Gregor Krzmanc
PhD Student
Department of Physics, Stanford University
Advised by Benjamin Nachman and Kazuhiro Terao
gregork@stanford.edu
gregor.science

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

Gregor Krzmanc, Vinicius Mikuni, Benjamin Nachman, Callum Wilkinson

Keywords
Foundation models, Pretraining, Transfer Learning, Neutrino-nucleus Interactions

Overview

We investigate whether foundation models could be used to speed up training and enhance sensitivity for particle physics studies across detectors and energy scales. We apply pre-trained and randomly-initialized transformer-based models for classification and regression tasks on MINERvA Open Data.

Key Finding

Our results suggest that particle-level foundation models pretrained on collider physics datasets acquire inductive biases that generalize across energy scale, detector technology, and underlying physics.

Setup

1. Preprocess MINERvA Experiment Open Data in the same format as used for training particle physics foundation models: Represent each reconstructed object (muons, photons, prongs, blobs) as a token with kinematic features (energy, momentum) + additional object-dependent features (dE/dx, vertex position and timing, PID)

2. Define classification and regression benchmark tasks relevant to constraining neutrino-nucleus interaction models

3. Compare training dynamics of pretrained vs. randomly-initialized models

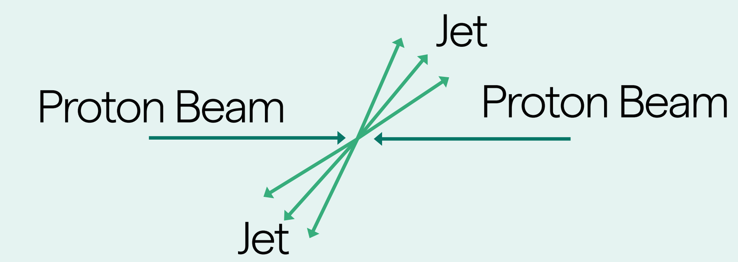
Stanford University Human-Centered Artificial Intelligence

Stanford Data Science

Datasets

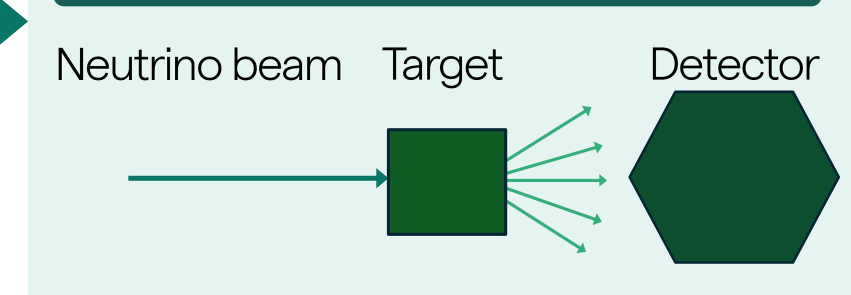
Pretraining

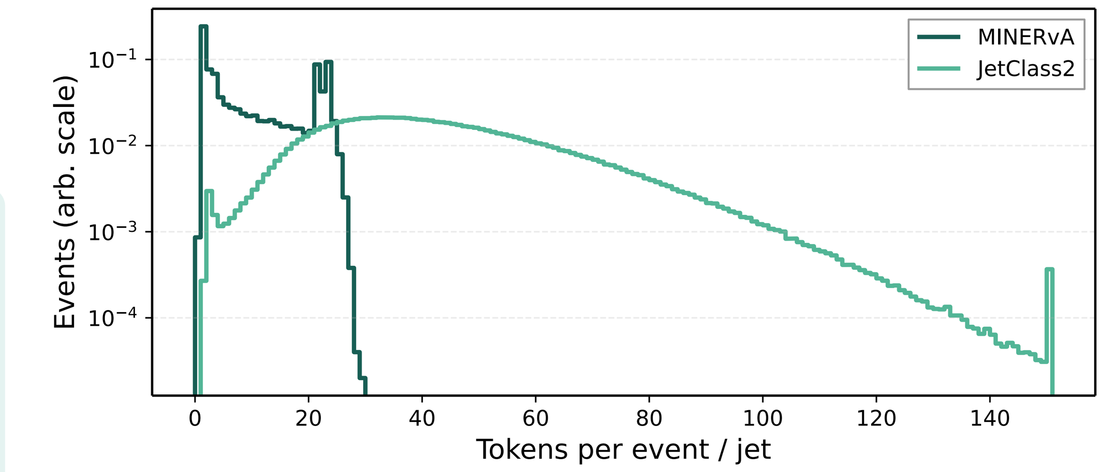
- 100 GeV to TeV jets
- O(50) tokens (particle candidates) per jet
- 1B jets

