
# CEVA-CMS Collaboration

## Fast Inference Single Shot Detection

Status Check: October 2019

adrianalan.pol@cern.ch

# Workspace

- Gitlab
  - https://gitlab.cern.ch/adpol/ceva-cms-jet-ssd
- Drive
  - Models
  - Slides
  - Notes
  - etc.

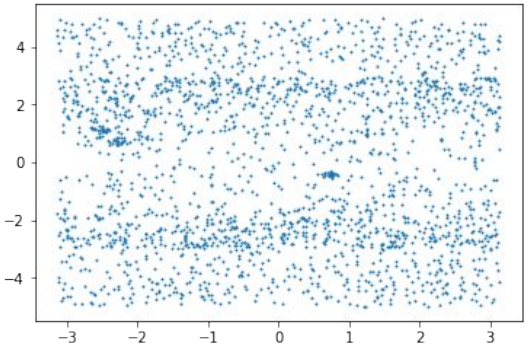
Mail me your usernames

# Generator/Dataset

- Done several iterations on dataset generation.
- Correction from last meeting: 4380973 events (not 5M).
- Simulations not perfect e.g. missing files or events inside files.
- Expected ~0.5TB in h5 (previously npy - too big).
  - Now just b and h jets.
- Split: train (top, bottom, higgs, W, q?), test (tth). We still miss q jets.

# Generator/Dataset: Open Issues

- Crystal to pixel mapping:
  - Is one crystal equal to one pixel?
- Based on the current mapping <0.1% of the events end up with 2 of the crystals in 1 pixel.
  - Incorrect mapping?

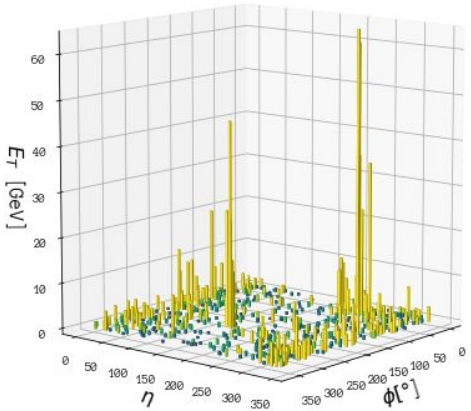


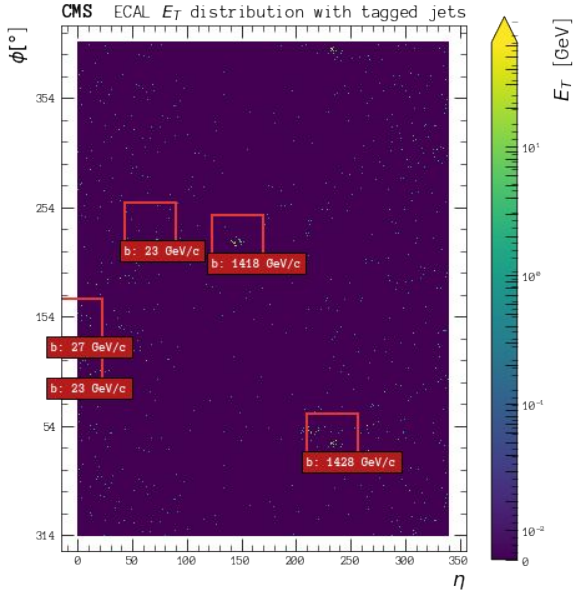
# Generator/Dataset Summary

| Type (classes) | No Events (images) | No Jets (objects) | μ [jets/image] | minPT [GeV/c] |
|---|---|---|---|---|
| bb | 769973 | 3738320 | 4.86 | 20.9 |
| tt | 940000 | 1945845 | 2.07 | 432.7 |
| hh | 990000 | 2111569 | 2.13 | 312.4 |
| WW | 911000 | 2056696 | 2.26 | 201 |
| qq | 0 | 0 | 0 | 0.48 |
| tth | 770000 | 4165924 | 5.41 | |

