

Event Reconstruction with Deep Learning

Amir Farbin^{*†}

University of Texas Arlington (US)

E-mail: afarbin@uta.edu

The recent Deep Learning (DL) renaissance has yielded impressive feats in industry and science that illustrate the transformative potential of replacing laborious feature engineering with automatic feature learning to simplify, enhance, and accelerate raw data processing. This document overviews current attempts to apply Deep Learning to Event Reconstruction in High Energy Physics experiments.

*38th International Conference on High Energy Physics
3-10 August 2016
Chicago, USA*

^{*}Speaker.

[†]A footnote may follow.

1. Deep Learning

Artificial Neural Networks (NNs) are a biologically inspired means of computation, most commonly used in High Energy Physics for probabilistic inference, e.g. signal versus background classification. A common application is particle identification (PID), where underlying physical properties of particles are exploited to distinguish them from one another. Current NN-based PID techniques use single hidden-layer NNs trained on physics observables, e.g. energy loss (dE/dx), Cherenkov angle, or shower shape, that are painstakingly reconstructed from the raw electronic readout of particle detectors. In addition to classification, the Universal Approximation Theorem, which states that single-hidden layer (or shallow) NNs can approximate continuous functions, indicates that NNs are well suited for regression problems, though there is no guarantee that the parameters of such networks can be learned.

Deep Neural Networks (DNNs), characterized by multiple successive hidden layers, have long been recognized [1] for their potential for modeling high level abstractions and therefore efficiently encapsulating complicated tasks. But training DNNs has been difficult, primarily due to the Vanishing Gradient Problem [2]: the fact that gradient of early layer parameters are generally small, making training these layers through minimization of a cost function prohibitively inefficient. In recent years, this problem has been mitigated by new techniques, the emergence of large datasets that enable training longer, and the evolution of commodity Graphics Processing Units (GPUs) from fix function to full programmability, providing the computing power to train on large samples.

The world is now in the midst of a renaissance in Machine Learning and Artificial Intelligence, known as Deep Learning. Billion-neuron multi-layer artificial neural networks trained on raw and sometimes unlabeled data, can now recognize objects, detect human emotion and intent, play video games, translate between languages, generate mathematical proofs, sometimes better than humans and most importantly with minimal engineering. Developed in University Machine Learning labs, DL has led Google, Facebook, and other industry-leading companies to rethink everything, building Artificial Intelligence teams, software, processors, and cloud services, and demonstrating impressive feats.

Beyond better performance, new capabilities demonstrated by Deep Learning drive the unprecedented excitement across various fields and domains. Deep Neural Networks (DNNs), can learn features from raw data, eliminating need for expensive hand coded feature engineering. They can find these features in unlabeled data opening new tools for analyzing complex and poorly understood data. They can generate complex data starting with only examples, enabling simulation without a model and providing speed when there is a model. They can filter noise and compress data, find anomalies, and solve problems we do not know how to solve algorithmically.

The first application of DL in HEP demonstrated that (fully-connected) DNNs out-perform shallow networks in SUSY and Higgs signal versus background classification, deriving new “features” from 4-vectors that were not encapsulated in traditional observables [3]. While application of DNNs to high level data such as 4-vectors, will likely improve the sensitivity of searches and the precision of measurements, application of DNNs to low level data, perhaps even raw data, will likely have greater impact. Compared to the laboriously engineered reconstruction algorithms, DNNs can not only provide better algorithmic performance, but can also reduce the time and cost of algorithm development while yielding faster algorithms that run efficiently on the newest many-

core processors and high throughput coprocessors such as GPUs. For decades HEP relied on the independence of particle physics events (e.g. collisions) to parallelize their workflows at the highest level by running many independent instances of their software. The transition to in-event parallelization necessary for data parallel workflows, which better suite new many-core processors and co-processors, has proven to be difficult. In contrast, the “code” of DNN-based algorithms are encapsulated in weights that turn input data to results via linear algebra and other highly parallelized operations that are already optimally implemented in Deep Learning frameworks and libraries from the co-processor manufacturers.

