GEP qg Tagger

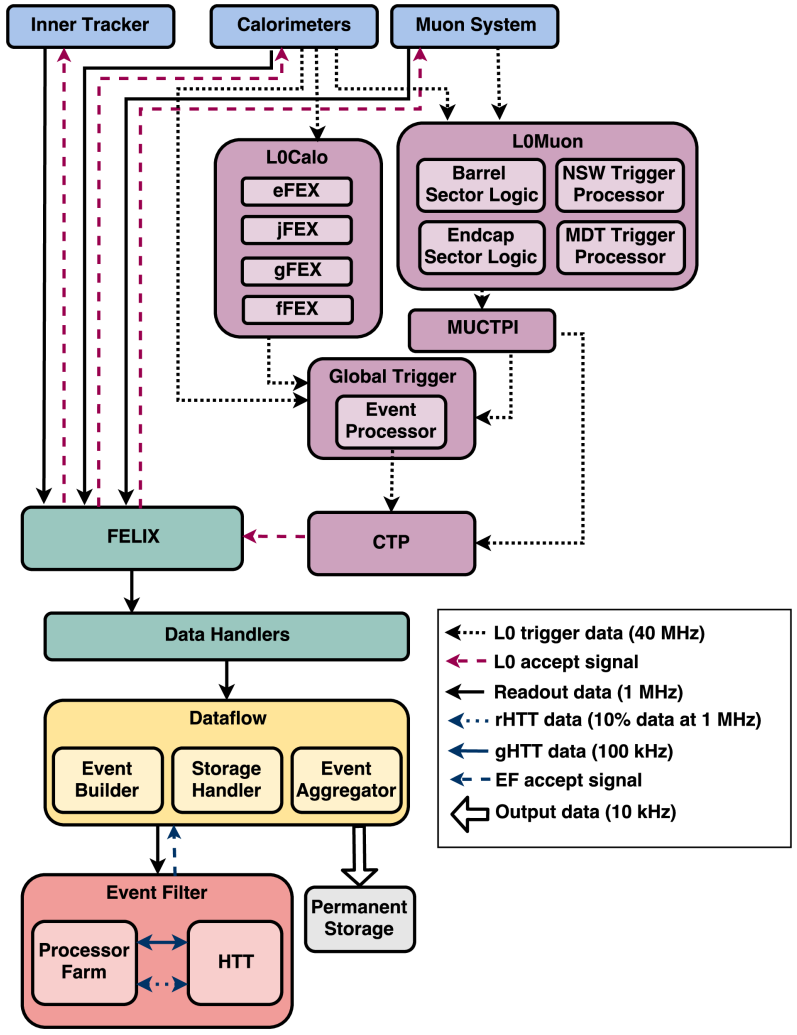
Santosh Parajuli

Southern Methodist University

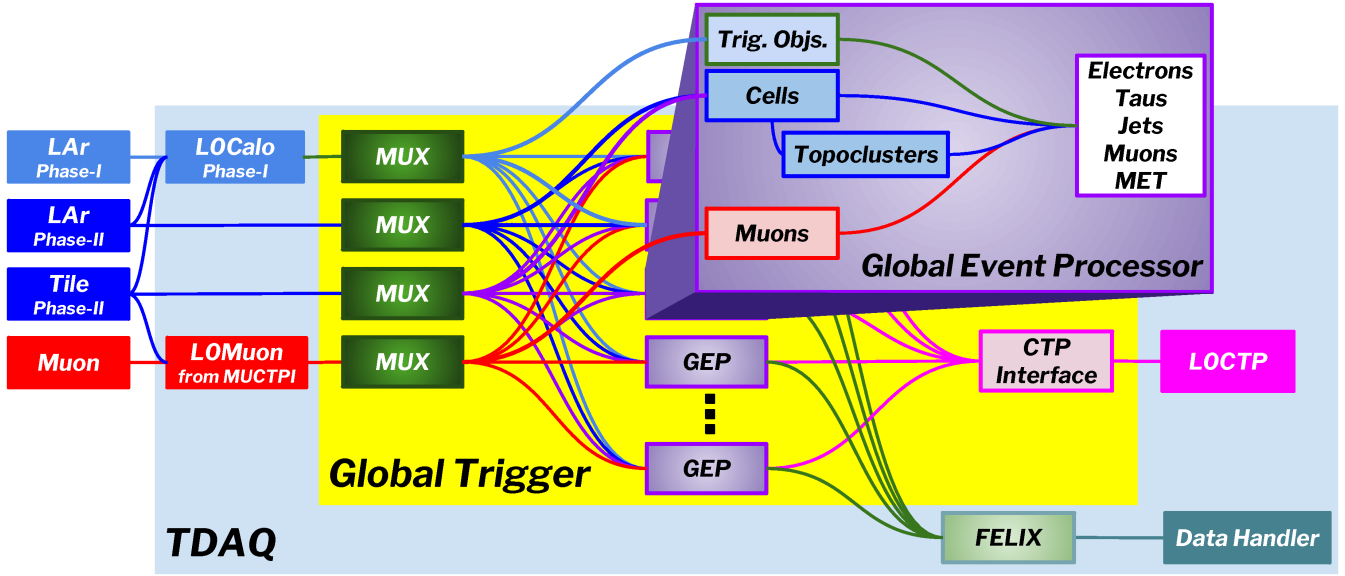
1 March 2023

SMU®

Global Trigger and Global Event Processer(GEP) in ATLAS



TDAQ System after the Phase-II upgrade (baseline design)



Schematic view of the Global Trigger System

Quark/Gluon Tagger Algorithm for GEP

- Binary Classifier Implementation, CNN Model.
- GEP Simulation Sample is used.

