

Deep Learning usage by Large Experiments

Benjamin Nachman

Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94704, USA

E-mail: bpnachman@lbl.gov

Abstract. The modern machine learning revolution holds great promise for high energy physics. Enormous detectors have been designed to search for rare signals in order to expose the smallest distance scales in nature. Modern tools like deep neural networks will be required to fully explore the sub-nuclear universe. Using the radiation pattern inside high energy jets as an example, we show how neural networks are being used in new ways to produce a class of techniques that were not previously possible.

1. Introduction

Advances in machine learning have led to a “deep learning revolution” that has resulted in many tools and techniques for quickly and reliably training sophisticated neural networks with many layers. Complex neural networks can learn directly from low-level features that reside in the high-dimensional datasets common in large experiments. New ideas for deep neural networks (DNN) are being developed at a fast pace, but Ref. [1–3] are rather comprehensive introductions/reviews.

In many applications to large experiments like ATLAS and CMS at the Large Hadron Collider (LHC), DNNs are ‘simply’ replacing shallower learning algorithms. For example, the ATLAS experiment has recently used deep neural networks to combine a small number of high-level features for highly Lorentz boosted boson, top quark, and bottom quark tagging [5–7]. Unsurprisingly, performance is essentially the same as training shallower classifiers with the same inputs. Deep neural networks are most useful when the relevant dimensionality of the inputs is very large, $\sim \mathcal{O}(100)$ or bigger.

Two areas from large experiments in the high energy physics community making use of high dimensional data with deep neural networks are neutrino physics and jet physics. Neutrino physics detectors require a large volume to compensate for the small interaction rate. The data from liquid argon time projection chambers (e.g. NO ν A, DUNE [8], MicroBooNE [9]) or vats of liquid scintillator surrounded by photomultiplier tubes (e.g. JUNO [10]) can be represented as a series of images with translationally invariant features. This makes convolutional neural networks (CNNs) a natural choice for this setting and there are now many proposed and applied techniques of CNN’s to neutrino physics. Jets are collimated sprays of particles resulting from high energy quark and gluon production. The radiation pattern inside jets contains a wealth of information about the particle that initiated them. There are many proposals for using deep learning with this radiation pattern; see Ref. [4] for a recent review. The remainder of this

talk focuses on representing jets as an image [20]¹ to illustrate the application of deep neural networks for classification [21–26], regression [28], and generation [29–31].

2. The Jet Image and Special Relativity

Jet images are a two-dimensional fixed representation of the radiation pattern inside a jet. Most applications with jet images use a gray-scale digital image where the pixel intensity corresponds to the momentum deposited in a particular region of the detector. Figure 1 illustrates how the energy pattern in the cylindrical coordinates of a typical collider physics detector are unrolled to form a two-dimensional jet image. Unlike natural images of scenery or faces, jet images do not have any clear sharp features, do not have smooth pixel intensities, and are mostly empty (low occupancy). This is a challenge for applying standard techniques out-of-the-box.

