Cross-Domain Transfer with Particle Physics Foundation Models: From Jets to Neutrino Interactions

Gregor Krzmanc[1*] **Vinicius Mikuni**[2] **Benjamin Nachman**[1,3] **Callum Wilkinson**[4]
[1]Stanford University [2]Nagoya University
[3]SLAC National Accelerator Laboratory [4]Lawrence Berkeley National Laboratory
*gregork@stanford.edu

Abstract

Future AI-based studies in particle physics will likely start from a foundation model to accelerate training and enhance sensitivity. As a step towards a general-purpose foundation model for particle physics, we investigate whether the OmniLearned foundation model pre-trained on diverse high-Q^2 simulated and real pp and ep collisions can be effectively transferred to a few-GeV fixed-target neutrino experiment. We process MINERvA neutrino–nucleus scattering events and evaluate pre-trained models on two types of tasks: regression of available energy and binary classification of charged-current pion final states ($\mathrm{CC1}\pi^{\pm}$, $\mathrm{CCN}\pi^{\pm}$, and $\mathrm{CC1}\pi^{0}$). At the 3M-parameter scale, pre-trained OmniLearned-small consistently outperforms similarly sized scratch-trained models at matched compute budget and training steps. These results suggest that particle-level foundation models acquire inductive biases that generalize across energy scale, detector technology, and underlying physics, pointing toward a paradigm of detector-agnostic inference in particle physics.

1 Introduction

Foundation models are revolutionizing the way we do particle physics, with a growing number being proposed to directly operate on particle physics data [1–13]. One foundation model that has been demonstrated to solve challenges across experiments is OmniLearned [3, 9, 10], which is pre-trained on particle-level jet representations from simulated and real pp and ep collider events.

Beyond collider physics, OmniLearned has shown transferability to a variety of tasks. Near transfers have been studied over data fidelities (fast/full simulation), between experiments, and across collision systems (pp and ep). Far transfers have been studied with cosmological structure formation [14] and molecular dynamics [15]. This work explores a "medium" transfer from high-energy colliders (TeV scale, $\mathcal{O}(10^2)$ particles per event, hermetic detectors) to a few-GeV fixed-target neutrino experiment ($\mathcal{O}(1\text{–}10)$ reconstructed objects per event, asymmetric geometry). Collider jets are QCD-dominated [16]; few-GeV neutrino–nucleus scattering spans multiple regimes and nuclear effects [17–24].

A major, multi-decade, experimental effort is underway focused on making precision neutrino oscillation measurements using few-GeV accelerator neutrino beams [25, 26]. A key challenge for oscillation experiments is to disentangle the neutrino oscillation phenomenon — which varies as a function of neutrino energy and distance traveled — from uncertainties in both the energy distribution of neutrinos produced by the accelerator neutrino beam and the rate at which neutrinos interact. Those uncertainties are large for current accelerator neutrino experiments [27–31] and may be limiting for next-generation experiments such as DUNE [32] and Hyper-Kamiokande [33].

The MINERvA experiment at Fermilab was specifically designed to address this challenge [34, 35], using a finely segmented scintillator tracking calorimeter and a variety of nuclear targets. MINERvA took data in the NuMI accelerator neutrino beam [36] in a variety of configurations [37, 38]. MINERvA has published extensive few-GeV neutrino–nucleus cross-section measurements [39–50] and has released data, simulation, and analysis tools publicly, which enabled this work.

This paper is organized as follows. Section 2 describes the regression and classification tasks that we use for studying transferability of OmniLearned to neutrino data from MINERvA. The machine learning models, including our baselines, are introduced in Sec. 3. Numerical results are presented in Sec. 4. The paper ends with conclusions and outlook in Sec. 5.