1) qcd-multijet sample-> ntuple with the branches,

```
[b'AntiKt10LCTopoLeadJets_partonTruthLabel',
 b'AntiKt10LCTopoLeadJets_pt',
 b'AntiKt10LCTopoLeadJets_nConstitutents',
 b'AntiKt10LCTopoLeadJets_eta',
 b'AntiKt10LCTopoLeadJets_phi',
 b'AntiKt10LCTopoLeadJets_clus_calE',
 b'AntiKt10LCTopoLeadJets_clus_calEta',
 b'AntiKt10LCTopoLeadJets_clus_calPhi',
```

- Parton truth labels to classify leading R=1.0 jets as quark and gluon.

- Using R=1.0 jets.
- No pT reweighting applied. (But the images are normalized)

Motivation for R=1.0 jets Quark/Gluon Tagger Algorithm for GEP

- For massive particles that are sufficiently boosted, it is advantageous to reconstruct their hadronic decay products as a single radius (large-R) jet.

- Hadronically decaying W , Z and H bosons and top quarks, which is distinct from the radiation pattern of a gluon-initiated jet.

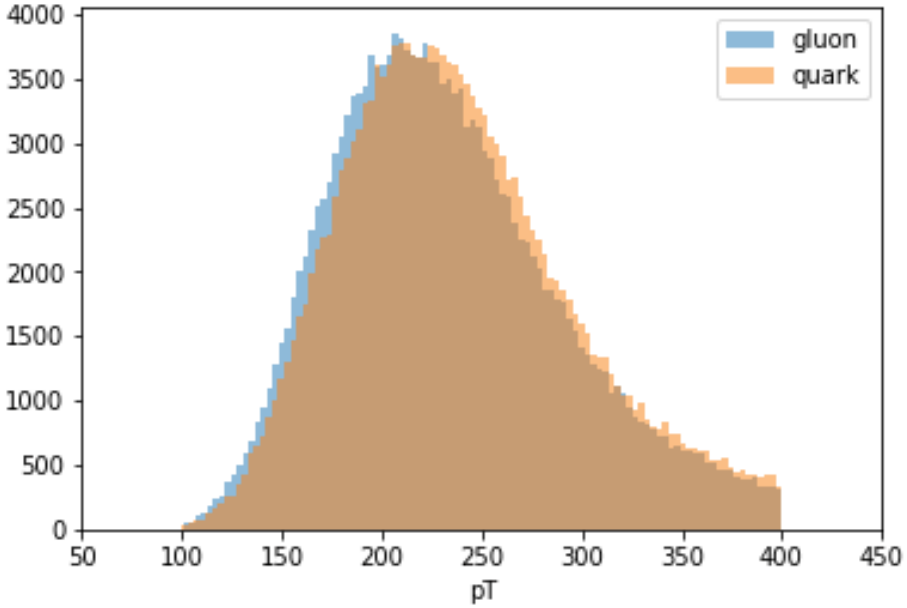
- Searches for heavy resonances.

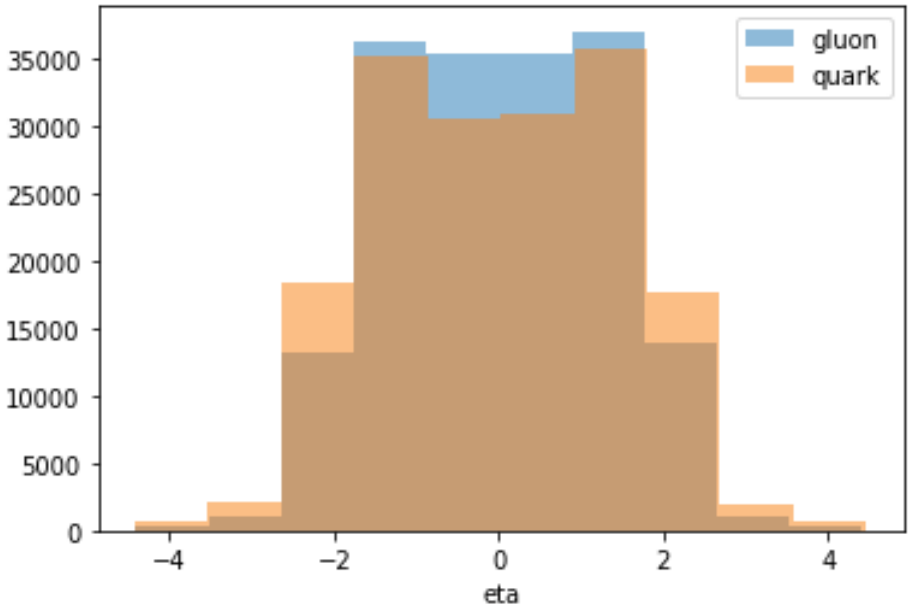
- Searches for dark matter particles.

