Semi-Visible Jets

David Lai

Overview

- current progress (kinematic plots)
- brief overview of autoencoder paper
- high level features

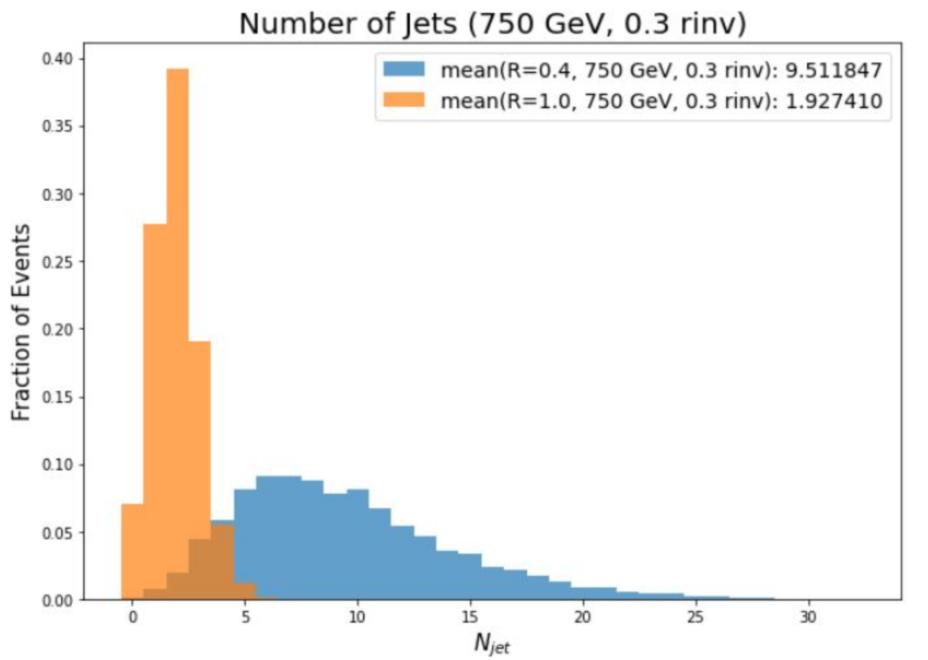
Progress

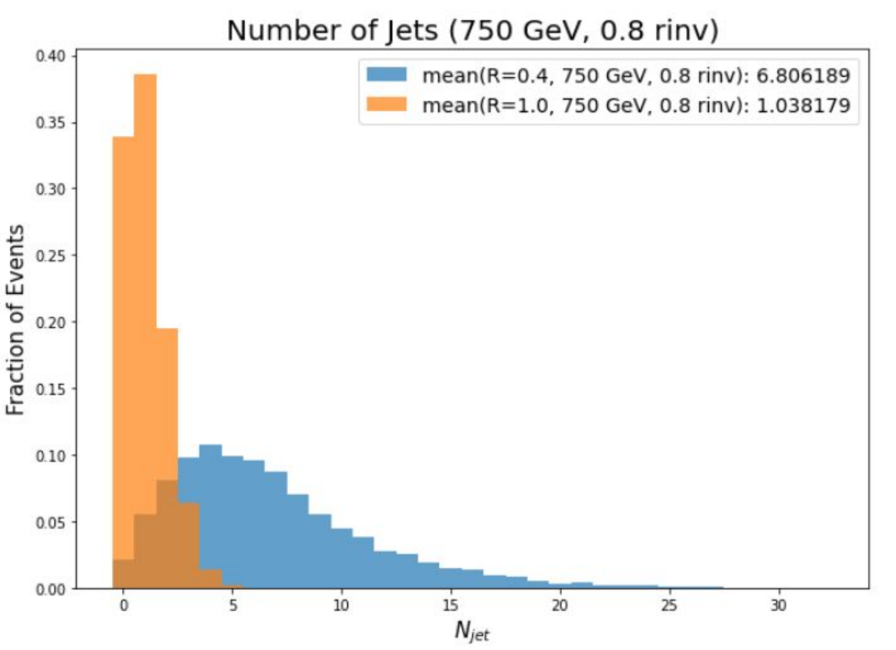
- Download/rucio get DAOD files from Joe's Dataset (successful)
- Converting DAOD to nTuple files (successful)
  - 750 GeV & 0.3 rinv, 750 GeV & 0.8 rinv, 1500 GeV & 0.3 rinv, 1500 GeV & 0.8 rinv.
- Make Selection Cuts (complete)
- Plot Kinematic Plots (complete)
- read the paper [Autoencoders for Semi Visible Jet Detection] (working)
  - It details some nice variables -> an input to a jet tagger.

