

Deep Learning Acceleration of the Boosted Higgs Program and HEP Computing

Nhan Tran, Wilson Fellow
 Fermi National Accelerator Laboratory
 Phone: (267) 629-9892, Email: ntran@fnal.gov
 Year Doctorate Awarded: 2011

Number of times previously applied: 1

Topic Area: Experimental Research at the Energy Frontier in High Energy Physics
 DOE National Laborator Announcement: LAB 19-2019

The discovery of the Higgs boson is a triumph for the field of particle physics and completes the Standard Model (SM). Despite the Higgs discovery, fundamental questions about our universe remain such as the nature of dark matter, the hierarchy problem, and the matter-antimatter asymmetry. Proposed solutions to these problems can be addressed by exploring new physics coupled to the Higgs boson – the so-called *Higgs portal*. This proposal uniquely probes the Higgs portal by exploring its behavior in extreme parts of phase space highly sensitive to new physics contributions. Further, this proposal will explore solutions to outstanding big data computing challenges at the Large Hadron Collider (LHC) which are applicable to many high energy physics experiments. Computing resources will not scale to meet future needs of the LHC and new innovative paradigms are needed. The catalyst for this research program is deep learning techniques which extend our physics sensitivity and improve our computational efficiency by large factors. The proposal consists of two main aspects: expanding and improving searches for new physics via anomalous Higgs boson couplings at high transverse momentum p_T with the CMS experiment at the LHC; and accelerating computationally-heavy simulation and reconstruction algorithms through new deep learning computing paradigms that fit well with the HEP distributed computing standard.

I. BOOSTED HIGGS

In the current LHC era, since the discovery of the Higgs boson, it is important both to use the Higgs boson as a lamppost and to unearth hidden possible signatures of new physics. The PI has recently pioneered conceptually novel methods to study the Higgs [1] ($gg \rightarrow H \rightarrow b\bar{b}$) and search for light hidden hadronic dark sector particles [2] (light $Z' \rightarrow qq$). The key to both is to search for highly boosted, highly collimated particle jet signatures underneath overwhelming SM backgrounds. These seminal searches marked the beginning of a completely new physics program at the LHC. One of the keys to this program is jet substructure methods which are used to classify these interesting physics signatures. The PI will expand and accelerate this emergent physics program to unlock new initial and final states by deploying deep learning methods that are well-suited to these complex, high-dimensionality jet substructure classification tasks.

The original analysis deployed a tagging method that used traditional Boosted Decision Tree multivariate methods to identify the Higgs to $b\bar{b}$ candidate. However, newer tagging methods using deep learning techniques developed by the PI and his team have been developed which demonstrate both a significant performance improvement and the ability to distinguish $H \rightarrow b\bar{b}$ and $H \rightarrow c\bar{c}$ [3]. With these new methods, we plan to achieve 3σ evidence for the boosted Higgs to $b\bar{b}$ signature and pioneer a first search for the boosted $H \rightarrow c\bar{c}$ process at the same time. The measurement of the Higgs coupling to charm quarks, $H \rightarrow c\bar{c}$, is a largely experimentally unconstrained coupling but theoretically well-known and therefore could be a place where new physics could be lurking. Our goal is to improve on current limits which are at 100 times the Standard Model prediction. Additionally, we plan to observe the $Z \rightarrow c\bar{c}$ process in a single jet for the first time at the LHC. This observation would be a major milestone in proving the viability of the technique to identify $H \rightarrow c\bar{c}$.

Exploring the Higgs boson produced at very high p_T is not simply a measurement program, but a unique and very sensitive probe of new high scale physics contributions to the Higgs boson [4]. For example, the dominant gluon fusion production mode is the most sensitive way to explore new heavy particles coupling to the Higgs-top coupling loop through large deviations to the SM expectation as the Higgs p_T spectrum increases. Further, by selecting on the production mode of the boosted Higgs, we can also probe different anomalous couplings to the Higgs and do a full taxonomy of the boosted Higgs kinematics to disentangle the source of potential new physics contributions. For example, from previous work by the PI, the VH associated production and vector boson fusion production channels at very high p_T have been shown to have powerful sensitivity to small CP violating Higgs couplings [5]. Through deep learning extensions of the boosted Higgs program and differentiation of the Higgs production mode, we can achieve unprecedented sensitivity to the Higgs in an extreme part of phase space as a powerful probe of new physics couplings to the Higgs. The evolution of this program on the timescale and dataset of LHC Run 3 will be a very exciting period. We plan to demonstrate, for the first time, evidence of boosted $H \rightarrow b\bar{b}$ in potentially multiple production channels, and an observation of the $Z \rightarrow c\bar{c}$ process in a single jet.

