

Convolutional Neural Networks in High Energy Physics for Particle Classification

Claudio Barros

University at Buffalo

May 15, 2021

Outline

- 1 Particle Kinematics
- 2 Jets and Jet Clustering
- 3 Neural Networks
- 4 Convolutional Neural Networks
- 5 Results
- 6 Conclusion

Particle Kinematics

- Experiments at the Large Hadron Collider (LHC) involve collisions between particles at relativistic speeds, with a center-of-mass energy on the order of $\sqrt{s} = 13 \text{ TeV}$.
- By colliding hadrons we can probe their inner structure in the hopes of discovering new physics beyond the standard model.

Overall view of the LHC experiments.

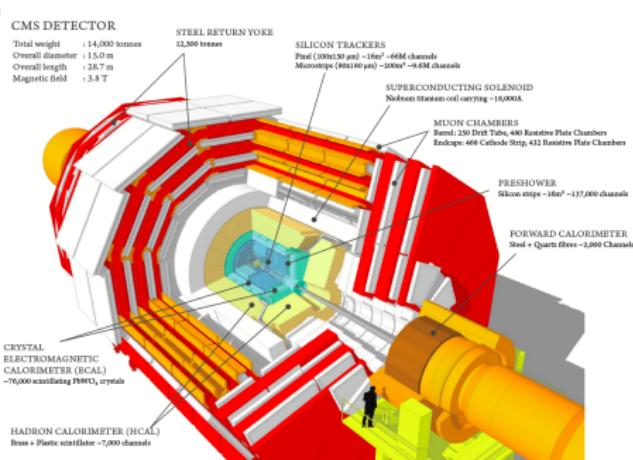
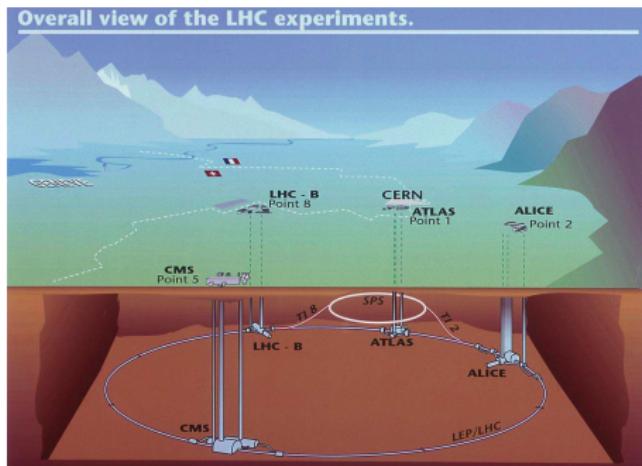


Figure 1: Aerial view of the LHC, and side view of the CMS detector [1,2].

We can start by understanding the kinematics of the particles at play. Let's consider the effect of boosting a particle with four-momentum $p^\mu = (E, p_x, p_y, p_z)$ along its longitudinal component (i.e. z-axis).

- The components along the z-axis, (E, p_z) , are Lorentz not invariant.
- The components orthogonal to the z-axis, (p_x, p_y) , are Lorentz invariant.
- Thus, boosting from a rest frame transforms the longitudinal components of p_z as [3]:

$$\begin{pmatrix} E \\ p_z \end{pmatrix} = \exp \begin{pmatrix} 0 & w \\ w & 0 \end{pmatrix} \begin{pmatrix} m \\ 0 \end{pmatrix} = \begin{pmatrix} \cosh w & \sinh w \\ \sinh w & \cosh w \end{pmatrix} \begin{pmatrix} m \\ 0 \end{pmatrix}$$

Where w is the rapidity and m is the particle mass. The rapidity and Lorentz factor are related through $\gamma = (1 - \beta^2)^{-1/2} = \cosh w$.

Putting all this together, we can choose the transverse momentum p_T , mass m , and azimuthal angle ϕ as our invariant variables, and the rapidity w as our transforming variable.

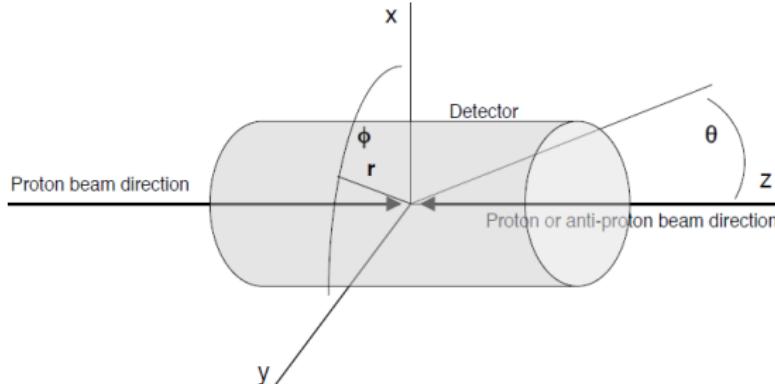


Figure 2: ATLAS and CMS coordinate system. The azimuthal angle ϕ circles the z-axis, while θ is measured from z. The LHC center points along y [4].