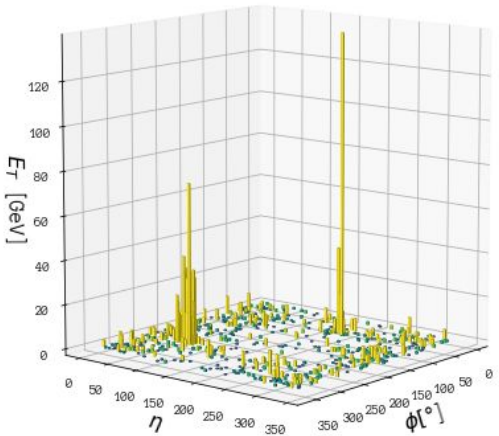
Train Test

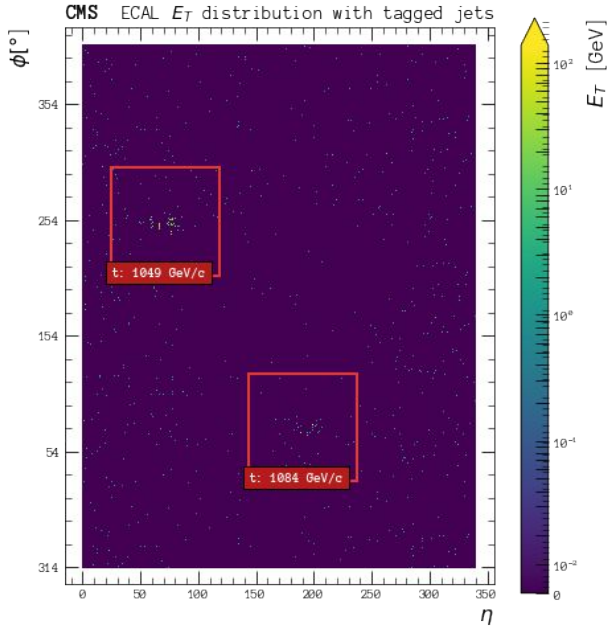
# b jets


$E_T$ [GeV]
60
50
40
30
20
10
0
0
50
100
150
200
250
300
350
$\eta$
350
300
250
200
150
100
50
0
$\phi$[°]

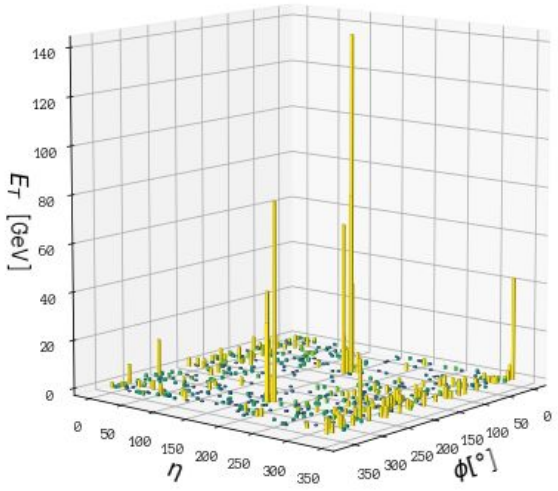

CMS ECAL $E_T$ distribution with tagged jets
$\phi$[°]
354
254
154
54
314
b: 23 GeV/c
b: 1418 GeV/c
b: 27 GeV/c
b: 23 GeV/c
b: 1428 GeV/c
0
50
100
150
200
250
300
350
$\eta$
$E_T$ [GeV]
$10^1$
$10^0$
$10^{-1}$
$10^{-2}$
0

