Progress Report of the Integration of Jet-Tagging Neural Networks into the CMS Level 1 Trigger

Anthony Vizcaíno Aportela*
University of California San Diego
(CMS)
(Dated: May 11, 2023)

The planned upgrade to the High Luminosity Large Hadron Collider (HL-LHC) presents significant challenges for the real-time event selection process due to the increased collision rate and data volume. The Compact Muon Solenoid (CMS) experiment's Level 1 (L1) Trigger system plays a critical role in selecting events of interest and requires the development and implementation of faster and more efficient algorithms to cope with the demanding conditions. This paper presents a plan to develop and integrate a jet-tagging neural network into the CMS L1 Trigger system in the context of the HL-LHC upgrade, as well as a debrief of progress and preliminary results. Building on the lessons learned from previous studies involving b-tagging and tau-tagging neural networks, we aim to optimize the jet-tagging algorithm for the real-time event selection process using specialized tools like the hls4ml Python library. We propose to apply optimization techniques to meet the stringent timing requirements and validate the performance of the integrated jet-tagging algorithm using simulated collision data.

I Introduction

The Large Hadron Collider (LHC) at CERN has been at the forefront of high-energy physics research, providing invaluable insights into the fundamental particles and forces that govern the universe. The Compact Muon Solenoid (CMS) is one of the major experiments at the LHC, responsible for collecting and analyzing data from proton-proton collisions. The CMS Level 1 (L1) Trigger system plays a critical role in selecting events of interest in real-time from the vast amount of data generated by these collisions.

With the planned upgrade to the High Luminosity LHC (HL-LHC) [1], the collision rate and the volume of data generated will significantly increase, posing new challenges for the L1 Trigger system. The HL-LHC is expected to operate at a luminosity that is approximately 5–7 times higher than the current LHC, leading to a larger number of simultaneous interactions and a higher rate of potentially interesting physics events. In this context, the development and implementation of faster and more efficient algorithms for event selection in the L1 Trigger system becomes crucial to maximize the physics potential of the CMS experiment.

The CMS data flow management comprises radiation-hard application-specific integrated circuits (ASICs), field-programmable gate arrays (FPGAs), and other computational resources. At the heart of this system is the L1 Trigger [2], which selects relevant events out of the vast amount of data generated by the LHC collisions. The L1 Trigger identifies 1 event out of 400 in a 12.5 μs window, ensuring the retention of events of interest for further analysis. This real-time event selection is crucial for reducing the data volume to a manageable level for the subsequent High-Level Trigger (HLT) and offline analyses.

One promising approach to address this challenge is the integration of deep neural networks (DNNs) into the L1 Trigger system. These DNNs can be tailored to efficiently identify and tag specific physics objects, such as jets, in the collision data, thereby enhancing the real-time event selection capabilities of the CMS experiment. To implement these DNNs on field-programmable gate arrays (FPGAs), which are at the core of the L1 Trigger system, specialized tools like the hls4ml workflow [3, 4] can be employed to translate the trained models into FPGA-compatible hardware description languages (HDLs).

In this study, we propose a similar approach for the integration of a jet-tagging algorithm using deep neural networks. We also debrief progress made towards this goal. Jet-tagging algorithms are essential for identifying and characterizing the jets produced in high-energy collisions. By integrating a jet-tagging neural network into the L1 Trigger, we aim to enhance the detector's capabilities in managing and processing the immense data flow generated by the LHC and improve the identification and analysis of various physics phenomena. This research holds significant potential for advancing our understanding of the fundamental particles and forces in the universe.

II Jet-Tagging Neural Network Integration Plan

To integrate the jet-tagging neural network into the CMS L1 Trigger, we plan to follow a similar approach used for the b-tagging and tau-tagging neural networks, using the hls4ml workflow [5]. The integration process will include the following steps:

* aaportel@ucsd.edu

II.a Develop the jet-tagging neural network architecture and train the model using appropriate datasets

We have designed a neural network architecture tailored to the jet-tagging problem, taking into consideration the unique characteristics of the L1 Trigger system and the requirements for real-time event selection. The model is trained using simulated collision data containing various jet types, allowing the neural network to learn the distinctive features of each jet category.

