
Survey of Machine Learning Techniques for High Energy Electromagnetic Shower Classification

Michela Paganini

Yale University & Lawrence Berkeley National Laboratory
michela.paganini@yale.edu

Luke de Oliveira

Vai Technologies & Lawrence Berkeley National Laboratory
lukedeo@vaitech.io

Benjamin Nachman

Lawrence Berkeley National Laboratory
bnachman@cern.ch

Abstract

Correctly identifying the nature and properties of particles produced at the Large Hadron Collider is a crucial task for fully exploring sub-nuclear length scales in search of new physical phenomena. This work addresses the active research problem of object identification in the context of elementary particles interacting with a multi-layer, heterogeneously-segmented electromagnetic calorimeter. This work is positioned within the physics literature trend of replacing domain-specific feature engineering with the design and augmentation of deep learning approaches from the computer vision literature. We propose a multi-stream DenseNet architecture that takes advantage of lower level detector information. We provide performance and parameter-efficiency comparisons with other computer vision approaches, as well as standard classification techniques based on shower shapes – engineered features that describe geometric properties of the particle cascades. Experiments are conducted on public simulated datasets to provide useful benchmarks for future technical improvements. The DenseNet-style architecture achieves state-of-the-art performance on both $e^+ - \gamma$ and $e^+ - \pi^+$ classification tasks.

1 Introduction

Treating calorimeters as digital cameras has had a long history in high energy particle physics [34, 16]. Calorimeter cells can be treated as pixels in a camera and energy deposited can be interpreted as pixel intensity. Recently, deep neural networks have revolutionized image processing, with significant improvement over traditional techniques on a variety of tasks. Many of these modern techniques have already been applied to high energy physics in the context of *jet images* [18] for classification [19, 14, 15, 10, 23, 26, 1, 11], regression [25], and generation [20] as well as in the context of neutrino identification and classification [31, 12, 9, 32] in liquid argon time projection chambers.

The jet image generation work has recently been extended to particle showers in a longitudinally segmented calorimeter [29] using Generative Adversarial Networks (GANs). The adversary to the generative network is a classifier network that learns to distinguish fake from real images. This classifier takes as input calorimeter images and performs binary classification; it is therefore natural

Table 1: Specifications of the calorimeter layers

Layer Number	Depth in z direction (mm)	$N_{\text{cells},x}$	Width of cell in x direction (mm)	$N_{\text{cells},y}$	Width of cell in y direction (mm)
0	90	3	160	96	5
1	347	12	40	12	40
2	43	12	40	6	80

to ask how well this architecture performs when applied to the task of particle identification. In this paper, we explore the classification of particles in a longitudinally segmented calorimeter using the techniques developed in Ref. [29] as well as other ideas from modern computer vision.

Traditional techniques for identifying particles in a longitudinally segmented calorimeter rely on a small number of *shower shapes*. These engineered features of the three-dimensional shower profile are powerful tools for identifying and calibrating photons and electrons [3, 2, 5, 6] as well as extracting pointing information for photons [4, 7]. Our goal is to show how much one can gain from using modern machine learning techniques, treating calorimeter region around one particle shower as digital image with multiple layers. Unlike a typical RGB image, longitudinally segmented calorimeter images are sparse, without smooth features or sharp edges, and have a causal relationship between layers so it is not sufficient to treat each layer independently. For these reasons, state-of-the-art image processing techniques must be adapted to fit for this application domain.

This paper presents an application of deep neural network techniques for electron, photon, and pion identification and regression in a longitudinally segmented electromagnetic calorimeter. Similar efforts, using similar datasets, exist within the high energy physics community. We provide a set of baselines to help reduce the search space towards optimal solutions.

2 Dataset

We utilize a public dataset [30, 29] composed of 500,000 e^+ , 500,000 π^+ , and 400,000 γ showers induced by the electromagnetic and nuclear interactions that the incident and secondary particles undergo as they propagate through the designed electromagnetic calorimeter.

The geometry of the detector, built from a modified version of the GEANT4 B4 example, consists of a cubic section along the radial (z) direction of $V = 480 \text{ mm}^3$ of an ATLAS-inspired electromagnetic calorimeter, at a distance of $z_0 = 288 \text{ mm}$ from the origin. The volume is segmented along its radial dimension into three layers of depth 90 mm, 347 mm, and 43 mm, each composed by flat alternating layers of lead (absorber) and LAr (active material) of thickness 2 mm and 4 mm, respectively. Each of the three sub-volumes has a different resolution, with voxels of dimensions summarized in Table 1.