Cut Flow Table

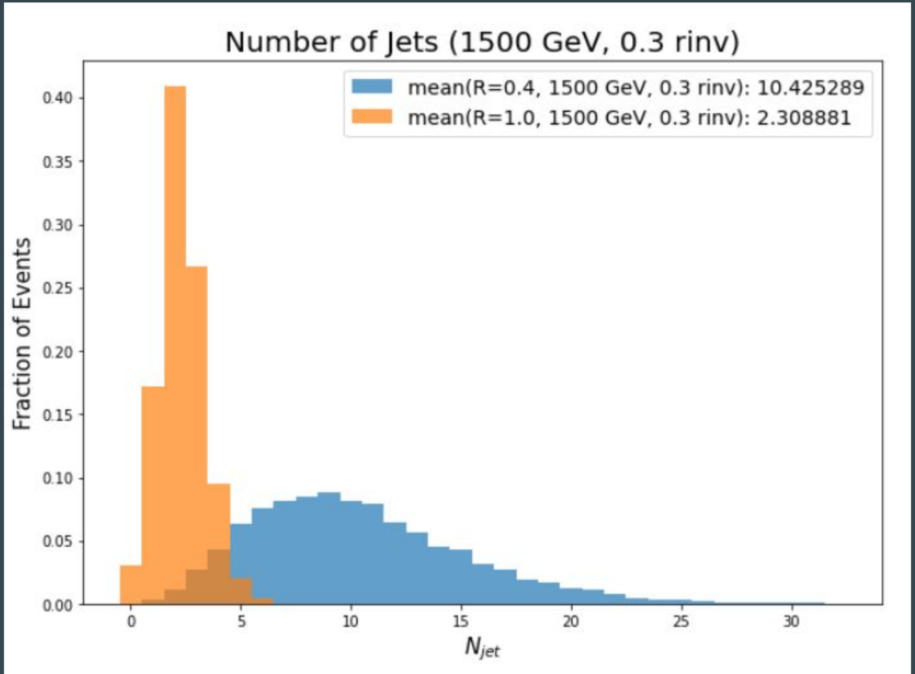
	Selection Cut (750 GeV, 0.3 rinv)	R=0.4 (750 GeV, 0.3 rinv)	R=1.0 (750 GeV, 0.3 rinv)	Selection Cut (750 GeV, 0.8 rinv)	R=0.4 (750 GeV, 0.8 rinv)	R=1.0 (750 GeV, 0.8 rinv)	Selection Cut (1500 GeV, 0.3 rinv)	R=0.4 (1500 GeV, 0.3 rinv)	R=1.0 (1500 GeV, 0.3 rinv)	Selection Cut (1500 GeV, 0.8 rinv)	R=0.4 (1500 GeV, 0.8 rinv)	R=1.0 (1500 GeV, 0.8 rinv)
0	Input Event Size	9960	9960	Input Event Size	9953	9953	Input Event Size	9965	9965	Input Event Size	9960	9960
1	Number of Jet >= 2	9865	6494	Number of Jet >= 2	9187	2743	Number of Jet >= 2	9921	7946	Number of Jet >= 2	9422	4024
2	Jet PT > 25 GeV	70	6483	Jet PT > 25 GeV	108	2735	Jet PT > 25 GeV	72	7935	Jet PT > 25 GeV	120	4015
3	|eta| < 2.5	47	4970	|eta| < 2.5	84	2016	|eta| < 2.5	60	6463	|eta| < 2.5	89	3081

