

Improved vertex finding in the MINERvA passive target region with convolutional neural networks and Deep Adversarial Neural Network

Anushree Ghosh*

Universidad Tecnica Federico Santa Maria - Departamento de Fisica Casilla 110-V, Avda. Espana 1680, Valparaiso, Chile

E-mail: anushree.ghosh@usm.cl

The precise understanding of the nuclear dependence of the neutrino-nucleus cross section is very important in its own right, and in turn helps with the accurate measurement of the neutrino oscillation parameters. We study these issues by examining the event kinematics and cross section ratios between different passive targets at MINERvA. There is an ongoing study of the A dependent nuclear effects in MINERvA; this measurement hinges on identification of the target nucleus, thereby demanding accurate reconstruction of the event vertex. Vertex reconstruction is usually done with tracking-based algorithms. However, the performance of this method suffers when there are tracks created by secondary interactions or decays, or when significant shower activity occurs near the vertex region. Here, we present an alternative vertex finding method in the nuclear target region using machine learning. We use an application of deep learning based on convolutional neural networks (CNNs). Additionally, we explore the use of Deep Adversarial Neural Networks (DANNs) in minimizing possible biases coming from the use of simulated data for the algorithm's training sets.

*The 19th International Workshop on Neutrinos from Accelerators-NUFACT2017
25-30 September, 2017
Uppsala University, Uppsala, Sweden*

*Speaker.

1. Introduction

The MINERvA experiment (Main INjector ExpeRiment for n-A) is a neutrino nucleus scattering experiment located at Fermi National Accelerator Laboratory which is designed to precisely measure the cross sections of different neutrino interaction on various nuclear targets. It uses the NuMI (Neutrinos at Main Injector) beam, [1] which gives an intense flux of neutrinos or antineutrinos, created by the decay of mesons produced in proton-carbon collisions at 120 GeV. The MINERvA detector, described in details [2], consists of a fine-grained tracker, a passive target region, downstream and circumferential calorimetry, and MINOS near detector as muon spectrometer. The central tracking region of the MINERvA detector is constructed from 120 planes of parallel triangular strips of plastic scintillator, arranged almost perpendicular to the beam axis. Each strip contains a wavelength-shifting fiber, which delivers light generated by charged particles to photomultiplier tubes. The detector, surrounded by an electromagnetic calorimeter and a hadronic calorimeter, consists of several nuclear targets, including water, liquid helium, iron, lead and carbon. The most upstream part of the detector includes five inactive targets, numbered from upstream to downstream, each separated by four active tracking modules. Target 4 is made of lead whereas targets 1, 2, and 5 are constructed of steel and lead plates joined together. Target 3 has graphite, steel, and lead plates. The Magnetized MINOS near detector sits downstream of MINERvA which serves as a muon spectrometer.

The precise understanding of the nuclear dependence of the neutrino-nucleus cross section is very important to reduce the uncertainty associated with it, which in turn helps the accurate measurement of the neutrino oscillation parameters. This can be understood through the event kinematics and the measurement of the cross sections as ratios between different passive targets at MINERvA. The recent change of the NuMI beam line to the Medium Energy configuration has increased both the intensity and average neutrino energy thereby greatly improving the projected sensitivities for nuclear and hadronic structure analyses utilizing MINERvA's multiple nuclear targets. This study requires the precise identification of the target nucleus, thereby demanding the accurate reconstruction of the vertex.

2. Traditional vertex reconstruction and challenges in medium energy

In traditional reconstruction method, the vertex is identified by reconstructing tracks using a variety of pattern recognition algorithms [2]. Generally, tracking is done by first identifying the longest track in an event, called "anchor track" and then recursively searching its beginning and ending points for other tracks. The origin of the anchor track is the first candidate for the primary vertex. However, the traditional track-based vertexing can fail for deep inelastic scattering (DIS) events with high hadronic shower near the vertex as the shower activity may obscure the vertex location or tracks created by secondary interactions or decays might bias the vertex location. In case of the target analyses, missing the vertex by a few centimeters may cause a significant problem. Also, with the increased beam energy, there is an increase of events with mis-reconstructed vertices and the general strategy of rejecting events with these mis-reconstructed vertices may reduce the statistics gained.