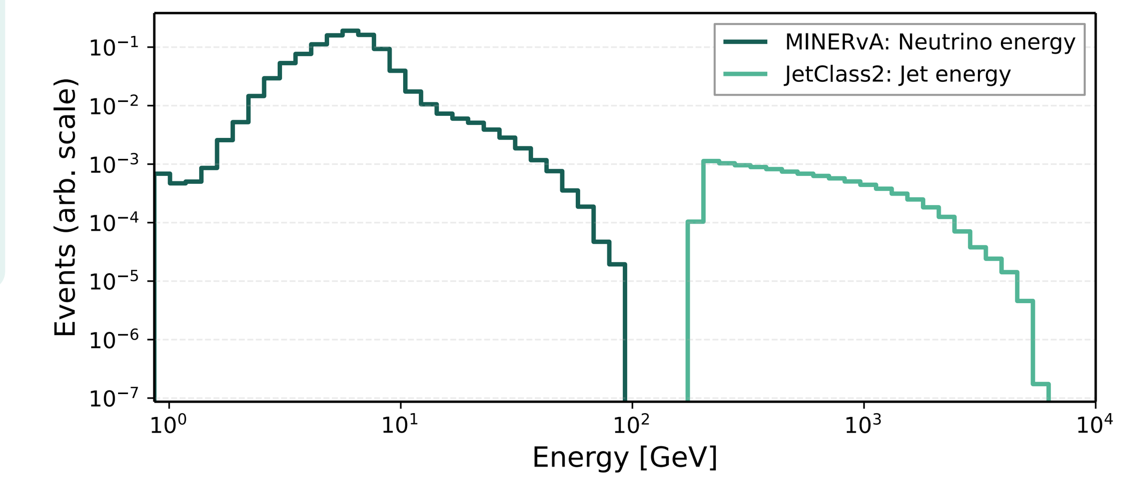


Fine-tuning

- Few-GeV events at the MINERvA Experiment
- O(10) tokens per event
- 6M events







Models

OmniLearned & HyperScale

(trained with event-level and particle-level features)

- Trained from scratch
- Pretrained on 1B jets

ViT-like Transformer

(trained with event-level and particle-level features)

- Trained from scratch

MLP

(trained only with event-level features)