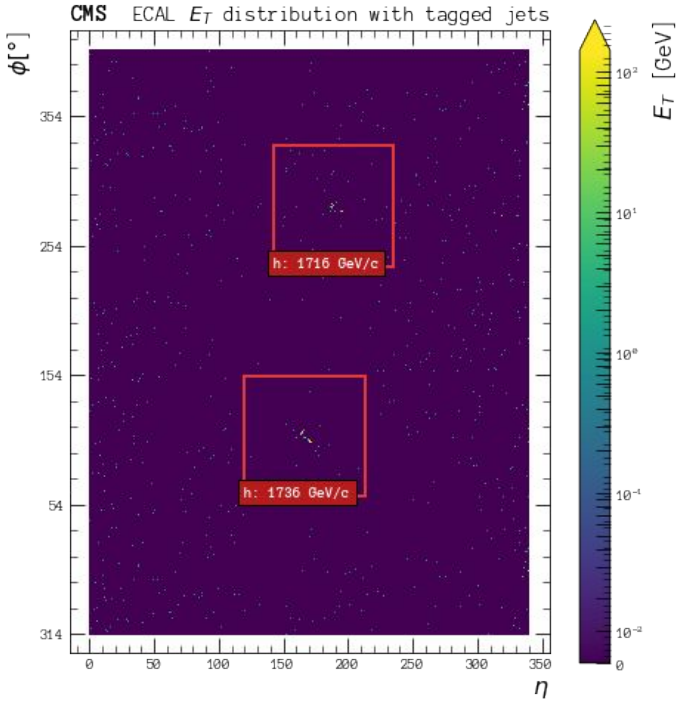
# t jets


$E_T$ [GeV]
120
100
80
60
40
20
0
0 50 100 150 200 250 300 350
$\eta$
0 50 100 150 200 250 300 350
$\phi$[°]

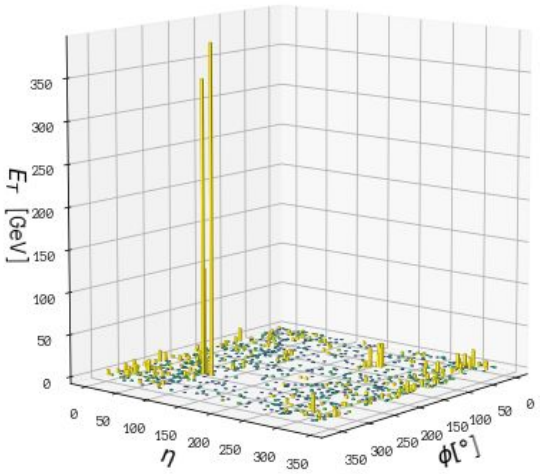

CMS ECAL $E_T$ distribution with tagged jets
$\phi$[°]
354
254
154
54
314
t: 1049 GeV/c
t: 1084 GeV/c
0 50 100 150 200 250 300 350
$\eta$
$E_T$ [GeV]
$10^2$
$10^1$
$10^0$
$10^{-1}$
$10^{-2}$
0

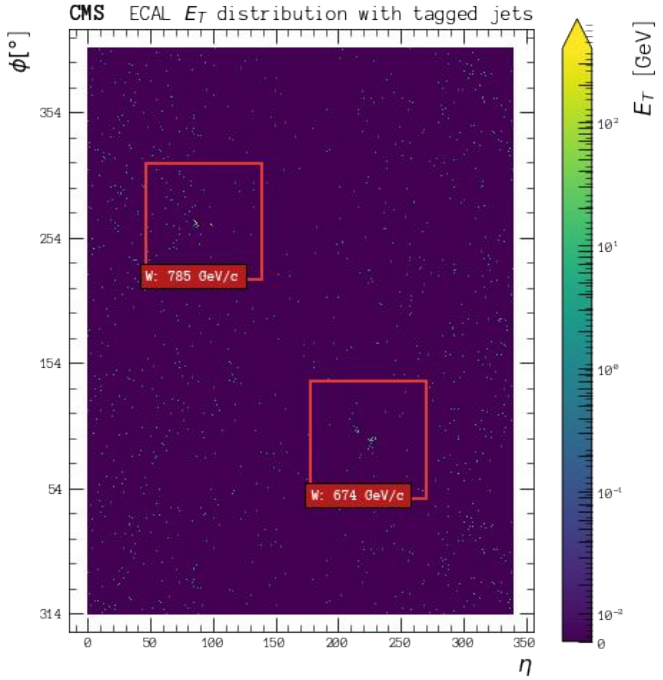
# h jets


$E_T$ [GeV]
140
120
100
80
60
40
20
0
0
50
100
150
200
250
300
350
$\eta$
350
300
250
200
150
100
50
0
$\phi$[°]


CMS ECAL $E_T$ distribution with tagged jets
$\phi$[°]
354
254
154
54
314
h: 1716 GeV/c
h: 1736 GeV/c
0
50
100
150
200
250
300
350
$\eta$
$E_T$ [GeV]
$10^2$
$10^1$
$10^0$
$10^{-1}$
$10^{-2}$
0