Events in MINERvA can be represented as images and the vertex finding can be represented as an image classification problem that is well suited for a machine learning application. Hence, we use the deep convolutional neural networks to localize neutrino interaction vertices in the MINERvA nuclear target regions in search of better vertex reconstruction method.

3. Convolutional neural networks (CNNs)

Neural networks is a computing system, which can learn from observational data. Typically, a neural network is made of three type of layers, namely, Input layer , hidden layer, output layer. The layers are made of a number of interconnected 'nodes', which contain an 'activation function'. The nodes of the input and output layers are connected by weighted connections. Neural networks become "deep" when the outputs of one network become inputs of another, creating many hidden layers in the network. Convolutional neural networks (CNNs) [3] are a particular type of deep neural network that employ a structure inspired by the biological structure of the eye, and they are particularly well-suited to image recognition and computer vision problems. In a CNN, the inputs of each layer, typically an image at the beginning of the network, are mapped into outputs by convolving a kernel with the inputs thereby extracting a set of features from the input image. By using the same weights for the kernel across the input, the network is able to achieve the translationally invariant features. Stacking multiple layers with many kernels per layer allows a CNN to learn an optimal set of kernels for extracting features that describe the input image. As an image progresses through the layers of a CNN, the features extracted grow in semantic complexity [4]. In this way, CNN learns the important features of a dataset without a great deal of "by-hand" feature engineering. In addition to the data layers where the image is read into the neural network, the Neural Networks also utilizes convolutional, pooling, fully-connected, dropout and softmax layers. The final features extracted are then fed into a softmax loss layer which compares these features to the target label or value and assigns the cost using a cross-entropy function, which is subsequently minimized.

There are many frameworks which have been developed to train and apply deep neural networks. For this work, we employ Theano [5] and we employ Caffe [6] to study domain adversarial neural nets (see Section 5).

We present two strategies for analyzing events in the target region - first we perform a "large" segment hybrid classification-localization exercise and classify events as being in a particular target or in the scintillator between the targets . Here, we label the nuclear target region as 11 segments, where the odd segments refer to the passive target and the even segments refer to the scintillator trackers between the targets as well as the regions upstream and downstream of the nuclear target region. We also perform a classification-localization based on individual planes in the detector where we reconstruct the exact plane containing the interaction. In this case, we continue to treat the problem as a "classification" problem with each passive target also counting as its own plane. Here we label the nuclear target region as 67 segments, where each of the segment number refers to the scintillator plane and passive targets in MINERvA. The segment number will then mapped into the plane number in the detector space, where each passive target will occupy at least 2 planes.

4. Improvements to physics analysis

Figures 1 and 2 show the difference in the performance directly in the analysis. Here we use the “67 segments” version of the network. The plots show the event distributions as a function of plane number and separated by the type of target from where the vertex is originated. The green color represents signal events. In the track based vertex reconstruction (fig 1(a) and 2(a)), there are more signals leaking into the planes upstream and downstream of passive target. In order to regain these signals we apply wider cut to define the passive target events as shown by arrow in the figures. On the other hand, Deep Convolutional Neural Network (DCNN) based vertex reconstruction shows cleaner separation between plastic upstream of target and downstream of target compare to former method. DCNN predicts more signal events correctly and there is almost no signal leakage to the neighbouring planes. Hence, we can apply a tighter cut for our target events by only taking events that are predicted to be coming from the target.