From this geometry we can see the following relations [5]:

$$p_z = |\mathbf{p}| \cos \theta$$

$$p_x = p_T \cos \phi$$

$$p_T = |\mathbf{p}| \sin \theta$$

$$p_y = p_T \sin \phi$$

$$m_T^2 = p_T^2 + m^2$$

$$p_T = \sqrt{p_x^2 + p_y^2}$$

Our full set of coordinates transform as $(p_T, w, \phi, m) \rightarrow (p_T, w + w_z, \phi, m)$.

Using the relations $E = \gamma m$ and $|\mathbf{p}| = \gamma m\beta$, we see that $\frac{|\mathbf{p}|}{E} = \beta$. Combining this with $\tanh w = \beta$ gives us a useful relation for the rapidity:

$$\tanh w = \frac{|\mathbf{p}|}{E} \quad \text{so} \quad w = \operatorname{artanh} \frac{|\mathbf{p}|}{E} = \frac{1}{2} \ln \frac{E + |\mathbf{p}|}{E - |\mathbf{p}|}$$

If we now define w in terms of p_z we can see clearly that $w \rightarrow 0/\infty$ as $p_z \rightarrow 0/E$, which helps give some physical intuition:

$$w = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}$$

Unfortunately, w is not always a useful parameter since it requires knowledge of E and p_z . A practical quantity in the relativistic limit (i.e. $E = |\mathbf{p}|$) is the *pseudo-rapidity* η :

$$w = \frac{1}{2} \ln \frac{1 + p_z/|\mathbf{p}|}{1 - p_z/|\mathbf{p}|} = -\ln \sqrt{\frac{1 - \cos \theta}{1 + \cos \theta}} = -\ln \left(\tan \frac{\theta}{2} \right)$$

We define this quantity as the pseudo-rapidity $\eta \equiv -\ln(\tan(\theta/2))$. Unlike the rapidity, differences the in pseudo-rapidity are not generally invariant, so we restrict our usage of η to the relativistic limit.



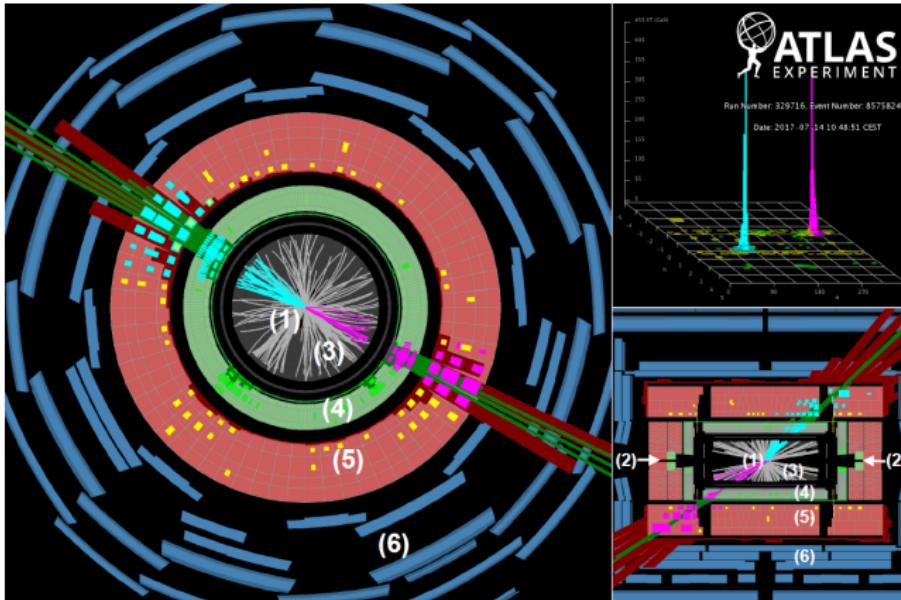


Figure 3: A proton-proton collision recorded by ATLAS depicting a two jet creation event. The upper right-hand side contains calorimeter clusters of transverse energies in the (η, ϕ) plane, while the lower graph depicts a longitudinal view; the z-direction points into the page [5].

Jets and Jet Clustering

A jet can be informally thought of as a collimated flow of particles. A concrete algorithm for reconstructing them is as follows [6]:

- ① An event at the LHC constitutes as a set of initial particles to be clustered.
- ② Find the pair (m, n) in this set with the minimal inter-particle distance d_{mn} defined as:

$$d_{mn} = \min(p_{Tm}^{2\lambda}, p_{Tn}^{2\lambda}) \Delta R_{mn}^2$$

where $\lambda = (-1, 0, 1)$ depending on which algorithm is used, and $\Delta R_{mn} = \sqrt{\Delta\eta_{mn}^2 + \Delta\phi_{mn}^2}$ is the particle separation in the $\eta - \phi$ (or the $w - \phi$) plane.