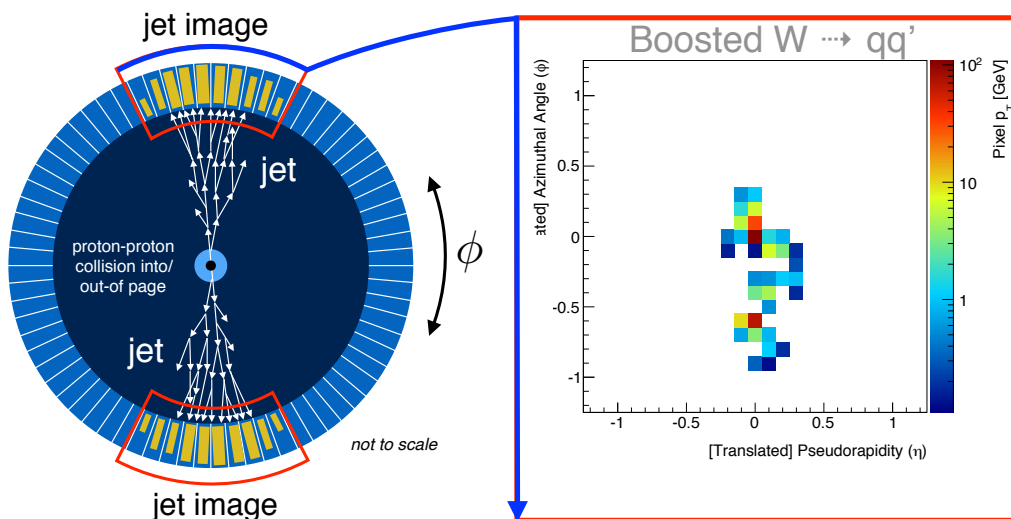


Figure 1. Left: a transverse view of a particle collision where the incoming particles are into/out of the page. Jets are collimated sprays of particles that are measured as localized high energy deposits. The cylindrical geometry is unrolled to form a jet image; Right: a typical jet image from a Lorentz boosted W boson decay. The pixel intensity is the momentum transverse to the incoming particles (p_T). Image reproduced from Ref. [30].

Despite the differences with natural images, jet images are a powerful way to represent the radiation pattern inside a jet because one can directly visualize the physical information that is available for learning. For example, Fig. 2 shows two jet images where a clear physical story can describe their differences and builds a foundation for applying sophisticated techniques for exploiting the differences between the images.

The first step in any image-processing analysis is image pre-processing. The symmetries of spacetime are known symmetries of jet images. Removing them during pre-processing can help machines to learn faster and smarter, but must be used with caution. For example, Fig. 3 shows how jet image rotations (which are not proper rotations in space) can distort the information content of a jet image. Other standard pre-processing steps like image normalization are known to distort the information inside a jet image [21]. Some pre-processing can nonetheless be useful and is an opportunity to inject domain knowledge.

¹ There are also many proposals for non-image-based techniques. See e.g. [11–19].

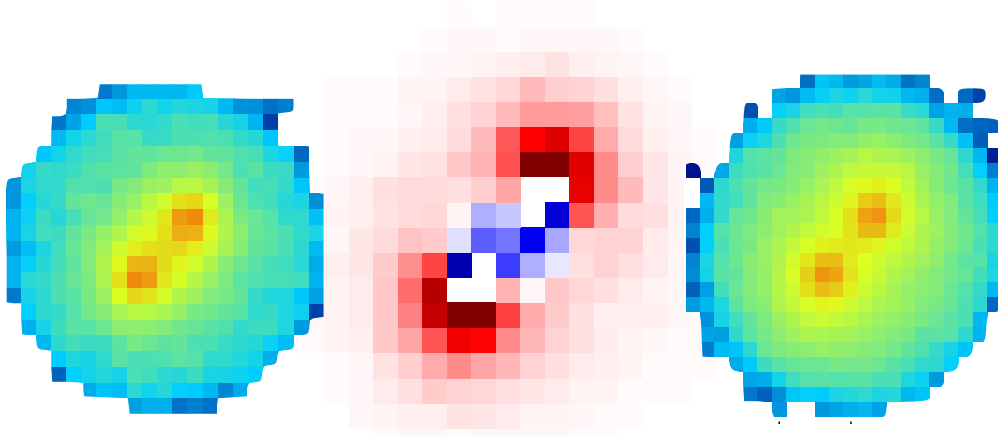


Figure 2. The average over many jet images from a boosted Higgs boson decay into two bottom quarks (left), a gluon³ splitting into two bottom quarks (right), and the difference between the left and right images (center). The two energy lobes in the left and right images represent the ‘subjets’ formed from nearby quarks. The Higgs boson does not participate in the strong force and therefore the quarks it produces have equal and opposite strong-force-charge; they thus radiate like a dipole. In contrast, the gluon has net strong force charge and therefore radiates much more away from the two energy lobes.