2. Feature Learning

Particle physics detectors record the energy deposits left by the relatively long-lived particles emerging from high-energy collisions in hundreds of millions of instrumented detector elements, resulting in data that can be analogous to a highly granular image of the particle interactions. Reconstruction algorithms interpret the raw data, for example voltages from every channel of a detector component (or sub-detector), under different particle hypothesis, for example electron or muon, and produce particle candidates with a measurements of the 4-vector and other particle properties.

Event reconstruction typically consists of a series of algorithms that identify energy deposits from noise (known as feature extraction), recognize patterns of deposits consistent with a particle trajectory or interaction (commonly referred to as pattern recognition), combine deposits and measure properties within each sub-detector (e.g. clustering in calorimeters or track fitting in tracking detectors), and finally combine information from each sub-detector to form particle candidates. Developing reconstruction is an time consuming and expensive feature engineering task. For example, ATLAS’s reconstruction was developed over a course of two decades at an estimated cost of \$250 million.

Convolutional Neural Networks (CNNs) are a class of DNNs that are inspired by the visual cortex and are capable of learning features in audio (1D), images (2D), or video (3D). Each convolutional layer scans a large number of small tensors across the data, using a convolution operation to detect and note the presence of features in feature maps that are often subsequently reduced in resolution by pooling to promote shift invariance. Starting in 2014, Deep Convolution Neural Networks (CNNs) have exponentially improved performance on ImageNet, a one-million image classification challenge[4], to now super-human performance[5].

The parameters of the CNN’s convolution kernels, which represent feature detectors, are learned during training. In images, the kernels in the earlier layers respond to changes in color, while in later layers they are triggered by higher level features such as eyes in face images or wheels in car images. In the final layers kernels identify full objects such as faces or cars. In many ways, this hierarchy parallels that of HEP event reconstruction algorithms. And since raw data from many types of detectors can be naturally represented as images, CNNs are the best candidates for DNN-based reconstruction algorithms. The first application of CNNs has been in classification tasks, such as particle identification, in “imaging” detectors such as Time-Projection Chambers (TPCs), Cherenkov Imaging detectors, and high granularity Calorimeters.

3. Neutrino Detectors

Since neutrinos only interact weakly, experiments detect them by instrumenting large volumes of target material such as water, oil, or Liquid Argon and looking Cherenkov light, ionization, or particle showers that emerge from the neutrino interaction. A large class of neutrino measurement only require identification of the neutrino flavor (typically electron or muon neutrino, ν_e or ν_μ) and estimation of its energy. When the neutrino interacts via a charge current (CC), the emerging lepton indicates the flavor. But when the interaction is mediated by neutral current (NC), the neutrino flavor is not determinable, but photons from neutral pion (π^0) decays can be mistaken for electrons (e) from and charged pions (π^\pm) can be mistaken for muons (μ). Therefore excellent e/π^0 and μ/π^\pm is a critical requirement in neutrino experiments.

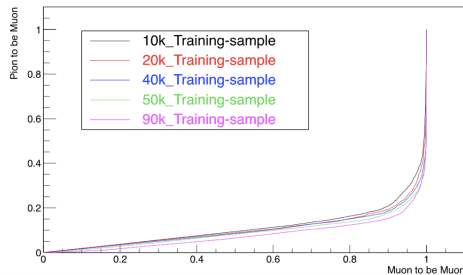
3.1 Nova Experiment

The first application of CNN in HEP was in the Nova experiment [6], whose far detector is composed of filled liquid scintillator bars arranged in layers that alternate between horizontal (X) and vertical (Y) orientations. They built a Siamese network with two parallel branches each passing one orientation through 3 inception modules (from GoogleNet[7]) before merging into a final inception layer. While the neutrino classification performance of their CNN was comparable to their traditional technique in most cases, they were able to achieve 40% better electron efficiency for the same background rate, an improvement that can significantly enhance their neutrino oscillation analyses. Inspecting their CNN, they found features that clearly detected hadronic or muonic activity.