- ③ Find the single particle (i) with the minimal particle-Beam distance:

$$d_{iB} = p_{Ti}^{2\lambda} R^2$$

where R is the radius of the jet.

- ④ If $d_{iB} < d_{mn}$, particle i is classified as a jet and removed from the set. Otherwise, (m, n) are combined into an object k which is re-added to the set. Repeat until the set is empty.

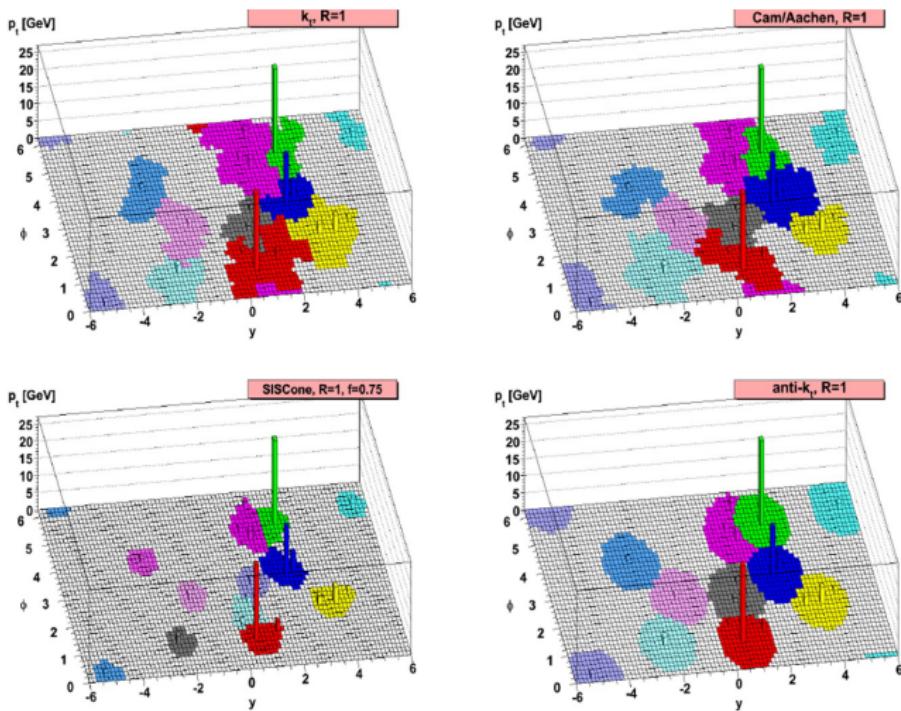


Figure 4: Jet reconstruction algorithms for different values of λ and with $R = 1$. Clockwise from top left: k_T algorithm for $\lambda = 1$, Cambridge/Aachen algorithm for $\lambda = 0$, and anti- k_T algorithm for $\lambda = -1$. The circular shape of anti- k_T is to be noted, which contributes to its widespread usage in LHC experiments [7].

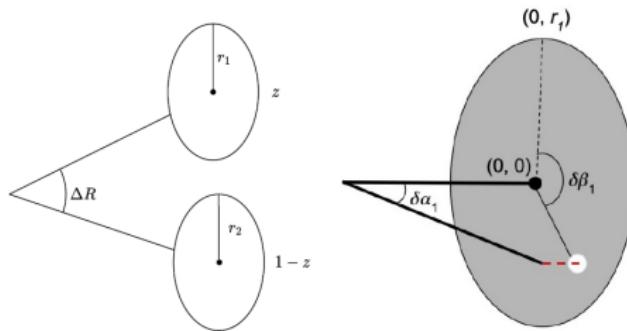


Figure 5: A toy di-jet event (left) depicting jet substructure (right). Each sub-jet is comprised of 10 particles [8].

We will adopt a simplistic toy model as our basis of jet production. Here a jet splits into two sub-jets of radii r_1 and r_2 , with normalized momenta z and $1 - z$ respectively. The sub-jet angles $\delta\alpha_i$ and $\delta\beta_i$ are analogous to the LHC jet angles θ and ϕ , while $\Delta R = \theta$ since we will focus on jets with $p_z \ll E$ so that $\eta = 0$ [9].

A final but crucial point to consider is the source of a jet. Due to the high energies of colliding protons, a myriad of resonances are produced which can be divided into two categories:

- A heavy, possibly BSM, resonance which decays into lighter W/Z/H bosons, which then decay hadronically to produce jets. These are the signal sources.
- Everything else. This includes QCD jets that originated from partons; these are the background jets.

The signal/background jets follow Gaussian/exponential distributions respectively, a well known empirical fact [10,11]. Identifying the signal jets (i.e. the resonant anomaly) against a noisy background is the key goal of this project.

