IRIS-HEP Fellows Program Project Proposal

The usage of Deep Learning for QCD background estimation to measure differential cross section of events with a photon and heavy flavour jets using the data collected by the CMS experiment

Applicant: Andrii Len

Mentor: Dr. Ece Asilar

Project duration: 12 Weeks

Proposed start date: 27 June 2022

Introduction and Project Description

After the Higgs Boson [1, 2] discovery in 2012 the Standard Model (SM) [3] of Particle Physics has been completed. One of the following natural steps would be to measure cross section of various SM processes. The aim of this project is to support the first time measurement of the differential cross section of photon production in association with b quarks in hard parton-parton interaction at 13 TeV using the data provided by CMS Experiment at LHC in Run 2 [4, 5]. Measuring the cross section ratio of events with 2 b quarks to 1 b quarks provides an important test of perturbative Quantum Chromodynamics predictions in addition to the parton density functions of b quarks and gluons. Having the last cross section measurement of this process from ATLAS experiment at 8 TeV [6] allows CMS physicists to perform the analysis urgently but at the same time robustly. The analysis will cover the entire data-taking period with a total integrated luminosity of 137 fb^{-1}. The challenge of the analysis is the existence of the irreducible Quantum Chromodynamics multi-jet events as the largest background. It is crucial to obtain a reliable separation of signal and background events with minimal reliance on Monte Carlo event generators. The fact that we have very few discriminating variables between useful signal and background makes the task a difficult one, so we need deep learning [7, 8] to solve this problem.

The main purpose is to train the algorithm by using two energy regions, in one of which only background photons are presented and in the other one the signal of photons we are interested in are added to background. The focus of the present project is to find optimal deep learning models to be used for the separation of signal and background events. The project involves investigating the models using Monte Carlo events as well as validation of the models using the data from side band regions. As a final step of the project, these models will be used to predict the QCD background contribution in the signal regions. As a result of the analysis, precise measurement of differential cross section will guide the field on the choice of Parton Distribution Functions and Monte Carlo event generators.

Software Deliverables

In this work, we will use Python, especially the Coffea and Numpy packages to reformat the data sets into ones that can be fed to the deep learning algorithm. For deep learning, we will use Keras and TensorFlow. At the end of the project, the best performing models will be delivered in terms of H5 and JSON file formats.

Preliminary Timeline

Week 1-2 Literature scan and learning information on CMS experiment, event format, fundamental particles, cross section measurements.

Week 3-5 Getting familiar with Linux environments. Learning how deep learning works and practicing Python and Coffea.

Week 6-8 Start learning how to use deep learning technically. Model building, training, testing, predicting.

Week 9-10 Hyperparameter search. Training models.

Week 11 Obtain the predictions from MC and data regions, and compare them with previous results.

Week 12 Summarising the results as a short report.

References

[1] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[2] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[3] F. Halzen and A. Martin, "Quarks and leptons: an introductory course in modern particle physics", Physics textbook, Wiley, 1984.

[4] CMS Collaboration, "The CMS experiment at the CERN LHC", JINST 3 (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[5] R. Bruce et al., "Baseline LHC machine parameters and configuration of the 2015 proton run", doi:10.5170/CERN-2015-002.100, arXiv:1410.5990..

[6] ATLAS Collaboration, "Measurement of differential cross sections of isolated-photon plus heavy-flavour jet production in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector", Phys. Lett. B 776 (2018) 295–317, doi:10.1016/j.physletb.2017.11.054, arXiv:1710.09560

[7] Schmidhuber, J. Deep learning in neural networks: An overview. Neural Networks 2015, 61, 85117.

[8] LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436.