

Track ML

Adrian Alan Pol
Cesare Calabria
Valentin Volkl
Adriano Di Florio

May 25nd – Third Patatrack Hackaton – Final Scram

The aim – TrackML challenge

Start to work together on the challenge

<https://www.kaggle.com/c/trackml-particle-identification>

Featured Prediction Competition

TrackML Particle Tracking Challenge

High Energy Physics particle tracking in CERN detectors

\$25,000 Prize Money

CERN · 301 teams · 3 months to go (2 months to go until merger deadline)

Overview Data Kernels Discussion Leaderboard Rules Team My Submissions Submit Predictions

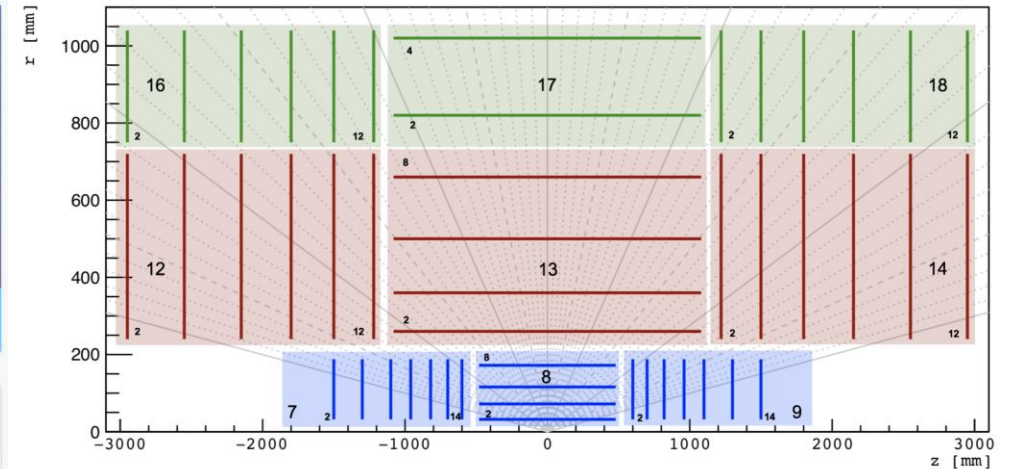
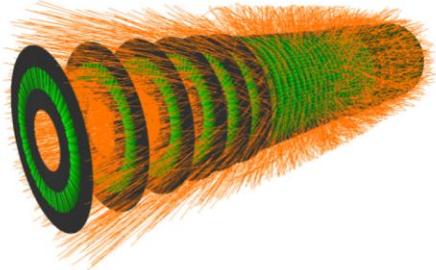
Overview

Description

To explore what our universe is made of, scientists at CERN are colliding protons, essentially recreating mini big bangs, and meticulously observing these collisions with intricate silicon detectors.

While orchestrating the collisions and observations is already a massive scientific accomplishment, analyzing the enormous amounts of data produced from the experiments is becoming an overwhelming challenge.