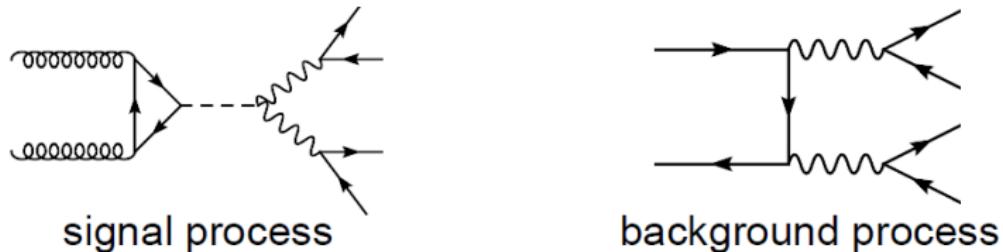


Figure 6: Feynman diagrams for Higgs production and a background process [12].

Neural Networks

To successfully differentiate between signal and background jets we shall make use of a Neural Network (NN) with binary classification (i.e. a multi-layer perceptron). As the name implies, the basic unit of a NN is a neuron, in analogy to a biological brain.

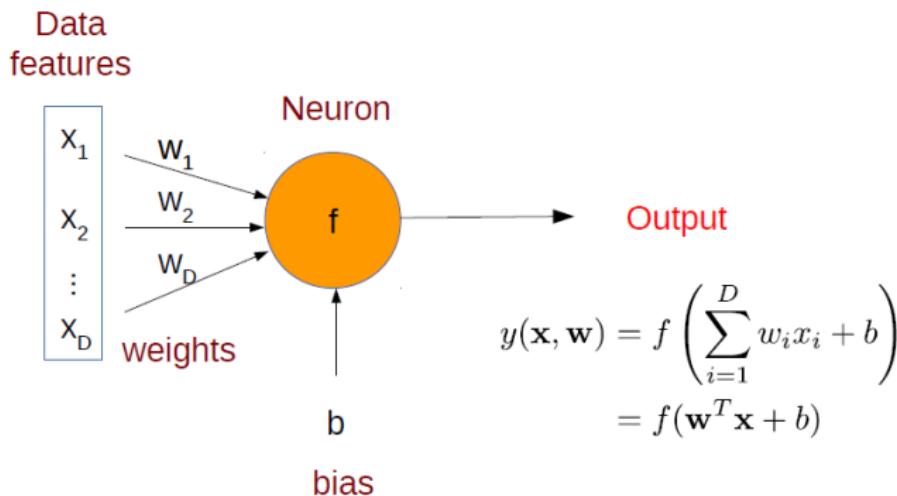


Figure 7: Diagram of a simple perceptron algorithm on a neuron. The activation function can be taken to be $f(x) = \text{ReLU}(x)$ [13].

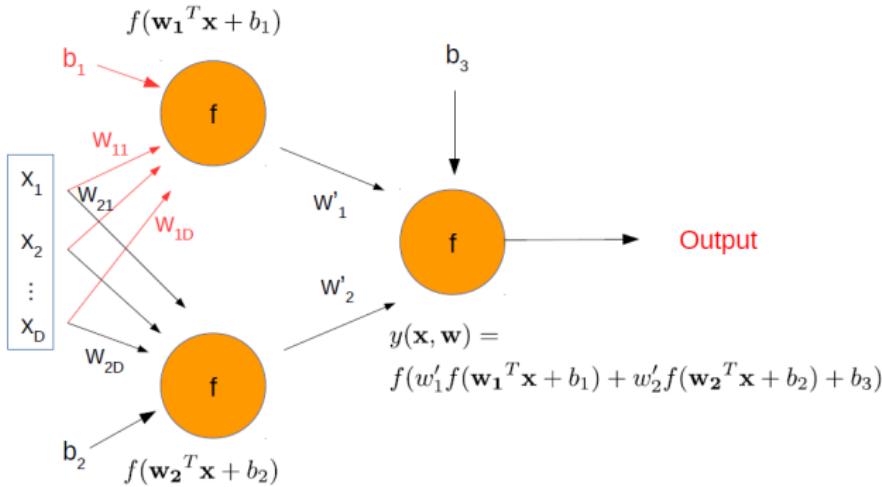


Figure 8: Multi-Layer Perceptron (MLP) with 1 hidden layer.

We must train this MLP in order for our NN to learn. This can be accomplished in three steps: **1)** Complete a forward pass through the network and compute the loss function **2)** Perform a back-propagation and compute the gradient of the loss function w.r.t. the weights and the biases for each layer, and **3)** Update all weights and biases, and repeat the previous steps [14].

