Exploring Raw HEP Data using Deep Neural Networks

Evan Racah[1], Wahid Bhimji[1], Seyoon Ko[2], Peter Sadowski[3], Craig Tull[1], Sang-yun Oh[4], Lisa Gerhardt[1], Prabhat[1]

[1] Lawrence Berkeley National Laboratory, [2] Seoul National University, [3] University of California, Irvine, [4] University of California, Santa Barbara,
http://arxiv.org/abs/1601.07621 Contact: wbhimji@lbl.gov

Introduction

High Energy Physics has made use of artificial neural networks for some time. Recently, however, there has been considerable development outside the HEP community, particularly in deep neural networks for the purposes of image recognition. We describe the deep-learning infrastructure at NERSC, and analyses built on top of this. These are capable of revealing meaningful physical content by transforming the raw data from particle physics experiments into learned high-level representations using deep convolutional neural networks (CNNs), including in unsupervised modes where no input physics knowledge or training data is used.

Here we describe in detail a project for the Daya Bay Neutrino Experiment showing both unsupervised learning and how supervised convolutional deep neural networks can provide an effective classification filter with significantly better accuracy than other machine learning methods. These approaches have significant applications for use in other experiments triggers, data quality monitoring or physics analyses.

Deep learning at NERSC

'Cori' supercomputer:

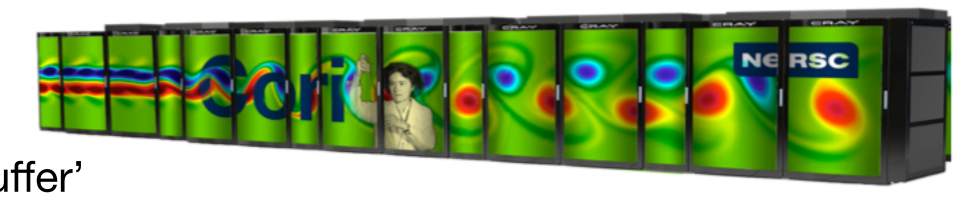
https://www.nersc.gov/users/data-analytics/data-analytics/deep-learning/

Phase 1: 1630 Haswell Nodes (available now)

Phase 2: 9300 Knights-Landing (KNL) Nodes (coming later this year)

28 PB Lustre Filesystem

1.5 PB NVRAM-based 'Burst Buffer'



Deep Learning Frameworks:

Theano - for flexibility in method development

Keras / Lasagne - Theano based but higher-level for ease of use

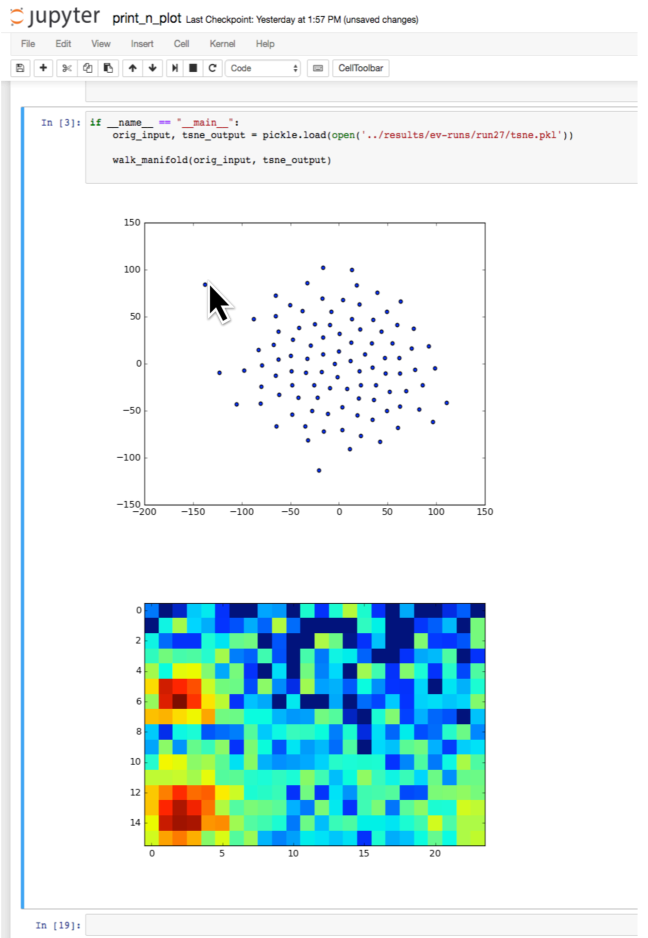
Caffe - including IntelCaffe with performance highly optimised for KNL

