Convolutional neural networks for energy and vertex reconstruction in DUNE

Ilsoo Seong
Department of Physics and Astronomy
University of California Irvine

Lars Hertel
Department of Computer Science
University of California Irvine

Julian Collado
Department of Computer Science
University of California Irvine

Lingge Li
Department of Computer Science
University of California Irvine

Nitish Nayak
Department of Physics and Astronomy
University of California Irvine

Jianming Bian
Department of Physics and Astronomy
University of California Irvine

Pierre Baldi
Department of Computer Science
University of California Irvine

Abstract

Measuring neutrino CP violation and mass hierarchy is currently one of the biggest challenges in particle physics. The DUNE neutrino experiment is the next-generation flagship neutrino program in the US designed to solve these problems. The DUNE detector uses liquid argon time projection chamber (LArTPC) technology, considerably improving the spatial resolution, neutrino detection efficiency and background rejection. However, reconstructing neutrino events with DUNE presents many challenges due to missing energy caused by argon impurities, nonlinear detector energy responses, invisible energy, hadron identities (mass), and overlaps between lepton and hadron interactions. One way of approaching this problem is using machine learning to reconstruct the neutrino events from pixel map images of interactions in the detectors. Here we present a regression convolutional neural network with a custom architecture to reconstruct neutrino energy and interaction vertices. For neutrino energy, we show considerable performance improvements in Monte Carlo simulations, compared with previous traditional energy reconstruction methods and initial results in interaction vertices.

1 Introduction

Neutrinos are extremely weakly interacting subatomic particles of considerable interest due to many strange properties that could provide a window into understanding the fundamental laws of physics. A particularly curious phenomenon is the oscillation between different types of neutrinos (flavors), which prove they have mass. A large experimental physics program[1, 2] has been set up to study the parameters that affect this oscillation probability, as a function of neutrino energy.

Neutrino interactions are rare and complicated making these measurements quite challenging. The next-generation experiment, DUNE is uniquely positioned to overcome these challenges by using liquid argon time projection chamber (LArTPC) technology, considerably improving the spatial

Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada.

resolution, neutrino detection efficiency and background rejection[3]. The goal of the experiment is to measure the oscillation probability between ν_μ and ν_e flavors. To measure the oscillation parameters it is crucial to accurately reconstruct the neutrino energies and interaction points from the finely observed details of the interaction. However, reconstructing neutrino events with DUNE presents many challenges due to missing energy caused by argon impurities, nonlinear detector energy responses and overlapping particle trajectories. Here we describe a convolutional neural network approach to reconstruct neutrino energies and interaction vertices at DUNE with a custom architecture suited towards this kind of regression problem. We estimate ν_e and ν_μ energies and ν_e interaction vertices with these models. These techniques have been very successful in particle physics due to their capacity to work with very complex data and the fact that many events can be represented as images[4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. Neural networks have been previously used at NOvA for flavor tagging, oscillation physics, electron neutrino energy and electron energy prediction[15, 16, 17] and have also been used in other neutrino experiments.[18, 19, 20, 21, 22].

2 The DUNE experiment

DUNE, located in South Dakota, will consist of 4×10kT modules filled with Liquid Argon (LAr) to detect neutrinos from a high-power neutrino beam from Fermilab, near Chicago, 1300km away. Charged particles produced from neutrino interactions, traverse and ionize the LAr producing electrons that drift horizontally towards the anode by an electric field. In the single-phase detector modules, the anode consists of two induction wire planes (U, V) oriented at 60° to each other and a collection plane, oriented vertically (Z). The wire planes measure the number of electrons and where they are drifting from the detector. Combined with the time of electron drift, they enable a 3D reconstruction of the neutrino interaction by looking at the digitized charge readout in the U-T, V-T and Z-T projections. The cartesian coordinates, (X, Y and Z) are obtained from a transformation of the 3 projections, where the time of electron drift, T is mapped to X and the combination of U and V mapped to Y, with Z denoting the third dimension.