Convolutional Neural Networks

A Convolutional Neural Network (CNN), is an algorithm that allows for the processing of images or data through a series of connected layers. Its most relevant features are:

- The hidden layers, composed of multiple convolutional filters that serve to reduce the dimensionality and extract important small-scale features from the image.
- The classification section, which takes as input the output of the hidden layers, pieces them together, and then makes a classification decision (i.e. signal or background jet).

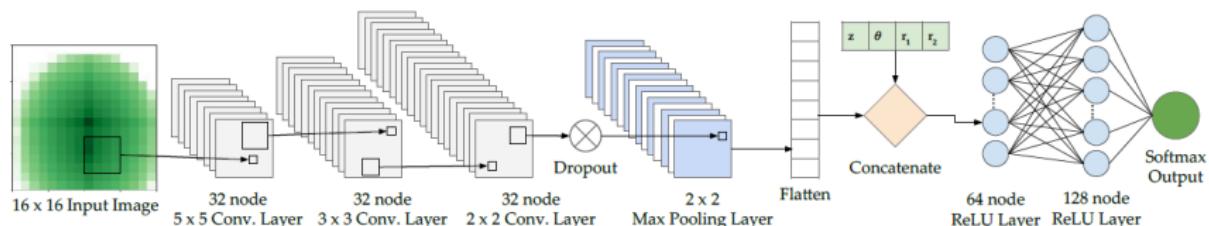


Figure 9: The architecture of a 2D CNN whose input is the image of a signal or a background jet [8].

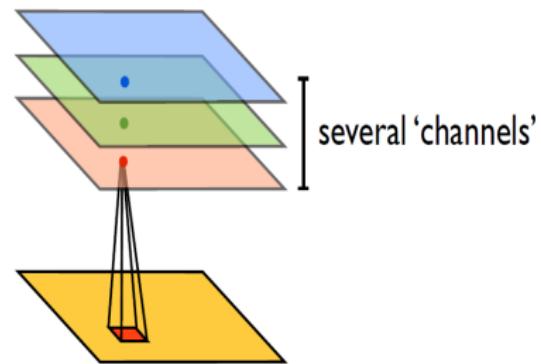
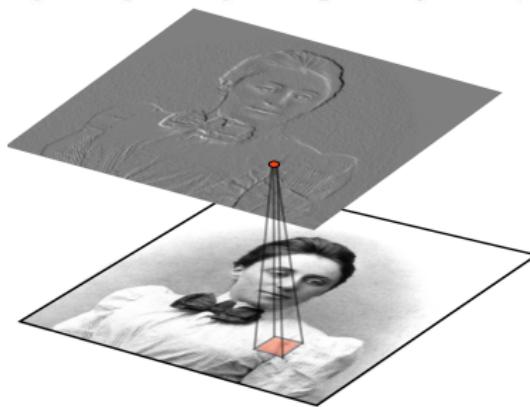
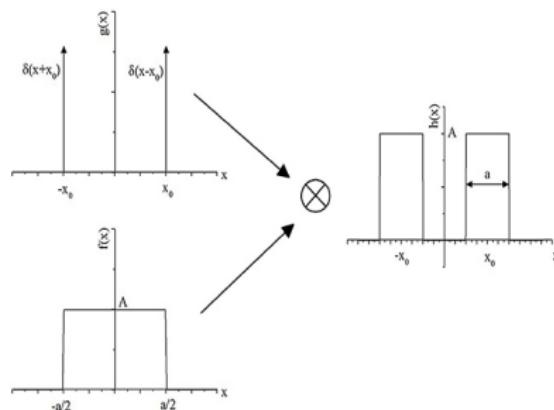
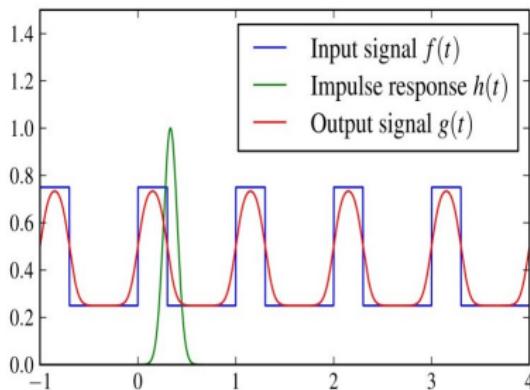


Figure 10: Top: Convolutions between various analytical functions. Bottom: Convolution of an image with different filtering kernels [15,16].

The case of a 1D CNN is identical to that of a 2D CNN, the difference being that now the input is a list containing the kinematic parameters defined in Fig. 2 and Fig. 5. Note that it is also divided into a hidden layer region (the convolutional and pooling layers), and a classification region (the flattening, ReLU, and softmax output layers).

