

Context-Enriched Identification of Particles with a Convolutional Network for Neutrino Events

F. Psihas,¹ E. Niner,² M. Groh,³ R. Murphy,³ A. Aurisano,⁴
A. Himmel,² K. Lang,¹ M. D. Messier,³ A. Radovic,⁵ and A. Sousa⁴

¹*Department of Physics, University of Texas at Austin
Austin, Texas 78712, USA*

²*Fermi National Accelerator Laboratory
Batavia, Illinois 60510, USA*

³*Department of Physics, Indiana University
Bloomington, Indiana 47405, USA*

⁴*Department of Physics, University of Cincinnati
Cincinnati, Ohio 45221, USA*

⁵*Department of Physics, College of William & Mary
Williamsburg, Virginia 23187, USA*

Particle detectors record the interactions of subatomic particles and their passage through matter. The identification of these particles is necessary for in-depth physics analysis. While particles can be identified by their individual behavior as they travel through matter, the full context of the interaction in which they are produced can aid the classification task substantially. We have developed the first convolutional neural network for particle identification which uses context information. This is also the first implementation of a four-tower siamese-type architecture both for separation of independent inputs and inclusion of context information. The network classifies clusters of energy deposits from the NOvA neutrino detectors as electrons, muons, photons, pions, and protons with an overall efficiency and purity of 83.3% and 83.5%, respectively. We show that providing the network with context information improves performance by comparing our results with a network trained without context information.

I. INTRODUCTION

Experiments in high-energy physics employ a variety of detection technologies and analysis techniques to study the properties of subatomic particles. Such techniques are vital for interpreting raw data as particle interactions from which physical quantities can be extracted. The performance of reconstruction algorithms largely dictates the physics potential of particle physics experiments. In the past several decades, reconstruction techniques have evolved to adapt to the growing complexity of detector technologies and the increasing volume of data, from the hand-scanning of bubble chamber plates [1] to the application of deep learning algorithms on large datasets [2].

Classification of interactions, or events, in detector data is often similar to image recognition in that topological features are useful discriminators for both tasks. Detector data are very image-like in that, in many cases, they can be mapped into two-dimensional arrays of values preserving spatial information [3]. In recent years, deep neural networks (predominantly convolutional, adversarial, and recurrent neural networks) have been employed in experiments across the field for tasks such as classification [4], simulation [5], and energy reconstruction [6], among others [2]. Convolutional Neural Networks (CNNs) have proven useful for detector data analysis given their ability to exploit topological information, and are increasingly common in the field. In neutrino experiments, CNNs are most commonly used for classification tasks [4, 7–9], though other applications, including image segmentation, have been tested on simu-

lations [10].

In experiments measuring neutrino oscillations, the observables of interest are the flavor and the energy of incident neutrinos, e.g. [11], among other quantities. Identification of the neutrino flavor requires discrimination between charged or neutral current interactions and separation from backgrounds, which include muons, electrons, photons, and hadrons from cosmogenic sources. Being neutral particles, neutrinos themselves deposit no energy as they traverse the detectors. Instead, the neutrino flavor can be determined by identifying the leptonic charged particle produced in the neutrino interaction. In addition, identifying all the final-state particles allows for cross section measurements of specific interaction types, a topic of great interest in the field [12], and enables in-depth studies of the neutrino events, leading to improvements in energy reconstruction.

Much like in image recognition, the context within which an object is depicted can aid the classification task. In particle identification, the context is the remainder of the event activity which corresponds to all particles produced in the interaction. This context provides information about the physics of the interaction, such as conservation of energy and momentum, conservation of quantum numbers, nuclear effects, particle multiplicity, etc. By providing context to the network, it is encouraged to learn from physical quantities implying conservation laws as well as the relative topology of the particle with respect to the accompanying activity.

We present the first implementation of a CNN for particle identification in neutrino detector data which uses

context information. This is also the first implementation of a CNN for pattern recognition which uses a four-tower siamese-type architecture [13] for including context information and independent two-dimensional views.

We have developed a CNN for classification of individual final state particles originating from neutrino interactions. We evaluate the performance of the technique on simulations from the NOvA experiment, but this technique is extendable to similar detector technologies. The classification of particles in NOvA events and our choices for a network architecture are described in Section II. We discuss the performance and utilization of this network in Section III.

II. PARTICLE CLASSIFICATION

The NuMI Off-Axis Electron Neutrino Appearance (NOvA) experiment is a long-baseline neutrino oscillation experiment. NOvA's primary physics goal is to measure neutrino oscillations by detecting neutrino interactions in two detectors, a near detector at 1 km and a far detector at 810 km from the muon neutrino (ν_μ) beam source [14]. In the far detector, a fraction of the ν_μ s will have oscillated into electron neutrinos (ν_e) or tau neutrinos (ν_τ) with a probability dependent on the energy of the neutrino and other physical parameters associated with this phenomenon [15]. Making precise measurements of neutrino oscillation probabilities requires both identification of the flavor and measurement of the incoming neutrino energy, which in turn rely on the ability to study particle interactions in matter.

The NOvA detectors are sampling calorimeters whose detection mechanism consists of collecting the scintillation light produced by charged particles within them. The detectors are made of alternating vertical and horizontal planes filled with scintillator material [16] such that a view of the energy deposits, or hits, from the vertical or horizontal planes is equivalent to a two-dimensional top or side view of the activity in the detector.

Figure 1 shows hits from particles produced in neutrino interactions in the NOvA far detector. The energy deposited by charged particles along their path can be used to reconstruct their trajectory, or track. The amount of energy deposited per unit distance (dE/dx) along tracks is an example of a useful discriminator for particle identification given the well-known behavior of particles as they travel through matter [17]. The events depicted are particles produced at NOvA's energy range, with a peak neutrino energy of 2 GeV. Different particles, with characteristically varying energy depositions, may be produced in neutrino interactions. At these energies, the signature of a muon is a long straight trajectory with a uniform dE/dx profile, corresponding to the minimum ionizing energy of the particle. Proton tracks tend to be significantly shorter than muon tracks, and are characterized by large energy deposits near the end of the track. While proton, charged pion, and muon tra-

jectories can look similar, the scattering behavior of each can be substantially different, causing deviations in their trajectories that are well described [17]. Electron and photon signatures are characterized by electromagnetic showers which broaden along the initial portions of their path.