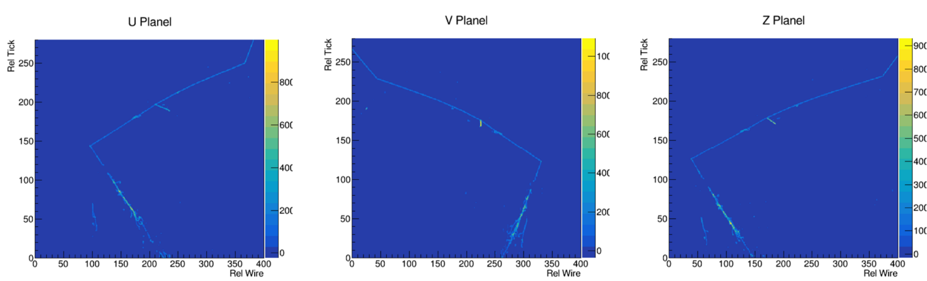


Figure 1: Pixel Maps for ν_e interactions in the three projections

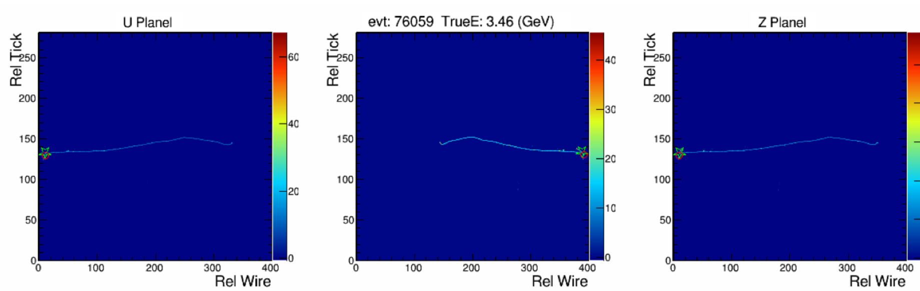


Figure 2: Pixel Maps for ν_μ interactions in the three projections

2.1 Data Samples

We create the dataset using the standard DUNE simulation chain[23] for neutrino interactions. The charge readout from drifting electrons is deconvoluted from detector effects by subtracting the noise profile and dividing out the electronics response in Fourier space. The charge pulses at each wire-plane (hits) are then disambiguated and fit with a gaussian shape to measure the total charge and the peak time for each hit. Clusters of hits are collected in each of the 2D projections and then combined into a 3D representation by matching the collection of hits in each projection.[23]

The traditional kinematic approach to reconstructing neutrino energies involves calibrating the total measured charge readout (ADC) from the leptonic and hadronic portion of the neutrino interaction to the expected true energy of the particle in the simulation. This is done separately because the leptonic and hadronic portions have quite different topologies and charge deposition profiles. The results are then summed to give an estimate of the neutrino energy.

For our purposes, we use an image representation of the entire interaction directly and feed it into our model. This side steps the complicated reconstruction required in charge calibration and topology identification, which may not reach the desired accuracy. These high resolution images are coarsened by merging different time ticks and wire-planes for computational feasibility in training. For ν_μ interactions where the leptonic portion is characterized by very long μ-tracks resulting in extremely large images (6720 ticks$\times$2800 wires), we merge 7 wires and 24 time ticks, resulting in a 280$\times$400-pixel image. We use 625000 events for training and testing the estimation of ν_μ energy. For ν_e interactions, where the leptonic portion is a much shorter electron shower, we merge 6 time ticks, also resulting in a 280$\times$400-pixel image. These images contain 90% of all the hits in the interaction on average. We use 930000 events to train and test our model for ν_e energy and vertex regression. Sample images for ν_e and ν_μ can be seen in figure 1 and 2 respectively. We only use samples which are fully contained inside the fiducial volume of the detector.

To find the interaction vertex, even a 280$\times$400-pixel image is quite large, so proceed in two stages. The first stage estimates the wire-plane and time tick closest to the vertex, which is computationally feasible due to the integer target variable. The second stage uses a smaller 24$\times$40-pixel image around the output of the first stage to estimate the interaction vertex as a continuous variable. Here we use the same examples as in the ν_e energy training.

For all the datasets above, the charge readout in each pixel is normalised between 0 and 1, by a constant scaling of 500. Different normalisation methods were found to not affect the result in any significant way.