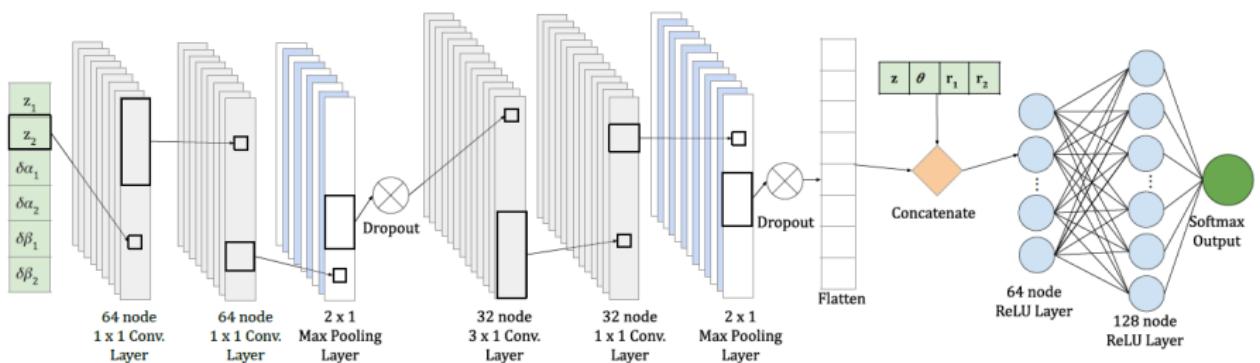
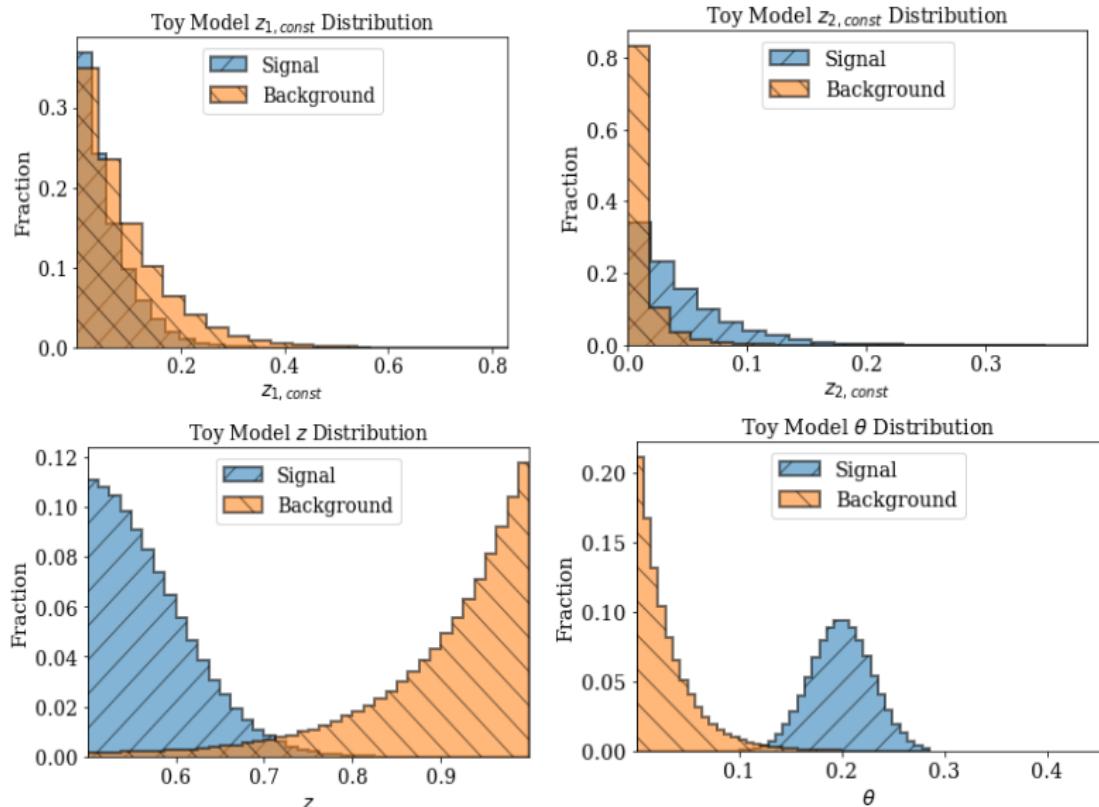


Figure 11: The architecture of a 1D CNN that takes a list as an input.

Results

We start by summarizing some of the outcomes of our 1D CNN algorithm.



The data gathered from the simulation of one million LHC events is fed into our NN, whose task is to differentiate signal from QCD jets. For such a simplistic model, it is not surprising that the CNN is able to perform with nearly 100% efficiency. Please visit the companion repository to examine the code used to generate these plots [17].