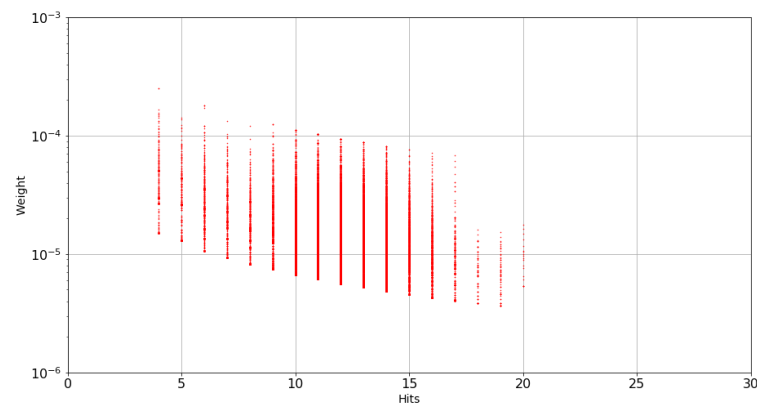
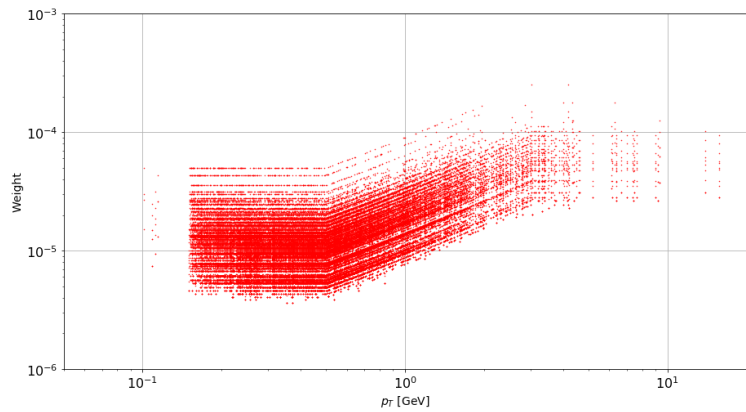
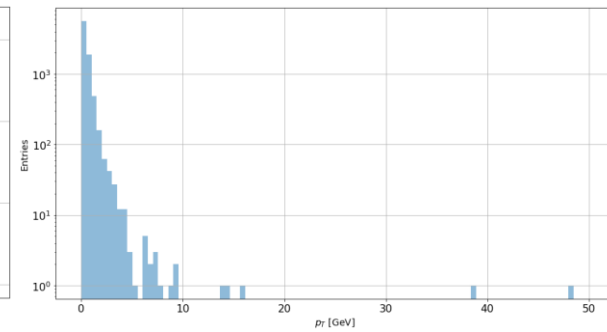
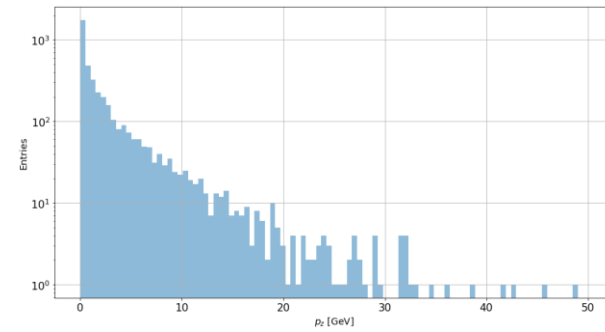
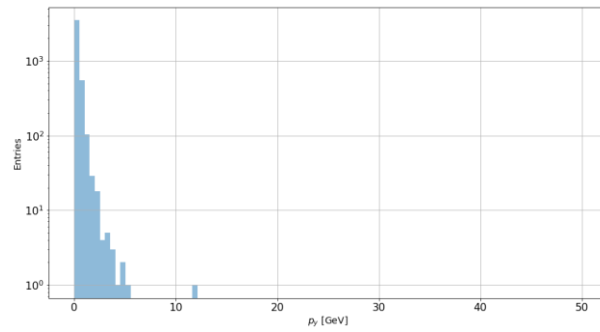
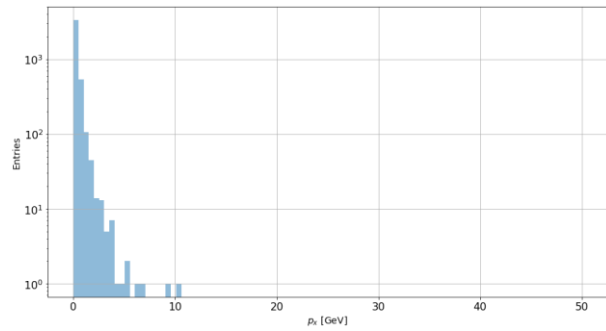
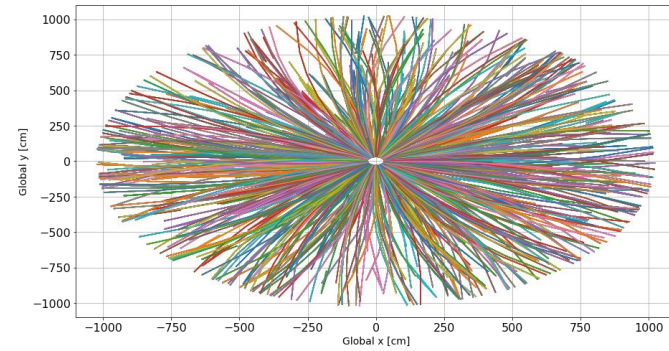
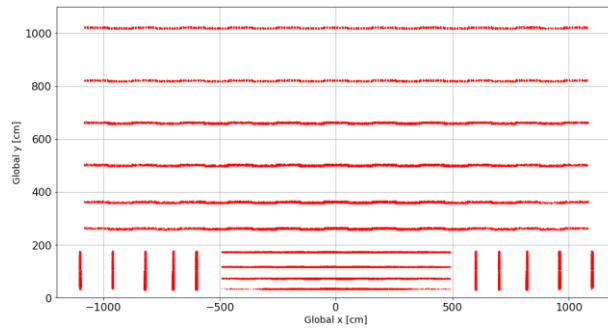
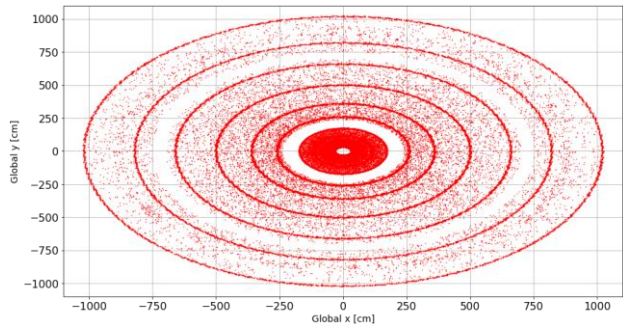
Event rates have already reached hundreds of