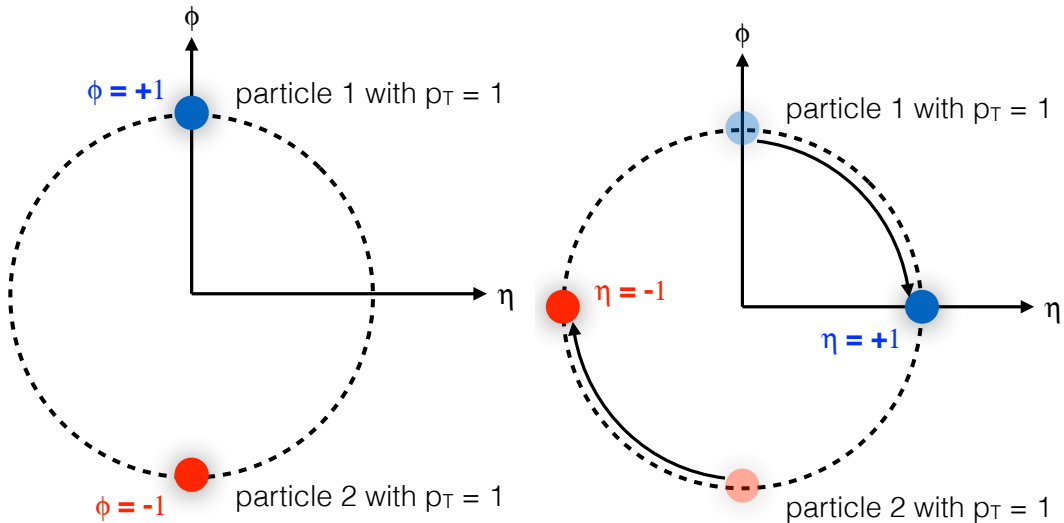


Figure 3. The jet mass of an image is derived from the four-vector sum of the pixel intensities, each treated as massless. The jet mass is a powerful feature for distinguishing interesting jets from generic jets. The images shown here have a simplified ‘jet’ with only two hit pixels. Rotating the image by $\pi/4$ does not preserve the image mass if the pixel intensities are left unchanged. This introduces a rotation-dependent smearing to the jet mass that can reduce the discrimination potential of a classifier.

³ Technically, this is implemented as a Higgs boson decaying into two bottom quarks where the color strings are ‘by-hand’ distorted to make this Higgs act like a gluon in its radiation pattern. In this way, the Higgs kinematics are the same between the two samples. Due to color-connections between the gluon-like Higgs and the beam remnants, there is a dependence on the rest of the event for the right plot but not the left plot.

3. Classification

The most basic machine learning task is differentiating between a small number of categories (classification). Machine learning algorithms can be trained to identify the particle that originated the jet by analyzing the radiation pattern in the jet image. Figure 4 shows a Receiver Operating Characteristic (ROC) for the binary classification task of distinguishing jets from Lorentz boosted W bosons and generic quarks/gluons. The classifier is trained directly on the pixel intensities without any other guidance (aside from pre-processing). A reoccurring theme is that the machine learning algorithm can match or out-perform existing techniques. Understanding what the machine has learned beyond existing algorithms is a very important and active subject of research.

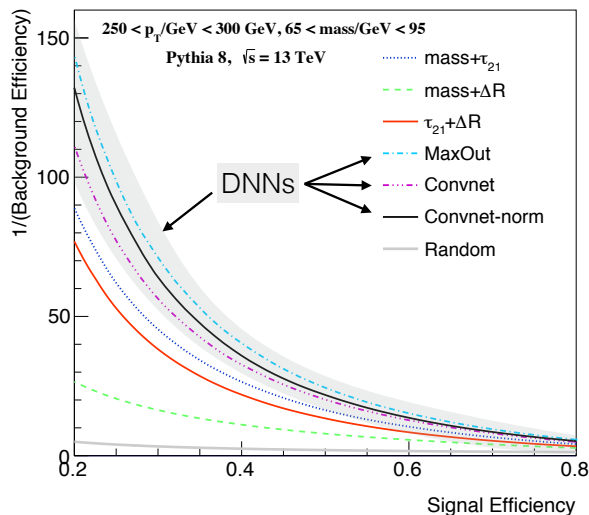


Figure 4. An ROC curve that shows the tradeoff between the probability to correctly identify a Lorentz boosted W boson jet and the probability to mis-classify a generic quark/gluon jet. Various deep neural network architectures are shown in the shaded region. The other observables are known (simple) features used in existing techniques for jet classification. Reproduced from Ref. [21].

One opportunity that is also a key challenge of deep neural networks is their ability to utilize subtle features within a jet. Traditional classification techniques in high energy physics use fully supervised learning where labeled examples are produced from simulation. However, subtle features are difficult to accurately model and therefore the classifier trained on simulation may be sub-optimal when applied on (unlabeled) data. There are now multiple proposals for training with data directly [32–34]. One proposal based on weak supervision is shown in Fig. 5.