Some Basic Plots for quark and gluon jets

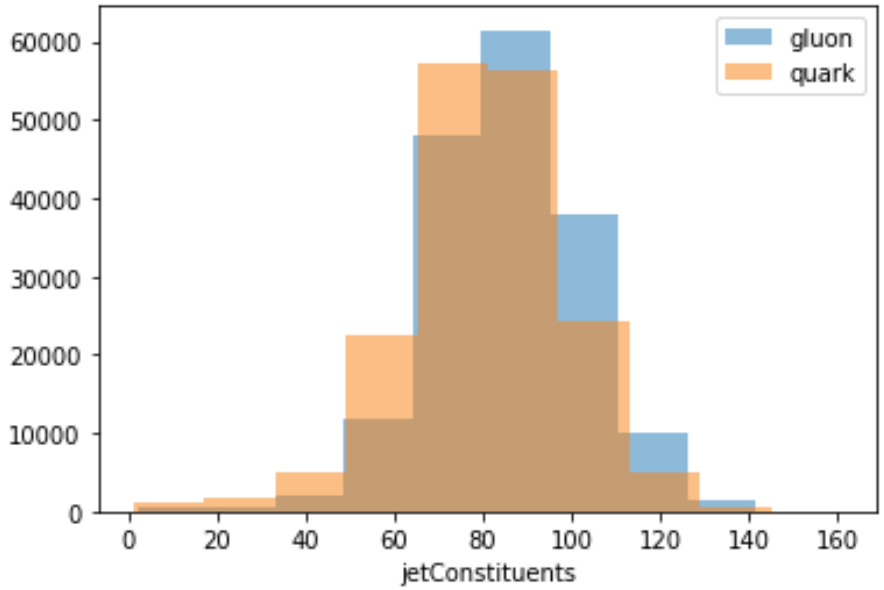
- For Quark: AntiKt10LCTopoLeadJets_partonTruthLabel >=1 and <=5 are selected.
- For Gluon: AntiKt10LCTopoLeadJets_partonTruthLabel = 21 is selected.
- pT >= 200 GeV to pT =< 250 GeV events are selected.
- 60,334 signal(quark jets) and 60,334 Background (gluon jets) events are used here. Total-> 120,668 events

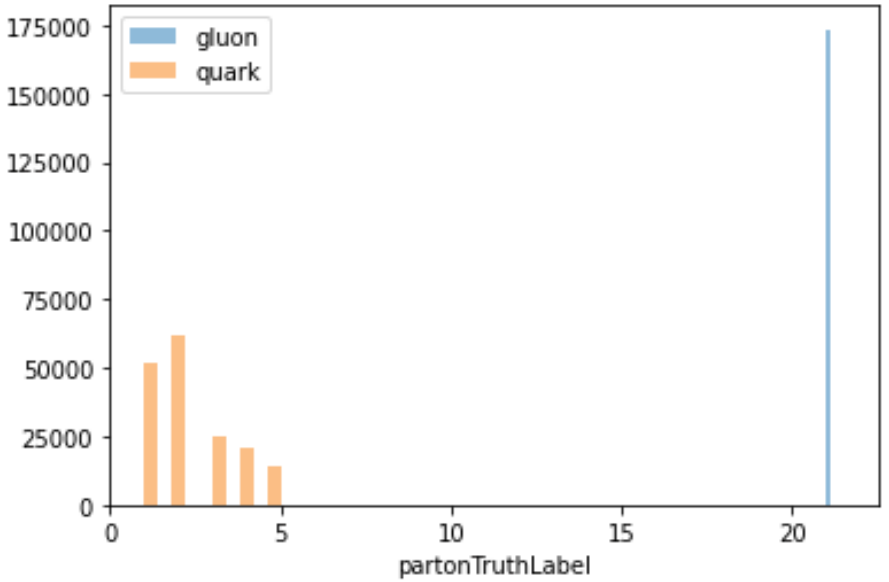
Some Basic Plots for quark and gluon jets(pT: 100-400 GeV)