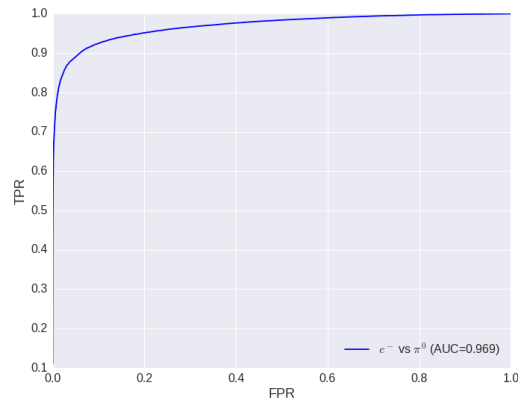
3.2 LArTPC Experiments

Liquid Argon Time Project Chamber (LArTPC) is the detector technology employed in the currently running MicroBooNE experiment and the upcoming short and long baseline neutrino programs. LArTPC allows tracking, calorimetry, and particle identification in one detector, and is expected to yield 80% neutrino energy with 1% fake rate, roughly twice the efficiency of previous neutrino experiment. In LArTPC detectors, charged particles from the neutrino interaction ionize the liquid argon leaving tracks, while electromagnetic and hadronic showers produce cones of ionizing secondary particles. An electric field drifts the resulting ionization electrons to planes of parallel wires at one end of the detector. Since the velocity of the drifting electrons are nearly constant, their arrival time on wires is proportional to the distance to initial ionization location. Therefore a simple plot of the digitized voltage on every wire versus time produces a detailed image of tracks and showers that are often easily identifiable by the human eye. Unfortunately, automatic reconstruction of LArTPC events has proven to be difficult and has yet to be demonstrated. The first generation of LArTPC experiments, Argoneut and ICARUS, had small data samples and therefore relied on humans to perform the pattern recognition.

The first application of CNN to LArTPC detectors was performed by Amir Farbin (UTA) using simulated data from LArIAT testbeam experiment. LArIAT has two planes each consisting of 240 wires 4 mm apart and oriented at ± 60 deg from the horizontal. The ADCs sample the voltage of 4096 times in an event, yielding 240 by 4096 measurements for each of the two planes. As an early demonstration, the GoogleNet image classification network was fed with images that stored



(a) Efficiency for selecting real muons as muons versus selecting real charged pions as muons. These plot demonstrate that for muons that larger training samples yield better performance.



(b) Efficiency for selecting real electrons as electrons versus selecting real neutral pions as electrons.

the two wire planes of the detector as different color intensities. The time dimension was reduced to 256 by summing N contiguous samples and selecting the 256 wide region with largest charge deposit. Comparing the CNN performance for different values of N indicated $N = 6$ was a good compromise between loss of resolution and ability to fully capture particles in the 256 sample window. Performance was observed to improve with larger training samples (see figure a), which were limited to 100,000 events (10% reserved for evaluation) for the early studies. These studies yielded 6% fake rate for 90% efficiency for electron versus photon classification, 15% fake rate for 90% efficiency for muon versus charge pion classification, 5% (12%) fake rate for 80% efficiency for charge current versus neutral current classification for electron (muon) neutrinos. Farbin's effort has now evolved to a collaboration with computer scientists Pierre Baldi and Peter Sadowski at UCI to perform particle identification and energy reconstruction in LArIAT experiment. This work is currently in progress. As an example, fFigure b shows performance of electron versus neutral pion separation.

The MicroBooNE experiments has also investigated the application of CNNs to LArTPC [9]. They compared the performance of out-of-the-box networks such as AlexNet [10], GoogLeNet [7], Faster-RCNN [11], Inception-ResNet-v2 [12], and ResNet [13]. Beyond particle and neutrino identification, they explored using semantic segmentation to identify and separate detector regions with single particles, neutrinos, and cosmuics.

4. Neutrinoless Double Beta Decay

The NEXT experiment aims to search for neutrinoless double beta decay ($0\nu\beta\beta$) using high pressure xenon (HPXe) Gas TPC, read out by SiPMs to produce 3-D images. NEXT relies on topological signatures to separate signal two electron events from single electron background events with kinetic energy comparable to the end-point of the $0\nu\beta\beta$ decay. For their CNN studies [8], NEXT projected their 3D data into three 2D images which were fed to GoogleNet as a combined image with three color intensity. They found that CNNs out perform the traditional technique by factor of 1.2 to 1.6, with potential for further improvement. They also used CNNs to help optimize

detector design by comparing different detector granularities and relative contribution of physics processes to degrading performance.