4. Regression

A natural generalization of classification is to cases when there are (possibly infinitely) many classes. Jet image-based regression techniques have recently been applied for image de-noising.

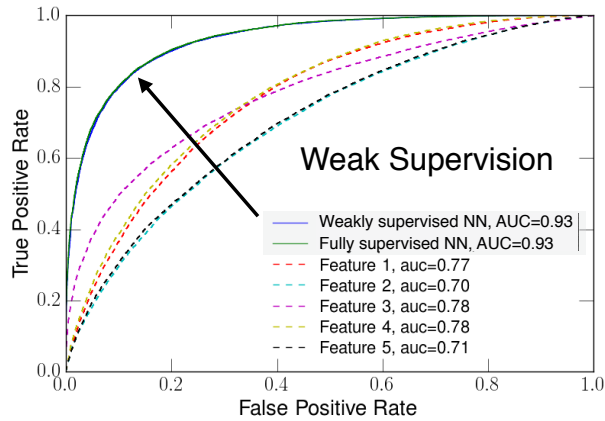


Figure 5. An ROC curve that shows how the performance of a traditional fully supervised learner with per-example labels can be matched with a technique that only knows the class proportions. Reproduced from Ref. [32].

At the LHC, many proton-proton collisions happen simultaneously, but only (at most) one is interesting. Using information from different detectors, a noise image can be predicted from the various measurements. The noise jet image can be subtracted from the original noisy image, after which traditional or DNN-based classification can proceed. This is illustrated in Fig. 6. Extensions, modifications, and variations of this idea are currently under study by the ATLAS and CMS collaborations for estimating detector distortions to the signal pixel intensity and not just the noise (‘jet energy scale’).

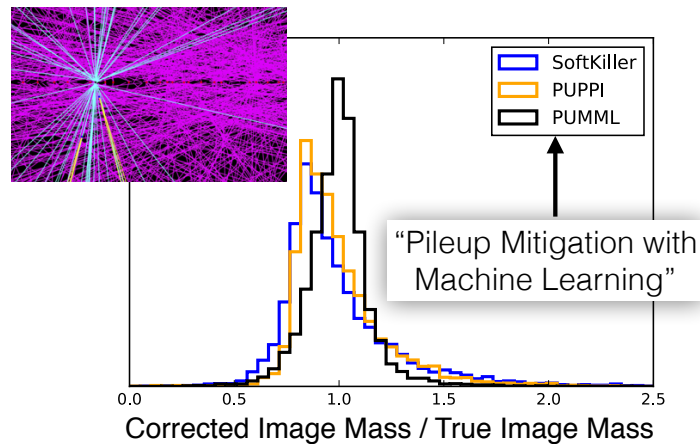


Figure 6. Top left: At the LHC, many collisions occur at the same time. Only one is of interest and the particles produced by the other collisions are a source of noise. Applying a jet-image-based CNN for de-noising (‘PUMML’) is very effective at correcting the jet mass back to the value without noise. In comparison, existing techniques (‘Softkiller’ and ‘PUPPI’) introduce a bias and have a worse resolution. Reproduced from Ref. [28].

5. Generation

Rounding out the applications of neural networks is generation, i.e. producing new images which can be used for subsequent classification or inference. One technique that has gained attention in high energy physics is the adversarial neural network (GAN) [35]. Neural networks are simply (compositions of) functions, so a neural network generator is a mapping between noise and structured images. In the GAN setup, this function from noise to structured images is complemented by a classifier network that tries to distinguish generated images from real images. The two networks are iteratively optimized and when the classifier is maximally confused, the generator will be effective at producing realistic looking images. High energy physics analysis relies heavily on physics-based simulation, but these simulators can be very slow (up to $\mathcal{O}(\text{min}/\text{event})$). An exciting possibility is to use GANs to replace or augment physics-based simulations in order to speed them up. Figure 7 shows how the jet mass distribution can be learned by a GAN even though it was only shown examples of raw pixel intensities during training. One can naturally extend the jet image technology to multiple causally connected layers as is typical of a longitudinally segmented calorimeter. This extended geometry is shown in Fig. 9. Calorimeter simulation is a bottleneck for simulation at the LHC and speeding it up with a high-fidelity neural network simulator has the potential to save immense resources and empower a broad set of physics analyses.