2 Tasks

We evaluate models and transfer learning on one regression task and three binary classification tasks:

- **Available hadronic energy regression.** To constrain low-energy nuclear models, we regress event-level *available hadronic energy*, $E_{\mathrm{available}}$. We define it as proton kinetic energy[1] plus total energy of photons, pions, electrons, and charged kaons; other mesons and baryons are neglected. Excluding poorly modeled neutrons and nuclear binding effects makes this quantity more reliably inferable from detector observables.
- **Binary event classification.** We define three positive classes, each a separate task: $\mathrm{CC1}\pi^{\pm}$, $\mathrm{CCN}\pi^{\pm}$, and $\mathrm{CC1}\pi^{0}$. These correspond to charged-current (CC) events with exactly one charged pion ($\pi^{\pm}$), $N \geq 1$ charged pions, and exactly one neutral pion (π^{0}) with no charged pions. For each task, the negative class is all remaining events. All three are derived from a single multi-class model.

These tasks capture key few-GeV reconstruction challenges relevant to both oscillation and cross-section analyses. Crucially, they are defined using *visible* quantities that an ideal detector could reconstruct, reducing explicit dependence on the simulation input model. Available hadronic energy is the closest measurable proxy to true energy transfer to the nucleus and contributes to oscillation energy estimators (with reconstructed leptonic energy). Classifying events into different final states also helps disentangle contributions from interaction mechanisms and nuclear effects.

3 Methods

Dataset. We use simulated MINERvA Medium Energy Forward Horn Current (FHC) events from the standard Monte Carlo playlist 1A (both helium and water targets empty). Events are preprocessed into nested per-event tensors with truth labels and global features. The data are split into training, validation, and test sets, stratified by interaction type. The training dataset 1A contains 6M events, while the validation and test sets each contain 700k events.

Event representation. Each event is a variable-length sequence of *tokens*, one per reconstructed object: up to two photons, up to one muon, and variable numbers of directional *prongs* and calorimetric *blobs*. Continuous inputs follow the convention of inputs used during pre-training of OmniLearned models [10]: beam-axis (η, ϕ), $\log p_{\mathrm{T}}/\mathrm{GeV}$, and $\log E/\mathrm{GeV}$, and a discrete particle type. Each token has 5 further continuous features, and each event has 15 global continuous features summarizing the event-level properties.

Models. Particle tokens (and global event-level feature vectors) are processed with a transformer encoder. We compare the following model families:

- **MLP (global features):** A lightweight baseline ($\sim$284k parameters) with hidden width 128, 4 residual blocks, and sigmoid linear unit activation, operating only on the 15 global event-level features. This serves as a proxy for classical event reconstruction methods since only physics-engineered features are used.
- **Transformer:** A ViT-style [51] point-cloud transformer with additive positional encoding from (η, ϕ), evaluated at xsmall (800k parameters) and small (3M parameters) scales.

[1] Including antiprotons for completeness, though they are vanishingly rare at few-GeV energies.

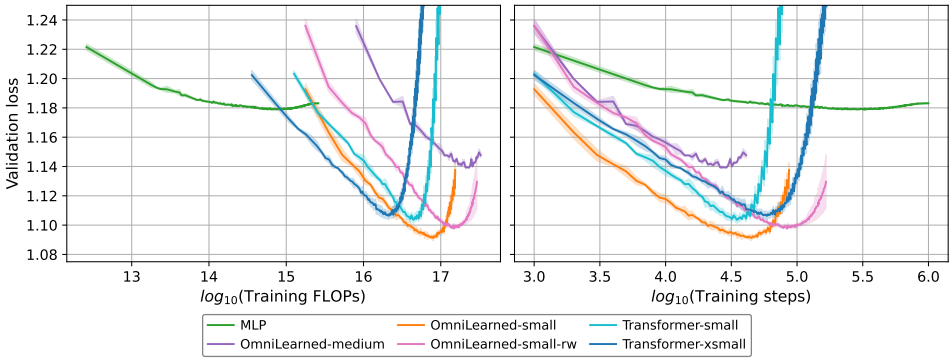

Figure 1: Classification validation loss on playlist 1A as a function of cumulative training FLOPs (left) and training steps (right), with logarithmic horizontal axes. Pre-trained OmniLearned-small improves faster and to a lower plateau than the same-size models trained from scratch.