II.b Implement the trained jet-tagging neural network in the hls4ml library

The hls4ml library is a versatile tool for translating deep learning models into FPGA-compatible hardware description languages (HDLs). We utilize this workflow to implement the trained jet-tagging neural network into the L1 Trigger system, ensuring that the model's performance is optimized for the real-time event selection process.

II.c Apply optimization techniques to meet timing requirements and improve system performance

To ensure that the jet-tagging neural network meets the stringent timing requirements of the L1 Trigger system, we will apply or have already experimented in applying the following three optimization techniques, as demonstrated in previous studies:

1. Area constraint: By constraining the neural network's placement on the FPGA, we can optimize the timing performance by reducing the propagation delay between components.

2. Explicit instantiation: Vivado HLS tends to prefer more explicit instructions, which can improve timing performance. By explicitly writing down every instantiation, we can further optimize the system's performance.

3. Quantization: By adjusting the fixed-point representation of the neural network's weights and activations, we can optimize resource usage and meet the timing requirements of the L1 Trigger system. Quantization can be done during or after training by either truncating parameters after training or by enforcing low precision during training. Both of these come at the cost of some model performance.

II.d Integrate the optimized jet-tagging neural network into the L1 Trigger system and validate its performance

Once the jet-tagging neural network has been optimized using the techniques mentioned above, we will integrate it into the L1 Trigger system. We will then inject simulated collision data and analyze the output results to validate the performance of the integrated jet-tagging algorithm. The validation process will involve comparing the jet-tagging performance to that of existing algorithms and assessing the potential improvements in the identification and analysis of various physics phenomena.

III Preliminaries

In this section, we provide preliminary results on the development of the jet-tagging neural network, its quantization, and the conversion to HLS using the hls4ml workflow. The results demonstrate that our jet-tagging model and its quantized version have comparable performance, and they have been successfully converted into an HLS implementation suitable for integration into the CMS L1 Trigger system.

III.a Training the Jet-Tagging Neural Network Model

We have trained a jet-tagging neural network model using Keras, a popular deep learning library. The model was developed considering the unique requirements of the L1 Trigger system and the need for real-time event selection. We used simulated collision data containing various jet types for training, allowing the neural network to learn the distinctive features of each jet category. The model achieved high accuracy in identifying and tagging different jet types, indicating its potential for enhancing the real-time event selection capabilities of the CMS experiment.

III.b Quantizing the Jet-Tagging Neural Network Model

To optimize the resource usage and meet the timing requirements of the L1 Trigger system, we quantized the trained jet-tagging neural network model using QKeras [6], a library for quantizing Keras models. The quantized model utilized fixed-point representation for its weights and activations, resulting in reduced model size and faster inference times while maintaining comparable performance to the original model.

