DIANA Fellowship Proposal: Optimizing fast convolutional neural networks for identifying long-lived particles in the CMS high-granularity calorimeter

Particles with long lifetimes are an important possibility in the search for new phenomena, and often appear in Beyond the Standard Model theories, notably in models that describe the elementary particle nature of dark matter. When produced at the CERN Large Hadron Collider (LHC), these long-lived particles (LLPs) have a distinct experimental signature: they can decay far from the primary proton-proton interaction but within a detector such as ATLAS or CMS, or even completely pass through the detector before decaying.

The data at the ATLAS and CMS experiments are collected using triggers that select events in real time, reducing the event rate from the 40 MHz bunch crossing rate down to about 1 kHz that can be written to disk. Since most triggers assume that the particles originate from the proton-proton interaction vertex and are not displaced, dedicated triggers are necessary for detecting displaced particles from LLP decays.

Due to the higher occupancy with up to 200 proton-proton interactions per bunch crossing, in particular in the forward region, a new High-Granularity Calorimeter (HGCal) will be installed in CMS endcaps for the High-Luminosity LHC. The HGCal will be the first imaging calorimeter in a running experiment at a high-energy collider, which generates many new opportunities, such as using it for a pattern-recognition-based trigger for displaced particles. Furthermore, deep neural networks of limited size can be deployed on field-programmable gate arrays (FPGAs) using dedicated tools such as HLS4ML, and can therefore now be included directly in the CMS first-level (L1) hardware trigger.

We have shown that it is possible to directly use neural network based pattern recognition to trigger on distinct calorimeter signatures from displaced particles, such as those that arise from the decays of LLPs. The study was performed for a high granularity forward calorimeter similar to the CMS HGCal. We showed that a simple convolutional neural network (CNN) that could in principle be deployed on dedicated fast hardware can efficiently identify showers from displaced particles down to low energies, while providing a low trigger rate.

The goal of this project is to optimize the CNN so that it can be deployed on FPGAs within the CMS HGCal operational requirements. The student will work under the supervision of lead developers and HLS4ML experts at CERN to identify which aspects of the network increase the latency and will then identify and implement strategies to improve the performance. Candidates for this position should have a good working knowledge of python. Experience with other programming languages, such as C++, and with machine learning frameworks, such as TensorFlow/Keras, will be considered favorably. The anticipated duration of the project is the three month period May - July, 2021 or June - August, 2021 although there is some flexibility related to the exact start and finish dates.

Juliette Alimena (CERN) will supervise the student. A timeline describing the work plan and deliverables is provided on the next page.

Timeline

- weeks 1-2 become familiar with LLPs, the HGCal, the L1 trigger, FPGAs, TensorFlow/Keras, and docker containers;
- weeks 3-4 learn how to train models for the CNN trigger on GPUs and interact with the code development system;
- weeks 5-6 weeks 5-6 working with experts, define a strategy for systematically identifying problem areas and improving performance;
- weeks 7-8 begin to execute the systematic studies defined in weeks 5-6 and mitigate the most egregious problems.
- weeks 9-11 carefully document prior work; continue to identify and mitigate problems;
- weeks 12-13 define a road-map for future work; complete documentation.

At the end of the project, the student will present his/her work at an IRIS-HEP topical meeting.