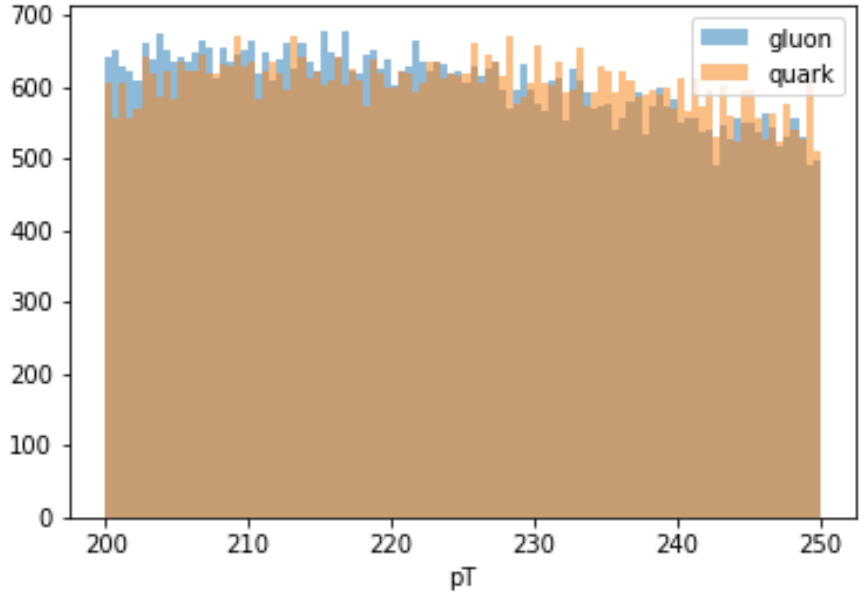


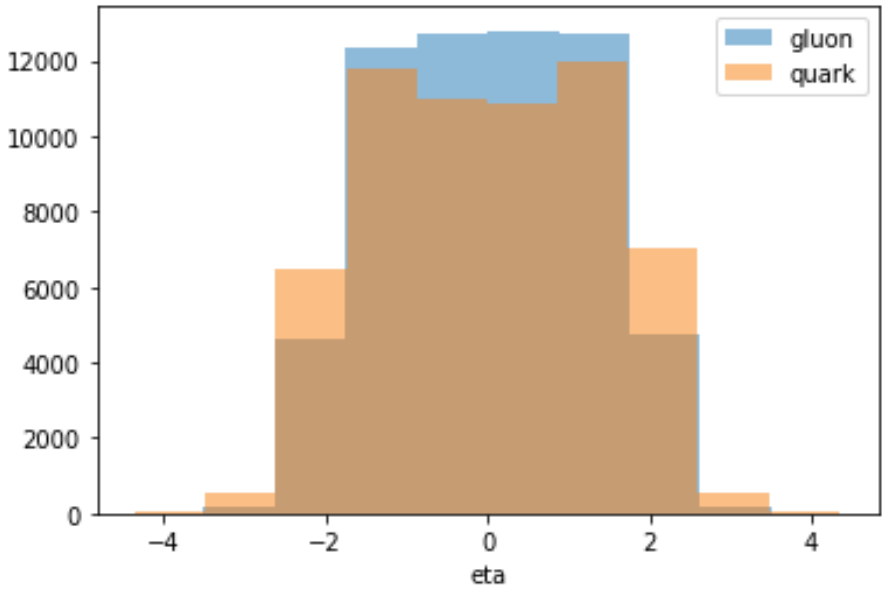
173,461 events*2



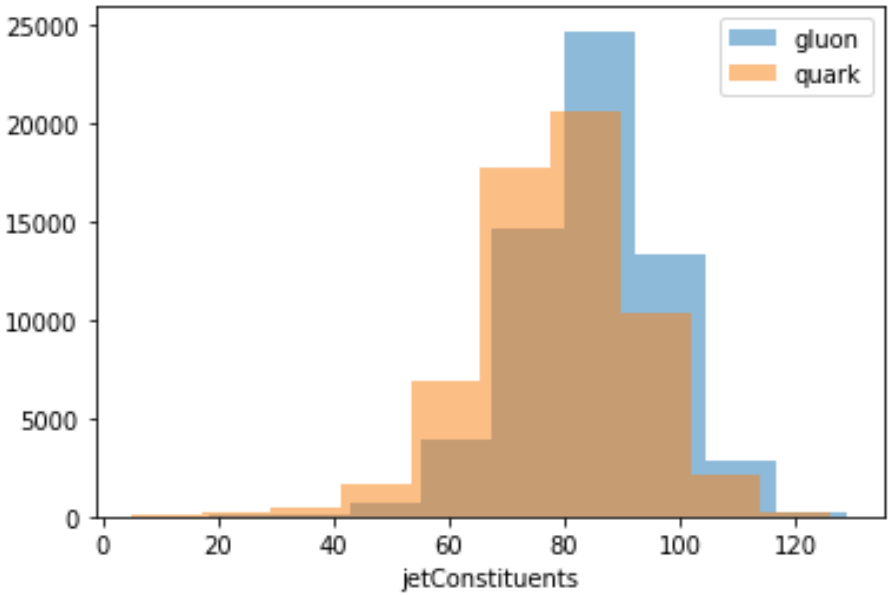


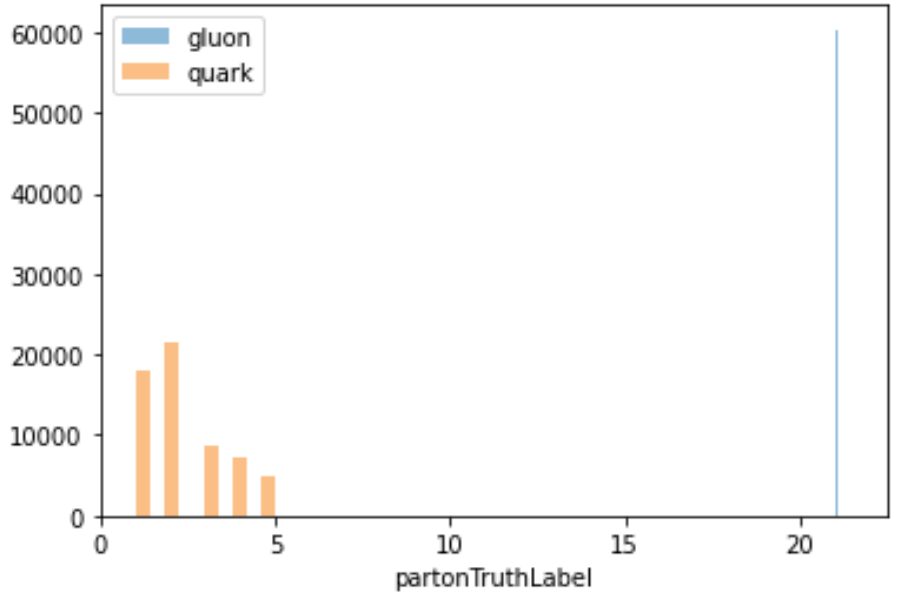
Some Basic Plots for quark and gluon jets(pT: 200-250 GeV)