Visual discrimination of different particles is often enhanced by contextual information from the whole interaction. In the case of photons and electrons, both initiate electromagnetic showers which appear similar in these detectors. However, they can be distinguished by the separation of the shower at the interaction vertex, the point where the neutrino interaction occurred. Electrons will deposit energy immediately from the vertex while photons will travel some distance, characterized by the radiation length of the medium, before producing an electron-positron pair and initiating a shower.

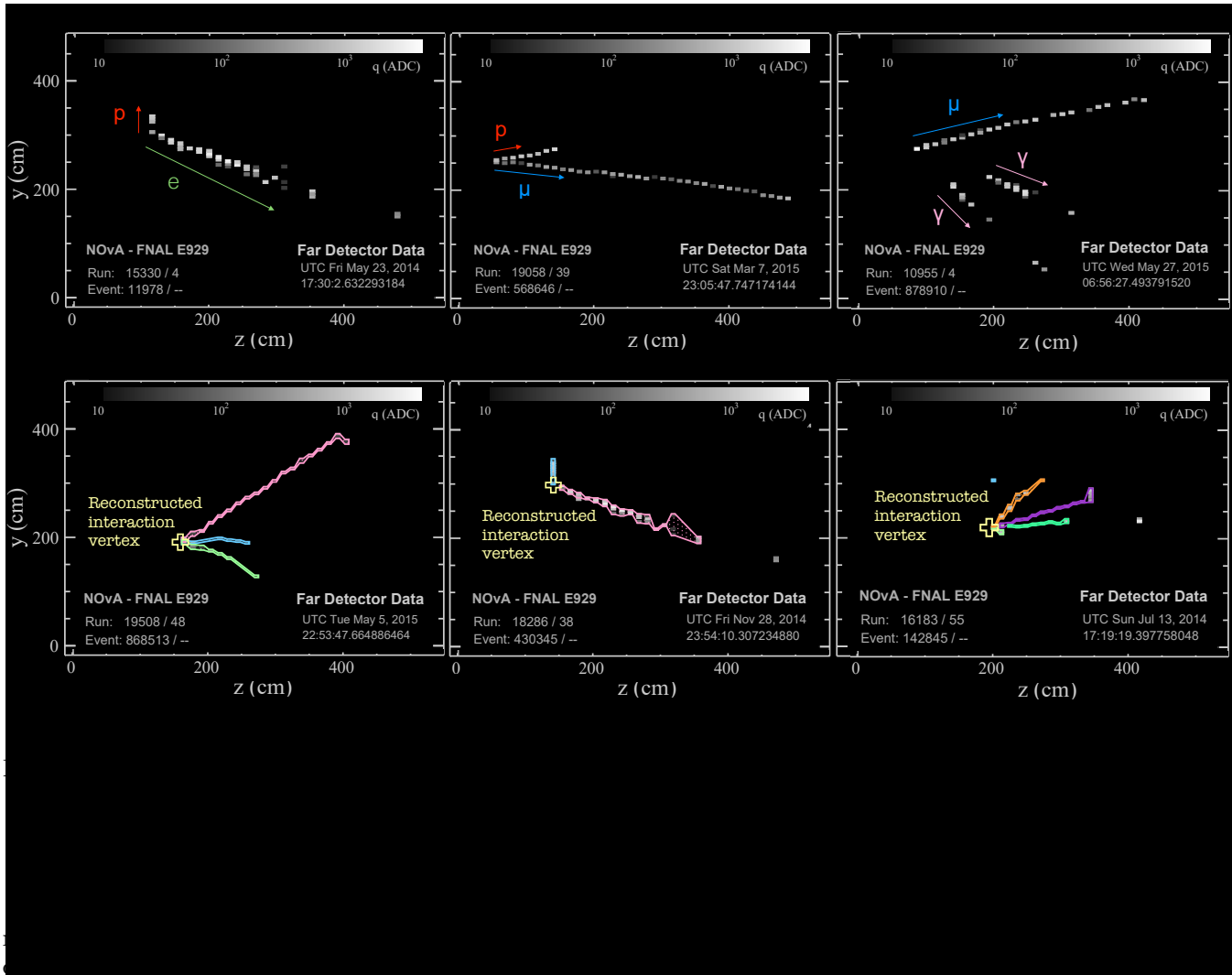
Our network classifies particles isolated by the NOvA reconstruction algorithms [18], which aim to separate the contributions from each individual particle into a cluster. Figure 1 (bottom) shows the reconstructed clusters in NOvA events. Often, particle trajectories overlap in the same detector cell which causes individual clusters to include hits produced by multiple particles. The average purity of the clusters varies by particle type, being typically much lower for short clusters, closest to the interaction vertex, which correspond largely to pions and protons. Despite the inefficiencies present in the clustering, a sample of particles built from fully simulated neutrino interactions was used as a training sample as they bear a greater resemblance to detector data than what could be accomplished with simulated single particles, where the clusters would contain no contamination.

For particle classification, we employ the siamese tower structure to separate the top and side views of the cluster [4], and we incorporate two additional towers, one for each view of the complete interaction, thus including the full event information as context. Several characteristics of the clusters in the event, like opening angle and distance from the interaction vertex, are contextual indicators of their identity and have been used successfully in the past as part of event selection. The intention of adding full context images to the input is to provide the network with this capability as well.

The context of an image has been shown to be critical for object detection tasks [19] and context-enriched learning has been previously implemented for image recognition [20]. The four-tower architecture implemented in this work is a new approach at context-enriched classification, and the first instance of context-enriched deep learning algorithms for particle physics.

A. Training Methodology

We use a sample of simulated neutrino interactions in the NOvA Far Detector for training, testing, and evaluation. While raw data from the NOvA experiment is



in reference [23].

Hits within the interactions are clustered within regions of space spanning some angle with respect to the reconstructed interaction vertex using a fuzzy K-means algorithm [24], and matched between the top and side views by comparing energy depositions along the z -axis using a Kuiper test [25]. The construction of clusters is constrained by the spatial and temporal resolution of the detector and subject to inefficiencies in the algorithm, which impacts the ability to contain the detected energy deposited by a single particle uniquely. Inefficiencies in vertex finding and separating overlapping particles also contribute to reducing the quality, the completeness and purity, of the reconstructed clusters.