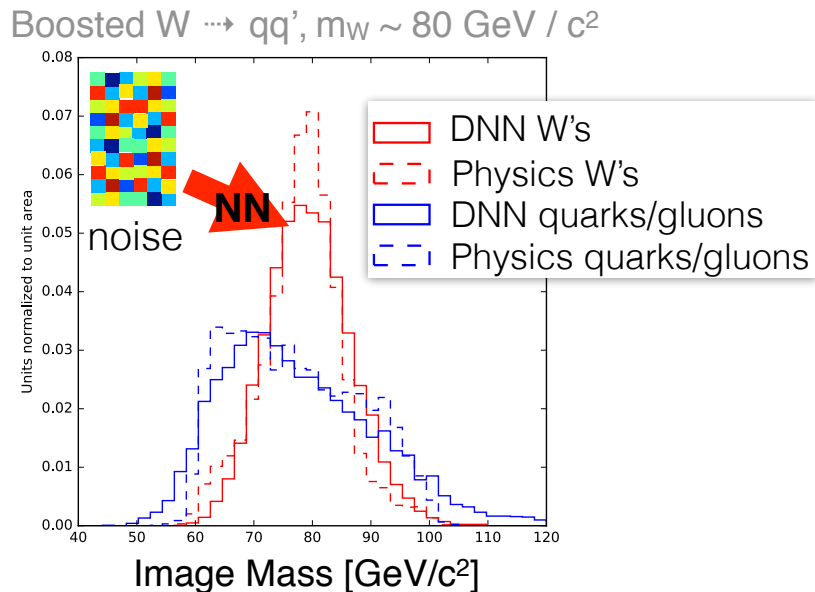


Figure 7. The jet mass distribution for signal (Lorentz boosted W bosons) and background (generic quarks and gluons) for a physics-based-simulator as well as a GAN trained on simulation images. Reproduced from Ref. [30].

6. Conclusions and Outlook

Modern machine learning tools have attracted widespread interest in the high energy physics community. DNN classification, regression, and generation are powerful tools to fully exploit the

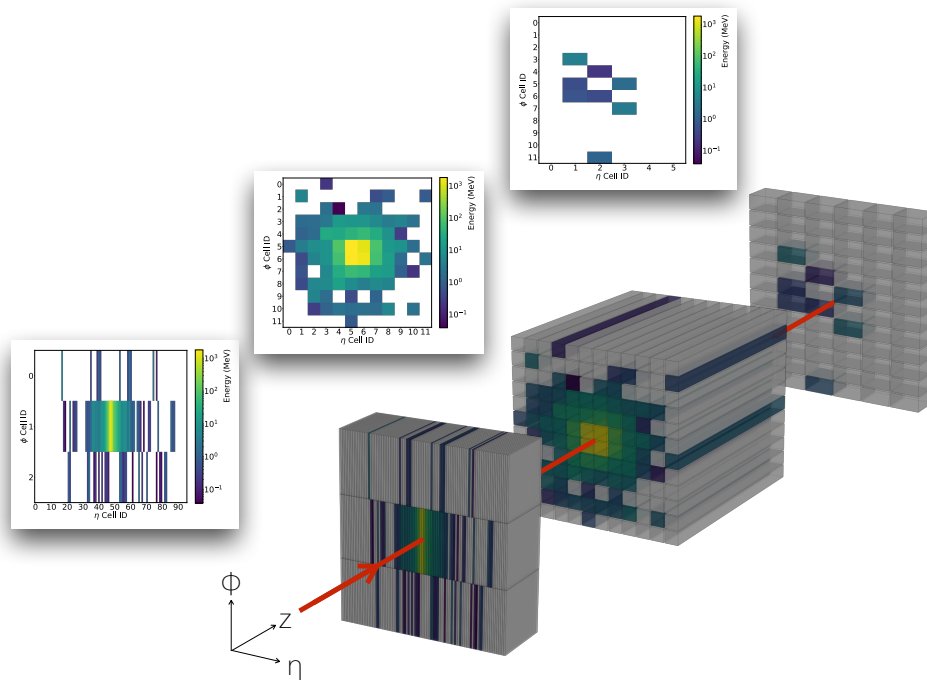


Figure 8. The setup of a three-layer calorimeter. The energy deposition in each layer resembles a grayscale jet image. Reproduced from Ref. [29].

physics program at the LHC and other large experiments with high-dimensional data. Some of these algorithms have already been studied using data or full detector simulations at the LHC. Figure 9 shows a CMS collaboration comparison of various techniques, including one developed by CMS in another context [18, 19]. For the tasks studied so far, it appears that any modern tool with access to the full radiation pattern has the same (excellent) performance. Taking these techniques to data and understanding exactly what the machines are learning will add robustness to the approaches and may even help us to learn something new about nature.

