



Contribution ID: 121

Type: **Parallel Presentation**

Empowering AI Implementation: The Versatile SLAC Neural Network Library (SNL) for FPGA

Thursday, November 21, 2024 1:45 PM (15 minutes)

SLAC has developed a library based framework that enables the deployment of machine learning (ML) models on Field Programmable Gate Arrays (FPGAs) located at the edge of the data chain, near the instrumentation. It's called the SLAC Neural Network Library (SNL), utilizes Xilinx's High-Level Synthesis (HLS) and offers an API inspired by Keras for TensorFlow. By adopting a streaming data approach, SNL optimizes the data flow between neural network layers, minimizing the need for buffering and achieving high frame rates with low latency, critical for real-time applications in experimental environments.

A key feature of SNL is its ability to re-load neural network weights and biases after training without the need for re-synthesis. This allows for rapid updates to the deployed models, enhancing adaptability and performance in dynamic environments. Additionally, the framework supports network quantization, which helps optimize the use of FPGA digital signal processing (DSP) and memory resources, crucial for maximizing efficiency in resource-constrained edge computing scenarios.

Primary authors: DAVE, Abhilasha (SLAC National Lab); RUSSELL, J.J. (SLAC National Accelerator Laboratory (US)); HERBST, Ryan (SLAC National Accelerator Laboratory (US))

Presenter: DAVE, Abhilasha (SLAC National Lab)

Session Classification: RDC 05 - Trigger and DAQ Parallel Session

Track Classification: RDC Parallel Sessions: RDC5: Trigger and DAQ