Convolutional Neural Networks for Electron Neutrino and Electron Shower Energy Reconstruction in the NOvA Detectors

Lars Hertel
Department of Statistics
University of California, Irvine
Irvine, CA 92617
lhertel@uci.edu

Lingge Li
Department of Statistics
University of California, Irvine
Irvine, CA 92617
linggel@uci.edu

Pierre Baldi
School of Information and Computer Sciences
University of California, Irvine
Irvine, CA 92617
pfbaldi@uci.edu

Jianming Bian
Department of Physics and Astronomy
University of California, Irvine
Irvine, CA 92617
bianjm@uci.edu

Abstract

Neutrinos can help us further our current understanding of the fundamental laws of the universe. The NOvA experiment aims to study the oscillation of neutrinos. We developed a convolutional neural network based energy estimator for electron neutrinos and electron showers in the NOvA detectors to improve experiment analysis. Our method achieves state-of-the-art performance on electron neutrino and electron shower energy regression compared to traditional methods.

1 Introduction

Neutrinos are nearly massless fundamental particles. They rarely interact with matter and are difficult to detect. Their strange properties make them useful for understanding the fundamental laws of physics. The NOvA experiment at Fermilab focuses on measuring neutrino oscillation, mass hierarchy and CP violation. In this work, we aim to improve energy reconstruction for neutrino oscillation experiments in NOvA[4]. This is also important for the experiment T2K[2] and the next-generation neutrino oscillation experiments DUNE[3], Hyper-K[1], JUNO[12], and PINGU[9] which rely on similar analysis methods.

Neutrino oscillation is a function of the neutrino energy. It is therefore important to reconstruct the energy of the neutrino in an observed interaction. Most of the energy deposited in the detector during an electron neutrino interaction comes from an electron shower. For this reason, it is also important to reconstruct the electron shower energy. In this article, we describe a convolutional neural network (CNN) based energy estimator to reconstruct electron neutrino energy recorded in NOvA. We use the same model architecture to train another estimator for electron shower energy in those interactions.

2 The NOvA Experiment

NOvA is a neutrino experiment optimized to observe the oscillation of muon neutrinos to electron-neutrinos. NOvA uses a 14-kt liquid scintillator detector in Ash River, Minnesota. This is used to detect the oscillated muon neutrino beam. The beam is produced 810 km away at Fermilab. The

Workshop on Deep Learning for Physical Sciences (DLPS 2017), NIPS 2017, Long Beach, CA, USA.

NOvA detectors are fine grained, highly active tracking calorimeters. They consist of plastic (PVC) extrusions filled with liquid-scintillator. A wave shifting fiber is used to detect excitation from particles in the liquid-scintillator. Each cell extends the full width or height of the detector. Extrusions are assembled in alternating layers of vertical and horizontal extrusions. Measurements from a cell do not give information about the location in the cell. For this reason, hits caused by neutrino events are recorded in an x-z view and a y-z view hit map instead of a 3D image. The far detector has a total of 344,064 cells.

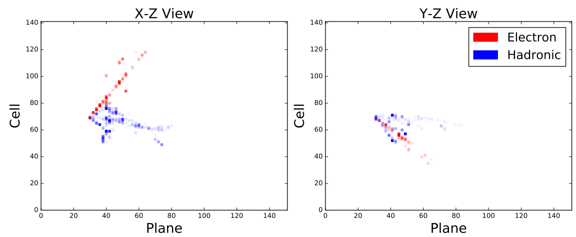

Figure 1: Shown are $x - z$ and $y - z$ views of a typical electron neutrino interaction example. Red shows the hits caused by the electron, blue shows hits from the hadronic component. The intensity of each pixel reflects the magnitude of the deposited energy.