3 Models

3.1 Energy Regression

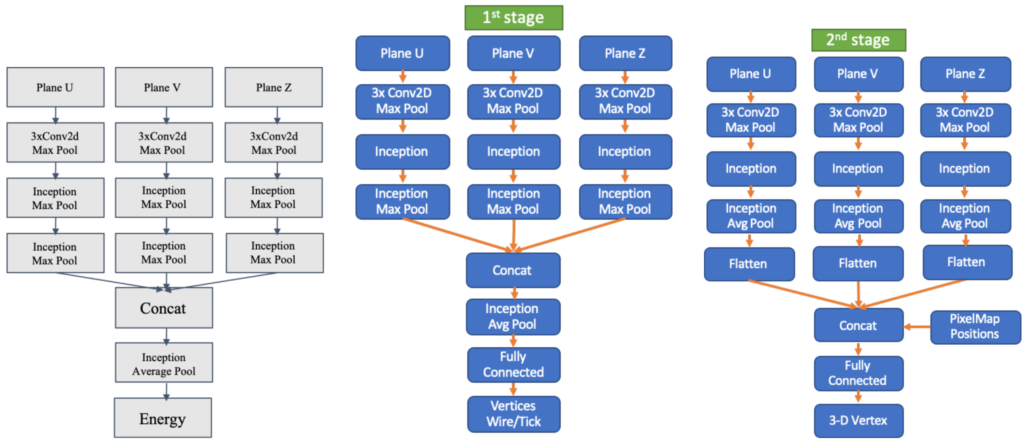
The model consists of three towers, one per image plane (U, V, Z) and can be visualized in figure 3. Each tower has a 7x7 convolutional layer followed by a 3x3 max pooling, both with 2x2 strides. Next a 1x1 convolution and a 3x3 convolutional layer with 1x1 stride is used followed by a 3x3 max pooling layer with 2x2 stride. Then two Inception blocks[24] and a further max pooling layer is used at the end of the tower. Afterwards the towers are concatenated and passed through another Inception block with an average pooling layer and a linear layer for calculating the energy. All convolutional layers use rectified linear units[25]. The models are trained using Adam[26] with learning rate 0.001 and batch size 16. The learning rate was multiplied by 0.25 if no improvement was found on the validation loss after 4 epochs. Adding dropout[27] or L2 norm for regularization did not improve performance compared to the base model. The same architecture and hyper-parameters were used for ν_μ energy. All models were trained using Keras [28] with a Tensorflow [29] backend.

Each event in the training is weighted to flatten the input energy distribution in order to remove any bias towards the energy distribution. The model was trained using a mean scaled error loss using these event-wise weights.

$$L(W, \{x_i, y_i\}_{i=1}^{n}) = \frac{1}{\sum_{i=1}^{n} \sqrt{w_i}} \sum_{i=1}^{n} \sqrt{w_i} \left| \frac{f_W(x_i) - y_i}{y_i} \right| \tag{1}$$

3.2 Vertex Regression

For the first stage of the interaction vertex model, the architecture and hyper-parameters are the same as in the energy prediction except the last layer is substituted with the wire-plane and time tick closest



(a) ν_e and ν_μ (b) Vertex stage one (c) Vertex stage two

Figure 3: Neural network architectures for energy and vertex regression.

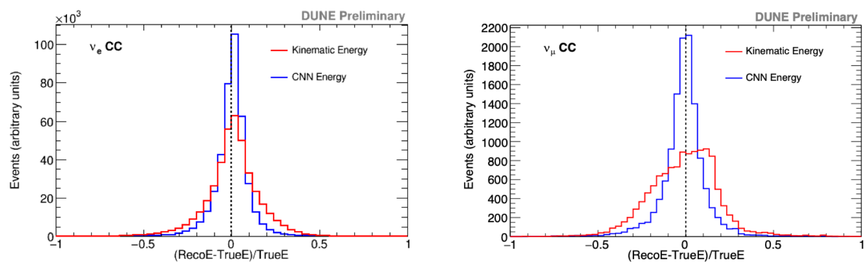


Figure 4: Energy regression results

to the true target vertex position instead of the energy. In the second stage the architecture is also the same except that average pooling is used after the inception blocks instead of max pooling and the output is flattened and concatenated with the digitized positions predicted by the first stage. The final layer then predicts the 3-dimensional cartesian coordinates (X,Y,Z) of the vertex position. The architecture for both stages can be seen in figure 3

Since the time ticks (representing the X coordinate) are merged by a factor of 5 in the pixel images, the resolution in X is recovered by weighting the loss function in that dimension by the same factor.

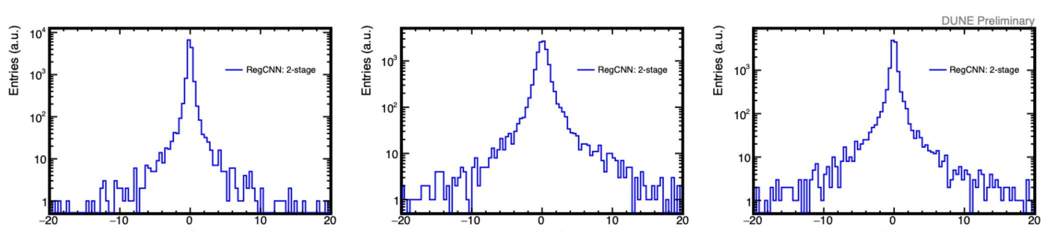


Figure 5: Vertex regression results

4 Results and Future Work

Figure 4 shows that our model energy estimate outperforms the kinematic method significantly. The resolution from a gaussian fit to the distribution for our model is 7.2% and 12.5% compared to 13.1% and 19.0% from the kinematic model, for ν_e and ν_μ respectively. This represents an improvement of

$\sim$40% over traditional techniques. From Figure 5, the resolution of the vertex coordinates is 0.98cm, 1.98cm and 1.67cm for X, Y and Z respectively, which is quite promising.