5. Large Hadron Collider

The LHC presents a much more challenge environment than neutrinos experiments for application DNN to raw data. Nonetheless, efforts are now underway to apply DNNs to Calorimetry and Tracking. Beyond improving performance achieving better performance, the hope is that DNNs will help alleviate the computing challenges of the High Luminosity LHC (HL-LHC) which will come online in the next decade. Computing costs for HL-LHC are estimated to be at least a factor of 5 more than current LHC, and likely much more due to the stalling of Moore's Law.

5.1 Calorimetry

The ATLAS Electromagnetic and Hadronic calorimeters produce 3D images of variable granularity in η , ϕ versus depth of energy deposits. Particles, such as photons, are identified by their characteristic shower profiles, with their energies determined via a weighted fit of layer-wise deposits calibrated to test beam and Z decays. Any improvement, for example in photon identification or energy resolution, can dramatically effect searches and measurements, for example producing narrower Higgs peaks with less background. Several factors give hope that significant improvements can be achieved with more sophisticated techniques. For example, energy reconstruction in the LAr Electromagnetic calorimeter currently does not use shower shape information and is not correcting for variations in the LAr calorimeter's characteristic accordion structure. Similarly the hadronic calibration does not use sampling information. Such effects would be naturally exploited by a CNN.

Another potentially high impact of DNNs to calorimetry is fast showering. Full Geant4 shower simulation in the ATLAS calorimeter takes of order of an hour. Fast shower techniques such as shower libraries or high dimensional binning of shower observables generally suffer from intractable memory requirements. DNNs may provide a much more powerful technique. Starting with examples only, Generative DNNs have been demonstrated to generate new images of faces, furnished rooms, or text in style of a specific author (e.g. Shakespeare).

5.2 Tracking

Pattern recognition rate in particle tracking scales quadratically with hits in the tracking detector. As a result, tracking in HL-LHC events, where every event will consist of 200 proton collisions, is one of biggest challenges for the HL-LHC. In addition, some tracking and vertex finding at 40 MHz bunch-crossing might be required for the trigger. While some are investigating dedicated hardware, such as GPUs, FPGAs, or associated memory, a group of ATLAS and CMS physicists are hoping that by presenting the HL-LHC tracking problem as a Machine Learning challenge (TrackingML) with a prize, solutions arise that scales better with number of hits. One source of inspiration is DeepMind's AlphaGo [14] artificial intelligence agent, which was able to assess positions in the game of Go by looking at the whole board with a DNN instead of performing a look-ahead tree search.

References

- [1] Ivakhnenko, Alexey (1965). *Cybernetic Predicting Devices*. Kiev: Naukova Dumka.
- [2] Hochreiter et al., "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," In S. C. Kremer and J. F. Kolen, editors, *A Field Guide to Dynamical Recurrent Neural Networks*. IEEE Press, 2001.
- [3] P. Baldi, P. Sadowski and D. Whiteson, *Nature Commun.* **5**, 4308 (2014) doi:10.1038/ncomms5308 [arXiv:1402.4735 [hep-ph]].
- [4] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei.
- [5] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. *ArXiv e-prints*, December 2015.
ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.
- [6] 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.
- [7] Christian Szegedy et al. Going Deeper with Convolutions. arXiv 1409.4842, 2014.
- [8] J. Renner *et al.* [NEXT Collaboration], arXiv:1609.06202 [physics.ins-det].
- [9] MicroBooNE Collaboration, "Convolutional Neural Networks Applied to Neutrino Events in a Liquid Argon Time Projection Chamber", MicroBooNE-NOTE-1019-PUB.
- [10] Alex Krizhevsky et al. Imagenet Classification with Deep Convolutional Neural Networks. NIPS 25, pages 1106–1114, 2012.
- [11] Ren, Shaoqing and He, Kaiming and Girshick, Ross and Sun, Jian. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks NIPS 28 pages 91-99, 2015.
- [12] Szegedy et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. arXiv:1602.07261v1, 2016.
- [13] Kaiming He et al. Deep Residual Learning for Image Recognition. arXiv:1512.03385, 2015.
- [14] *Nature* 529, 484–489 (28 January 2016) doi:10.1038/nature16961.