IRIS-HEP Fellowship Proposal

Omar Alterkait, University of Colorado, Boulder
Duration: December 2020 - May 2021
Area: Innovative Algorithms

Proposal

The prospect of the HL-LHC and the high pileup associated with it has made track reconstruction much more challenging. Following track reconstruction, an offline algorithm that is able to distinguish real tracks (tracks associated with a single charged particle) from fake tracks (tracks not associated with a single charged particle) is beneficial. The CMS software contains a DNN for this task; however, this is based on the current detector.

For the phase 2 upgrade of the CMS detector, the outer tracker is quite different which means that this task needs to be reevaluated. The phase 2 algorithm currently in use employs older track reconstruction algorithms that rely on a series of selection criteria to select tracks. The goal of my project is to develop a more efficient post track reconstruction algorithm using deep learning to classify reconstructed tracks as either real or fake. I plan to use existing phase 2 relvals to start the process and additional MC can be generated if needed. My advisor for this project is Professor Kevin Stenson

Deep learning has been having a mini-revolution every couple of months. The purpose of my project is to try to incorporate some of these state of the art techniques and develop a high efficiency track classification algorithm. This part of the project will be done mainly in Python. I will use the Awkward Array and Uproot python packages to access and handle the data, and use Pytorch to build and train the model. I am currently taking an applied deep learning which should be very helpful in the development of this project. I plan to attend the CMS Data Analysis School in January.

I will start off by spending some time to analyze the current post-track reconstruction algorithm used. After that, I will develop the deep learning model, and try and optimize the model as much as possible. Finally, I will implement the deep learning model into C++, and incorporate it into the CMS software. I also plan on doing an Honors Thesis on this project.

Timeline

I will work halftime over the next six months

- **December - January**: Analyze the the current post-reconstruction algorithm used along with the reconstructed data
- **February - March**: Develop and Train the Deep Learning model
- **March - May**: Implement the Machine Learning Code into C++ and incorporate it in CMS software.