Further improvements to the training are being planned. For example extending the ν_μ samples with uncontained events where the μ tracks exit the detector, which traditional techniques find extremely challenging to estimate. Also, using sliding pixel maps across the μ-track could further improve the resolution on ν_μ events due to its topology. Finally the vertex model will be compared to other techniques currently being developed.

Our work shows using convolutional neural networks for ν_μ and ν_e energy estimation beats state of the art approaches by a large margin and shows promising results for vertex estimation. Using these techniques show a promising path towards furthering the physics program of the DUNE experiment.

Acknowledgments

We wish to acknowledge a hardware grant from NVIDIA. This material is based upon work supported by the National Science Foundation under grant number 1633631.

References

[1] M. A. Acero et al. New constraints on oscillation parameters from ν_e appearance and ν_μ disappearance in the NOvA experiment. *Physical Review D*, 98(3):032012, 2018.

[2] K. Abe et al. Measurement of neutrino and antineutrino oscillations by the T2K experiment including a new additional sample of ν_e interactions at the far detector. *Physical Review D*, 96(9):092006, 2017.

[3] Babak Abi, R Acciarri, MA Acero, M Adamowski, C Adams, D Adams, P Adamson, M Adinolfi, Z Ahmad, CH Albright, et al. The dune far detector interim design report volume 1: Physics, technology and strategies. *arXiv preprint arXiv:1807.10334*, 2018.

[4] Daniel Guest, Julian Collado, Pierre Baldi, Shih-Chieh Hsu, Gregor Urban, and Daniel Whiteson. Jet flavor classification in high-energy physics with deep neural networks. *Physical Review D*, 94, 07 2016.

[5] Luke de Oliveira, Michael Kagan, Lester Mackey, Benjamin Nachman, and Ariel Schwartzman. Jet-images – deep learning edition. *Journal of High Energy Physics*, 2016, 11 2015.

[6] James Barnard, Edmund Dawe, Matthew Dolan, and Nina Rajcic. Parton shower uncertainties in jet substructure analyses with deep neural networks. *Physical Review D*, 95, 09 2016.

[7] Leandro Almeida, Mihailo Backovic, Mathieu Cliche, Seung Lee, and Maxim Perelstein. Playing tag with ann: Boosted top identification with pattern recognition. *Journal of High Energy Physics*, 2015, 01 2015.

[8] Patrick Komiske, Eric Metodiev, and Matthew Schwartz. Deep learning in color: towards automated quark/gluon jet discrimination. *Journal of High Energy Physics*, 2017, 12 2016.

[9] Pierre Baldi, Kevin Bauer, Clara Eng, Peter Sadowski, and Daniel Whiteson. Jet substructure classification in high-energy physics with deep neural networks. *Physical Review D*, 93, 03 2016.

[10] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. Searching for exotic particles in high-energy physics with deep learning. *Nature communications*, 5:4308, 07 2014.

[11] Peter Sadowski, Balint Radics, Ananya, Yasunori Yamazaki, and Pierre Baldi. Efficient antihydrogen detection in antimatter physics by deep learning. *Journal of Physics Communications*, 1, 06 2017.

[12] Peter Sadowski, Pierre Baldi, Edison Weik, Daniel Whiteson, Edward Goul, and Andreas Søgaard. Decorrelated jet substructure tagging using adversarial neural networks. *Physical Review D*, 96, 03 2017.

[13] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. Calogan: Simulating 3d high energy particle showers in multi-layer electromagnetic calorimeters with generative adversarial networks. *Phys. Rev. D*, 97, 2018.

[14] Luke de Oliveira, Michela Paganini, and Benjamin Nachman. Learning particle physics by example: Location-aware generative adversarial networks for physics synthesis. *Comput Softw Big Sci*, 2017.

[15] A. Aurisano, A. Radovic, D. Rocco, A. Himmel, Mark Messier, E. Niner, G. Pawloski, F. Psihas, Alexandre Sousa, and P. Vahle. A convolutional neural network neutrino event classifier. *Journal of Instrumentation*, 11, 04 2016.