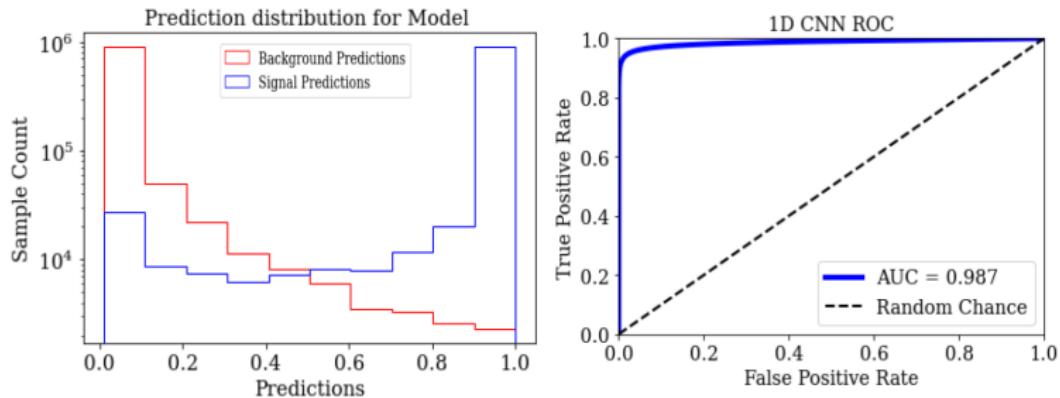
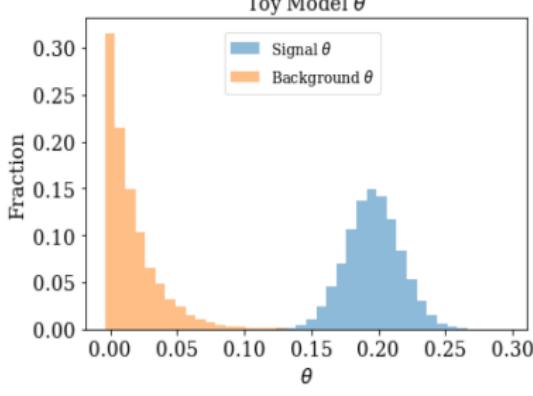
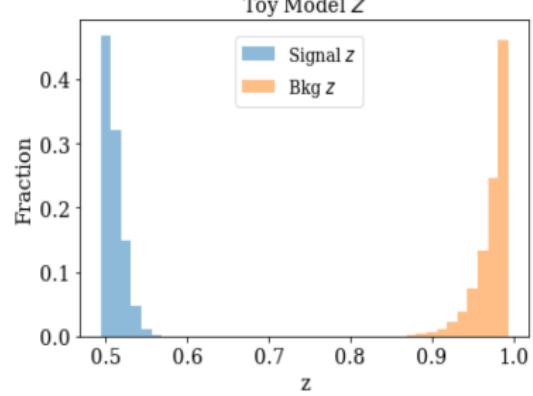
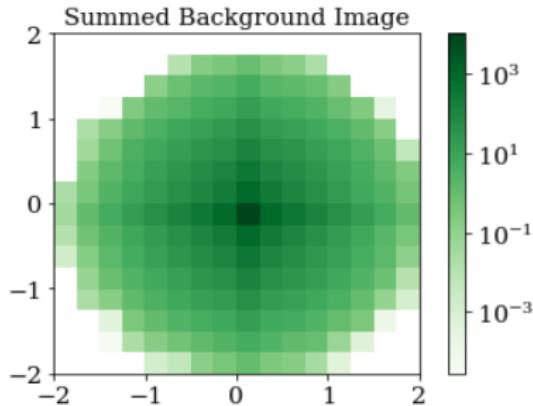
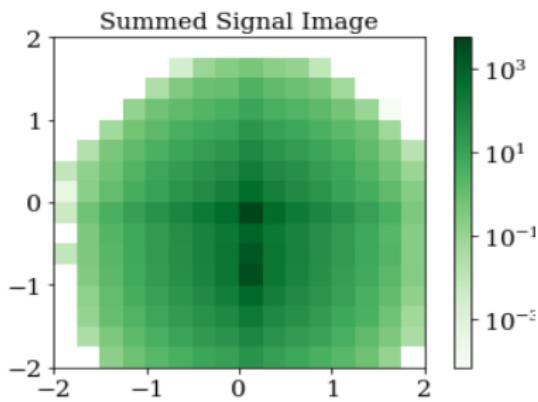


Figure 12: Left: Prediction distribution model showing a clear separation between signal and background. Right: ROC curve with an Area Under Curve (AUC) value nearly equal to unity. This once again shows that our CNN acts as a nearly perfect binary classifier.

Below are the 16×16 images that were used as inputs for our 2D CNN.



Once again, it is immediately apparent that our 2D CNN is acting as a nearly perfect classifier. The wide separation in the prediction distribution model demonstrates that the CNN correctly identifies a signal event and rejects a QCD event nearly every time.

