FERMILAB-PUB-XX-YYY-E

Prepared for submission to JINST

Demonstration of FPGA-accelerated machine learning inference for computing challenges in physics

[a] *Fermi National Accelerator Laboratory, Batavia, IL 60510, USA*
[b] *Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Abstract: Resources required for high-throughput computing in large scale particle physics experiments face challenging demands both now and in the future. The growing exploration of machine learning algortihms in paritcle physics offer new solutions to simulation, reconstruction, and analysis. We explore the possbility that applications of machine learning simultaneously also solve the mounting computing challenges. By accelerating machine learning algorithms on dedicated hardware, we can speed up particle physics tasks with similar and often improved performance. We perform a proof-of-concept study of FPGA (Field Programmable Gate Array) hardware-accelerated machine learning using Project Brainwave by Microsoft Azure to accelerate image classification tasks by a factor of 10^3 or more over traditional CPU inference. By employing machine learning acceleration as a web service within the experimental software framework of the CMS experiment at the CERN LHC, we demonstrate a heterogeneous compute solution for particle physics experiments that requires minimal modification to the current computing model. The image classifications tasks are adapted for jet identification in the CMS experiment and event classification in the Nova neutrino experiment at Fermilab.

Contents

1 Introduction

2 ML co-processor acceleration as a service

3 Performance and examples

3.1 Core system performance

3.2 Nova transfer learning example

3.3 LHC transfer learning example

3.4 C++ experimental software integration

4 Summary and Outlook

Acknowledgements

We would like to thank ...
XXXX are supported by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. XXXX is supported by a Massachusetts Institute of Technology University grant.

References