powerful way to visualize the physics information learned in deep representations is the *deep correlation jet-image*, constructed as the Pearson correlation coefficient of the pixel intensities with the deep neural network output, as shown in Figure 8. This image shows how discriminating information is contained within the network. In particular, one finds that the the localization and size of the second leading p_T subjet (located at the bottom of the image) is highly correlated with high CNN activations corresponding to signal W jets. This is consistent with our expectation for soft QCD gluon emission compared with hard emissions for the case of W jets. Broad blue regions around and outside the leading two subjets are indicative of gluon radiation expected in QCD jets.

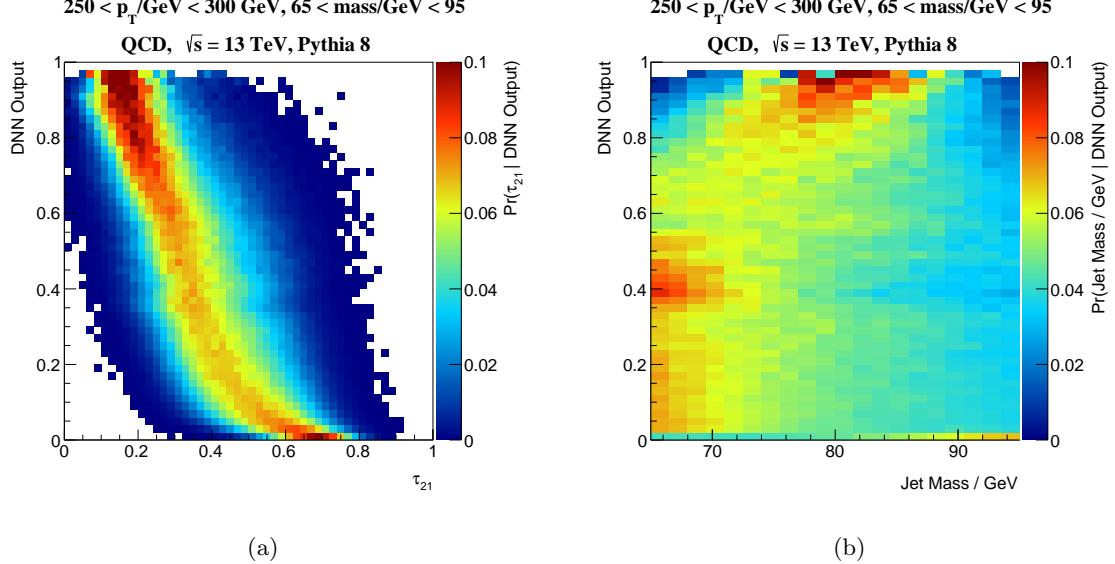


Figure 7. CNN output as a function of τ_{21} (a) and mass (b).

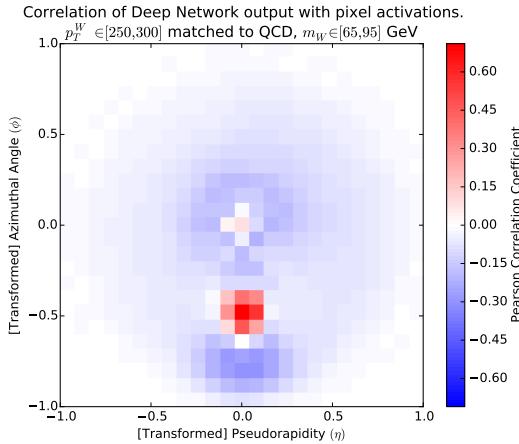


Figure 8. Deep correlation jet-image computed as the linear correlation between the CNN output and the pixel intensity. Red regions are signal-like, whereas blue regions are background-like

In order to understand what new physics information DNN are learning beyond jet mass and τ_{21} , we considered a highly *restricted phase space* so that τ_{21} and mass does not vary significantly over this reduced phase space. Figure 9 shows the ROC curves for the deep neural networks and the individual jet substructure variables for a subset of jets with very narrow values of p_T , mass, and τ_{21} . The small range of τ_{21} between 0.19 and 0.21 selects both signal and background jets that have a very similar (and significant) 2-prong structure. In other words, this restrictive selection ensures that signal and background jets look almost identical in terms of jet mass and n-subjettiness. As expected, jet mass and τ_{21} provide almost no discrimination power in Figure 9, whereas both CNN and MaxOut networks are still able to distinguish between signal and background within this restrictive phase space, confirming that they are learning new information beyond mass and τ_{21} .

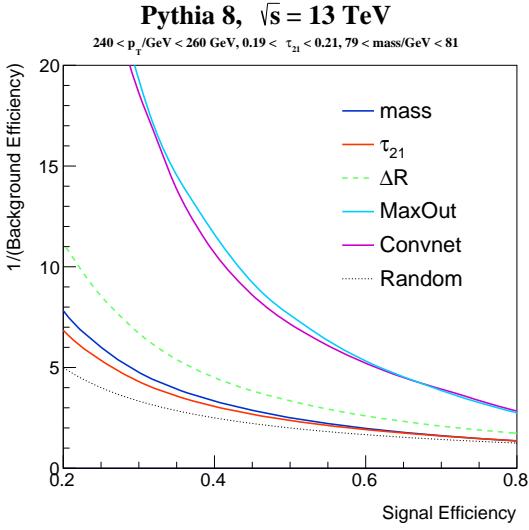


Figure 9. ROC curves in a highly restrictive phase space of $79 \text{ GeV} < m < 81 \text{ GeV}$ and $0.19 < \tau_{21} < 0.21$.

