Contribution ID: **104**   Type: **Lightning 5 min talk + poster**

Real-Time AI-Based Data Selection in LArTPC Experiments Using Accelerated FPGA Platforms

Tuesday 15 October 2024 15:55 (5 minutes)

The demand for machine learning algorithms on edge devices, such as Field-Programmable Gate Arrays (FP-GAs), arises from the need to process and intelligently reduce vast amounts of data in real-time, especially in large-scale experiments like the Deep Underground Neutrino Experiment (DUNE). Traditional methods, such as thresholding, clustering, multiplicity checks, or coincidence checks, struggle to extract complex features from large data volumes. In contrast, certain machine learning algorithms offer more efficient, accurate, and power-efficient processing, making real-time analysis feasible and minimizing the need for costly offline data processing. We designed 2D convolutional neural networks (2DCNNs) to effectively detect rare events and reject background noise, demonstrating the viability of CNN-based algorithms for this application. Modern tools like hls4ml and HLS have streamlined the deployment of these models on FPGA hardware. The deployment of this model on Xilinx Alveo U250/U55c accelerator cards has demonstrated promising performance, comfortably meeting resource budget and latency targets. This talk will showcase the potential for expanding the model to classify a wider range of signals with greater precision, along with the FPGA optimizations we have adopted to make it suitable for DUNE.

Focus areas

Author: MALIGE, Akshay

Presenter: MALIGE, Akshay

Session Classification: Lighting talks