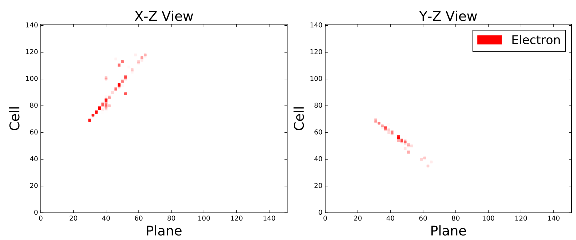

Figure 2: Diagram of the $x - z$ and $y - z$ views of electron shower input examples. The intensity of each pixel reflects the amount of energy deposited in the cell.

2.1 Sample

The standard NOvA simulation is used for the CNN training, validation and testing. The NOvA detector simulation chain is described in [5]. The ν_e CC event reconstruction begins with clustering cell hits by space-time coincidence. The procedure collects hits from a single neutrino interaction (slice). The slices then serve as the foundation for all later reconstruction stages [7]. For each slice, a modified Hough transform is used to identify prominent straight-line features. Then the lines are tuned in an iterative procedure until they converge to the slice's interaction vertex. Prong clusters are then reconstructed using a Fuzzy K-means algorithm where the interaction vertex serves as seed [13].

The ν_e CC interactions are tagged by true neutrino interaction information in the simulation. Each input to the regression CNN for ν_e CC energy includes two separate pixel maps in x-z and y-z detector views. The cell closest to the reconstructed interaction vertex in each view is chosen as the reference cell. Cells are contained within -30 to 120 planes in the z-direction. In the x(y)-direction cells are included within -70 to +70 cells. Therefore, the size of the input image to the CNN is 151×141 in each view. The image size is thus large enough to contain the entire ν_e CC interaction while avoiding hits caused by noise and cosmic rays in the same slice. Keeping the image size small also reduces computational complexity during training. Calibrated cell energy is used as the intensity of each pixel in the input pixel maps. Finally, energy is position dependent in the detector. This feature is accommodated by using x-, y- and z-coordinates of the reconstructed vertex as additional inputs of the to the fully connected layers of the CNN. For the training of the ν_e CC energy regression CNN we use the complete interaction as input. An example of such an input is shown in figure 1. For training we use 612868 interaction samples of this type.

For the electron shower energy CNN, cell hits from the reconstructed shower are used as the input. Figure 2 shows one example of this. The image size for the shower CNN is the same as the ν_e CC

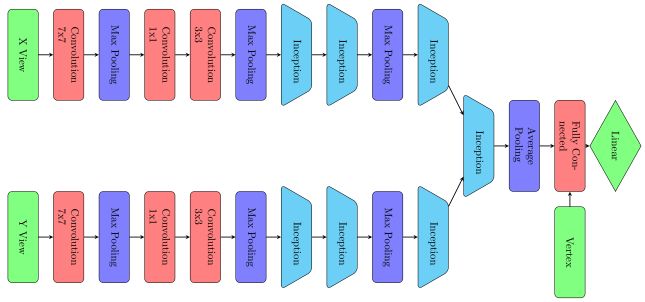

Figure 4: The CNN architecture used for energy regression.

CNN. We use 667000 shower examples including their reconstructed vertices for training of the shower CNN.

3 Methods

3.1 Data Processing

The pixel values of images are typically normalized before CNN training. The purpose of this is to increase numerical stability and gradient quality. Here most image pixels are zero and the non-zero pixel values tend to be small. We apply three normalization methods: mean zero unit variance standardization, log transformation and constant scaling. The three methods produce similar results. Therefore, a constant scaling factor of 100 is chosen after visual inspection of the input spectrum for ν_e and electron. During training, data are processed in batches.

3.2 CNN Details

The model inputs consist of $x - z$ view and $y - z$ view images as well as the reconstructed vertex position. The ν_e energy model takes in event images while the shower energy model takes in shower images. For the output, the ν_e energy model returns the predicted neutrino energy while the shower energy model returns the predicted shower energy. Our CNN utilizes a siamese network structure where each sub-network processes an image from one view. Features are not shared between sub-networks to encourage learning from both views. Each sub-network has three Inception modules [14] to simultaneously extract features of different dimensions. A CNN with another Inception module on top of the sub-networks produces the final output after the vertex position is added. Each convolutional layer including those in the Inception modules have 32 filters; the fully connected layer has 200 units. The network architecture is shown in diagram 4. The architecture builds on the work done in [6]. In our experiments, using additional Inception modules does not improve model performance. This is expected because our images are sparse relative to natural images.