60,334 events*2



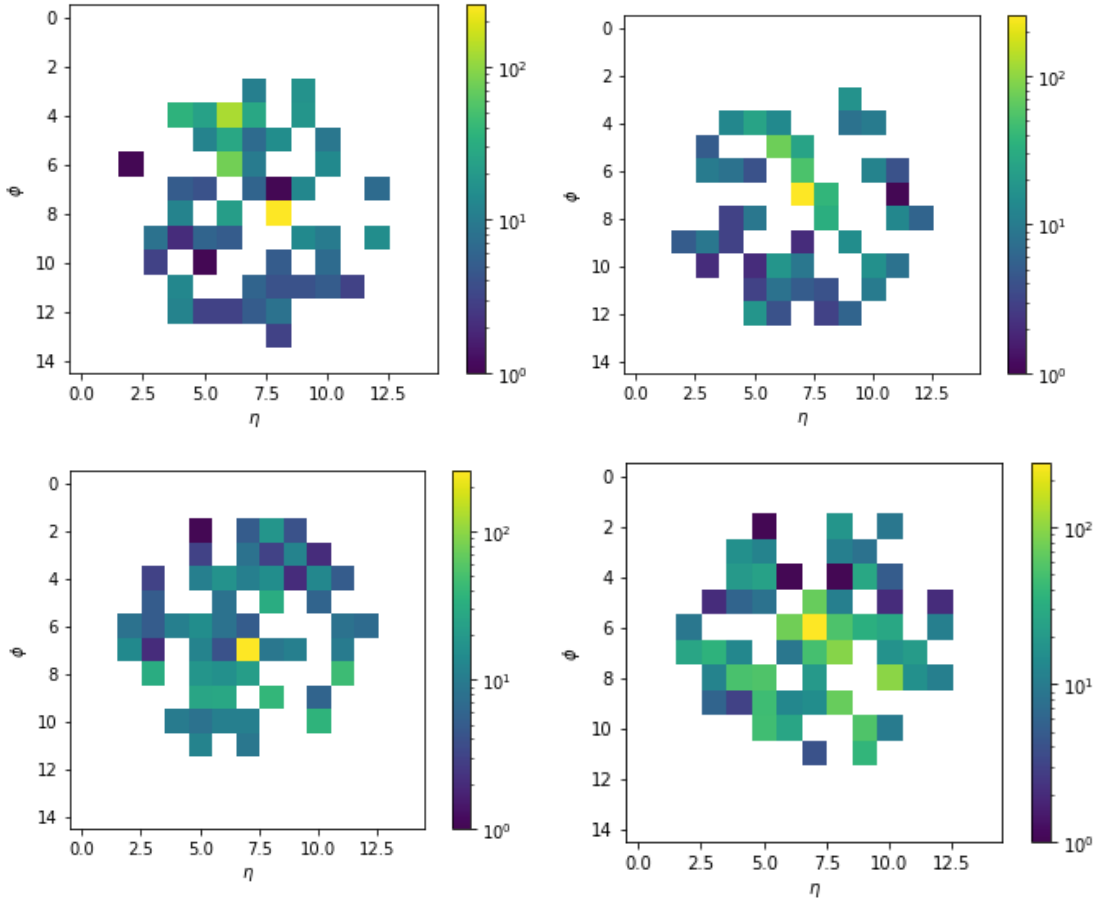


Making the Jet Image in eta-phi space

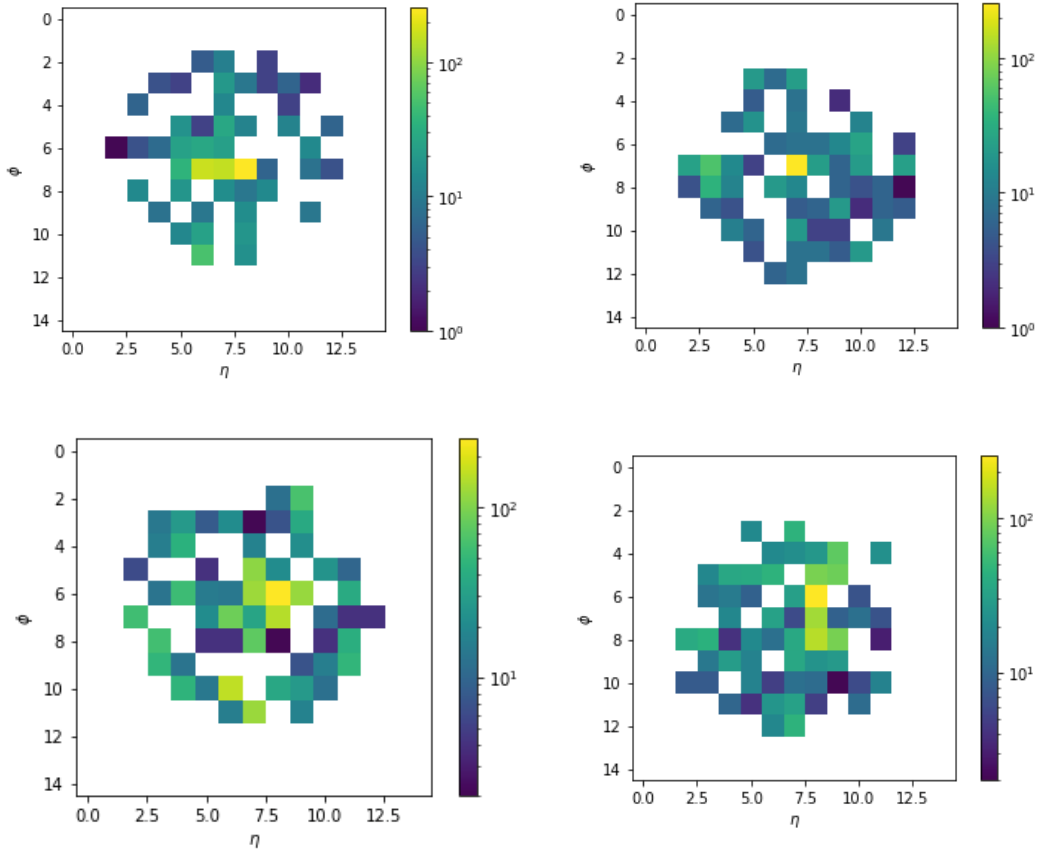
- Topological cluster inputs from calorimeter are used to make the jet image.
- The constituents inside a jet are translated in eta and phi so that the jet's center is located at center in eta-phi space.
- 15 * 15 pixel image in eta-phi space.
- The intensity of each pixel is the total cluster energy with in the pixel.
- Pixel value is then normalized by dividing it by the value of the hottest (maximum) pixel in the image. This scaling ensures that the pixel values of the entire image are between 0 and 1. Then, the pixel values are scaled to a range between 0 and 255, this is done by multiplying each pixel value by 255.

- https://github.com/sparajul/fastmachinelearning/blob/main/preProcessJetImages_santosh.py ->Prepare Image
- https://github.com/sparajul/fastmachinelearning/blob/main/prepareTestTrain_santosh.py. ->Prepare test/train
- https://github.com/sparajul/fastmachinelearning/blob/main/TrainCNN.ipynb -> Training/hls4ml