- Trained from scratch

Tasks and Results

Classification

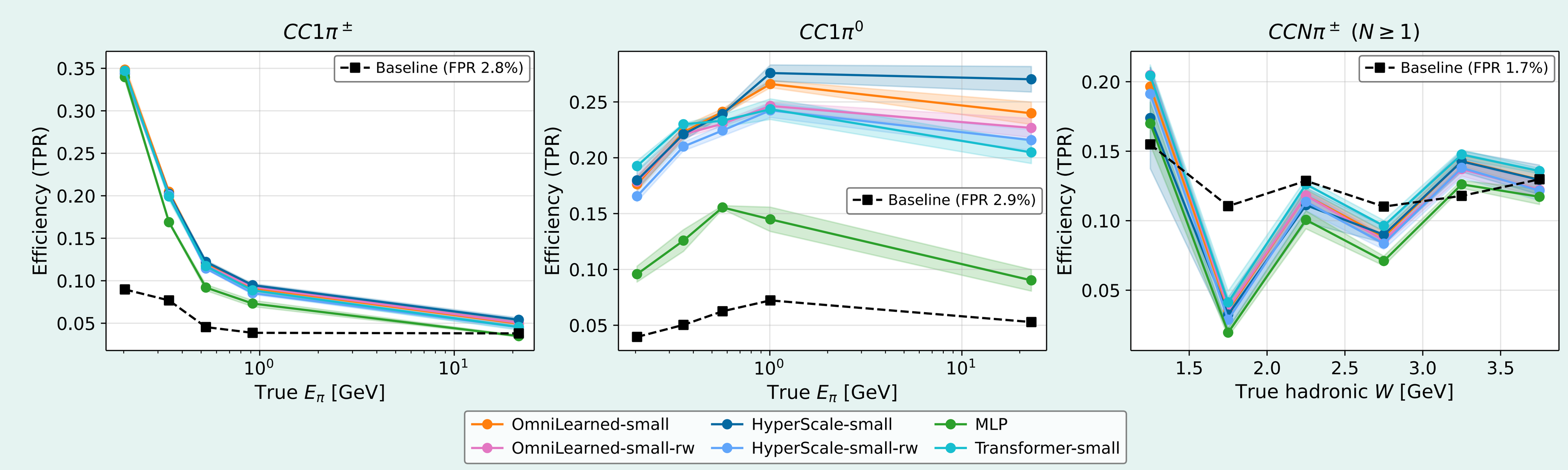
Three separate **signal** definitions for charged-current (CC) events:

$CC1\pi^{\pm}$ $CC1\pi^{0}$ $CCN\pi^{\pm}$

Background:

NC (no muon in final state)

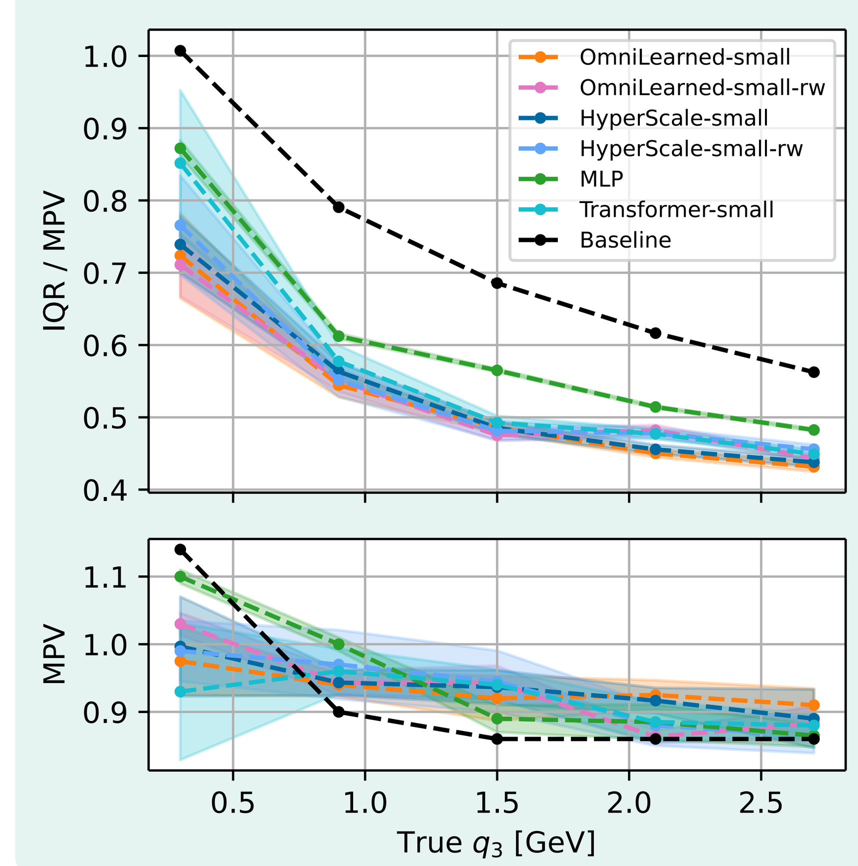
CC (all classes except signal class)



