Tasks for Prospective GSoC Applicants for the *CMS and End-to-End Deep Learning Project @ ML4SCI Umbrella Organization*

Below are the tasks that we will use to evaluate prospective students for the E2E projects. After completing the first 2 common tasks, please thoroughly complete the specific third task for your project of interest and (optionally) other third tasks if you would like to also be considered for additional E2E projects at the same time (this may increase your chances of success, but make sure you don't do it at the expense of the specific project you are interested in)

Note: please work in your own github branch (i.e. NO PRs should be made). Send us a link to your code when you are finished and we will evaluate it.

Common Task 1. Electron/photon classification
Datasets:
https://cernbox.cern.ch/index.php/s/AtBT8y4MiQYFcgc (photons)
https://cernbox.cern.ch/index.php/s/FbXw3V4XNyYB3oA (electrons)

Description: 32x32 matrices (two channels - hit energy and time) for two classes of particles electrons and photons impinging on a calorimeter

Please use a deep learning method of your choice to achieve the highest possible classification on this dataset (**we ask that you do it both** in Keras/Tensorflow **and** in PyTorch). Please provide a **Jupyter notebook** that shows your solution. The model you submit should have a ROC AUC score of at least 0.80.

Common Task 2. Deep Learning based Quark-Gluon Classification

Datasets: https://cernbox.cern.ch/index.php/s/hqz8zE7oxyPjvsL

Description 125x125 matrices (three channel images) for two classes of particles quarks and gluons impinging on a calorimeter.
For description of 1st dataset please refer to the link provided for the dataset.

Please use a Convolutional Neural Network (CNN) architecture of your choice to achieve the highest possible classification on this dataset (**in your preferred choice** of framework for example: Tensorflow/Keras **or** Pytorch). Please provide a **Jupyter notebook** that shows your solution.

- ***(If applying for the End-to-End Deep Learning Regression for Measurements with the CMS Experiment project then only perform the following additional task.)***

Regression

Please train a model to estimate (regress) the mass of the particle based on particle images using the provided dataset.

DataSet Description: 125x125 image matrices with name of variables: ieta and iphi, with 4 channels called X_jet (Track pT, DZ and D0, ECAL). Please use at least ECAL and Track pT channels and 'am' as the target feature. Please train your model on 80% of the data and evaluate on the remaining 20%. Please make sure not to overfit on the test dataset - it will be checked with an independent sample.

Datasets:

Link to dataset: https://cernbox.cern.ch/index.php/s/94ihPUHBY5XSVII

- ***(If applying for the Graph Neural Networks for End-to-End Particle Identification with the CMS Experiment project then only perform the following additional task.)***

Graph Neural Networks

Description:

- Choose 2 Graph-based architectures of your choice to classify jets as being quarks or gluons. Provide a description on what considerations you have taken to project this point-cloud dataset to a set of interconnected nodes and edges.
- Discuss the resulting performance of the 2 chosen architectures.

Datasets (Same as in Task 2): https://cernbox.cern.ch/index.php/s/hqz8zE7oxyPjvsL

- ***(If applying for the Vision Transformers for End-to-End Particle Identification with the CMS Experiment project then only perform the following additional task.)***

Vision Transformers

Description:

- Train a Transformer model of your choice on the dataset below to achieve the performance closest to your CNN model's performance in Task 1.
- Discuss the resulting performance of the 2 chosen architectures.

Datasets (Same as in Task 1):

  - **https://cernbox.cern.ch/index.php/s/AtBT8y4MiQYFcgc (Photons)**
  - **https://cernbox.cern.ch/index.php/s/FbXw3V4XNyYB3oA (Electrons)**

- ***(If applying for the End-to-End Deep Learning Reconstruction for CMS Experiment project, then only perform the following 2 additional tasks.)***

Setup CMS Software Framework (CMSSW) on your local machine (with docker)

Hint:
Use CERN CentOS7 + CMSSW docker images from here:
https://hub.docker.com/r/clelange/cmssw (standalone image, tag 10_6_8_patch1 suggested) or
https://hub.docker.com/r/clelange/cc7-cmssw-cvmfs (required packages downloaded on-the-fly from CERN's SW repository cvmfs)

Some information about the CMSSW docker images can be found here:
https://github.com/clelange/cmssw-docker/blob/master/README.md
(especially the section "Running containers")

In the latter case (cvmfs-based) you can setup a CMSSW development area with, e.g.:
cmsrel CMSSW_11_0_1; cd CMSSW_11_0_1/src/; cmsenv
Additional CMSSW packages can be downloaded using, e.g.:
git cms-addpkg DataFormats/TestObjects

Code can be compiled typing:
scram b
And run with:
cmsRun [configfile]

Show that you can apply your model that you have trained in task 1 and 2 on the subset of the same electron/photon and quark/gluon dataset but by using CMSSW **(Please do not make any PRs to CMSSW, work only within your own branch)**

Hint: You can take a look at the following classes/inference examples in CMSSW:

https://github.com/cms-sw/cmssw/tree/master/PhysicsTools/MXNet
https://github.com/cms-sw/cmssw/tree/master/PhysicsTools/ONNXRuntime
https://github.com/cms-sw/cmssw/tree/master/RecoBTag
https://github.com/cms-sw/cmssw/tree/master/RecoBTag/ONNXRuntime
https://github.com/cms-sw/cmssw/tree/master/RecoBTag/TensorFlow

---

Test Submission Instructions: Please send us your CV and a link to all your completed work (github repo, Jupyter notebook + pdf of Jupyter notebook with output) to ml4-sci@cern.ch with Evaluation Test: E2E in the title.