# W jets


$E_T$ [GeV]
350
300
250
200
150
100
50
0
0 50 100 150 200 250 300 350
$\eta$
350 300 250 200 150 100 50 0
$\phi$[°]

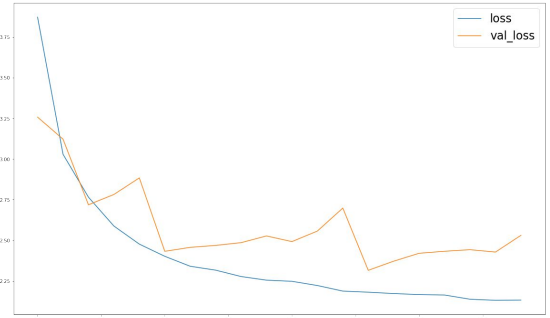

CMS ECAL $E_T$ distribution with tagged jets
$\phi$[°]
354
254
154
54
314
W: 785 GeV/c
W: 674 GeV/c
0 50 100 150 200 250 300 350
$\eta$
$E_T$ [GeV]
$10^2$
$10^1$
$10^0$
$10^{-1}$
$10^{-2}$
0

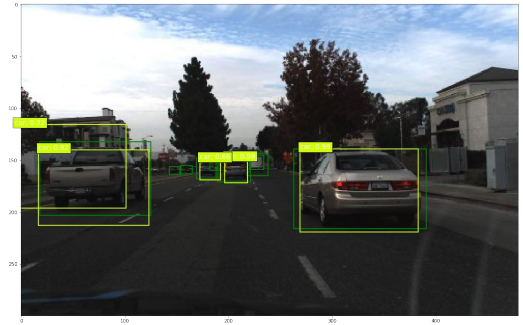
# SSD 7: The Base Model

- 7 layers:
  - Convolutions: 5x5 or 3x3 stride 1x1
  - BatchNormalization
  - Activations: ELU
  - Max Pooling: 2x2 stride 1x1
- 4 predictor layers:
  - Convolutions: 3x3 stride 1x1
- Changed 5x5 filter to 3x3
- Number of parameters: 178 884