Neon - optimised for high performance and parallel implementations

TensorFlow - ease of use and flexibility in addition to a large, growing community

Interactive computing:

Frameworks available within a deeplearning kernel on the NERSC Jupyter service (ipython.nersc.gov): allows for interactivity and scaling up to runs on Cori



Other HEP NN projects at NERSC

Ice Cube: Applying machine learning techniques to improve astrophysical neutrino detection to probe the origins of these particles

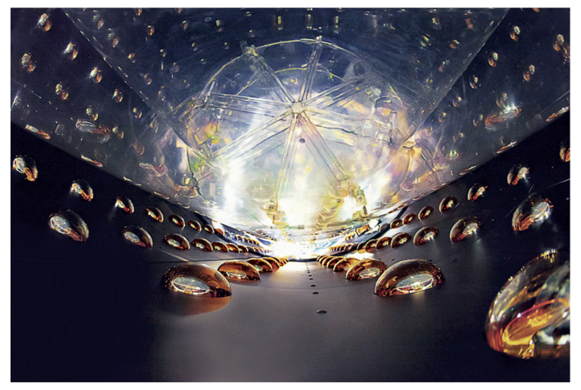
ATLAS: (HL-LHC) Tracking using recurrent NNs (such as LSTMs)

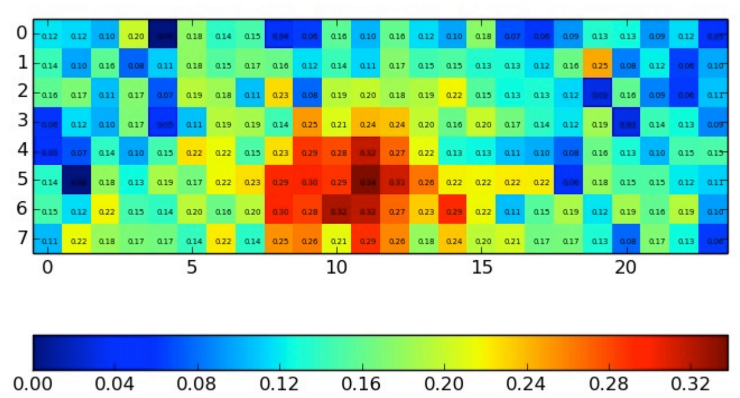
Calorimeter: using semi-supervised CNNs on calorimeter clusters and cells for multi-jet new physics analyses

Probabilistic Programming: coupling inference to detector simulations for new physics detection

Cosmology image analysis: using CNNs to find clusters of galaxies and filaments in large cosmology simulations

Use Case: Daya Bay





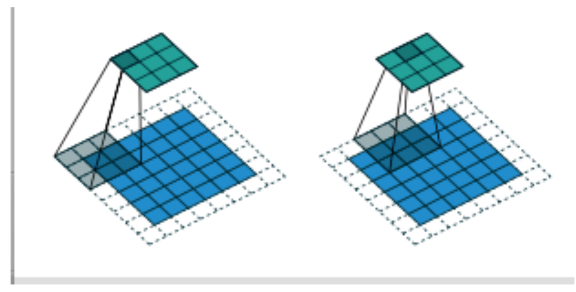
The Daya Bay experiment conducts precision measurements of reactor neutrinos. The experiment is comprised of several antineutrino detectors (shown above left) comprised of 192 PMTs in a cylindrical arrangement. Backgrounds include Muons, instrumental effects, such as so-called 'Flashers', caused by misfiring photomultiplier tubes; and rarer backgrounds, such as the radioactive decay of 'Lithium-9', that can look very like signal events.

Separating out signal from background currently relies on simple cuts on physics-motivated features. These may miss new sources of background (such as 'Flashers' that weren't known before construction) and introduce large-systematic uncertainties on determination of signal-like backgrounds such as that caused by Lithium decay. Deep learning could help improve sensitivity, identify new unexpected sources of background; and determine structure in the signal as well as in the different types of background.