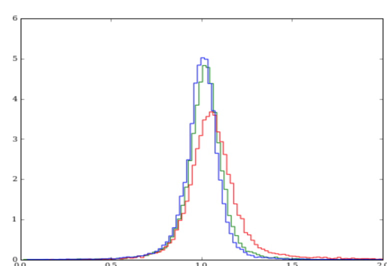

Figure 3: The plot compares resolutions using different loss functions. The blue curve is mean absolute scaled error, the green curve is mean absolute error, and the red curve is mean squared error.

Since the energy distributions are right skewed, we are interested in the energy resolution $\frac{E_{reco}}{E_{true}}$. To evaluate model performance, we analyze histograms of the resolution on the test set and their corresponding root mean square (RMS). As training loss we investigate the mean squared error and mean absolute error. Given the same resolution, we can expect larger errors for larger energies; the mean squared error would give more weight to high energy examples. This effect is reduced yet still existent in the mean absolute error. We consider scaling the error by the true value. Using a

	Baseline	L2			FC L2	Dropout	
		1e-4	1e-5	1e-6	1e-5	0.1	0.2
Val Loss	0.08444	0.08538	0.08397	0.08398	0.08429	0.08458	0.08696

Table 1: Table showing validation loss 3.2 for the ν_e network. L2 refers to L2-weight-penalty on all convolutional layers, FC L2 refers to L2-weight-penalty on the fully connected layer and Dropout refers to Dropout applied to the fully connected layer.

squared scaled error results in extremely large loss values for small energies. We therefore use the mean absolute scaled error given by:

$$L = \frac{1}{n}\sum_{i}^{n}\left|\frac{o_i - t_i}{t_i}\right| \tag{1}$$

where o_i is the CNN output, t_i the target energy and n the batch size. It can be seen that this optimizes a shifted version of the energy resolution: $\frac{E_{reco}-E_{true}}{E_{true}} = \frac{E_{reco}}{E_{true}} - 1$. Figure 3 shows validation performances for mean absolute scaled error, mean absolute error and mean squared error. We also investigate the effect of regularizers on the regression CNN. We consider an L2 weight-penalty on the convolutional layer weights or the fully connected layer weights and dropout. Best validation losses are shown in table 1. It can be seen that small values of L2 weight penalty on the convolutional layers improves validation performance. Results presented in section 4 refer to the baseline model.

The models are trained with stochastic gradient descent using the ADAM algorithm [11]. We choose a batch size of $n = 128$ and an initial learning rate of $1e-3$ with learning rate reductions of 10 when the validation loss shows no improvement larger than 0.004 for 4 epochs. Models are trained for 100 epochs or until the validation loss does not increase by at least 0.001 for 5 epochs. The weights from the epoch with the best validation loss are kept. All models are implemented in Keras [8] with Tensorflow backend. For testing and production, the model weights are transferred to Caffe [10] which is integrated in the NOvA software framework.

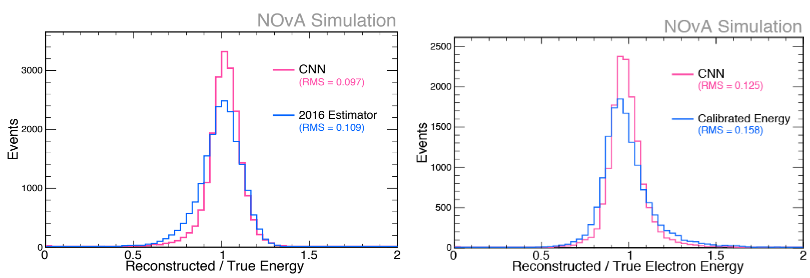

Figure 5: Plot of results for electron neutrino energy (left) and electron shower energy (right). For each plot the traditional reconstruction method is shown in blue and the pink curve is the CNN based method. A Gaussian curve is fitted to each histogram and its standard deviation is given as RMS values for each method.