A selection was applied to the events prior to training, requiring the ends of all clusters to be away from the edges of the detector, ensuring the use of fully contained events. For each particle category, a threshold was chosen for the purity of the clusters, to balance a representative sample of realistic input data with clear identities of the clusters. A purity of 0.5 or more was used for muons, electrons, and photons and a purity of 0.35 or more was used for pions and protons. The efficiency and purity of all clusters for each type of particle can be seen in

used for training and evaluation to reduce total evaluation time on NOvA data. Particle trajectories which are longer than 5 m come predominantly from muons (over 95%), whose characteristic pattern of energy deposition can be easily identified with other techniques.

The final sample consisted of 2.95 million particles made up of 9% muons, 33% electrons, 9% pions, 29% protons, and 20% photons. There is no distinction in the sample between a particle and its antiparticle. The sample was split, 80% used for training and the remaining 20% used to evaluate the algorithm's performance.

The network architecture is a simplified tower structure compared to GoogLeNet [26] and to our previous implementation [4]. The three consecutive inception modules per tower in the original architecture were reduced to one per tower [27], reducing the number of required computations with a roughly equivalent accuracy as the original classifier. The kernel size was optimized for our running time requirements, and the width of the network was reduced by a factor of two, without significant losses in performance. These improvements were also applied to the version of the Convolutional Visual Network (CVN) used by NOvA in its latest analysis [11].

The model was optimized using stochastic gradient de-

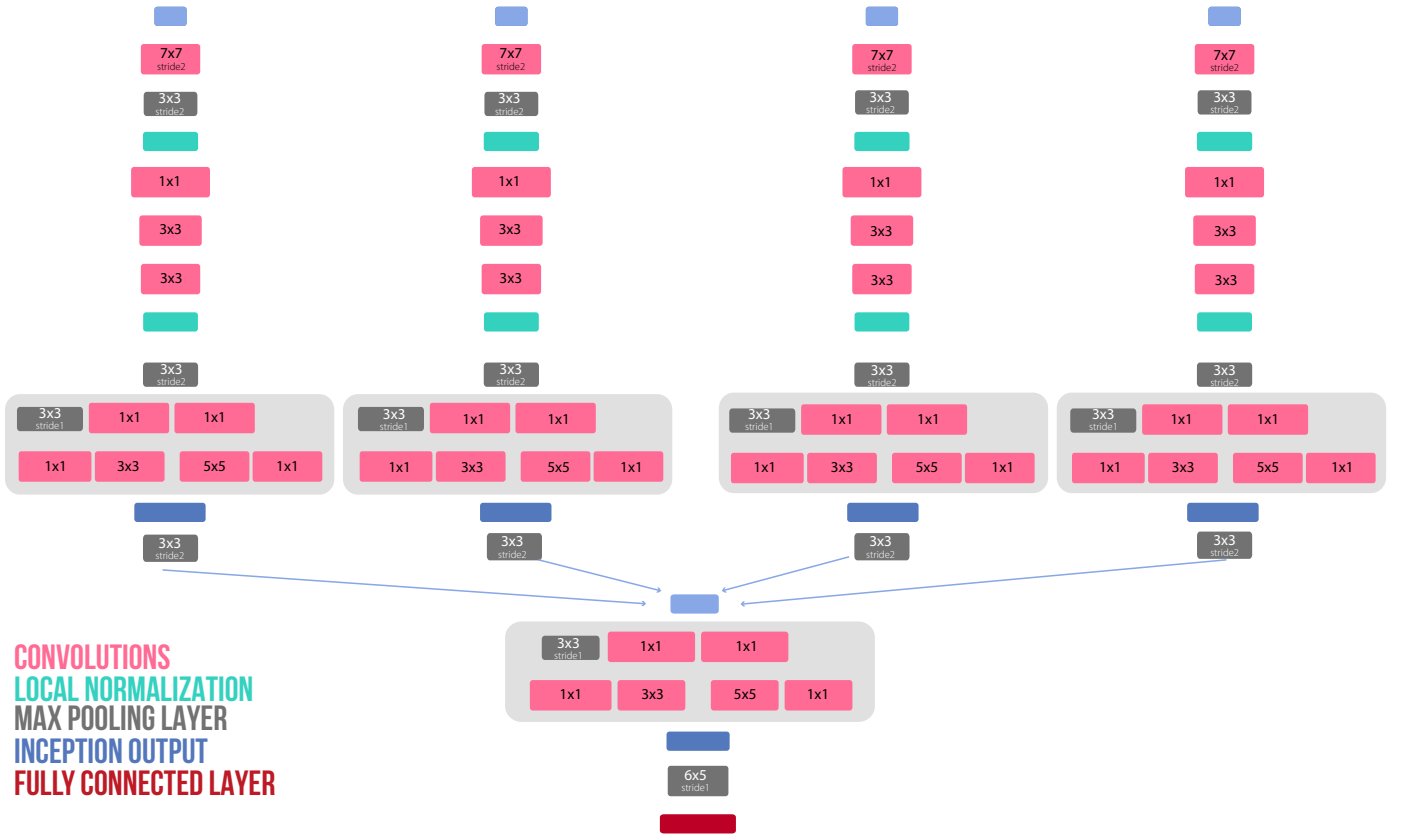


FIG. 2: A diagram of the CNN architecture for particle classification. The network has four towers for the top and side views of the event. Two of the four towers are for the views of the particle of interest. The other two towers are for the views of the entire interaction. Highlighted in light gray are inception modules.

scent with categorical cross entropy as the loss function. Figure 4 shows the loss and accuracy, a standard test of training stability. The training and test losses agree indicating no signs of over-training.

The final layer of the network is a dense layer with five nodes, one for each training category, using a softmax activation function:

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}} \quad (1)$$

where y_i are the outputs from the network before the activation function and $S(y_i)$ are the corresponding softmax scores, which yield a normalized set of outputs.

The network was trained using Caffe [28] 1.0 on four NVIDIA Tesla K20 GPUs on the BigRed II computing cluster at Indiana University. The training spanned a total of 700,000 iterations with a batch size of 64. After 300,000 iterations, the learning rate was reduced by a factor of 10.

Evaluation of NOvA data and simulations are produced in a ROOT [29] format compatible with the *art* framework [30]. Model evaluation was incorporated to the NOvA analysis infrastructure with the

use of Caffe C++ API. The source code and pre-trained models for the two network architectures used in this work are available on <https://github.com/FernandaPsihas/ParticleTowerCNN>.