II. DEEP LEARNING ACCELERATION

The LHC and CMS are simultaneously preparing for a high luminosity HL-LHC upgrade, starting in the mid-2020s, with increasing collision rates and extremely complex environments. Datasets will grow by a factor of 50 during the HL-LHC era. Further, detectors will have up to 10 times more readout channels and information and the collision environment complexity will increase with 5 times more instantaneous luminosity. At the same time, our computing capabilities cannot scale to meet these demands as single CPU performance has stalled from the breakdown of Moore’s Law and Dennard Scaling. Industry solutions to these big data challenges involve a large investment in heterogeneous computing deploying specialized coprocessor hardware in which traditional CPUs are coupled to more efficient computational hardware such as GPUs (NVidia), FPGAs (Microsoft, AWS, Intel), and ASICs (Google, Apple). At the core of the new hardware solutions is an emphasis on deep learning algorithms and instruction sets. The literature on specialized hardware has demonstrated factors of 50-100 speedup over CPUs. In this sense, machine learning becomes a *universal language* in which to express algorithms that can be implemented on more efficient computing platforms.

The key to the revolution in HEP is two-fold: first, migrating traditional physics algorithms into machine learning algorithms and second, integrating specialized hardware into the HEP computing model in a feasible way. The first aspect dictates that our most computationally challenging problems are explored with machine learning algorithms – large scale detector simulation and algorithms such as tracking and clustering. Efforts are on-going in this direction (TrackML challenge, the HepTrkX/ExaTrkX effort, Generative Algorithms on physics simulation) including an effort by the PI to apply geometrical deep learning to generic clustering tasks. The second aspect is less explored in HEP and the focus of this proposal. The PI leads a pioneering effort [6] in this direction to explore how to optimally integrate heterogeneous computing into HEP. It builds on his expertise using FPGAs and ASICs for ultrafast event processing in the CMS trigger system [7]. The HEP computing model relies on globally-distributed (grid) computing built on a wide variety of computing hardware, and it will be very challenging and costly to require coprocessors at every experimental computing center. By exploring industry tools from Xilinx, AWS, Google, and Microsoft, a first proof-of-concept studies has been performed to demonstrate the viability of heterogeneous computing *as a web service* where deep learning inference can be off-loaded to dedicated cloud resources in an efficient way.

This research program proposes to build off these promising studies to explore the challenges in scaling this approach to the HEP computing scale. There are three aspects to be studied in this program to demonstrate the viability of this approach. First, the proof-of-concept studies need to be extended to work on the scale of a computing farm and the worldwide grid. We will quantify the scalability of costs, bandwidth and computing resources is important to contrast against more traditional computing architectures. Second, because this is relatively new technology, we need to explore how industry software tools develop to match HEP needs. For example, we will work on generically translating HEP-style deep learning algorithms to specialized hardware. There are new tools coming on the market and our studies can provide valuable feedback in the development of these tools. Finally, the third aspect is the translation and development of the HEP algorithms into machine learning algorithms. This proposal will couple closely with the PI's work on developing for generic clustering algorithms by translating those onto specialized hardware.

This research program is well-suited to the varied and broad expertise at Fermilab. The Scientific Computing Division includes many core originators and developers of software frameworks used by many neutrino experiments and the CMS experiment. They will be key in understanding the challenges to interface heterogenous hardware with experimental software and developing the necessary tools. To scale up these tools, additional computing experts on grid and distributed parallel computing will be invaluable. For the specialized hardware translation, electrical engineering will be required to continue developing machine learning algorithms on specialized chips and understand the limitations of various technologies. It is the combination of machine learning expertise, also a great resource at Fermilab, electrical engineering, core software development, and large scale computing resources and experts which make Fermilab the ideal place to execute this research program. The goal of this proposal is to understand the landscape of available deep learning technologies, including their fully integrated computing pipeline, and how they are best integrated into the current and future HEP computing paradigms. Improving our computing efficiency by large factors is the only way we will be able to sustain the simulation, reconstruction, and analysis of HL-LHC data as we move to exabyte scale datasets and all available options must be explored. The lessons learned for the LHC will have a very large impact on the way that computing evolves for other big data experiments such as DUNE.

