Snowmass LOI: Scalable, End-to-End Optimizable Data Reconstruction and Physics Inference Techniques for Large-scale Particle Imaging Neutrino Detectors

C. Adams[1], A. Aurisano[2], R. Berner[3], J. Bian[4], J. M. Conrad[5], M. Del Tutto[6], P. de Perio[7], L. Domine[8], F. Drielsma[8], J. Hewes[2], G. Karagiorgi[9], M. Mooney[10], N. Prouse[7], X. Qian[11], H. A. Tanaka[8], K. Terao*[8], Y.-T. Tsai[8], K. V. Tsang[8], T. Usher[8], Z. Vallari[12], and T. Wongjirad[13]

[1]Argonne National Laboratory
[2]University of Cincinnati
[3]University of Bern
[4]University of California, Irvine
[5]Massachusetts Institute of Technology
[6]Fermi National Accelerator Laboratory
[7]TRIUMF
[8]SLAC National Accelerator Laboratory
[9]Columbia University
[10]Colorado State University
[11]Brookhaven National Laboratory
[12]California Institute of Technology
[13]Tufts University

Thematic Areas: (check all that apply □/■)
■ (NF1) Neutrino Oscillations
■ (NF2) Sterile Neutrinos
■ (NF3) BSM
□ (NF4) Neutrino from natural sources
□ (NF5) Neutrino properties
■ (NF6) Neutrino Interaction Cross Sections
□ (NF7) Applications
□ (NF8) Theory of Neutrino Physics
□ (NF9) Artificial Neutrino Sources
□ (NF10) Neutrino Detectors

Abstract

Particle imaging detectors such as Time Projection Chambers (TPC) and Water Cherenkov (WC) detectors are widely used in the current and future key neutrino programs including Deep Underground Neutrino Experiment (DUNE), Short Baseline Neutrino (SBN) program, Super-Kamiokande, Hyper-Kamiokande, and neutrino-less double beta decay experiments such as NEXT. These detectors face unique challenges including poorly understood nuclear models, limited calibration methods for understanding the detector response, and inefficient inference of neutrino oscillation parameters to name a few. Recent advancements in Computer Vision (CV) and Machine Learning (ML) are directly relevant to these research programs and have been adopted to boost physics output. More adaptation of those techniques, emerging exascale computing facilities, and R&D of domain-specific, scientific ML techniques are expected within the next decade. In this letter, we briefly describe ML and statistical methods to address key challenges and propose a community-wide initiative to support, and guide the R&D program to maximize the impact in experimental neutrino physics. This effort requires building of a collaborative ecosystem within neutrino physics as well as interconnections with a wider scientific research community.

*Corresponding Author: Kazuhiro Terao (kterao@slac.stanford.edu)

Introduction

Machine Learning (ML) techniques are widely used across the field of high energy physics [1]. Recently, methods from Computer Vision (CV) have been applied to data from particle imaging neutrino detectors including Time Projection Chambers (TPCs) and Water Cherenkov (WC) detectors [2, 3, 4, 5, 6, 7, 8]. Despite some successful applications, critical research challenges remain. How can we study nuclear models with ML methods? How do we interpret the output of ML models? What is the most efficient way to process millions of neutrino images with ML models? Can we develop a research ecosystem to incorporate emerging ML techniques and co-develop scientific ML applications with experts outside our domain? Future directions for developing scientific ML methods are discussed in this letter in an attempt to address those common questions across neutrino experiments including Deep Underground Neutrino Experiment (DUNE), Short Baseline Neutrino (SBN) program, Water Cherenkov (WC) detectors in general, and neutrino-less double-beta decay experiments such as NEXT. The letter is relevant for Neutrino Oscillations (NF1), Sterile Neutrinos (NF2), BSM (NF3), and Neutrino Interaction Cross Sections (NF6).