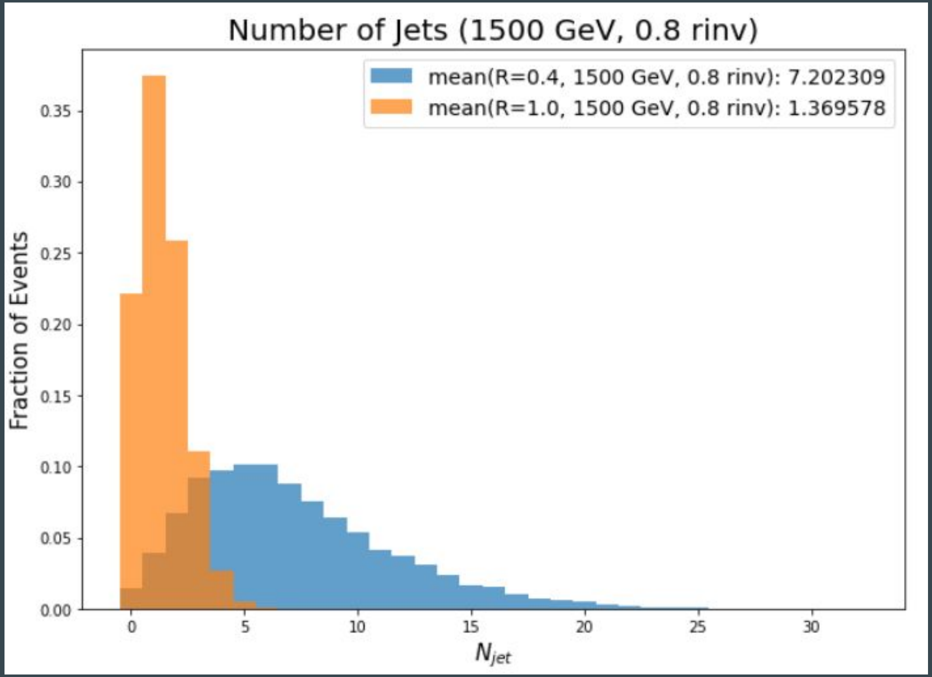
Number of Jets



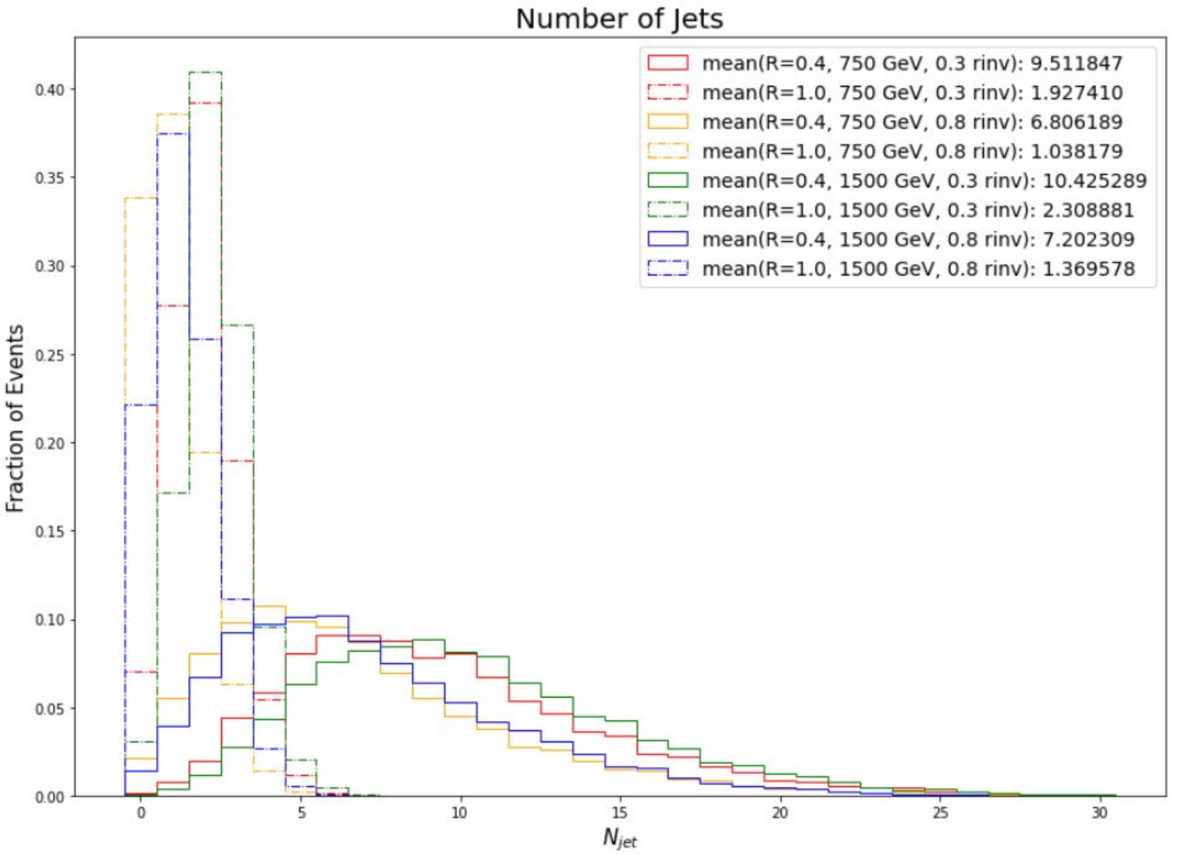