- $O(10k)$ particles, $O(100k)$ hits
- HL-LHC like environment (PU=200)

What has been done

Data exploration and visualization . . . like a lot of it



What has been done

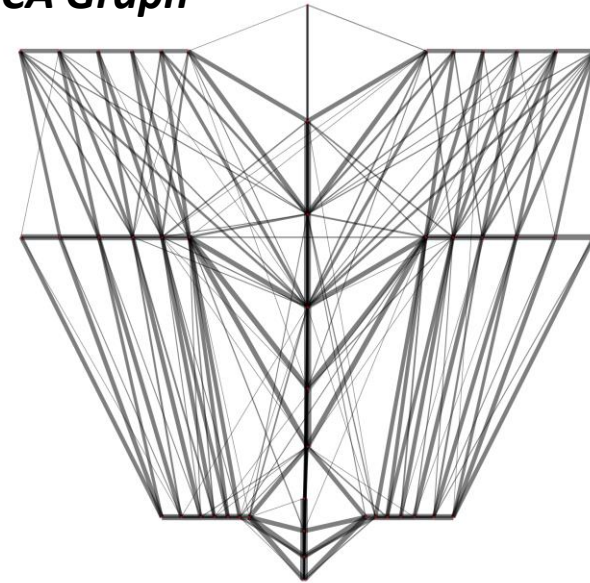
Started to share resources and know-how

Set up a strategy and start building from that

“Tracking-like” NN based workflow

1. Building seeds (*quadruplets*)
2. From *quadruplets* predicting next hit *position* with a RNN
3. Taking *all the hits* in a fiducial region around the *predicted position*
4. Reiterate until **no hit** found in the fiducial region
5. Clean tracks (DNN-like model)

CA Graph



- **TrikTrak CA for building triplets/quadruplets (running on c++ with pybindings) done**

Interfaced directly with the pandas trackML dataset

- **RNN for extending the seeds from 3rd or 4th to 4th or 5th hit done**

Tuned, rethought and is working kind of neatly. Need to tune the fiducial region to get a proper efficiency/duplicates ratio

- **Concatenation of RNNs to extend from nth hit to (n+1)th hit done**

<https://gitlab.cern.ch/adpol/trackml>

What we learned and what's next

- Better to focus on a simpler version of the challenge by now ignoring small issues, of course (e.g. multiple hits on the same layer)
- We got familiar with the dataset and the challenge
- Continue to work on the challenge in the next months as a group
- We **understood** that 20-30 people will finish all the pasta you prepare whatever the amount (lowerlimit ~ 5kg)

What's next

- Tuning CA/TrikTrak parameters and cuts to get the best balance efficiency/fake rate rejection
- How to properly propagate the RNNs
- How to properly select of hits with a fiducial region
- Train/test and submit