3 Method

The proposed DenseNet-style architecture is tested, along with a set of baseline architectures, on two classification tasks (e^+ versus γ , and e^+ versus π^+). The scope of this approach is to document both successful and unsuccessful attempts, and to inform the community on what techniques appear to be more promising and worth pursuing.

All neural networks are build using KERAS v2.0.6 [17] with TENSORFLOW v1.2.1 [8] as a back-end, and trained on an NVIDIA GeForce GTX TITAN X with the Adam [24] optimizer to minimize the cross-entropy between the predicted and target distributions. After a baseline hyperparameter scan, the learning rate is set to 0.001 for all networks on the $e^+ - \gamma$ task, and 0.0001 in the $e^+ - \pi^+$ task.

Six-Layer Fully-Connected Network on Shower Shapes The first baseline is a feed-forward neural network with 20 shower shape input variables. The architecture consists of five fully connected - LeakyReLU [27] - dropout [33] - batch normalization [22] blocks, with hidden representations of size 512, 1024, 2048, 1024, and 128 respectively, and a final one-dimensional output with sigmoid

activation. The network has a total of 4,873,985 trainable parameters, and the batch size is chosen to be 128.

Six-Layer Fully-Connected Network on Individual Pixel Intensities The network structure is identical to the one described above, except for the first layer that now receives as inputs the 504 calorimeter pixel intensities from a shower representation, as opposed to the 20 shower shape variables used in the previous benchmark. The network now has a total of 5,121,793 trainable parameters, and the batch size is chosen to be 128.

Three-Stream Locally-Connected Network Locally-connected layers have shown promising results compared to their convolutional counterpart in both classification and generation tasks [20, 19, 13] on high energy physics datasets, where domain-specific preprocessing techniques allow to rotate, crop, and center images with very high sparsity, dynamic range, and physical meaning associate to pixel intensities [18]. Unlike the case of natural images, this application domain has been shown to benefit from the location specificity of filters learned by locally-connected layers.

These were recently employed in the design of both the generator and the discriminator networks in Location-Aware Generative Adversarial Networks (LAGAN) [20], and their multi-stream evolution (CaloGAN) [29].

We draw inspiration from previous applications in generative modeling to test a similar design for the classification tasks presented in this work.

The network consists of three streams of LAGAN-style blocks, each aimed at processing images from one of the three calorimeter layers, and each containing a convolutional layer and three sets of locally-connected layers, batch normalization, and leaky rectified linear units. The features learned from the three streams provide different representations of individual showers, and are then concatenated and processed through a top fully-connected layer with a sigmoid activation. The network has a total of 17,525,697 trainable parameters, and the batch size is chosen to be 128.

Three-Stream Convolutional Network Although locally-connected layers were empirically found to work well with jet images centered at the origin [28], the advantage of using them over convolutional layers is expected to fade away as showers are produced at different incoming angles and positions. In fact, convolutional layers are designed to exploit feature translation invariance.

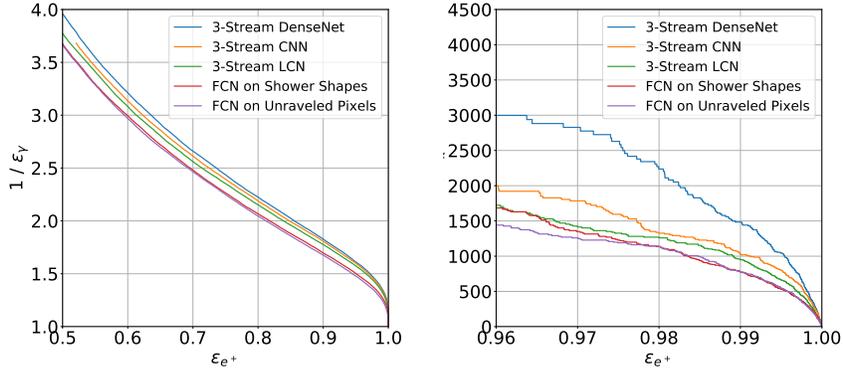
The architecture in the previous paragraph is modified by replacing all locally-connected layers with equivalent convolutional layers. The new network has a total of 7,434,881 trainable parameters, and the batch size is chosen to be 128.