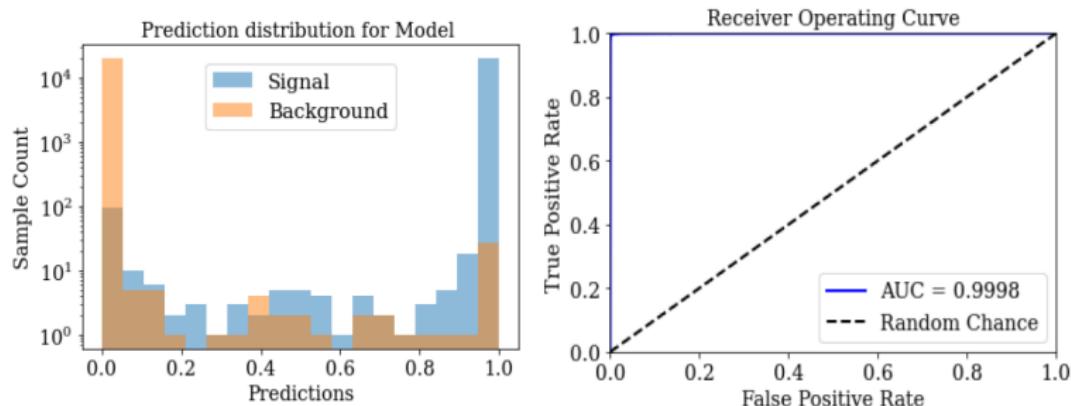


Figure 13: The AUC for a 2D CNN is even closer to unity than that of a 1D CNN.

Conclusion

- Through simple kinematic analysis we were able to demonstrate that particle jets can be fully described by a handful of parameters.
- These parameters, whether in list or image form, were then fed into a CNN to train it to correctly classify the origin of a jet.
- Although our naive toy model fails to capture some of the more complex and exotic aspects of high energy collisions, we were nevertheless able to arrive at positive outcomes. The "black box" nature of NN certainly lends to this.
- The methods discussed here can also be fully adapted to the problem of anomaly detection, where our goal is to extract a signal (particle resonance) from a broad noisy background.
- The next step to this project will be to implement Pythia/Coffea to generate better event simulations.

References

- [1] Z. L. Marshal, *A Measurement of Jet Shapes in Proton-Proton Collisions at 7.0 TeV Center-of-Mass Energy with the ATLAS Detector at the Large Hadron Collider*, CalTech, (PhD Thesis, 2010).
- [2] A. M. Paasch, *Towards an Improved Precision of Jet Mass Measurements in Decays of Boosted Top Quarks at CMS*, (PhD Thesis, 2020).
- [3] M. Peskin, and D. Schroeder. *An Introduction to Quantum Field Theory*, (Westview Press, 1995).
- [4] D. Baden, *Jets, Kinematics, and Other Variables A Tutorial for Physics With p-p (LHC/Cern) and p-p (Tevatron/FNAL) Experiments*, Int.J.Mod.Phys.A13:1817-1845, (1998).
- [5] S. Marzani, G. Soyez, and M. Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, Lecture Notes in Physics, (2020).

- [6] N. M. Sarda, *On anomaly detection in particle accelerators*, MIT, (Master's Thesis, 2020).
- [7] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_t jet clustering algorithm*, JHEP 04, (2008).
- [8] G. Agarwal, L. Hay, I. Iashvili, B. Mannix, C. McLean, M. Morris, S. Rappoccio, U. Schubert, *Explainable AI for ML jet taggers using expert variables and layerwise relevance propagation*, (2021). arXiv:2011.13466.
- [9] M. Dasgupta, A. Fregoso, S. Marzani, Simone, and G. P. Salam, *Towards an understanding of jet substructure*, JHEP09 (2013)029.
- [10] H. Pacey, *QCD and the Search for Exotic Physics*, QCD@LHC, (2019).
- [11] Y. Gao, C. Tian, M. Duan, B. Li, and F. Hu. *Transverse momentum distributions of identified particles produced in pp , $p(d)A$, and AA collisions at high energies*, Pramana – J. Phys., Vol. 79, No. 6, (2012).

- [12] K. Cranmer, *Practical Statistics for Particle Physics*, CERN School HEP, (2011).
- [13] J. Donini, *Introduction to Machine Learning Neural Networks*, Journees Machine Learning et Physique Nucleaire, (2019).
- [14] C. F. Madrazo, I. Cacha, L. L. Iglesias, and J. Lucas, *Applications of Convolutional Neural Networks for Image Classification to the Analysis of Collisions in High Energy Physics*, (2017). arXiv:1708.07034.
- [15] <https://www.ncbi.nlm.nih.gov/books/NBK546153/>
- [16] F. Marquardt, *Machine Learning and Quantum Devices*, arxiv:2101.01759, to appear in “Quantum Information Machines; Lecture Notes of the Les Houches Summer School 2019”, eds. M. Devoret, B. Huard, and I. Pop.
- [17] <https://github.com/ubsuny/AnomalyDetection-NN-raspi4>