III. CLASSIFICATION PERFORMANCE

The inputs to the network are two maps of the cluster without context and two maps with context each for the top and side view. The full architecture is shown in Figure 2. The weights are specific to each tower to encourage independent learning from each input. The resulting features from the four towers are concatenated before a final inception, pooling, and fully connected layers. To assess the effect of adding context to the input data, a particle-only network was trained using only the two views of the cluster for comparison.

The separation achieved for different particle types can be visualized from the two-dimensional mapping of events shown in Figure 5. This mapping is produced using t-Distributed Stochastic Neighbor Embedding (t-SNE) [31], a visualization algorithm which approximately maps similarity in high dimensional feature-space onto

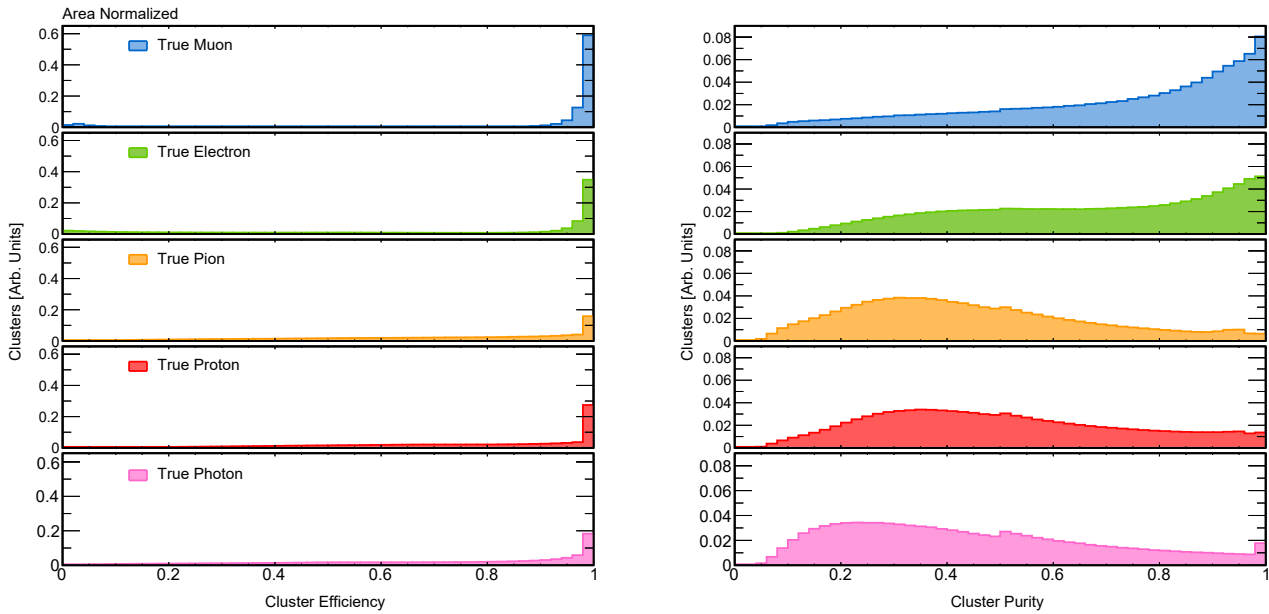


FIG. 3: The efficiency and purity of reconstructed clusters in the NOvA detectors. Left: Cluster efficiency, defined as the fraction of energy depositions from the particle associated with a cluster which are contained by the cluster. Right: Cluster purity, defined as the fraction of the energy contained in a cluster which comes from the particle it is associated with. Clusters are associated with whichever particle deposited the majority of the energy contained in each cluster. A high efficiency indicates that clusters tend to mostly contain the energy deposits from the particle in question. The considerably lower purity results from the overlap of contributions from multiple particles in detector cells contained in one cluster.

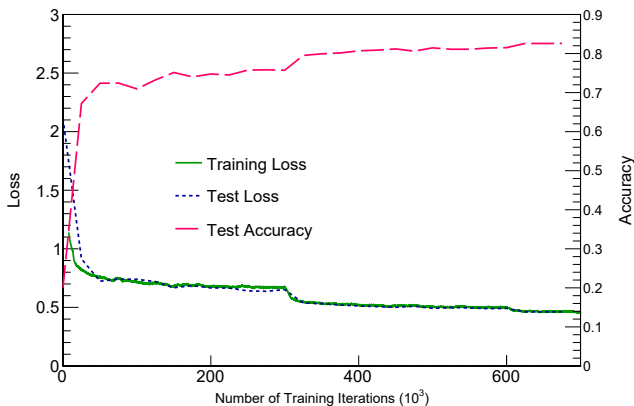


FIG. 4: The performance of the network as it is trained. The pink line depicts the accuracy of the network. The blue and green lines show the loss function of the test sample and training sample, respectively.

distance in two-dimensional space. The t-SNE visualizations represent the overlap in extracted features as proximity in space, providing information regarding the discrimination power of the network. Good separation is achieved for all particle types, but there are some expected regions of overlap, such as between electrons and photons.

One class of metrics used for evaluating the classification takes the largest softmax value as the simple prediction from the network. The top portion of Figure 6 shows the classification matrices for the networks simple prediction. The network achieves 93%, 75%, 93%, 65%, and 81% efficiency and 90%, 75%, 87%, 54%, and 89% purity for electrons, photons, muons, pions, and protons without any additional selections. The overall efficiency and purity of the classifier are 83.3% and 83.5% for all particles in the test sample. The figure shows worse performance in the distinction between protons and pions as well as electrons and photons, which is expected given the differences noted in Section II. We achieve separation of electrons and photons with only 6% contamination of photons in the electron-selected events, and 10% in the opposite case. It is important to emphasize that only the largest score has been used for identification, Figure 7 indicates further potential to select particles with higher selection purity using all scores.