Three-Stream DenseNet Densely Connected Convolutional Networks (DenseNets) [21] were introduced as an elegant solution to maximize information flow by reducing the path from input to output, in order to counter the vanishing gradient problem in very deep convolutional networks. DenseNets devise connections such that every layer receives as inputs the concatenated feature maps from every previous layer, and contributes its feature maps to every subsequent layer. These redundant connections favor feature reuse and persistence, to the point that the last classification layer will have at its disposal all of the features built by all previous layers in the network, therefore gaining access to different levels of feature representation. The network is, by design, very parameter efficient, with only 351,057 trainable parameters. To match the other benchmarks, the batch size is set to 128.

3.1 Experimental Results

We examine the performance of the binary classifiers described in Sec. 3 using receiver operating characteristic (ROC) curves (Fig. 1). The different efficiency ranges depicted on the axes of Figures 1(a) and 1(b) illustrate the difference in complexity between the two tasks: while charged pions are easier to separate from positrons and only the high signal efficiency range is displayed, photons share similar signatures in the electromagnetic calorimeter compared to positrons, yielding worse overall background rejection.

In both classification tasks, the DenseNet outperforms all other architectures and does so with one or two orders of magnitude fewer parameters. In the harder e^+ versus γ scenario, the relative



(a) ROC curves for $e^+ - \gamma$ classification (b) ROC curves for $e^+ - \pi^+$ classification

Figure 1: These performance plots illustrate the trade-off between maximizing the true positive ratio for positron identification (on the x -axis) and maximizing the background rejection, the inverse of the false positive ratio (on the y -axis). The five curves represent the performance of the following classifiers: in blue, the three-stream DenseNet; in orange, the three-stream convolutional network; in green, the three-stream locally-connected network; in red, the fully-connected network on shower shapes; in purple, the fully-connected network on individual pixel intensities.

Table 2: Percentage relative increase or decrease in γ rejection at five different e^+ efficiency working points compared to the baseline fully-connected network trained on shower shape variables

		e^+ efficiency				
		60%	70%	80%	90%	99%
Model	FCN on shower shapes	-	-	-	-	-
	FCN on unraveled pixels	-0.8%	-0.7%	-1.0%	-1.2%	-2.0%
	3-Stream Locally-Connected	+3.0%	+3.2%	+4.2%	+4.7%	+4.1%
	3-Stream Conv Net	+4.7%	+5.4%	+5.9%	+6.5%	+5.5%
	3-Stream DenseNet	+7.3%	+7.3%	+7.7%	+7.7%	+6.4%

Table 3: Percentage relative increase or decrease in π^+ rejection at five different e^+ efficiency working points compared to the baseline fully-connected network trained on shower shape variables

		e^+ efficiency				
		96%	97%	98%	99%	99.99%
Model	FCN on shower shapes	-	-	-	-	-
	FCN on unraveled pixels	-14.4%	-7.6%	+0.76%	+0.0%	-34.6%
	3-Stream Locally-Connected	+2.3%	+4.8%	+11.9%	+22.3%	-43.7%
	3-Stream Conv Net	+20.3%	+31.0%	+17.9%	+32.4%	-6.8%
	3-Stream DenseNet	+81.6%	+107.5%	+100.0%	+90.1%	+34.9%

performance differentials with respect to the shower shapes-based classifier are provided, for five different e^+ efficiency points, in Table 2. Similar results are provided in Table 3 for the e^+ versus π^+ classification task.

4 Conclusion

We benchmarked a range of machine learning methods on a particle identification task using a publicly available dataset. We highlighted unique properties of physical datasets that demand careful architecture design considerations. With domain specific evaluation constraints in mind, emphasis

was put on the computational and parameter efficiency of the various models. DenseNets were identified as highly a performant and efficient first solution.

Ongoing work on model interpretability identified critical regions in image space where the DenseNet and the baseline shower shapes-based tagger are not in agreement on the shower labels. Forthcoming work will provide evidence to conclude that the DenseNet is learning information beyond what is explained by the shower shapes, and it is correctly classifying these subsets of showers despite their apparent similarity to the opposite class. Further studies will be necessary to investigate what extra knowledge the DenseNet is relying on, and how this information can be used to augment the shower shapes.