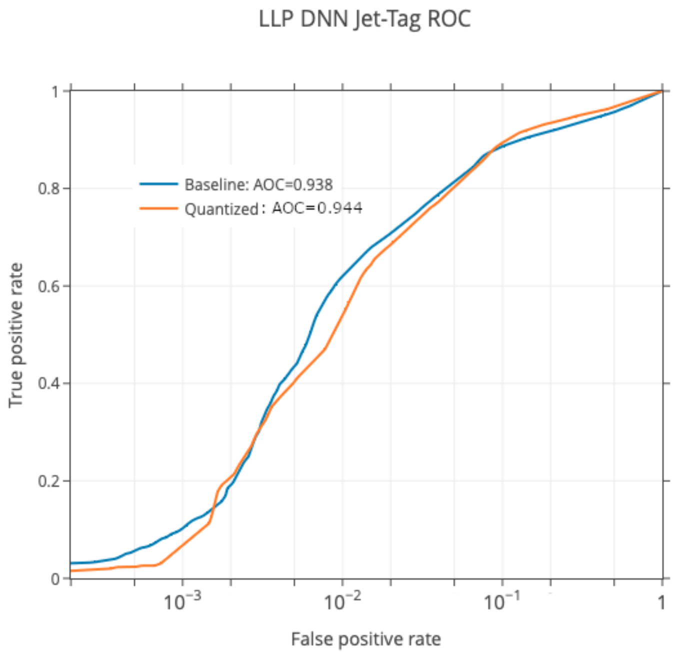

FIG. 1. ROC curves for the baseline (blue) and quantized (orange) keras models, with AUCs of .938 and .944 respectively.

III.c Converting the Models to HLS using hls4ml

We employed the hls4ml workflow to convert both the original Keras model and its quantized version into HLS implementations suitable for integration into the CMS L1 Trigger system. This process involved translating the trained deep learning models into FPGA-compatible hardware description languages (HDLs) and optimizing the performance of the models for the real-time event selection process. The successful conversion of both models to HLS demonstrates their potential for integration into the L1 Trigger system and the applicability of the hls4ml workflow in this context.

III.d Model Performance and Comparison

We evaluated the performance of the original and quantized jet-tagging models by analyzing their receiver operating characteristic (ROC) curves and comparing the area under the curve (AUC) values. The results, as seen in figure 1, indicate that both models exhibit comparable performance in terms of jet-tagging accuracy and efficiency. This finding suggests that the quantized model is suitable for integration into the L1 Trigger system, as it maintains a similar level of performance to the original model while providing the benefits of reduced resource usage and faster inference times.

IV Conclusion

In this study, we presented a plan to develop and integrate a jet-tagging neural network into the CMS Level 1 Trigger system in the context of the HL-LHC upgrade, building on the lessons learned from previous studies involving b-tagging and tau-tagging neural networks. Using specialized tools like the hls4ml workflow, we aim to optimize the jet-tagging algorithm for the real-time event selection process. By applying optimization techniques to meet the stringent timing requirements and validating the performance of the integrated jet-tagging algorithm using simulated collision data, we endeavor to enhance the real-time event selection capabilities of the CMS experiment. This will enable more accurate identification and analysis of jets produced in high-energy collisions at the LHC, ultimately advancing our understanding of the fundamental particles and forces in the universe. Our preliminary results demonstrate the feasibility of our approach and highlight the potential for significant contributions to high-energy physics research through the integration of neural networks into the L1 Trigger system.

[1] CERN, High luminosity lhc project, `https://hilumilhc.web.cern.ch/content/hl-lhc-project` (2023), accessed on May 9th, 2023.
[2] P. Das (CMS), An overview of the trigger system at the CMS experiment, Phys. Scripta **97**, 054008 (2022).
[3] J. Duarte *et al.*, Fast inference of deep neural networks in FPGAs for particle physics, JINST **13** (07), P07027, arXiv:1804.06913 [physics.ins-det].
[4] FastML Team, fastmachinelearning/hls4ml (2023).
[5] D. Hoang, A. Chambers, D. Rankin, and P. Harris, Implementing deep neural networks in cms level 1 trigger, `https://indico.cern.ch/event/1156222/contributions/5076141/attachments/2522824/4338292/DH_CMS_L1.pdf` (2021), presentation at the CMS L1 Trigger Group Meeting.
[6] C. N. Coelho, A. Kuusela, S. Li, H. Zhuang, T. Aarrestad, V. Loncar, J. Ngadiuba, M. Pierini, A. A. Pol, and S. Summers, Automatic heterogeneous quantization of deep neural networks for low-latency inference on the edge for particle detectors, Nature Mach. Intell. **3**, 675 (2021), arXiv:2006.10159 [physics.ins-det].

Acknowledgments

R. Marroquin Solares, Grad. Student; D. Diaz, Postdoc; J.M. Duarte, P.I.