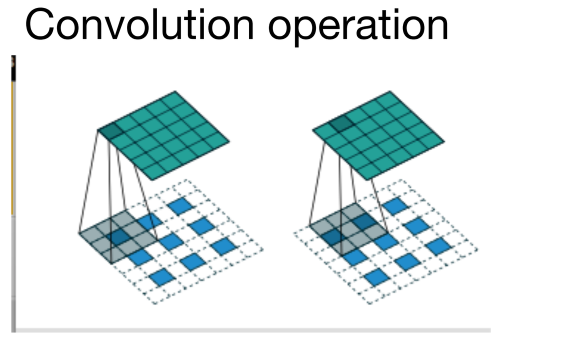
Methods

Convolutional Neural Network (CNN)

- Several convolutional and pooling layers followed by fully connected layers
- Use features learned from those layers to perform classification or regression tasks



Convolution operation



Deconvolution operation

Dumoulin, Vincent, and Francesco Visin. "A guide to convolution arithmetic for deep learning" *arXiv:1603.07285* (2016)

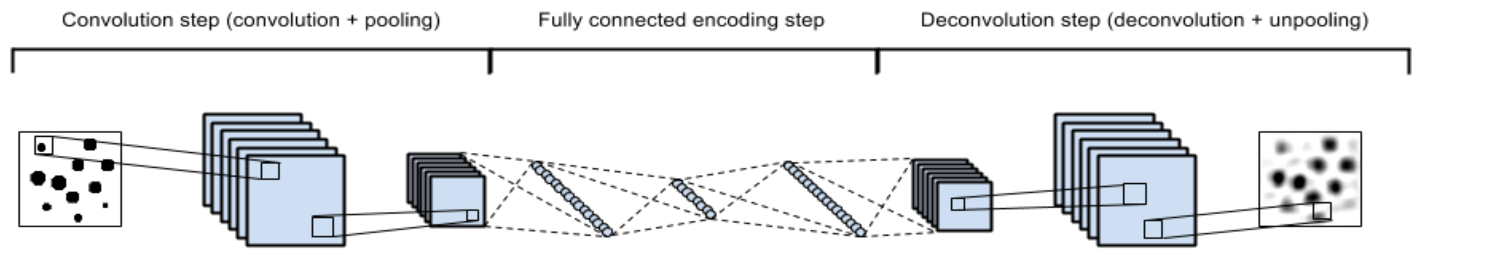
.

Why a convolutional neural network?

- Data captured by detectors can be modeled as 2-D image (top right)
- CNN captures our intuition about local structure and translational invariance

Unsupervised Learning:
What is convolutional autoencoder?

- A neural network where the target output is exactly the input.
- Consists of an encoder, layers that transform the input into a feature vector at the output of the middle layer (often called bottleneck layer or hidden layer), and a decoder, which usually contains several layers that attempt to reconstruct the hidden layer output back to the input.



https://swarbrickjones.files.wordpress.com/2015/04/conv_autoencoder.png

Why a convolutional autoencoder?

- Assume that most of the data clusters around a low dimensional manifold
- A regularized autoencoder attempts to discover the structure of this manifold by having the network to reconstruct the input despite some constraint, like smaller hidden layer dimensionality than that of input
- Forces the network to learn nonlinear, translation invariant and equivariant factors of variation between images - should differentiate types of events

Implementation

Architectures used are shown in the tables (right). Networks were implemented in *Neon*

Supervised: We use real data events labelled by the official collaboration analyses to train a CNN. Hyperparameter optimisation (for the number of filters and the size of the fully connected layer) are performed with *Spearmint*

layer	type	filter size	filters	stride	activation
1	conv	3×3	71	1	tanh
2	pool	2×2	1	2	max
3	conv	2×2	88	1	tanh
4	pool	2×2	1	2	max
5	fc	1×5	26	1	tanh
6	fc	1×1	5	1	softmax

Unsupervised: We learn a feature vector for the data using a convolutional autoencoder with no input physics labels.

We visualize these vectors using *t-SNE* to see which events cluster closer together.