The network without context was trained on the same sample of events and employed the same network architecture, but using only two towers for both views of the reconstructed cluster. The bottom portion of Figure 6 shows a comparison between the two networks. In this case, the positive values along the diagonal quantify the improvement in efficiency or purity with inclusion of context in the network. The improvement is more substantial for pions and photons, which gain up to 11% in ef-

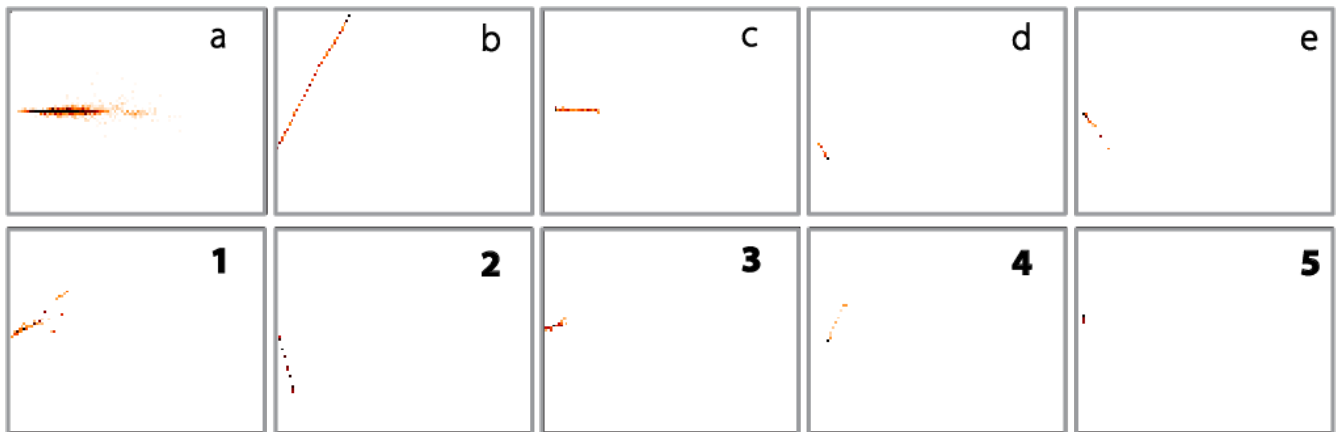
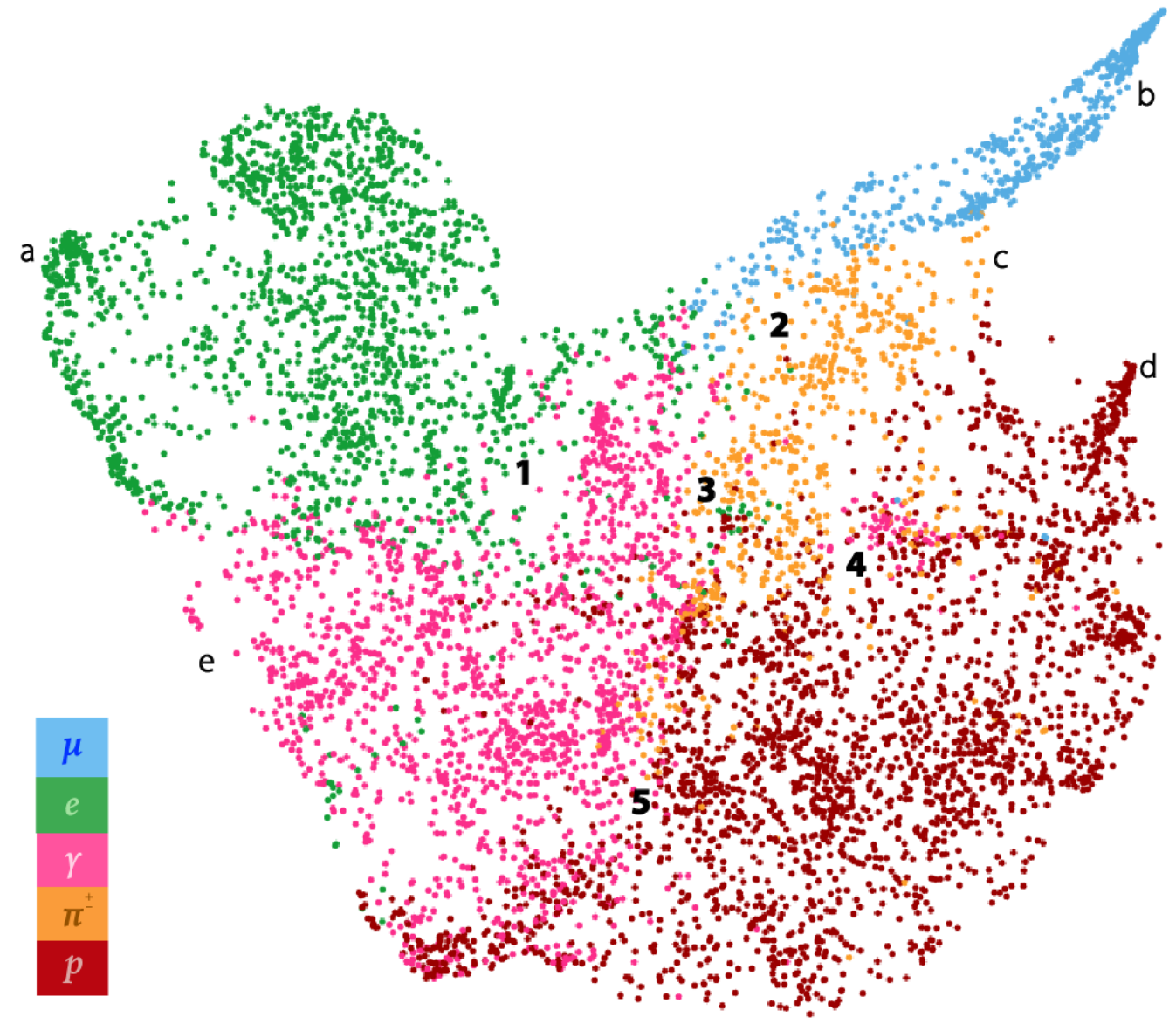


FIG. 5: Classified events in two dimensional probability space created with the t-SNE [31] technique from the classifier features. A clear separation can be seen between muons (blue) and electrons (green). The expected overlap is visible between pions (orange) and protons (red). Regions with good separation and example topologies are labeled a-e. Regions with poor separation and example topologies are labeled 1-5