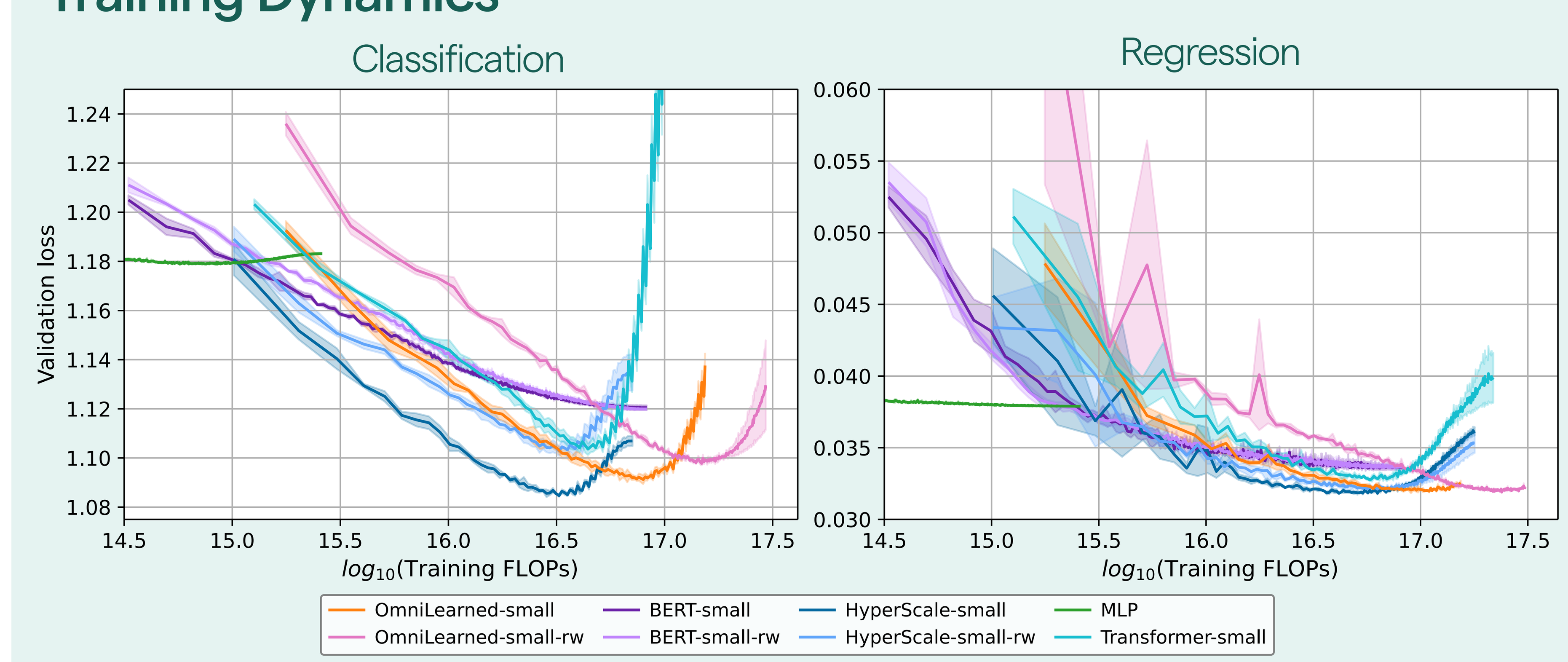
Regression

Regress available hadronic energy $E_{\mathrm{available}}$ for CC events:

Sum of energies of $p, \bar{p}, \pi^{\pm}, \pi^{0}, \gamma, e^{\pm}, K^{\pm}$



Training Dynamics



Key takeaways:

- Using event structure (vs. only global variables) improves performance (MLP vs. all other models)
- **Pre-trained models on jets** (OmniLearned-small / HyperScale-small) **lead to better performance at fixed compute** compared to the same randomly-initialized models (OmniLearned-small-rw / HyperScale-small-rw)