Figure 10 shows the deep correlation jet-images obtained in three small windows of τ_{21} . These images provide spacial information indicative of where in the image the networks are finding discriminating features beyond mass and n-subjettiness. A prominent feature in these images is the red region in-between the two leading p_T subjets in the cases of small and moderate τ_{21} values, related to the color flow pattern differences between W and QCD jets. Large values of τ_{21} , corresponding to jets that exhibit a strong one-prong structure, also show a particular radiation pattern not exploited in the two state-of-the-art jet substructure variables considered. These images can suggest further development of novel jet substructure variables that are able to capture these differences between signal and background.

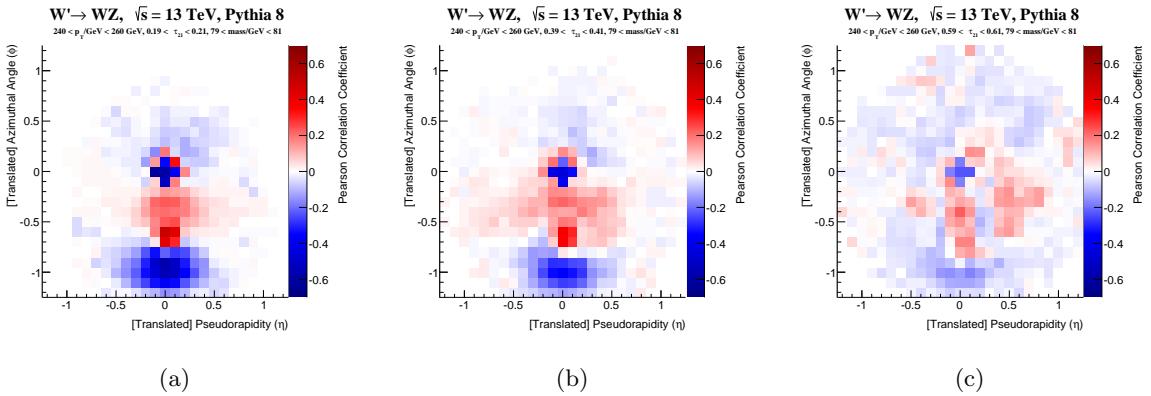


Figure 10. Deep correlation jet-image in a highly restrictive phase space of $79 \text{ GeV} < m < 81 \text{ GeV}$ and $0.19 < \tau_{21} < 0.21$ (a), $0.39 < \tau_{21} < 0.41$ (b), and $0.59 < \tau_{21} < 0.61$ (c).

8. Conclusions

This paper presents a new paradigm to study jets at the LHC. Representing jets as images enables the application of advanced computer vision and deep learning methods for jet classification, adding a new capability to understand LHC events and to design new, more powerful, classification methods. Jet-image deep neural networks classifiers significantly outperform state-of-the-art W boson jet classification methods as they are able to learn new discriminant features not captured in the physics motivated variables. Studies suggest that this additional information is related to the different color-flow patterns between W and QCD jets. While the DNN classifiers studied in this paper are very powerful, it was shown that they do not exploit all the information available for discrimination. Jet mass, in particular, is not fully learned by any of the DNN architectures considered. In addition to improved classification performance, DNN jet-image analysis provides new ways to visualize the information learned by the DNNs and to connect these visualizations with the physics of jets. The ideas presented in this paper have broad applicability and could be used for many different analyses where events can be represented as images both within and beyond high energy physics.

9. Acknowledgements

This work was supported by the Stanford Data Science Initiative and by the US Department of Energy (DOE) grant DE-AC02-76SF00515.

10. References

- [1] Cacciari M, Salam G P, and Soyez G 2008, *JHEP* **0804** 063
- [2] ATLAS Collaboration 2016, *JINST* **11** P04008
- [3] CMS Collaboration 2016, *J. Instrum.* **11** P01019
- [4] CMS Collaboration 2015, *J. Instrum.* **10** P08010
- [5] ATLAS Collaboration 2015, *ATL-PHYS-PUB-2015-022*
- [6] Russakovsky O, et. al. 2015, *arXiv:1409.0575* [cs.CV]
- [7] LeCun Y, Boser B, Denker J S, Henderson D, Howard R E, Hubbard W, et al. 1989, *Neural Comput.* **1** 541-551
- [8] Salam G P 2010, *Eur. Phys. J. C.* **67** 637-686
- [9] Thaler J, Tilburg K V 2011, *JHEP* **1103** 015
- [10] Abdesselam A, et. al. 2011, *Eur. Phys. J. C.* **71** 1661
- [11] Altheimer A, et. al. 2012, *J. Phys. G.* **39** 063001
- [12] Altheimer A, et. al. 2014, *Eur. Phys. J. C.* **74** 2792
- [13] Sjostrand T, Mrenna S, and Skands P Z 2008 *Comput. Phys. Commun.* **178** 852-867
- [14] Cogan J, Kagan M, Strauss E, Schwartzman A, 2015 *JHEP* **02** 118
- [15] Almeida L G, Backovic M, Cliche M, Lee S J, Perelstein M 2015, *arXiv:1501.05968* [hep-ph]
- [16] Thomson M A 1996, *Nuclear Instruments and Methods in Physics Research A* **382** 553-560
- [17] Oliveira L , Kagan M, Mackey L, Nachman B, Schwartzman A, 2015, *arXiv:1511.05190* [hep-ph]