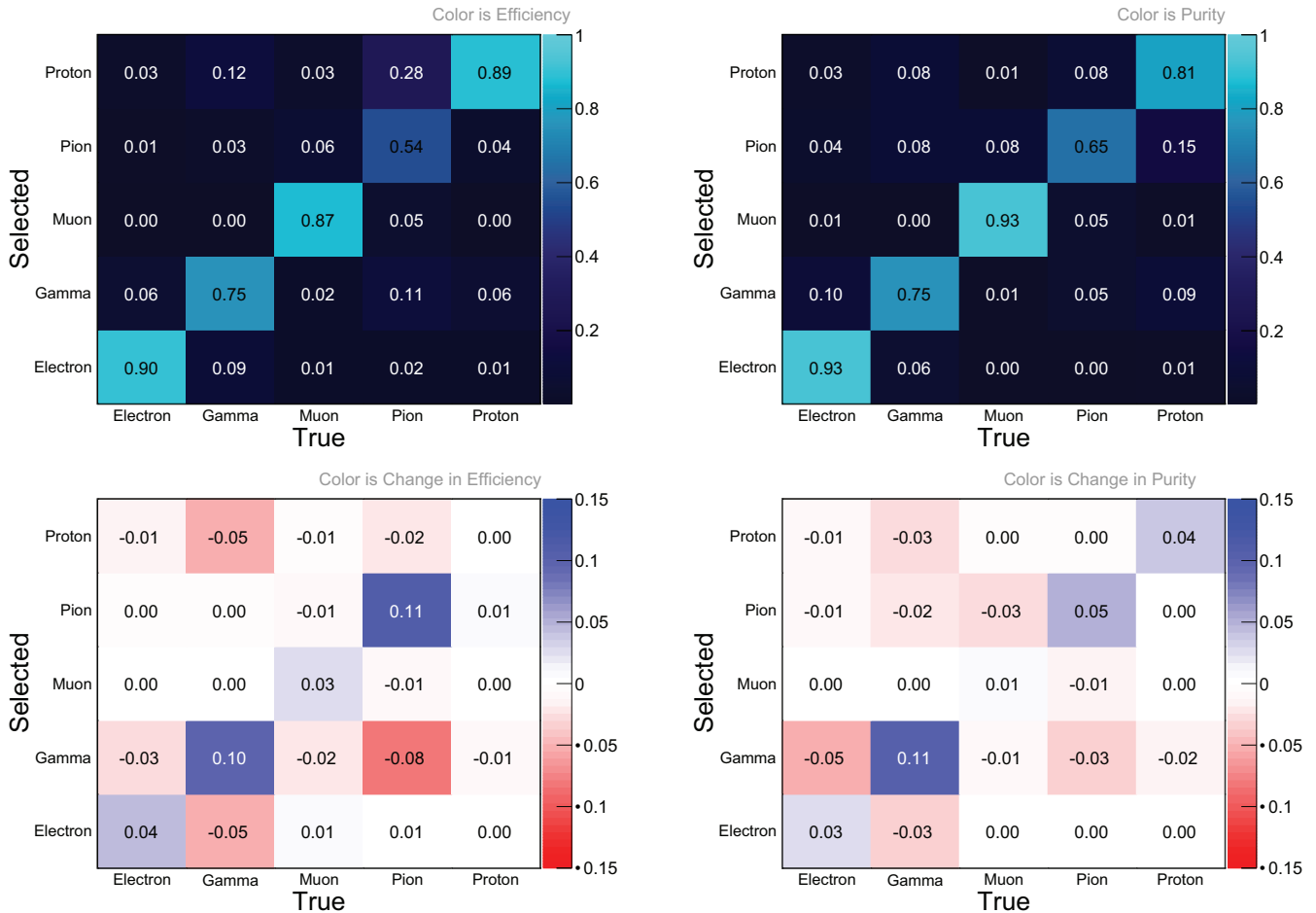


FIG. 6: Classification matrices for the context-enriched network and its performance compared to the particle-only network. Top left: The classification matrix, column-normalized. The diagonal shows the efficiency for each category and the off-diagonal shows how events are misidentified. Top right: The classification matrix row normalized. The diagonal shows the purity for each category and the off-diagonal shows the backgrounds for each type of particle. Bottom: The change in efficiency (left) and purity (right) obtained from the subtraction of the particle only matrices from those shown above. The positive numbers on the diagonal elements show the absolute improvement in efficiency (left) and purity (right) obtained from adding context.

efficiency for the simple selection. The photon-electron contamination improves by 5%, for electrons identified as photons, and 3%, for photons identified as electrons, and the misidentification of pions as photons is reduced by 8% in the presence of context information.

Alternatively to the simple prediction based on the largest softmax value, the network outputs can also be used by considering each of the outputs as a Particle Identification (PID) score and by choosing a selection threshold optimized for a specific figure of merit such as efficiency, purity, etc. Figure 7 shows the efficiencies of PID thresholds for each category. The performance on electrons and muons is superior to that for the other particles. For photons and pions, good separation can be achieved with PID values above 0.7.

The bottom portion of Figure 7 also shows the Receiver Operating Characteristic (ROC) curves for the classifica-

tion of each particle type. ROC curves show the discriminating power of the classifier when taking the softmax output as a PID score and by varying the PID signal selection threshold between 0 to 1. Separation between the curves corresponding to each training scheme indicates a difference in performance in terms of signal vs background efficiency. The classifier which uses context information outperforms the particle-only classifier in every category, though the difference is most significant for photon, pion, and electron identification.

Table I contains a summary of the metrics used to compare the network trained using context and the particle-only network. These results clearly show the improvement that adding context information provides to the task of classification.

More detailed studies are ongoing to assess opportunities for improvement from generation and selection of