Number of Jets

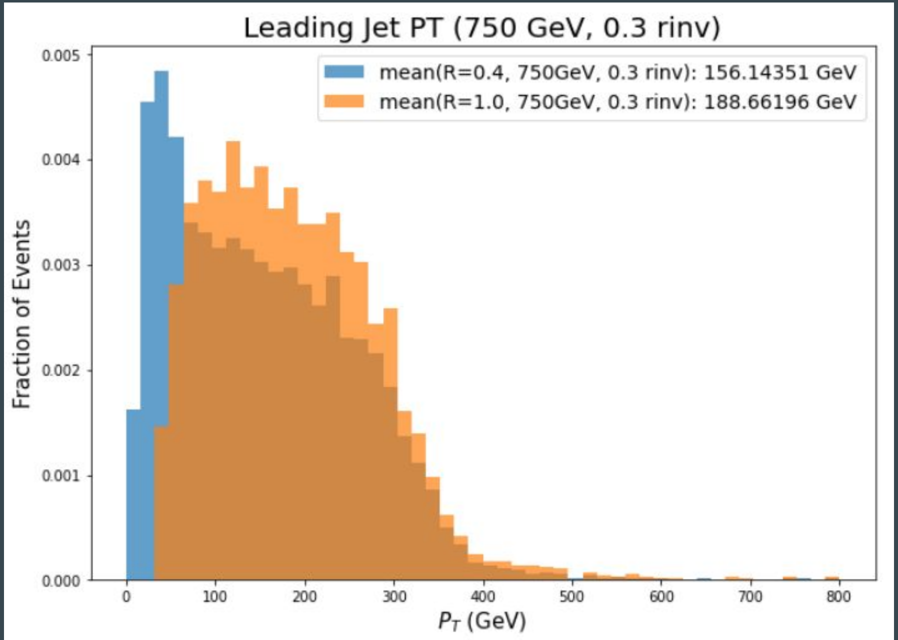