References

- [1] New Developments for Jet Substructure Reconstruction in CMS. 2017.
- [2] Morad Aaboud et al. Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data. *Eur. Phys. J.*, C76(12):666, 2016.
- [3] Morad Aaboud et al. Electron efficiency measurements with the ATLAS detector using 2012 LHC proton-proton collision data. *Eur. Phys. J.*, C77(3):195, 2017.
- [4] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.
- [5] Georges Aad et al. Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data. *Eur. Phys. J.*, C74(10):3071, 2014.
- [6] Georges Aad et al. Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data. *Eur. Phys. J.*, C74(7):2941, 2014.
- [7] Georges Aad et al. Search for nonpointing and delayed photons in the diphoton and missing transverse momentum final state in 8 TeV pp collisions at the LHC using the ATLAS detector. *Phys. Rev.*, D90(11):112005, 2014.
- [8] Martín Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [9] R. Acciarri et al. Convolutional Neural Networks Applied to Neutrino Events in a Liquid Argon Time Projection Chamber. *JINST*, 12(03):P03011, 2017.
- [10] Leandro G. Almeida, Mihailo Backovi, Mathieu Cliche, Seung J. Lee, and Maxim Perelstein. Playing Tag with ANN: Boosted Top Identification with Pattern Recognition. *JHEP*, 07:086, 2015.
- [11] ATLAS Collaboration. Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector. *ATL-PHYS-PUB-2017-017*, 2017.
- [12] A. Aurisano, A. Radovic, D. Rocco, A. Himmel, M. D. Messier, E. Niner, G. Pawloski, F. Psihas, A. Sousa, and P. Vahle. A Convolutional Neural Network Neutrino Event Classifier. *JINST*, 11(09):P09001, 2016.
- [13] P. Baldi, K. Bauer, C. Eng, P. Sadowski, and D. Whiteson. Jet substructure classification in high-energy physics with deep neural networks. *Phys. Rev.*, 93(9):094034, May 2016.
- [14] Pierre Baldi, Kevin Bauer, Clara Eng, Peter Sadowski, and Daniel Whiteson. Jet Substructure Classification in High-Energy Physics with Deep Neural Networks. *Phys. Rev.*, D93(9):094034, 2016.
- [15] James Barnard, Edmund Noel Dawe, Matthew J. Dolan, and Nina Rajcic. Parton Shower Uncertainties in Jet Substructure Analyses with Deep Neural Networks. *Phys. Rev.*, D95(1):014018, 2017.
- [16] R K Bock. Techniques of image processing in high-energy physics. 1996.

- [17] F. Chollet. Keras, 2015.
- [18] Josh Cogan, Michael Kagan, Emanuel Strauss, and Ariel Schwartzman. Jet-Images: Computer Vision Inspired Techniques for Jet Tagging. *JHEP*, 02:118, 2015.
- [19] Luke de Oliveira, Michael Kagan, Lester Mackey, Benjamin Nachman, and Ariel Schwartzman. Jet-images: deep learning edition. *JHEP*, 07:069, 2016.
- [20] Luke de Oliveira, Michela Paganini, and Benjamin Nachman. Learning particle physics by example: Location-aware generative adversarial networks for physics synthesis. *Computing and Software for Big Science*, 1(1):4, Sep 2017.
- [21] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [22] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15*, pages 448–456. JMLR.org, 2015.
- [23] Gregor Kasieczka, Tilman Plehn, Michael Russell, and Torben Schell. Deep-learning Top Taggers or The End of QCD? *JHEP*, 05:006, 2017.
- [24] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [25] Patrick T. Komiske, Eric M. Metodiev, Benjamin Nachman, and Matthew D. Schwartz. Pileup Mitigation with Machine Learning (PUMML). 2017.
- [26] Patrick T. Komiske, Eric M. Metodiev, and Matthew D. Schwartz. Deep learning in color: towards automated quark/gluon jet discrimination. *JHEP*, 01:110, 2017.
- [27] Andrew L. Maas, Awni Y. Hannun, and Andrew Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013.
- [28] Benjamin Nachman, Luke de Oliveira, and Michela Paganini. Pythia generated jet images for location aware generative adversarial network training, Feb 2017. Data set, DOI: 10.17632/4r4v785rgx.1.
- [29] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks. 2017.
- [30] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. Electromagnetic calorimeter shower images with variable incidence angle and position, Aug 2017. Data set, DOI: 10.17632/5fnxs6b557.2.
- [31] Evan Racah, Seyoon Ko, Peter Sadowski, Wahid Bhimji, Craig Tull, Sang-Yun Oh, Pierre Baldi, and Prabhat. Revealing Fundamental Physics from the Daya Bay Neutrino Experiment using Deep Neural Networks. 2016.
- [32] J. Renner et al. Background rejection in NEXT using deep neural networks. *JINST*, 12(01):T01004, 2017.
- [33] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958, January 2014.
- [34] M A Thomson. The Use of maximum entropy in electromagnetic calorimeter event reconstruction. *Nucl. Instrum. Methods Phys. Res., A*, 382(CERN-PPE-96-055):553–560. 14 p, Apr 1996.