[16] P. Adamson, L. Aliaga, D. Ambrose, Nikolay Anfimov, A. Antoshkin, E. Arrieta-Diaz, K. Augsten, A. Aurisano, C. Backhouse, M. Baird, Bindu Bambah, K. Bays, Bivash Behera, S. Bending, R. Bernstein, V. Bhatnagar, B. Bhuyan, J. Bian, Tyrone Blackburn, and R. Zwaska. Constraints on oscillation parameters from ν_e appearance and ν_μ disappearance in nova. *Physical Review Letters*, 118, 06 2017.

[17] Pierre Baldi, Jianming Bian, Lars Hertel, and Lingge Li. Improved energy reconstruction in nova with regression convolutional neural networks. *Phys. Rev. D*, 99:012011, Jan 2019.

[18] Evan Racah, Seyoon Ko, Peter Sadowski, Wahid Bhimji, Craig Tull, Sang-Yun Oh, Pierre Baldi, and Mr Prabhat. Revealing fundamental physics from the daya bay neutrino experiment using deep neural networks. *In International Conference on Machine Learning and Applications (ICMLA)*, 01 2016.

[19] J. Renner, A. Farbin, Javier Muñoz Vidal, Jose Benlloch-Rodríguez, A. Botas, Paola Ferrario, Juan Gómez-Cadenas, Vicente Alvarez, C.D.R. Azevedo, F.I.G.M. Borges, Sara Cárcel, J.V. Carrión, S. Cebrián, A. Cervera, Carlos Conde, José Díaz, M. Diesburg, R. Esteve, Luis Fernandes, and H. Yepes-Ramírez. Background rejection in next using deep neural networks. *Journal of Instrumentation*, 12, 09 2016.

[20] G.N. Perdue, A. Ghosh, M. Wospakrik, F. Akbar, D.A. Andrade, M. Ascencio, L. Bellantoni, A. Bercellie, M. Betancourt, G. F. R. Caceres Vera, T. Cai, M.F. Carneiro, J. Chaves, D. Coplowe, H. da Motta, G.A. Díaz, J. Felix, L. Fields, R. Fine, A. M. Gago, R. Galindo, T. Golan, R. Gran, J. Y. Han, D. A. Harris, D. Jena, J. Kleykamp, M. Kordosky, X.-G. Lu, E. Maher, W. A. Mann, C. M. Marshall, K.S. McFarland, A. M. McGowan, B. Messerly, J. Miller, J. K. Nelson, C. Nguyen, A. Norrick, Nuruzzaman Nuruzzaman, A. Olivier, R. Patton, M.A. Ramírez, R. D. Ransome, H. Ray, L. Ren, D. Rimal, D. Ruterbories, H. Schellman, C. J. Solano Salinas, H. Su, S. Upadhyay, E. Valencia, J. Wolcott, B. Yaeggy, and S. Young. Reducing model bias in a deep learning classifier using domain adversarial neural networks in the MINERvA experiment. *Journal of Instrumentation*, 13(11):P11020–P11020, nov 2018.

[21] S. Delaquis, M.J. Jewell, I. Ostrovskiy, M. Weber, T. Ziegler, J. Dalmasson, L. Kaufman, T. Richards, J.B. Albert, G. Anton, I. Badhrees, P.S. Barbeau, R. Bayerlein, D. Beck, Vladimir Belov, M. Breidenbach, Thomas Brunner, G.F. Cao, W.R. Cen, and O.Ya Zeldovich. Deep neural networks for energy and position reconstruction in exo-200. *Journal of Instrumentation*, 13:P08023–P08023, 08 2018.

[22] MicroBooNE collaboration, R. Acciarri, C. Adams, R. An, Jonathan Asaadi, M. Auger, L. Bagby, B. Baller, G. Barr, M. Bass, F. Bay, M. Bishai, A. Blake, T. Bolton, Leonard Bugel, L. Camilleri, D. Caratelli, B. Carls, R. Fernandez, and C. Zhang. Convolutional neural networks applied to neutrino events in a liquid argon time projection chamber. *Journal of Instrumentation*, 12, 11 2016.

[23] Michael James Wallbank. *Reconstruction and Analysis for the DUNE 35-ton Liquid Argon Prototype*. PhD thesis, Sheffield U., 2018.

[24] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2014.

[25] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR.

[26] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[27] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012.

[28] François Chollet et al. Keras. `https://keras.io`, 2015.

[29] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.