Quark and gluon Images



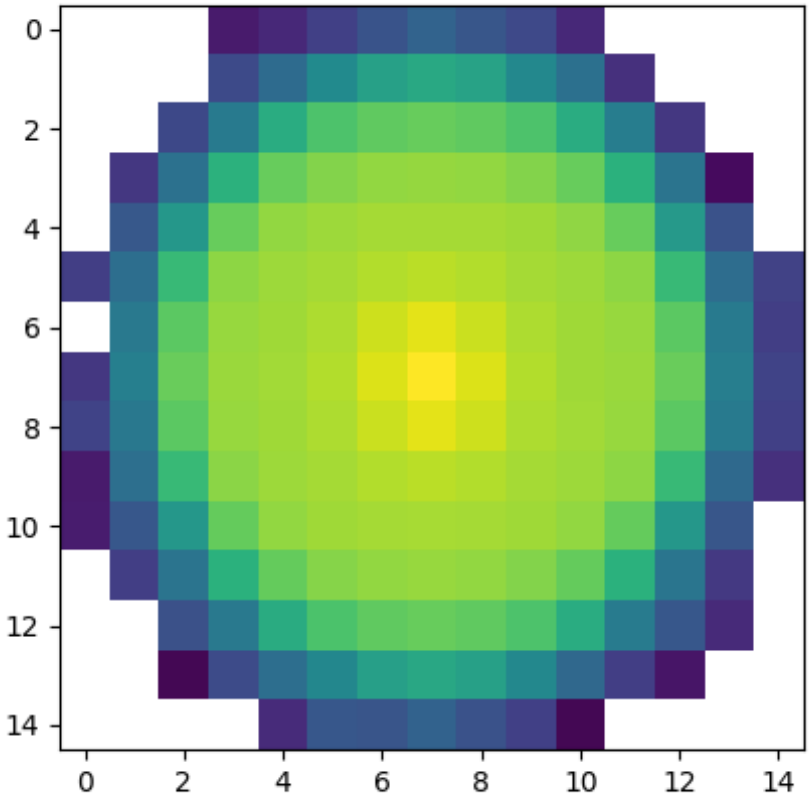
Quark Jets



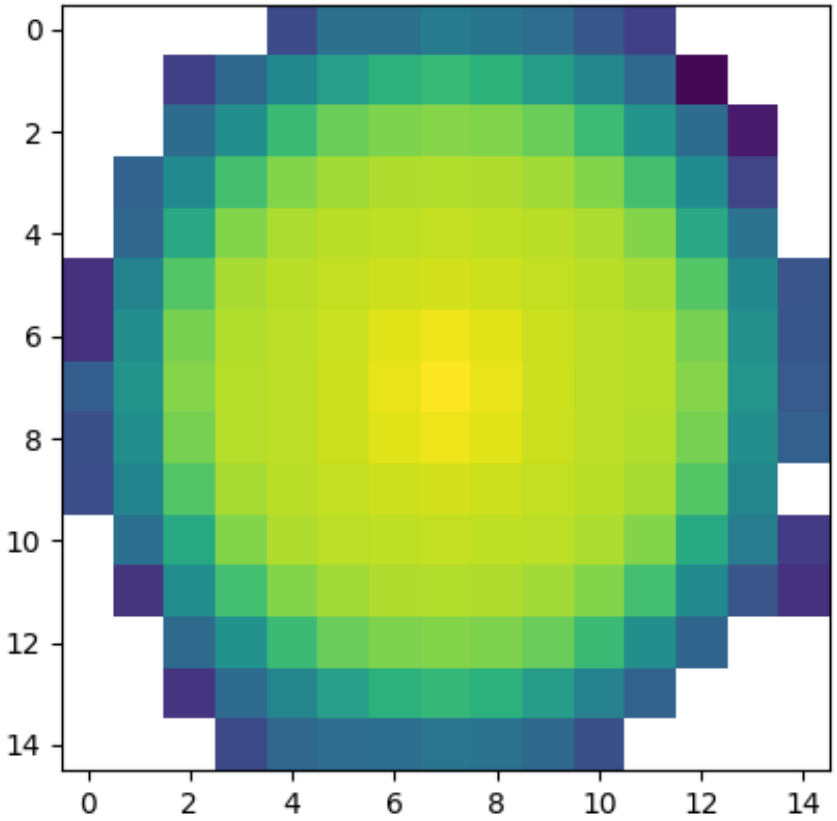
Gluon Jets

Comparison of Quark (left) and Gluon (right) jet images.

Average Images



Quark Jets



Gluon Jets

Comparison of Quark (left) and Gluon (right) jet images averaged over 60,334 jets.

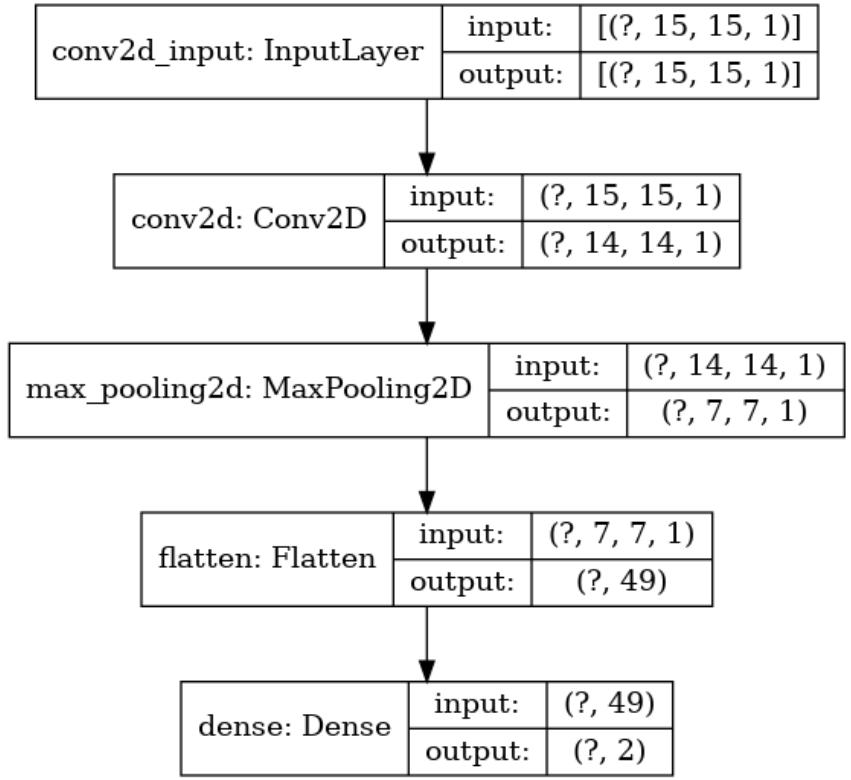
CNN Model and Performance:

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Flatten, Dropout, Activation, Conv2D, MaxPooling2D
import tensorflow as tf
model_cnn = Sequential()
model_cnn.add(Conv2D(1, (2, 2), input_shape=(15, 15,1), activation='relu'))
model_cnn.add(MaxPooling2D(pool_size=(2, 2)))


model_cnn.add(Flatten())
model_cnn.add(Dense(2, activation='softmax'))