- **OmniLearned:** The Particle Encoder Transformer v2 (PET2) [10] at small (3M parameters) and medium (53M parameters) presets. We use publicly available pre-trained checkpoints. For the small preset, we additionally train a randomly initialized variant (OmniLearned-small-rw) to isolate the effect of pre-training from architecture differences. For the medium preset, we freeze the backbone and train only the task-specific heads due to compute constraints; consequently, we treat medium results as exploratory and draw no conclusions about scaling from them.

Training. We train separate classification and regression models with Adam [52] (learning rate 10^{-4}, batch size 2048). Regression minimizes Smooth L1 [53] on $\log(1 + E_{\text{available}})$; classification minimizes weighted cross-entropy (class weights $\propto$ inverse frequency) with five classes ($CC1\pi^{\pm}$, $CC1\pi^{0}$, $CCN\pi^{\pm}$ with $N > 1$, other charged-current, neutral-current), from which task-specific signal/background labels are derived. We checkpoint every 1000 steps and keep the best validation loss; uncertainties and averages are computed over 4 independent runs.

4 Results

4.1 Computational efficiency

Figure 1 shows classification validation loss versus cumulative training FLOPs and versus optimizer steps on MINERvA Open Data playlist 1A. Pre-trained OmniLearned-small reaches lower loss than the similarly sized Transformer-small and randomly initialized OmniLearned-small-rw baselines at matched compute and at matched number of training steps.

4.2 Classification performance

Figure 2 summarizes tagging performance for $\mathrm{CC1}\pi^{\pm}$, $\mathrm{CC1}\pi^{0}$, and $\mathrm{CCN}\pi^{\pm}$ as true positive rate (TPR) at the false positive rate of a classical cut-based baseline, in bins of physically motivated kinematics: true pion energy and polar angle for the single-pion channels, and hadronic invariant mass W for $\mathrm{CCN}\pi^{\pm}$. Pre-trained models perform marginally better than the non-pretrained models over most bins, with the largest gains at low pion energy for $\mathrm{CC1}\pi^{\pm}$ and a similar trend for $\mathrm{CC1}\pi^{0}$ (where $\pi^{0} \to \gamma\gamma$ reconstruction is hardest). For $\mathrm{CCN}\pi^{\pm}$, gains are clearest at low W; at higher W all models can fall below the cut-based baseline, which we leave to future work.

4.3 Regression performance

Figure 3 summarizes available hadronic energy residuals on playlist 1A vs. three-momentum transfer q_3, using the interquartile range (IQR) of the residual distribution and the most probable value (MPV) of the reconstructed-to-true energy ratio. Pre-trained OmniLearned models achieve narrower IQR across q_3, with the clearest separation at low q_3 where nuclear effects are largest; MPV drifts slightly below unity at higher q_3 for all models, suggesting room to adjust the regression target or loss.

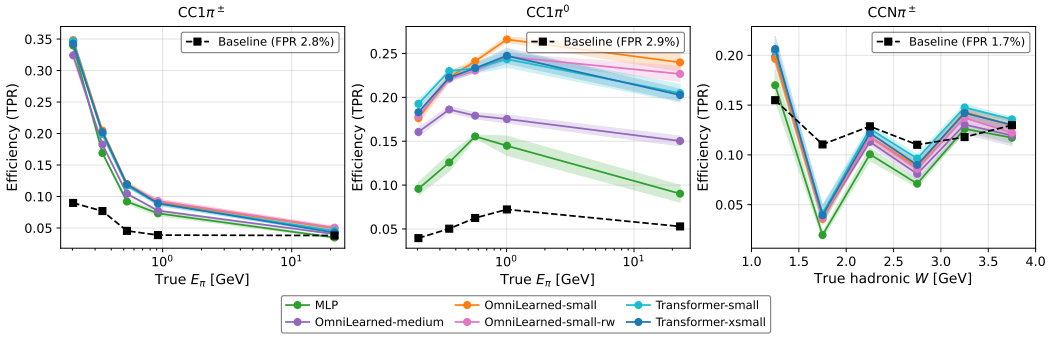

Figure 2: TPR at the baseline FPR on playlist 1A. Single-pion channels: bins in true pion energy and angle; negatives are the full background (all other events; fixed across bins). CCN$\pi^{\pm}$: bins in invariant mass of the hadronic system W.