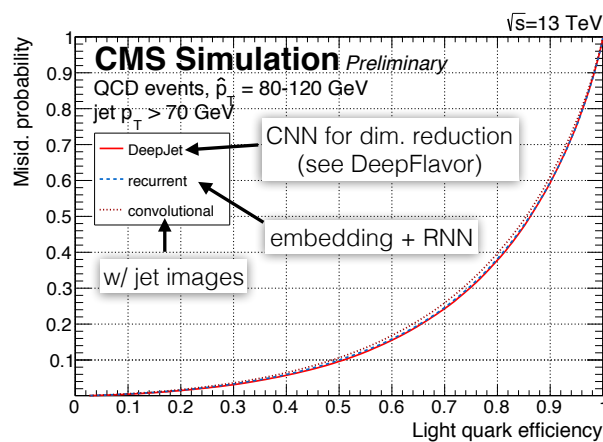


Figure 9. An ROC curve trading off the probability to correctly label a quark jet as such with the probability to mis-label a gluon-initiated jet as a quark jet. Three different DNN architecture types are compared. Reproduced from Ref. [27].

7. Acknowledgments

This work was supported by the U.S. Department of Energy, Office of Science under contract DE-AC02-05CH11231.

References

- [1] I. Goodfellow, Y. Bengio and A. Courville, ‘Deep Learning,’ (2016) MIT Press.
- [2] J. Schmidhuber, ‘Deep Learning in Neural Networks: An Overview,’ *Neural Networks* 61 (2015) 85.
- [3] Y. LeCun, Y. Bengio and G. Hinton, ‘Deep Learning,’ *Nature* 521 (2015) 436.
- [4] A. J. Larkoski, I. Moutl and B. Nachman, ‘Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning,’ [arXiv:1709.04464 [hep-ph]].
- [5] ATLAS Collaboration, ‘Machine Learning Techniques for the Identification of Hadronically Decaying W Bosons and Top Quarks in ATLAS,’ [ATL-PHYS-PUB-2017-004].
- [6] ATLAS Collaboration, ‘Performance of Top Quark and W Boson Tagging in Run 2 with ATLAS,’ [ATLAS-CONF-2017-064].
- [7] ATLAS Collaboration, ‘Optimisation and performance studies of the ATLAS *b*-tagging algorithms for the 2017-18 LHC run ,’ [ATL-PHYS-PUB-2017-013].
- [8] R. Acciarri *et al.* [MicroBooNE Collaboration], ‘Convolutional Neural Networks Applied to Neutrino Events in a Liquid Argon Time Projection Chamber,’ *JINST* **12**, no. 03, P03011 (2017). [arXiv:1611.05531 [physics.ins-det]].
- [9] A. Aurisano *et al.*, ‘A Convolutional Neural Network Neutrino Event Classifier,’ *JINST* **11**, no. 09, P09001 (2016). [arXiv:1604.01444 [hep-ex]]. see also F. Psihas *et al.*, these proceedings.
- [10] L. Wang *et al.*, ‘Muon reconstruction of JUNO experiment with convolutional neural networks’, these proceedings.
- [11] J. Pearkes, W. Fedorko, A. Lister and C. Gay, ‘Jet Constituents for Deep Neural Network Based Top Quark Tagging,’ [arXiv:1704.02124 [hep-ex]].
- [12] G. Louppe, K. Cho, C. Becot and K. Cranmer, ‘QCD-Aware Recursive Neural Networks for Jet Physics,’ [arXiv:1702.00748 [hep-ph]].
- [13] A. Butter, G. Kasieczka, T. Plehn and M. Russell, ‘Deep-learned Top Tagging using Lorentz Invariance and Nothing Else,’ [arXiv:1707.08966 [hep-ph]].
- [14] K. Datta and A. J. Larkoski, ‘Novel Jet Observables from Machine Learning,’ [arXiv:1710.01305 [hep-ph]].
- [15] K. Datta and A. Larkoski, ‘How Much Information is in a Jet?,’ *JHEP* **1706**, 073 (2017).
- [16] J. A. Aguilar-Saavedra, J. H. Collins and R. K. Mishra, ‘A generic anti-QCD jet tagger,’ [arXiv:1709.01087 [hep-ph]].
- [17] ATLAS Collaboration, ‘Identification of Jets Containing *b*-Hadrons with Recurrent Neural Networks at the ATLAS Experiment,’ [ATL-PHYS-PUB-2017-003]. see also M. Paganini *et al.*, these proceedings.