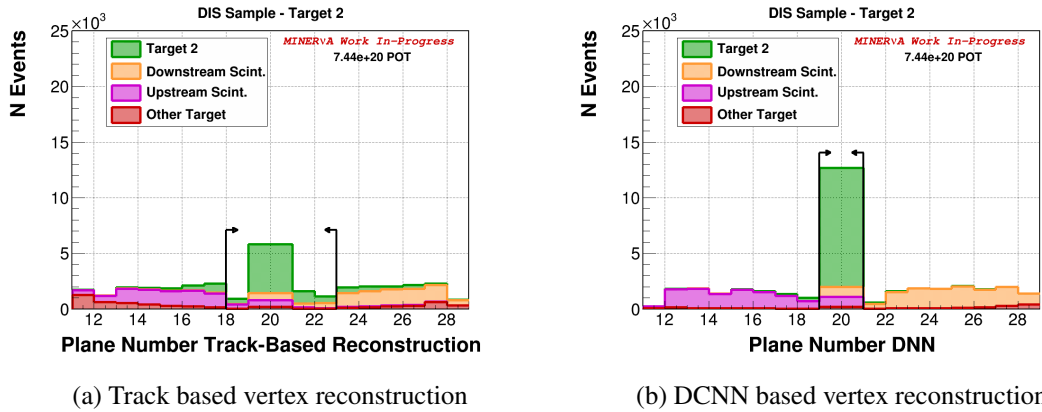


Figure 1: Source target material by reconstructed vertex location for Target 2. All the scintillator planes between Targets 1 and 2 are shown on the left side of the plots, and all the scintillator planes between Targets 2 and 3 are shown on the right side of the plots

5. Domain adversarial neural networks (DANNs)

A shortcoming of machine learning technique comes from the imperfect simulation model which is used to train the machine learning algorithms. Generally, the misclassification and misreconstruction are investigated by running the simulation several times with different parameters to account for possible differences between simulation and data which is a computationally challenging task.

We additionally study Domain Adversarial Neural Networks (DANNs) [7] where the trained model learns only those features that are present in both the simulation and the data and hence gives us a possible solution for the problem mentioned above.

DANN is operated by two loss functions where one is a regular feature classifier and the other is the domain classifier that attempts to distinguish events from *source* and *target* domains where source

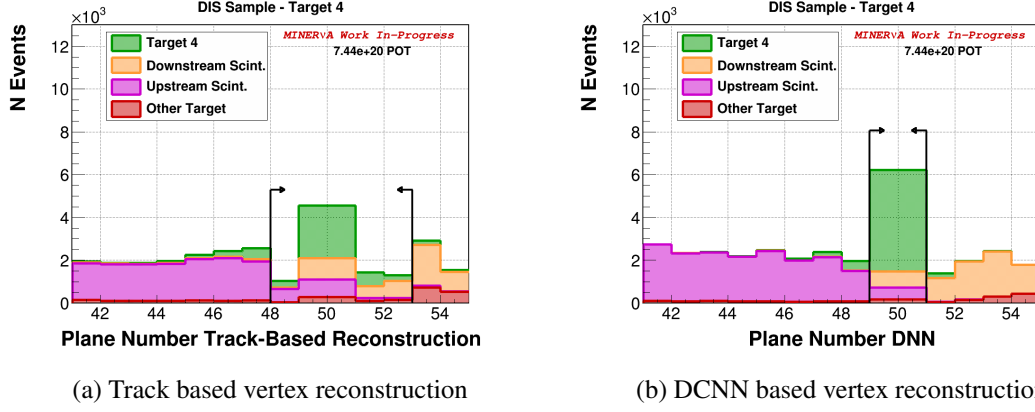


Figure 2: Source target material by reconstructed vertex location for Target 4. All the scintillator planes between Targets 3 and 4 are shown on the left side of the plots, and all the scintillator planes between Targets 4 and 5 are shown on the right side of the plots.

domain is our labeled Monte Carlo events and target domain is the unlabeled real data. Images from the source domain are used to train the classifier, while images from both the source and target domains are used to train the domain discriminator. The domain adaptation is achieved by minimizing the loss associated with the feature classifier and maximizing the loss associated with the domain classifier.