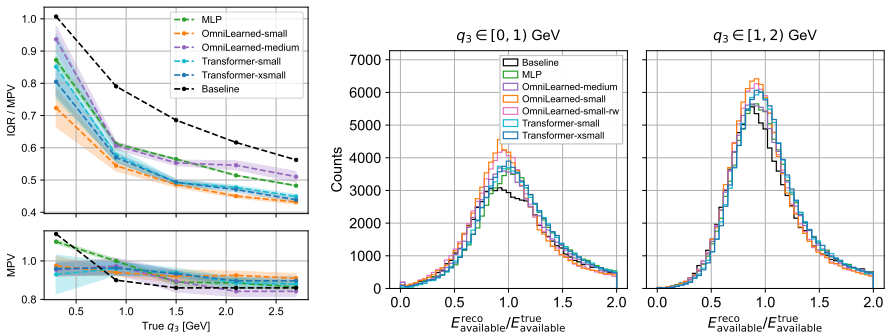

Figure 3: Available energy ($E_{available}$) reconstruction performance on playlist 1A: Interquartile range (IQR) and most probable value (MPV) of $E^{\text{reco}}_{\text{avail.}}/E^{\text{true}}_{\text{avail.}}$ for bins of q_3 (left) and distribution of $E^{\text{reco}}_{\text{avail.}}/E^{\text{true}}_{\text{avail.}}$ for two bins of q_3 (right).

5 Conclusions

We have shown that inductive biases learned during pre-training on collider physics jets transfer to the qualitatively different setting of a few-GeV fixed-target neutrino experiment. At the 3M-parameter scale, pre-trained OmniLearned-small outperforms comparably sized models trained from scratch, both in terms of final performance and compute efficiency. Despite a large domain shift—energy scale (TeV vs. few GeV), detector geometry and granularity, event multiplicity, and dominant physics (QCD jets vs. low-energy nuclear effects in weak interactions)—particle-token representations still carry useful signal. We interpret this as evidence that pre-training instills a useful geometric and kinematic prior — a bias toward attending to the relative positions and energies of sparse point clouds — that is broadly applicable across particle physics domains.

This supports using shared foundation models as a way to cut labeled-simulation demand when adapting to new experiments, i.e. a step toward *detector-agnostic inference* where the same backbone is fine-tuned with modest task-specific data.

Future work includes end-to-end fine-tuning of larger backbones (OmniLearned-medium was frozen here, likely limiting its performance, so we draw no scaling conclusions from those results) and pre-training mixtures that explicitly include neutrino-domain events, which may further help MINERvA-like settings and other detectors (e.g. MicroBooNE, DUNE).

Code Availability and Compute Requirements

Code: `https://github.com/gregorkrz/minerva-ml`. Data: `https://minerva.fnal.gov/opendata/`. Training one classification or regression run takes ∼6–8 h (OmniLearned-small) or ∼1–2 h (Transformer) on one NVIDIA A100 GPU.

References

[1] M. Feickert and B. Nachman, "A Living Review of Machine Learning for Particle Physics", arXiv:2102.02770.

[2] J. Birk, A. Hallin, and G. Kasieczka, "OmniJet-α: The first cross-task foundation model for particle physics", arXiv:2403.05618.

[3] V. Mikuni and B. Nachman, "Solving key challenges in collider physics with foundation models", *Phys. Rev. D* **111** (2025), no. 5, L051504, doi:10.1103/PhysRevD.111.L051504, arXiv:2404.16091.