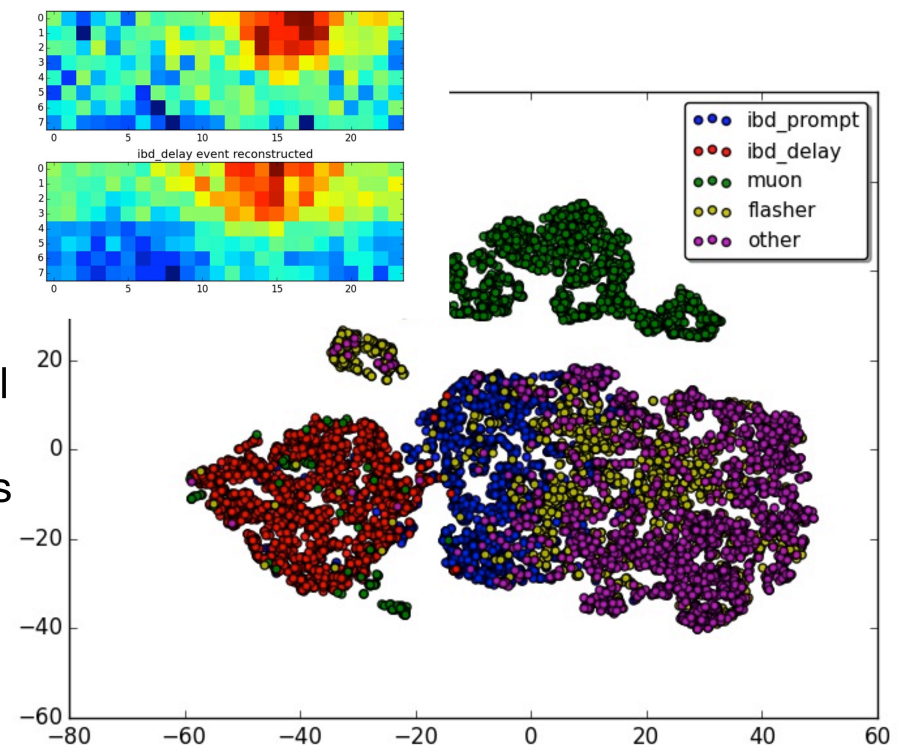
layer	type	filter size	filters	stride	pad	activation
1	conv	5×5	16	1	2x2	RELU [8]
2	pool	2×2	1	2	0	max
3	conv	3×3	16	1	1×0	RELU
4	pool	2×2	1	2	0	max
5	fc	2×5	10	1	0	RELU
6	deconv	2×4	16	2	0	None
7	deconv	2×5	16	2	0	None
8	deconv	2×4	1	2	0	None

Results

Supervised: Our deep CNN achieves > 97% classification accuracy across different classes of physics events (table right). As shown this is also significantly better than other machine learning approaches (k-NN and SVM) for this data.

Measure and Method	IBD prompt	IBD delay	Muon	Flasher	Other
F_1-score					
k-NN	0.775	0.954	0.996	0.784	0.806
SVM	0.846	0.962	0.996	0.895	0.885
CNN	**0.891**	**0.974**	**0.997**	**0.951**	**0.928**
Accuracy					
k-NN	0.950	0.990	0.998	0.891	0.896
SVM	0.966	0.992	0.998	0.947	0.938
CNN	**0.977**	**0.995**	**0.999**	**0.974**	**0.962**

Unsupervised: The deep autoencoder successfully identifies patterns of physics interest. The plot on the right shows t-SNE representation of learnt convolutional autoencoder (arbitrary axes) which clearly shows clustering of different signal and background events. Classes are labelled in plot but those labels are not used for training. Insert shows original and reconstructed image of one IBD delay event



Conclusions

We apply *unsupervised* convolutional neural nets to *raw* data from the Daya Bay experiment and have shown that the network can successfully learn patterns of physics relevance. Such unsupervised techniques could be used for a wide variety of particle physics experiments to aid in trigger decisions, in evaluating data quality, or to discover new instrument anomalies without having to engineer features.
We have also demonstrated the superiority of convolutional neural networks compared to other supervised machine learning approaches for running directly on raw particle physics instrument data. This offers the potential for use as fast triggers or in final analyses.
We are now focussing on more challenging backgrounds such as 'Lithium-9' that dominate the current systematic uncertainties from the experiment as well as developing new state-of-the-art methods incorporating denoising or variational autoencoders and semi-supervised approaches.