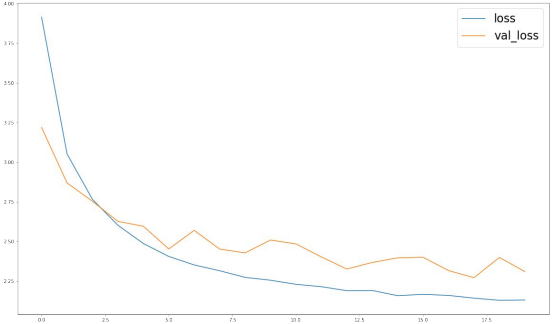
# SSD 7: Training loss

## Udacity

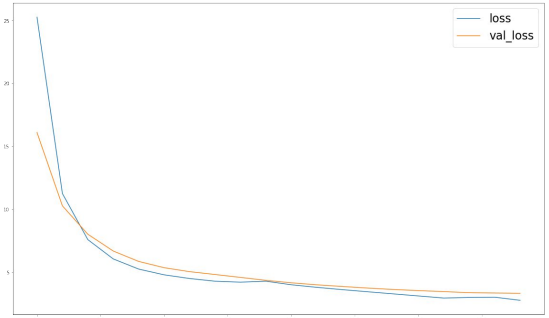


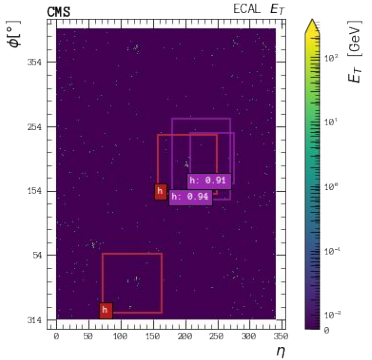



## Udacity (only 3x3)

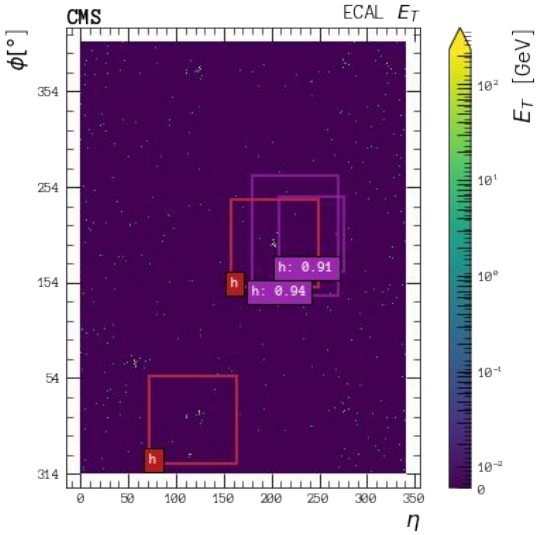


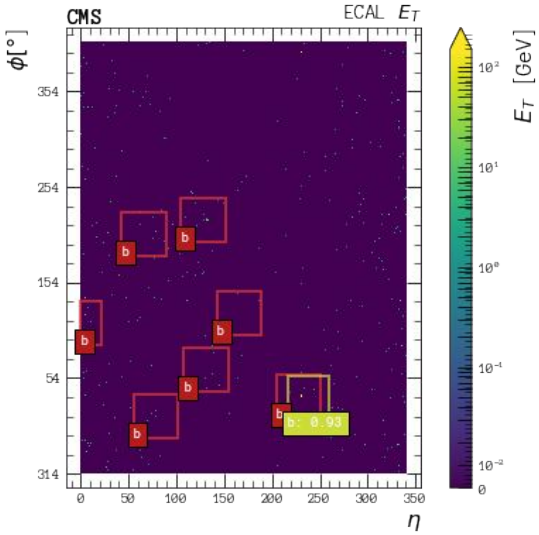

## Jets

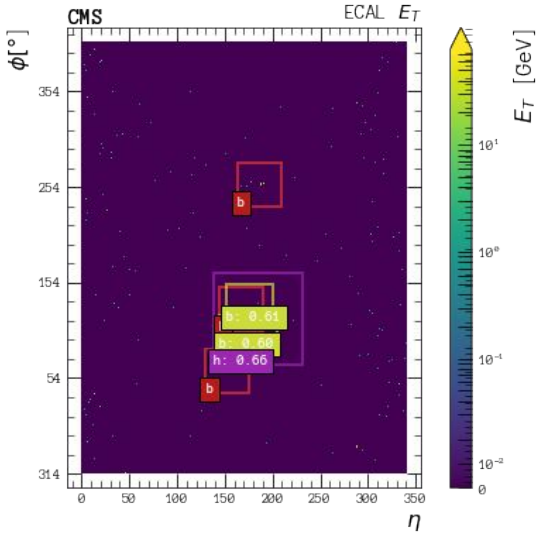


# SSD 7: First results



# Next Steps

- SSD with BNN
  - https://github.com/BertMoons/QuantizedNeuralNetworks-Keras-Tensorflow
- Find metric (and add evaluation script)
  - https://github.com/rafaelpadilla/Object-Detection-Metrics
- Settle on generator issues

# Things to discuss

- Collaboration with BGU
- AMLD 2020

*The Large Hadron Collider (LHC) operates at the remarkable rate of 40 MHz of proton-proton collisions. The volume of the data produced in collisions results in hundreds of exabytes of data per year, making the LHC one of the largest sources of data in the world today. Due to understandable storage constraints and other technological limitations (e.g. fast enough read-out electronics) the LHC experiments are required to reduce the number of recorded data. To this purpose, a set of algorithms is used to process and filter the incoming data stream while preserving the physics reach.*

*We are investigating fast inference deep neural networks (DNNs) solutions, to cope with an even higher number of collisions per second in the future and processing time constrains. We utilize dedicated digital signal processors (DSPs) provided by CEVA, the semiconductor company. To provide fast-enough operations we train the DNNs with weights and activations constrained to +1 or -1 (Binarized Neural Networks, BNNs). Our target application is a particle localization based on energy deposited in the detector data. We apply Single Shot Detection (SSD) method on sensor-based image-like representation of the data.*

*In this talk/poster we address our goals, challenges we face, details on the problem at hand and show first simulated results.*