# Compile model

model_cnn.compile(loss='categorical_crossentropy',
                  optimizer=tf.keras.optimizers.Adam(learning_rate=1.0e-4), metrics=['accuracy'])
history_cnn = model_cnn.fit(x_train, y_train, validation_split=0.2, epochs=150,
                            batch_size=256, shuffle=True, verbose=1)
model_cnn.summary()
```

1 Filter

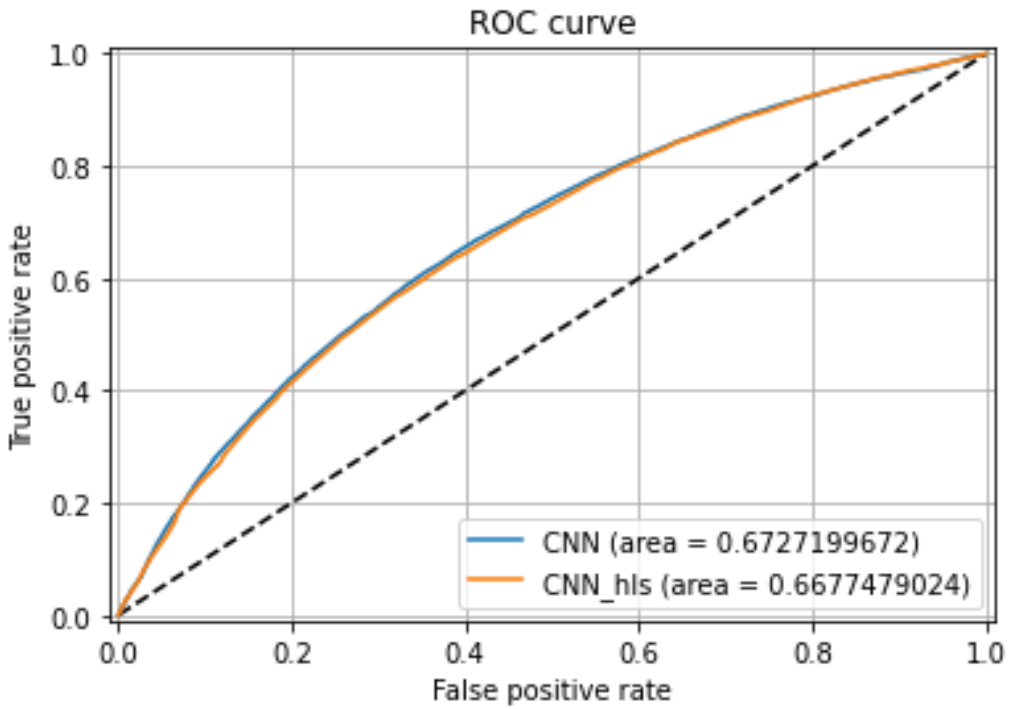
15*15 grayscale(1) image

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 14, 14, 1)         5
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 7, 7, 1)           0
_________________________________________________________________
flatten_1 (Flatten)          (None, 49)                0
_________________________________________________________________
dense_1 (Dense)              (None, 2)                 100
=================================================================
Total params: 105
Trainable params: 105
Non-trainable params: 0
```



CNN Model

HIS4ML and Performance:




```
SYNTHESIS REPORT:
================================================================
== Vivado HLS Report for 'myproject'
================================================================
* Date:           Wed Mar  8 16:46:27 2023

* Version:        2018.3 (Build 2405991 on Thu Dec 06 23:56:15 MST 2018)
* Project:        myproject_prj
* Solution:       solution1
* Product family: virtexuplus
* Target device:  xcvu13p-flga2577-2L-e
```

CNN Model/hls4ml Performance

HIS4ML and Performance:

```
{'EstimatedClockPeriod': '4.156',
 'BestLatency': '233',
 'WorstLatency': '233',
 'IntervalMin': '229',
 'IntervalMax': '229',
 'BRAM_18K': '5',
 'DSP48E': '85',
 'FF': '2035',
 'LUT': '4967',
 'URAM': '0',
 'AvailableBRAM_18K': '5376',
 'AvailableDSP48E': '12288',
 'AvailableFF': '3456000',
 'AvailableLUT': '1728000',
 'AvailableURAM': '1280'}
```

```
================================================================
== Performance Estimates
================================================================
+ Timing (ns):
    * Summary:
    +--------+-------+----------+------------+
    |  Clock | Target| Estimated| Uncertainty|
    +--------+-------+----------+------------+
    |ap_clk  |   5.00|     4.156|        0.62|
    +--------+-------+----------+------------+

+ Latency (clock cycles):
    * Summary:
    +-----+-----+-----+-----+----------+
    |  Latency  |  Interval | Pipeline |
    | min | max | min | max |   Type   |
    +-----+-----+-----+-----+----------+
    |  233|  233|  229|  229| dataflow |
    +-----+-----+-----+-----+----------+
```

Latency ~1.2 μs

hls4ml Performance

HIS4ML and Performance:

```
================================================================
== Utilization Estimates
================================================================
* Summary:
+-----------------+---------+-------+---------+---------+------+
|       Name      | BRAM_18K| DSP48E|    FF   |   LUT   | URAM |
+-----------------+---------+-------+---------+---------+------+
|DSP              |        -|      -|        -|        -|     -|
|Expression       |        -|      -|        0|       32|     -|
|FIFO             |        3|      -|      122|      186|     -|
|Instance         |        2|     85|     1907|     4713|     -|
|Memory           |        -|      -|        -|        -|     -|
|Multiplexer      |        -|      -|        -|       36|     -|
|Register         |        -|      -|        6|        -|     -|
+-----------------+---------+-------+---------+---------+------+
|Total            |        5|     85|     2035|     4967|     0|
+-----------------+---------+-------+---------+---------+------+
|Available        |     5376|  12288|  3456000|  1728000|  1280|
+-----------------+---------+-------+---------+---------+------+
|Utilization (%)  |    ~0   |   ~0  |    ~0   |    ~0   |     0|
+-----------------+---------+-------+---------+---------+------+
```

hls4ml Performance

THANK YOU!