[4] A. Hallin, "Foundation models for high-energy physics", in *2nd European AI for Fundamental Physics Conference*. 9, 2025. arXiv:2509.21434.

[5] P. Harris et al., "Re-Simulation-based Self-Supervised Learning for Pre-Training Foundation Models", arXiv:2403.07066.

[6] T. Golling et al., "Masked particle modeling on sets: towards self-supervised high energy physics foundation models", *Mach. Learn. Sci. Tech.* **5** (2024), no. 3, 035074, doi:10.1088/2632-2153/ad64a8, arXiv:2401.13537.

[7] M. Leigh et al., "Is Tokenization Needed for Masked Particle Modelling?", arXiv:2409.12589.

[8] J. Bardhan et al., "HEP-JEPA: A foundation model for collider physics using joint embedding predictive architecture", arXiv:2502.03933.

[9] V. Mikuni and B. Nachman, "Method to simultaneously facilitate all jet physics tasks", *Phys. Rev. D* **111** (2025), no. 5, 054015, doi:10.1103/PhysRevD.111.054015, arXiv:2502.14652.

[10] W. Bhimji, C. Harris, V. Mikuni, and B. Nachman, "Foundation model framework for all tasks involving jet physics", *Physical Review D* **113** (February, 2026) doi:10.1103/knmd-f5jm.

[11] D. Park et al., "FM4NPP: A Scaling Foundation Model for Nuclear and Particle Physics", arXiv:2508.14087.

[12] S. Young, Y.-j. Jwa, and K. Terao, "Particle trajectory representation learning with masked point modeling", *Mach. Learn. Sci. Tech.* **7** (2026), no. 2, 025023, doi:10.1088/2632-2153/ae47b8, arXiv:2502.02558.

[13] S. Young and K. Terao, "Panda: Self-distillation of Reusable Sensor-level Representations for High Energy Physics", arXiv:2512.01324.

[14] V. Mikuni, I. Elsharkawy, and B. Nachman, "Omnicosmos: Transferring particle physics knowledge across the cosmos", 2025. https://arxiv.org/abs/2512.24422.

[15] I. Elsharkawy, V. Mikuni, W. Bhimji, and B. Nachman, "Omnimol: Transferring particle physics knowledge to molecular dynamics with point-edge transformers", 2026. https://arxiv.org/abs/2601.10791.

[16] R. Horgan and M. Jacob, "Jet production at collider energy", *Nuclear Physics B* **179** (March, 1981) 441–460, doi:10.1016/0550-3213(81)90013-4.

[17] J. A. Formaggio and G. P. Zeller, "From eV to EeV: Neutrino Cross Sections Across Energy Scales", *Rev. Mod. Phys.* **84** (2012) 1307–1341, doi:10.1103/RevModPhys.84.1307, arXiv:1305.7513.

[18] L. Alvarez-Ruso, Y. Hayato, and J. Nieves, "Progress and open questions in the physics of neutrino cross sections at intermediate energies", *New J. Phys.* **16** (2014) 075015, doi:10.1088/1367-2630/16/7/075015, arXiv:1403.2673.

[19] U. Mosel, "Neutrino Interactions with Nucleons and Nuclei: Importance for Long-Baseline Experiments", *Ann. Rev. Nucl. Part. Sci.* **66** (2016) 171–195, doi:10.1146/annurev-nucl-102115-044720, arXiv:1602.00696.

[20] T. Katori and M. Martini, "Neutrino–nucleus cross sections for oscillation experiments", *J. Phys.* **G45** (2018), no. 1, 013001, doi:10.1088/1361-6471/aa8bf7, arXiv:1611.07770.

[21] NuSTEC Collaboration, "NuSTEC White Paper: Status and challenges of neutrino–nucleus scattering", *Prog. Part. Nucl. Phys.* **100** (2018) 1–68, doi:10.1016/j.ppnp.2018.01.006, arXiv:1706.03621.