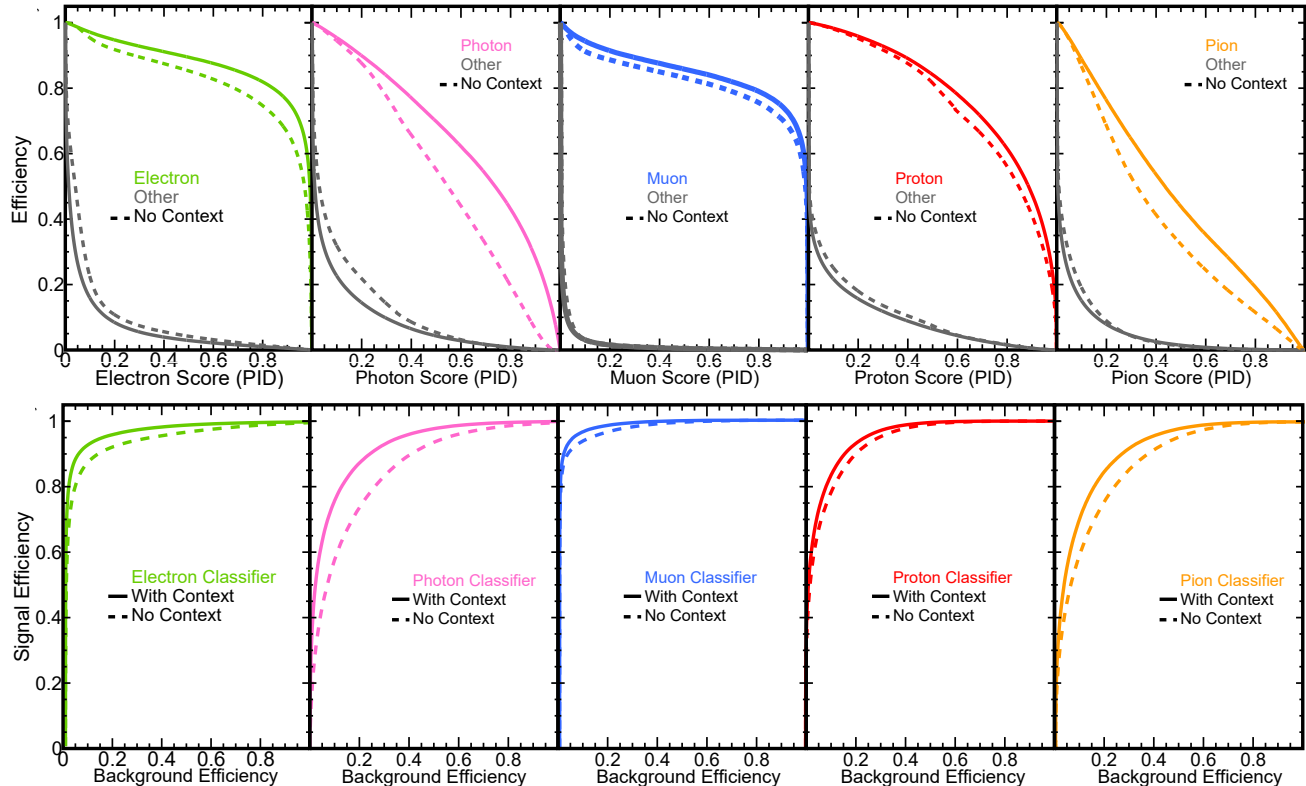


FIG. 7: Selection efficiencies of individual softmax outputs for each training category taken as PID scores. Top: Signal and background selection efficiency vs PID score for each of the five particle classifiers with its respective background. Bottom: Signal vs background efficiency curves as the PID signal threshold is varied. The solid lines depict the efficiencies for the four-tower network utilizing the interaction context. The dashed lines depict the efficiencies for the two-tower network that only utilizes the particle cluster.

Comparison Metric	Network	Electron	Photon	Muon	Pion	Proton
Background Efficiency for 90% Signal Efficiency	Particle & Context	3.2%	14.5%	1.1%	16.1%	9.4%
	Particle Only	8.0%	24.6%	2.2%	22.4%	12.1%
ROC Integral	Particle & Context	0.98	0.95	0.99	0.94	0.97
	Particle Only	0.97	0.91	0.99	0.92	0.96
Largest Score Selection Efficiency	Particle & Context	90%	75%	87%	54%	89%
	Particle Only	86%	74%	84%	43%	89%
Largest Score Selection Purity	Particle & Context	93%	75%	93%	65%	81%
	Particle Only	90%	64%	92%	60%	77%

TABLE I: Summary comparisons of the particle with context and the particle only networks. The largest score selection is the simple softmax selection shown in Figure 6. The ROC integrals and background efficiencies were obtained from Figure 7.

events for the training sample, bias reduction, and treatment of systematic uncertainties. Thus far, we have found no evidence of bias coming from simulated particle kinematics. Studies of detector response have also been implemented using this technique for the selection of photon pairs consistent with coming from neutral pion

decays. The event selection using this technique achieves a purity of 92%, an improvement of 60% for the same efficiency compared to previous methods [27].

This context-enriched CNN technique for particle classification has already been incorporated in NOvA's analyses, specifically in reconstructing the energy of electron

neutrino interactions [27], an essential part of NOvA's primary neutrino oscillation physics measurement. The energy in electron neutrino candidate events is reconstructed from electromagnetic activity by final-state electrons and photons, and from hadronic activity by final-state protons and pions. Each cluster is identified as electromagnetic or hadronic using this classifier. The energy resolution using this method is 11% [11], an improvement of 20% compared to previous methods. This context-enriched CNN technique is being explored further for selection of interaction final states [32] as well as energy reconstruction for muon neutrino candidate events [33].

IV. CONCLUSION

This work describes the first implementation of a deep learning algorithm which introduces interaction context to the problem of particle classification, where the context is introduced through the use of a four-tower, siamese, convolutional neural network architecture. We demonstrate the first application of the particle classifier based on this architecture to neutrino detector data. The overall efficiency and purity of the classifier for all particles in the test sample are 83.3% and 83.5%, respectively. The addition of context improves the performance of the classifier in all metrics, particularly in the categories most expected to benefit from contextual information. Improvements were found in both efficiency and purity for all particle types, in particular an 11% increase in the efficiency for simple selection (largest softmax value) of pions.

Detailed studies are ongoing to assess opportunities for improvement in sample composition, bias reduction, and treatment of systematic uncertainties. In particular,

studies involving tagged data from the upcoming NOvA test-beam program [34], and comparative studies with single particle simulations are being developed. Future studies in applications of CNNs to physics analyses are continuing, with focus on better understanding the impact of the simulated training samples in their performance and for analysis of particle data.

ACKNOWLEDGEMENTS

We thank the NOvA collaboration for the use of its software tools and simulation and for the review of this manuscript. This work was supported by the U.S. Department of Energy and the U.S. National Science Foundation. Fernanda Psihas's contributions were funded in part by CONACyT and the University of Texas system through the ConTex program. NOvA receives additional support from the Department of Science and Technology, India; the European Research Council; the MSMT CR, Czech Republic; the RAS, RMES, and RFBR, Russia; CNPq and FAPEG, Brazil; and the State and University of Minnesota. We are grateful for the contributions of the staff at the Ash River Laboratory, Argonne National Laboratory, and Fermilab. Fermilab is operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. DOE.

We are grateful to the Fermilab Scientific Computing Division for support and maintenance of the GPU cluster which was used in early stages of this work.

This research was supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute, and in part by the Indiana METACyt Initiative. The Indiana METACyt Initiative at IU was also supported in part by Lilly Endowment, Inc.