Interpretable ML Models and Data Reconstruction

There are two key elements to an algorithm's interpretability. The first is its architecture in which inductive bias and causal structure (e.g. the laws of physics) may be introduced [9]. The second is to produce a probabilistic output with uncertainty estimation beyond a single value estimation, which is common in typical neural network applications in the field. Both of these are critical topics in the research frontier of scientific ML R&D, which involves experts in both neutrino physics and ML.

An internal causal structure, for instance "a Michel electron must originate from the end of a muon trajectory", naturally leads to the design of a data reconstruction chain in which key physics observables are derived and a hierarchical dependency among them, or causality, is enforced. As such, the development of the data reconstruction chain coincides with highly interpretable, explainable ML models, and will be an important research frontier [10, 11, 12, 13, 14]. For reconstructed parameters, quantification techniques for uncertainties [15, 16] including both an intrinsic to the model and a propagated one from the input parameters will be necessary for full interpretability. Furthermore, a reconstruction chain that provides full details about a neutrino interaction including individual out-going particles with their type and kinematic information will be a critical tool to perform neutrino-nucleus cross-section analyses and improve our knowledge about nuclear models, which is needed for neutrino oscillation analysis.

Gradient-based Automated Optimization

Gradient-based optimization has been essential to the advancement of ML techniques. There are two key research areas with potential high impact on neutrino physics. The first is to make the data reconstruction chain differentiable, which is automatically achieved for a ML-based approach. This allows end-to-end optimization with minimal human intervention. A traditional approach in neutrino physics is to have multiple developers tune individual algorithms, then a software management team combine these to make the whole chain work. This process typically takes months or years and does not necessarily optimize the whole chain. This is mitigated within ML-based approaches where the optimization target and gradients are well defined, such that the whole chain can be optimized together.

The other research frontier is to restructure our simulation, which includes our physics models, using differentiable programming [17] so that we can directly infer physics parameters through gradient-based optimization methods [18, 19, 20]. This gives us the potential to tune detector physics parameters directly from real data, which may be important for detectors such as LArTPCs where calibration methods are limited. Being able to take a gradient also allows propagation of input systematic uncertainties to simulation output. Differentiable programming is expected to play a key role in neutrino physics.

Scalable ML Methods at High Performance Computing (HPC) Clusters

A major bottleneck in the application of machine learning techniques to neutrino detectors is the time taken to train, evaluate, adjust, and retrain models to achieve acceptable performance. Spending weeks to train a model is feasible for cutting edge ML research in industry but is prohibitive for application of these techniques to neutrino detectors. Several experiments and researchers have actively developed scalable techniques for accelerating training time of networks, including sparse convolutional techniques [10, 11, 12], sparse and scalable data distribution mechanisms [21, 22], and scalable model training via HPC systems such as Summit at Oak Ridge. Currently, use of these techniques requires expertise and experience in high performance computing and GPU programming. A particularly impactful development in neutrino detectors and ML is the lowering of the barriers of entry to HPC systems, and highly efficient and easy to use libraries for these techniques. Current experiments have shown training times reduced from ∼2 weeks to under an hour with these techniques [21].

After tne development of a trained model, efficient inference on data and simulation is essential. With the use of HPC systems, these models can efficiently leverage modern co-processors in a highly parallel way, as well as utilize the high bandwidth IO to process datasets in hours instead of months. Additionally, for the application of trained models on more inefficient systems, such as Open Science Grid, lower precision quantization techniques can be used to accelerate inference without loss of precision.

Mitigation of Domain Discrepancies

A notable challenge in neutrino experiments that seek to apply ML to their data is that models are typically trained based on large, high fidelity simulation datasets. While overtraining is easily avoided by generating more simulated data, domain-specific overfitting is a common pitfall that is harder to avoid. In this case, the model learns features of the simulation that are not present in the data. When applied to real data, the model makes incorrect and uninformed decisions.

Several techniques have been proposed to measure the domain discrepancy of a network during training time, and there are several tantalizing techniques in computer vision for mitigating the challenges of physics modeling when training neural networks. For example, an adversarial training method has been explored to mitigate different event generator domain discrepancies [23]. Using statistical tests of a network's intermediate activations, such as energy-distance, can yield a quantitative comparison of which networks will work when trained on simulation and applied to data [24]. Additionally, generative techniques such as Cycle-GAN [25] are extremely promising for learning, and then correcting, deficiencies in physics modeling in simulation.