III. OUTLOOK

In a post-Higgs-discovery era, the PI proposes a high impact research program that deploys deep learning to uniquely explore the Higgs portal and daunting computing challenges. The program includes an emerging suite of boosted Higgs measurements at the LHC enhanced by new final states using deep learning and the exploration of deep learning-dedicated hardware to solve future looming HEP computing challenges. Due to the impactful groundwork laid by the PI, the expertise that has been developed, and a broad range of Fermilab expertise in distributed computing, software, and electronics, this program is one that the PI is uniquely qualified to execute.

-
- [1] CMS Collaboration, Phys. Rev. Lett. 120, 071802 (2018).
 - [2] CMS Collaboration, Phys. Rev. Lett. 119, 111802 (2017).
 - [3] CMS Collaboration, CMS DP-2018/046 (2018).
 - [4] CMS Collaboration, <https://arxiv.org/abs/1812.06504> (2018).
 - [5] I. Anderson *et al.*, Phys. Rev. D 89, 035007 (2014).
 - [6] M. Liu *et al.*, <https://indico.fnal.gov/event/18104/session/8/contribution/16> (2018).
 - [7] J. Duarte *et al*, JINST 13 (07) (2018).

List of collaborators

- U.S. Collaborators and Co-editors:
 (not including Graduate and Postdoctoral Advisors and Advisees, listed below)
 Tulika Bose (Boston University), Sergei Chekanov (Argonne National Laboratory), Richard Cavanaugh (University of Illinois at Chicago), Timothy Cohen (University of Oregon), Sridhara Dasu (University of Wisconsin), James Dolen (Purdue Northwest University), Bertrand Echenard (California Institute of Technology), Steve Ellis (University of Washington), Robin Erbacher (University of California Davis), William Ford (University of Colorado), William Gary (University of California, Riverside), Eva Halkiadakis (Rutgers University), Kristian Hahn (Northwestern University), Philip Harris (Massachusetts Institute of Technology), David Hitlin (California Institute of Technology), Joseph Incandela (University of California Santa Barbara), Robert Johnson (University of California Santa Cruz), Ashutosh Kotwal (Duke University), Greg Landsberg (Brown University), Andrew Larkoski (Reed College), Owen Long (University of California, Riverside), Petar Maksimovic (Johns Hopkins University), Jeremiah Mans (University of Minnesota), Meenakshi Narain (Brown University), Timothy Nelson (SLAC National Laboratory), Salvatore Rappoccio (State University of New York Buffalo), Jeff Richman (University of California, Santa Barbara), Alexei Safonov (Texas A&M University), Philip Schuster (SLAC National Laboratory), Brian Shuve (Harvey Mudd College), Wesley Smith (University of Wisconsin), Kevin Stenson (University of Colorado), Jesse Thaler (Massachusetts Institute of Technology), Natalia Toro (SLAC National Laboratory), Keith Ulmer (University of Colorado)
- International Collaborators and Co-editors:
 (not including Graduate and Postdoctoral Advisors and Advisees, listed below)
 Torsten Akesson (Lund University), Matteo Cacciari (University of Paris, Diderot, France), Matthew Dolan (Melbourne University, Australia), Pietro Govoni (University of Milano-Bicocca, Italy), Andreas Hinzmann (Hamburg University, Germany), Roman Kogler (Hamburg University, Germany), Simone Marzani (University of Genoa, Italy), Kirill Melnikov (Karlsruhe Institute of Technology), Matthias Mozer (Karlsruhe Institute of Technology), Giovanni Petrucciani (CERN, Switzerland), Maurizio Pierini (CERN, Switzerland), Ruth Poettgen (Lund University), Gavin Salam (CERN, Switzerland), Alexander Schmidt (University of Aachen, Germany), Markus Schulze (Humboldt University of Berlin, Germany), Gregory Soyez (CEA Saclay, France),
- Graduate and Postdoctoral Advisors:
 Andrei Gristan (Johns Hopkins University)
- Graduate Student and Postdoctoral Associates:
 Owen Colegrove (University of California, Santa Barbara), Marat Freytsis (Institute for Advanced Study), Raffaele Gerosa (University of California, San Diego), Josh Hiltbrand (University of Minnesota), Clemens Lange (CERN), Cristina Mantilla Suarez (Johns Hopkins University), Omar Moreno (SLAC National Laboratory), Ian Mould (Lawrence Berkeley National Laboratory), Jennifer Ngadiuba (CERN), Marc Osherson (Rutgers University), Rishi Patel (University of Colorado), Dylan Rankin (MIT), Stanislava Sevova (Stanford University), Zhenbin Wu (University of Illinois at Chicago), Zijun Xu (SLAC National Laboratory), David Yu (Brown University)