To validate the success of this method, we compare simulated data sets that feature large perturbations to the underlying physics model such as by turning off and on the final state interactions (FSI) of the hadrons produced at the primary vertex as they propagate through the dense nuclear medium to exit the nucleus. In order to investigate the robustness of our network to FSI, we use two domains - simulation with FSI turned on and off. Only simulation using the LE flux was available with FSI turned off at the time of the study and we use the “11 segment” version of the network here. The studies shows that considering DANN partner, accuracy become higher and loss is reduced compare to DCNN. However, we should remember that the loss and accuracy can be affected by the biasness inherited by the network during training which is known as overfitting of the network. To check this overfitting, we force the model to overfit by using a very small training sample of size 1000 events and turning off the dropout layers in our network which helps the network to reduce the overfitting rate by randomly cutting off the connection between layers. We treat the simulation with FSI turned on as the source domain and simulation with FSI turned off as the target domain. We validate the network with on “FSI turned off” sample. Figures 3a and 3b shows that there is a delayed onset of overfitting when using a DANN compare to DCNN.

6. Conclusion

We see that DCNNs are advantageous in extracting the relevant features for a task such as finding the vertex in a neutrino charged current event and it shows significant improvement over traditional reconstruction method. To limit the model dependency and increase the robustness of the network, we study the Domain adversarial neural network. To test the effectiveness of DANN, we generate different Monte Carlo dataset by turning off and on the FSI model owing to have

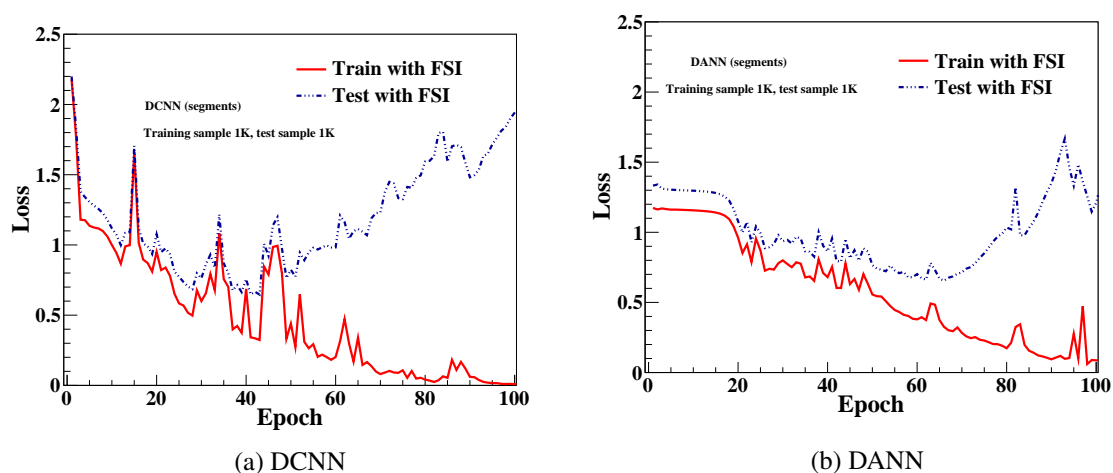


Figure 3: Overfitting test. Here the loss is plotted with number of epochs where epoch is defined as a single pass through the entire training set during the training a machine learning model.

event samples with different features. As a preliminary study, we show that considering a DANN partner the network helps to reducing the overfitting rate during training compare to than DCNN.

References

- [1] P. Adamson, et al., The NuMI Neutrino Beam, Nucl. Instrum. Meth. A806 (2016) 279-306. arXiv:1507.06690, doi:10.1016/j.nima.2015. 08.063
- [2] L. Aliaga, et al. [MINERvA Collaboration], "Design, Calibration, and Performance of the MINERvA Detector," Nucl.Instrum.Meth. A743 130-159 (2014).
- [3] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors, Nature 323 (1986) 533-536.
- [4] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012.
- [5] The Theano Development Team, Al-Rfou, et al., Theano: A Python framework for fast computation of mathematical expressions, ArXiv e- prints, arXiv:1605.02688.
- [6] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, arXiv preprint arXiv:1408.5093.
- [7] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, Journal of Machine Learning Research 17 (59) (2016) 1-35. URL <http://jmlr.org/papers/v17/15-239.html>