Public Datasets

Public datasets have been the drivers for multiple leaps in ML methods, including the ImageNet [26] for deep convolutional neural networks and the ShapeNet [27] for graph neural networks. Unfortunately, for the globally sparse, locally dense particle images in TPCs, as well as irregular and graph-like datasets from experiments like IceCube, there is no "industry standard" dataset for benchmarking models.

We propose a new suite of public datasets that moves the boundaries of these and other frontiers by using data samples from high precision particle imaging detectors used in high energy physics experiments. These science programs narrowly focus on specific physics goals and provide unique dataset features that are not available in industrial counterparts. An initiative has been made for simulation of LArTPC detectors [28], which has sparked several development efforts and collaborations across the field. We propose experimental collaborations to work together to produce and maintain such simulation samples. Furthermore, we consider the development of public datasets based on real experimental data an essential ingredient to driving cross-collaboration and fast dissemination of results in this area of ML.

Conclusion

Scientific ML R&D will remain an active area of research in the near future, and the field of neutrino physics continues to lead the development of interpretable, scalable ML techniques with a unique and strong tie to CV. Data reconstruction will enable ML methods to study neutrino-nuclear cross-section physics in full detail, which is critical for accurate neutrino oscillation measurements or the discovery of a hypothetical sterile neutrino. End-to-end pipelines provide automated optimization workflows and minimize or eliminate costly "by-hand" tuning by physicists. Exploring highly parallelizable ML algorithms will allow the neutrino community to utilize HPC facilities and scale our computing power by orders of magnitude. Through public datasets with common benchmark metrics, collaborative development across the community of neutrino physics and beyond should be enabled. We propose these future research directions to be discussed within the Neutrino Frontier of Snowmass 2021.

References

[1] Alexander Radovic, Michael Williams, David Rousseau, Michael Kagan, David Bonacorsi, Alexander Himmel, Adam Aurisano, Kazuhiro Terao, and Taritree Wongjirad. Machine learning at the energy and intensity frontiers of particle physics. *Nature*, 560, August 2018.

[2] 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. *Journal of Instrumentation*, 11(09):P09001–P09001, Sep 2016.

[3] Pierre Baldi, Jianming Bian, Lars Hertel, and Lingge Li. Improved energy reconstruction in NOvA with regression convolutional neural networks. *Physical Review D*, 99(1), Jan 2019.

[4] F. Psihas, E. Niner, M. Groh, R. Murphy, A. Aurisano, A. Himmel, K. Lang, M. D. Messier, A. Radovic, and A. Sousa. Context-enriched identification of particles with a convolutional network for neutrino events. *Physical Review D*, 100(7), Oct 2019.

[5] DUNE Collaboration. Neutrino interaction classification with a convolutional neural network in the DUNE far detector, 2020.

[6] R. Acciarri, C. Adams, R. An, J. Asaadi, M. Auger, L. Bagby, B. Baller, G. Barr, M. Bass, F. Bay, and et al. Convolutional neural networks applied to neutrino events in a liquid argon time projection chamber. *Journal of Instrumentation*, 12(03):P03011–P03011, Mar 2017.

[7] C. Adams, M. Alrashed, R. An, J. Anthony, J. Asaadi, A. Ashkenazi, M. Auger, S. Balasubramanian, B. Baller, C. Barnes, and et al. Deep neural network for pixel-level electromagnetic particle identification in the MicroBooNE liquid argon time projection chamber. *Physical Review D*, 99(9), May 2019.

[8] NEXT Collaboration. Background rejection in NEXT using deep neural networks. *Journal of Instrumentation*, 12(01):T01004–T01004, jan 2017.

[9] Peter W. Battaglia et al. Relational inductive biases, deep learning, and graph networks. *CoRR*, abs/1806.01261, 2018.