-
- [1] D. A. Glaser and D. C. Rahm, Characteristics of bubble chambers, *Phys. Rev.* **97**, 474 (1955).
 - [2] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, Machine learning at the energy and intensity frontiers of particle physics, *Nature* **560**, 41 (2018).
 - [3] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, Jet-images - deep learning edition, *JHEP* **07**, 069, arXiv:1511.05190 [hep-ph].
 - [4] 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, arXiv:1604.01444 [hep-ex].
 - [5] M. Paganini, L. de Oliveira, and B. Nachman, CaloGAN : Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks, *Phys. Rev.* **D97**, 014021 (2018), arXiv:1712.10321 [hep-ex].
 - [6] P. Baldi, J. Bian, L. Hertel, and L. Li, Improved Energy Reconstruction in NOvA with Regression Convolutional Neural Networks, (2018), arXiv:1811.04557 [physics.ins-det].
 - [7] R. Acciarri *et al.* (MicroBooNE), Convolutional Neural Networks Applied to Neutrino Events in a Liquid Argon Time Projection Chamber, *JINST* **12** (03), P03011, arXiv:1611.05531 [physics.ins-det].
 - [8] J. Renner *et al.* (NEXT), Background rejection in NEXT using deep neural networks, *JINST* **12** (01), T01004, arXiv:1609.06202 [physics.ins-det].
 - [9] G. N. Perdue *et al.* (MINERvA), Reducing model bias in a deep learning classifier using domain adversarial neural networks in the MINERvA experiment, *JINST* **13** (11), P11020, arXiv:1808.08332 [physics.data-an].
 - [10] C. Adams *et al.* (MicroBooNE), A Deep Neural Network for Pixel-Level Electromagnetic Particle Identification in the MicroBooNE Liquid Argon Time Projection Chamber, (2018), arXiv:1808.07269 [physics.ins-det].
 - [11] M. A. Acero *et al.* (NOvA), New constraints on oscillation parameters from ν_e appearance and ν_μ disappearance

- ance in the NOvA experiment, Phys. Rev. **D98**, 032012 (2018), arXiv:1806.00096 [hep-ex].
- [12] T. Katori and M. Martini, Neutrino-nucleus cross sections for oscillation experiments, J. Phys. **G45**, 013001 (2018), arXiv:1611.07770 [hep-ph].
- [13] J. Bromley, I. Guyon, Y. LeCun, E. Säcker, and R. Shah, Signature verification using a "siamese" time delay neural network, in *Proceedings of the 6th International Conference on Neural Information Processing Systems*, NIPS'93 (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993) pp. 737–744.
- [14] P. Adamson *et al.*, The NuMI Neutrino Beam, Nucl. Instrum. Meth. **A806**, 279 (2016), arXiv:1507.06690 [physics.acc-ph].
- [15] B. Pontecorvo, Neutrino Experiments and the Problem of Conservation of Leptonic Charge, Sov. Phys. JETP **26**, 984 (1968), [Zh. Eksp. Teor. Fiz.53,1717(1967)].
- [16] D. S. Ayres *et al.* (NOvA), The NOvA Technical Design Report 10.2172/935497 (2007).
- [17] H. Bichsel, D. E. Groom, and S. R. Klein, Passage of particles through matter, (2004).
- [18] M. Baird, J. Bian, M. Messier, E. Niner, D. Rocco, and K. Sachdev, Event Reconstruction Techniques in NOvA, in *Proceedings, 21st International Conference on Computing in High Energy and Nuclear Physics (CHEP 2015): Okinawa, Japan, April 13-17, 2015*, Vol. 664 (2015) p. 072035.
- [19] A. Dundar and I. Garcia-Dorado, Context augmentation for convolutional neural networks, CoRR **abs/1712.01653** (2017), arXiv:1712.01653.
- [20] K. D. Tang, M. Paluri, L. Fei-Fei, R. Fergus, and L. D. Bourdev, Improving image classification with location context, CoRR **abs/1505.03873** (2015), arXiv:1505.03873.
- [21] C. Andreopoulos *et al.*, The GENIE Neutrino Monte Carlo Generator, Nucl. Instrum. Meth. **A614**, 87 (2010), arXiv:0905.2517 [hep-ph].
- [22] S. Agostinelli *et al.* (GEANT4), GEANT4: A Simulation toolkit, Nucl. Instrum. Meth. **A506**, 250 (2003).
- [23] A. Aurisano, C. Backhouse, R. Hatcher, N. Mayer, J. Musser, R. Patterson, R. Schroeter, and A. Sousa (NOvA), The NOvA simulation chain, *Proceedings, 21st International Conference on Computing in High Energy and Nuclear Physics (CHEP 2015): Okinawa, Japan, April 13-17, 2015*, J. Phys. Conf. Ser. **664**, 072002 (2015).
- [24] J. Dunn, A fuzzy relative of the isodata process and its use in detecting compact, well-separated clusters, J. Cybern. **3**, 32 (1974).
- [25] E. D. Niner, *Observation of Electron Neutrino Appearance in the NuMI Beam with the NOvA Experiment*, Ph.D. thesis, Indiana U. (2015).
- [26] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going Deeper with Convolutions, (2014), arXiv:1409.4842 [cs.CV].
- [27] F. Psihas, *Measurement of Long Baseline Neutrino Oscillations and Improvements from Deep Learning*, Ph.D. thesis, Indiana U. (2018).
- [28] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, Caffe: Convolutional architecture for fast feature embedding, arXiv preprint arXiv:1408.5093 (2014).
- [29] R. Brun and F. Rademakers, Root - an object oriented data analysis framework, in *AIHENP'96 Workshop, Lausanne*, Vol. 389 (1996) pp. 81–86.
- [30] C. Green, J. Kowalkowski, M. Paterno, M. Fischler, L. Garren, and Q. Lu, The Art Framework, *Proceedings, 19th International Conference on Computing in High Energy and Nuclear Physics (CHEP 2012): New York, USA, May 21-25, 2012*, J. Phys. Conf. Ser. **396**, 022020 (2012).
- [31] L. van der Maaten and G. Hinton, Visualizing high-dimensional data using t-sne, (2008).
- [32] M. MUETHER, Status of the Neutrino-Induced Neutral Current Neutral Pion Production Cross Section Measurement from NOvA (2018).
- [33] M. BAIRD and E. SMITH, Reconstructing Neutrino Energies with the NOvA Detectors (2018).
- [34] K. LANG, The nova test beam program (2018).