[22] K. Mahn, C. Marshall, and C. Wilkinson, "Progress in Measurements of 0.1–10 GeV Neutrino–Nucleus Scattering and Anticipated Results from Future Experiments", *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 105–129, doi:10.1146/annurev-nucl-101917-020930, arXiv:1803.08848.

[23] N. Jachowicz and A. Nikolakopoulos, "Nuclear medium effects in neutrino- and antineutrino-nucleus scattering", *Eur. Phys. J. ST* **230** (2021), no. 24, 4339–4356, doi:10.1140/epjs/s11734-021-00286-8, arXiv:2110.11321.

[24] M. Sajjad Athar, A. Fatima, and S. K. Singh, "Neutrinos and their interactions with matter", *Prog. Part. Nucl. Phys.* **129** (2023) 104019, doi:10.1016/j.ppnp.2022.104019, arXiv:2206.13792.

[25] I. Esteban et al., "NuFit-6.0: updated global analysis of three-flavor neutrino oscillations", *JHEP* **12** (2024) 216, doi:10.1007/JHEP12(2024)216, arXiv:2410.05380.

[26] Particle Data Group Collaboration, "Review of particle physics", *Phys. Rev. D* **110** (2024), no. 3, 030001, doi:10.1103/PhysRevD.110.030001.

[27] T2K Collaboration, "The T2K Experiment", *Nucl. Instrum. Meth. A* **659** (2011) 106–135, doi:10.1016/j.nima.2011.06.067, arXiv:1106.1238.

[28] T2K Collaboration, "Results from the t2k experiment on neutrino mixing including a new far detector μ-like sample", *Phys.Rev.Lett.* **135** (2025) 261801, arXiv:2506.05889.

[29] NOvA Collaboration, "The NOvA Technical Design Report", *FERMILAB-DESIGN-2007-01* (10, 2007) doi:10.2172/935497.

[30] NOvA Collaboration, "Precision measurement of neutrino oscillation parameters with 10 years of data from the nova experiment", *Phys.Rev.Lett.* **136** (2026) 011802, arXiv:2509.04361.

[31] T2K, NOvA, R. Group Collaboration, "Joint neutrino oscillation analysis from the t2k and nova experiments", *Nature* **646** (2025) 818–824, doi:10.1038/s41586-025-09599-3, arXiv:2510.19888.

[32] DUNE Collaboration, "Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I Introduction to DUNE", *JINST* **15** (2020), no. 08, T08008, doi:10.1088/1748-0221/15/08/T08008, arXiv:2002.02967.

[33] Hyper-Kamiokande Collaboration, "Hyper-Kamiokande Design Report", arXiv:1805.04163.

[34] K. McFarland, "Minerva: a dedicated neutrino scattering experiment at numi", *Nuclear Physics B - Proceedings Supplements* **159** (September, 2006) 107–112, doi:10.1016/j.nuclphysbps.2006.08.073.

[35] MINERvA Collaboration, "Design, Calibration, and Performance of the MINERvA Detector", *Nucl. Instrum. Meth. A* **743** (2014) 130–159, doi:10.1016/j.nima.2013.12.053, arXiv:1305.5199.

[36] P. Adamson et al., "The NuMI Neutrino Beam", *Nucl. Instrum. Meth.* **A806** (2016) 279–306, doi:10.1016/j.nima.2015.08.063, arXiv:1507.06690.

[37] MINERvA Collaboration, "Neutrino Flux Predictions for the NuMI Beam", *Phys. Rev. D* **94** (2016), no. 9, 092005, doi:10.1103/PhysRevD.94.092005, arXiv:1607.00704. [Addendum: Phys.Rev.D 95, 039903 (2017)].

[38] MINERvA Collaboration, "Improved constraint on the MINERνA medium energy neutrino flux using ν^{-}e-$\rightarrow\nu^{-}$e- data", *Phys. Rev. D* **107** (2023), no. 1, 012001, doi:10.1103/PhysRevD.107.012001, arXiv:2209.05540.