[10] Laura Dominé and Kazuhiro Terao. Scalable deep convolutional neural networks for sparse, locally dense liquid argon time projection chamber data. *Phys. Rev. D*, 102:012005, Jul 2020.

[11] Laura Dominé, Pierre Côte de Soux, François Drielsma, Dae Heun Koh, Ran Itay, Qing Lin, Kazuhiro Terao, Ka Vang Tsang, and Tracy L. Usher. Point proposal network for reconstructing 3d particle endpoints with sub-pixel precision in liquid argon time projection chambers, 2020.

[12] Dae Heun Koh, Pierre Côte de Soux, Laura Dominé, François Drielsma, Ran Itay, Qing Lin, Kazuhiro Terao, Ka Vang Tsang, and Tracy Usher. Scalable, proposal-free instance segmentation network for 3d pixel clustering and particle trajectory reconstruction in liquid argon time projection chambers, 2020.

[13] Francois Drielsma, Qing Lin, Pierre Côte de Soux, Laura Dominé, Ran Itay, Dae Heun Koh, Bradley J. Nelson, Kazuhiro Terao, Ka Vang Tsang, and Tracy L. Usher. Clustering of electromagnetic showers and particle interactions with graph neural networks in liquid argon time projection chambers data, 2020.

[14] Xiangyang Ju, Steven Farrell, Paolo Calafiura, Daniel Murnane, Prabhat, Lindsey Gray, Thomas Klijnsma, Kevin Pedro, Giuseppe Cerati, Jim Kowalkowski, Gabriel Perdue, Panagiotis Spentzouris, Nhan Tran, Jean-Roch Vlimant, Alexander Zlokapa, Joosep Pata, Maria Spiropulu, Sitong An, Adam Aurisano, Jeremy Hewes, Aristeidis Tsaris, Kazuhiro Terao, and Tracy Usher. Graph neural networks for particle reconstruction in high energy physics detectors, 2020.

[15] Lingge Li, Nitish Nayak, Jianming Bian, and Pierre Baldi. Efficient neutrino oscillation parameter inference using Gaussian processes. *Phys. Rev. D*, 101:012001, Jan 2020.

[16] Sebastian Pina-Otey, Federico Sánchez, Vicens Gaitan, and Thorsten Lux. Likelihood-free inference of experimental neutrino oscillations using neural spline flows. *Physical Review D*, 101(11), Jun 2020.

[17] Atilim Gunes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research*, 18:1–43, 2018.

[18] Sergey Shirobokov, Vladislav Belavin, Michael Kagan, Andrey Ustyuzhanin, and Atılım Güneş Baydin. Black-box optimization with local generative surrogates, 2020.

[19] Filipe de Avila Belbute-Peres, Kevin A. Smith, Kelsey Allen, Joshua B. Tenenbaum, and Zico Kolter. End-to-end differentiable physics for learning and control. 2018.

[20] Corey Adams, Giuseppe Carleo, Alessandro Lovato, and Noemi Rocco. Variational Monte Carlo calculations of $\mathbf{A \leq 4}$ nuclei with an artificial neural-network correlator ansatz, 2020.

[21] Marco Del Tutto Corey Adams. Scalable, Distributed Machine Learning.

[22] Javier Duarte, Philip Harris, Scott Hauck, Burt Holzman, Shih-Chieh Hsu, Sergo Jindariani, Suffian Khan, Benjamin Kreis, Brian Lee, Mia Liu, and et al. FPGA-Accelerated Machine Learning Inference as a Service for Particle Physics Computing. *Computing and Software for Big Science*, 3(1), Oct 2019.

[23] 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, and et al. 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.

[24] Marco Del Tutto Marija Kekic, Corey Adams. Demonstration of background rejection using deep neural networks in the NEXT experiment.

[25] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2017.

[26] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[27] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An Information-Rich 3D Model Repository, 2015.

[28] Corey Adams, Kazuhiro Terao, and Taritree Wongjirad. PILArNet: Public Dataset for Particle Imaging Liquid Argon Detectors in High Energy Physics, 2020.