



Contribution ID: 113

Type: Talk

FPGA-accelerated machine learning inference for particle physics computing challenges

Tuesday, April 2, 2019 1:00 PM (25 minutes)

The growing exploration of machine learning algorithms in particle physics offers new solutions to simulation, reconstruction, and analysis. These new machine learning solutions often lead to increased parallelization and faster reconstruction times on dedicated hardware, here specifically Field Programmable Gate Arrays (FPGAs). We explore the possibility that applications of machine learning simultaneously also solve the increasing computing challenges. Employing machine learning acceleration as a web service, we demonstrate a heterogeneous compute solution for particle physics experiments that requires minimal modification to the current computing model. First results with Project Brainwave by Microsoft Azure, using the Resnet-50 image classification model as an example, demonstrate inference times of approximately 50 (10) milliseconds with our experimental physics software framework using Brainwave as a cloud (edge) service. We also adapt the image classifier, for example, physics applications using transfer learning: jet identification in the CMS experiment and event classification in the Nova neutrino experiment at Fermilab. Solutions explored here are potentially applicable sooner than may have been initially realized. We will also briefly present a short status on real-time machine learning inference on FPGAs for the hardware trigger.

Primary authors: DUARTE, Javier Mauricio (Fermi National Accelerator Lab. (US)); TRAN, Nhan Viet (Fermi National Accelerator Lab. (US))

Presenter: DUARTE, Javier Mauricio (Fermi National Accelerator Lab. (US))

Track Classification: 6: Architectures and techniques for fast track reconstruction