- [18] CMS Collaboration, ‘Heavy flavor identification at CMS with deep neural networks,’ [CMS-DP-2017-005].
- [19] CMS Collaboration, ‘CMS Phase 1 heavy flavour identification performance and developments,’ [CMS-DP-2017-013].
- [20] J. Cogan, M. Kagan, E. Strauss and A. Schwartzman, ‘Jet-Images: Computer Vision Inspired Techniques for Jet Tagging,’ *JHEP* **1502**, 118 (2015). [arXiv:1407.5675 [hep-ph]].
- [21] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, ‘Jet-images - deep learning edition,’ *JHEP* **1607**, 069 (2016). [arXiv:1511.05190 [hep-ph]].
- [22] P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, ‘Jet Substructure Classification in High-Energy Physics with Deep Neural Networks,’ *Phys. Rev. D* **93**, no. 9, 094034 (2016). [arXiv:1603.09349 [hep-ex]].
- [23] G. Kasieczka, T. Plehn, M. Russell and T. Schell, ‘Deep-learning Top Taggers or The End of QCD?,’ *JHEP* **1705**, 006 (2017). [arXiv:1701.08784 [hep-ph]].
- [24] J. Barnard, E. N. Dawe, M. J. Dolan and N. Rajcic, ‘Parton Shower Uncertainties in Jet Substructure Analyses with Deep Neural Networks,’ *Phys. Rev. D* **95**, no. 1, 014018 (2017). [arXiv:1609.00607 [hep-ph]].
- [25] P. T. Komiske, E. M. Metodiev and M. D. Schwartz, ‘Deep learning in color: towards automated quark/gluon jet discrimination,’ *JHEP* **1701**, 110 (2017). [arXiv:1612.01551 [hep-ph]].
- [26] ATLAS Collaboration, ‘Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector,’ [ATL-PHYS-PUB-2017-017].
- [27] CMS Collaboration, ‘New Developments for Jet Substructure Reconstruction in CMS,’ [CMS-DP-2017-027]. see also M. Stoye *et al.*, these proceedings.
- [28] P. T. Komiske, E. M. Metodiev, B. Nachman and M. D. Schwartz, ‘Pileup Mitigation with Machine Learning (PUMML),’ [arXiv:1707.08600 [hep-ph]]. See also E. M. Metodiev *et al.*, these proceedings.
- [29] M. Paganini, L. de Oliveira and B. Nachman, ‘CaloGAN: Simulating 3D High Energy Particle Showers in

- Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks,” [arXiv:1705.02355 [hep-ex]]. see also L. de Oliveira et al., these proceedings.
- [30] L. de Oliveira, M. Paganini and B. Nachman, “Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis,” [arXiv:1701.05927 [stat.ML]].
 - [31] S. Vallecorsa, GeantV Collaboration, “A Machine Learning tool for fast simulation in GeantV,” these proceedings.
 - [32] L. M. Dery, B. Nachman, F. Rubbo and A. Schwartzman, “Weakly Supervised Classification in High Energy Physics,” *JHEP* **1705**, 145 (2017). [arXiv:1702.00414 [hep-ph]].
 - [33] E. M. Metodiev, B. Nachman and J. Thaler, “Classification without labels: Learning from mixed samples in high energy physics,” [arXiv:1708.02949 [hep-ph]].
 - [34] G. Louppe, M. Kagan and K. Cranmer, “Learning to Pivot with Adversarial Networks,” [arXiv:1611.01046 [stat.ME]].
 - [35] I. Goodfellow, et al. “Generative Adversarial Networks,” [arXiv:1406.2661 [stat.ML]].