4 Results and Future Work

We introduced a Convolutional Neural Network based energy regressor for the NOvA detectors. The method shows promising results for electron neutrino energy and electron shower energy prediction. Figure 5 compares results for the CNN method against the previous best results. The taller peaks for electron neutrino energy and electron shower energy indicate smaller relative test errors. In particular, the CNN improves the RMS by 11% for ν_e CC energy reconstruction. For electron shower energy reconstruction the CNN shows an improvement of 21% over the current best method. We note that this method's testing is entirely based on simulation. Results are therefore preliminary. Our next step is to apply this method to muon neutrino energy and muon energy. The challenge in this is that muons have longer trajectories which require larger input images. Furthermore, we aim to combine our efforts with work at NOvA on event and particle classification. It is of particular interest to combine these models into one to reduce the computational cost during prediction.

References

[1] K Abe, T Abe, H Aihara, Y Fukuda, Y Hayato, K Huang, AK Ichikawa, M Ikeda, K Inoue, H Ishino, et al. Letter of intent: The hyper-kamiokande experiment—detector design and physics potential—. *arXiv preprint arXiv:1109.3262*, 2011.

[2] K Abe, J Adam, H Aihara, T Akiri, C Andreopoulos, S Aoki, A Ariga, T Ariga, S Assylbekov, D Autiero, et al. Measurement of neutrino oscillation parameters from muon neutrino disappearance with an off-axis beam. *Physical review letters*, 111(21):211803, 2013.

[3] Corey Adams, David Adams, Tarek Akiri, Tyler Alion, Kris Anderson, Costas Andreopoulos, Mike Andrews, Ioana Anghel, João Carlos Costa dos Anjos, Maddalena Antonello, et al. The long-baseline neutrino experiment: exploring fundamental symmetries of the universe. *arXiv preprint arXiv:1307.7335*, 2013.

[4] P Adamson, C Ader, M Andrews, N Anfimov, I Anghel, K Arms, E Arrieta-Diaz, A Aurisano, DS Ayres, C Backhouse, et al. First measurement of muon-neutrino disappearance in nova. *Physical Review D*, 93(5):051104, 2016.

[5] A Aurisano, C Backhouse, R Hatcher, N Mayer, J Musser, R Patterson, R Schroeter, and A Sousa. The nova simulation chain. In *Journal of Physics: Conference Series*, volume 664, page 072002. IOP Publishing, 2015.

[6] A Aurisano, A Radovic, D Rocco, A Himmel, MD Messier, E Niner, G Pawloski, F Psihas, A Sousa, and P Vahle. A convolutional neural network neutrino event classifier. *Journal of Instrumentation*, 11(09):P09001, 2016.

[7] Michael David Baird. *An analysis of muon neutrino disappearance from the NuMI beam using an optimal track fitter*. PhD thesis, Indiana University, 2015.

[8] François Chollet et al. Keras. `https://github.com/fchollet/keras`, 2015.

[9] The IceCube et al. Pingu sensitivity to the neutrino mass hierarchy. *arXiv preprint arXiv:1306.5846*, 2013.

[10] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678. ACM, 2014.

[11] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[12] Yu-Feng Li, Jun Cao, Yifang Wang, and Liang Zhan. Unambiguous determination of the neutrino mass hierarchy using reactor neutrinos. *Physical Review D*, 88(1):013008, 2013.

[13] Evan Niner. *Observation of electron neutrino appearance in the NuMI beam with the NOvA experiment*. PhD thesis, Indiana University, 2015.

[14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.