[39] MINERvA Collaboration, "Exploring neutrino–nucleus interactions in the GeV regime using MINERvA", *Eur. Phys. J. ST* **230** (2021), no. 24, 4243–4257, doi:10.1140/epjs/s11734-021-00296-6, arXiv:2107.02064.

[40] MINERvA Collaboration, "Measurement of inclusive charged-current ν_μ scattering on hydrocarbon at $\langle E_\nu \rangle \sim 6$ GeV with low three-momentum transfer", *Phys. Rev. D* **106** (2022), no. 3, 032001, doi:10.1103/PhysRevD.106.032001, arXiv:2110.13372.

[41] MINERvA Collaboration, "Simultaneous Measurement of Proton and Lepton Kinematics in Quasielasticlike $\nu\mu$-Hydrocarbon Interactions from 2 to 20 GeV", *Phys. Rev. Lett.* **129** (2022), no. 2, 021803, doi:10.1103/PhysRevLett.129.021803, arXiv:2203.08022.

[42] MINERvA Collaboration, "Simultaneous Measurement of Muon Neutrino $\nu\mu$ Charged-Current Single π+ Production in CH, C, H2O, Fe, and Pb Targets in MINERvA", *Phys. Rev. Lett.* **131** (2023), no. 1, 011801, doi:10.1103/PhysRevLett.131.011801, arXiv:2209.07852.

[43] MINERvA Collaboration, "Neutrino-Induced Coherent π+ Production in C, CH, Fe, and Pb at $\langle$E$\nu\rangle\sim$6 GeV", *Phys. Rev. Lett.* **131** (2023), no. 5, 051801, doi:10.1103/PhysRevLett.131.051801, arXiv:2210.01285.

[44] MINERvA Collaboration, "High-Statistics Measurement of Antineutrino Quasielastic-like scattering at $E_\nu \sim$ 6~GeV on a Hydrocarbon Target", *Phys. Rev. D* **108** (2023), no. 3, 032018, doi:10.1103/PhysRevD.108.032018, arXiv:2211.10402.

[45] MINERvA Collaboration, "Simultaneous Measurement of $\nu\mu$ Quasielasticlike Cross Sections on CH, C, H2O, Fe, and Pb as a Function of Muon Kinematics at MINERvA", *Phys. Rev. Lett.* **130** (2023), no. 16, 161801, doi:10.1103/PhysRevLett.130.161801, arXiv:2301.02272.

[46] MINERvA Collaboration, "Measurement of the multineutron $\nu^{-}\mu$ charged current differential cross section at low available energy on hydrocarbon", *Phys. Rev. D* **108** (2023), no. 11, 112010, doi:10.1103/PhysRevD.108.112010, arXiv:2310.17014.

[47] MINERvA Collaboration, "Measurement of electron neutrino and antineutrino cross sections at low momentum transfer", *Phys.Rev.D* **109** (2024) 092008, arXiv:2312.16631.

[48] MINERvA Collaboration, "Measurement of the A dependence of the $\nu\mu$ charged-current quasielasticlike cross section as a function of muon and proton kinematics at $\langle$E$\nu\rangle\sim$6 GeV", *Phys. Rev. D* **112** (2025), no. 5, 052005, doi:10.1103/h9vz-x4xf, arXiv:2503.15047.

[49] MINERvA Collaboration, "Measurement of charged-current ν_μ and $\bar{\nu}_\mu$ cross sections on hydrocarbon in a shallow inelastic scattering region", arXiv:2503.20043.

[50] MINERvA Collaboration, "Measurement of Inclusive Charged-Current $\bar{\nu}_\mu$ Scattering on C, CH, Fe, and Pb at $\langle E_{\bar{\nu}} \rangle \sim 6$ GeV with MINERvA", arXiv:2604.07091.

[51] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale", 2021. https://arxiv.org/abs/2010.11929.

[52] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", 2017. https://arxiv.org/abs/1412.6980.

[53] R. Girshick, "Fast r-cnn", doi:10.48550/arXiv.1504.08083. arXiv:1504.08083.