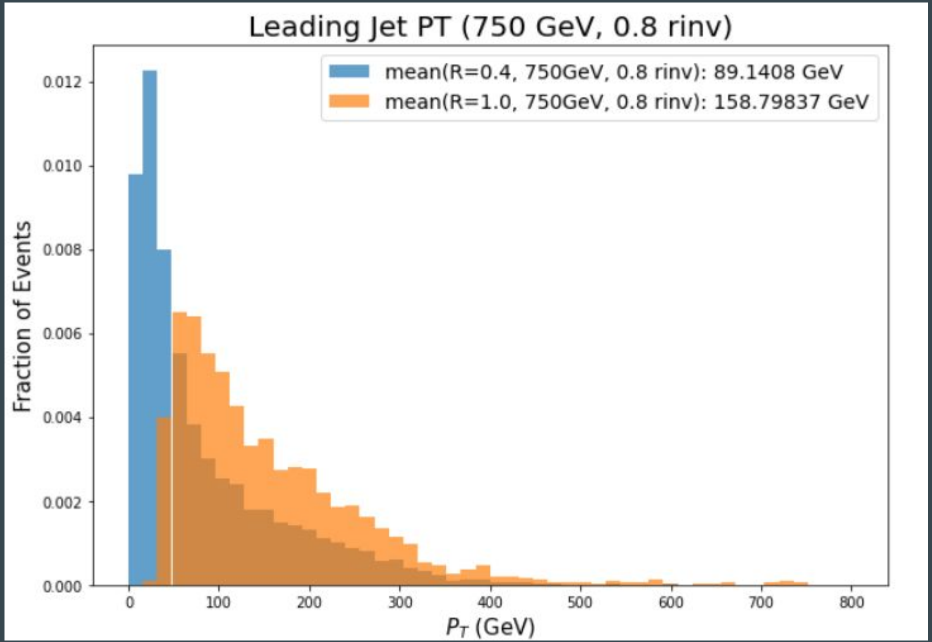


Number of Jets

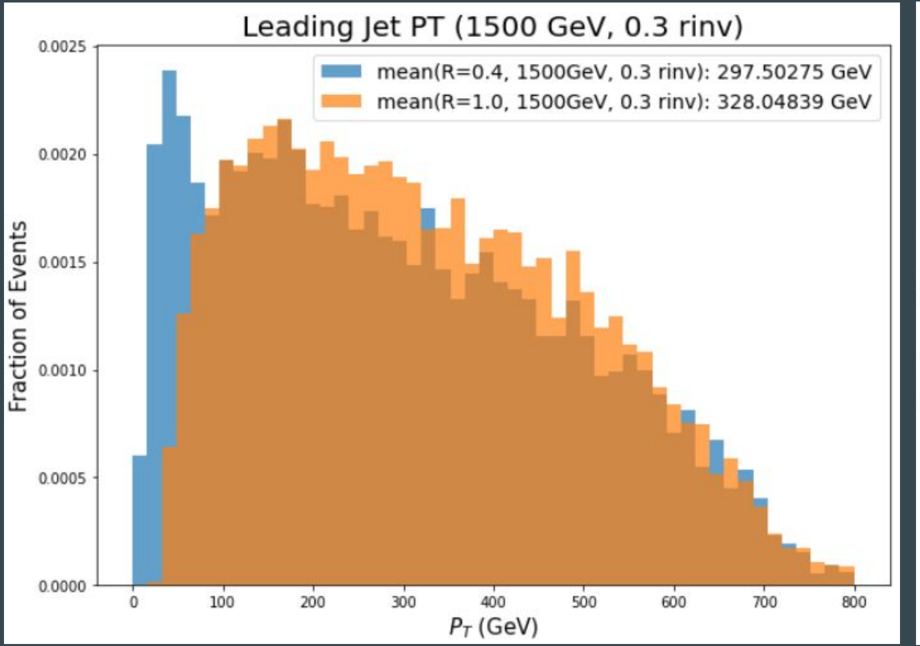


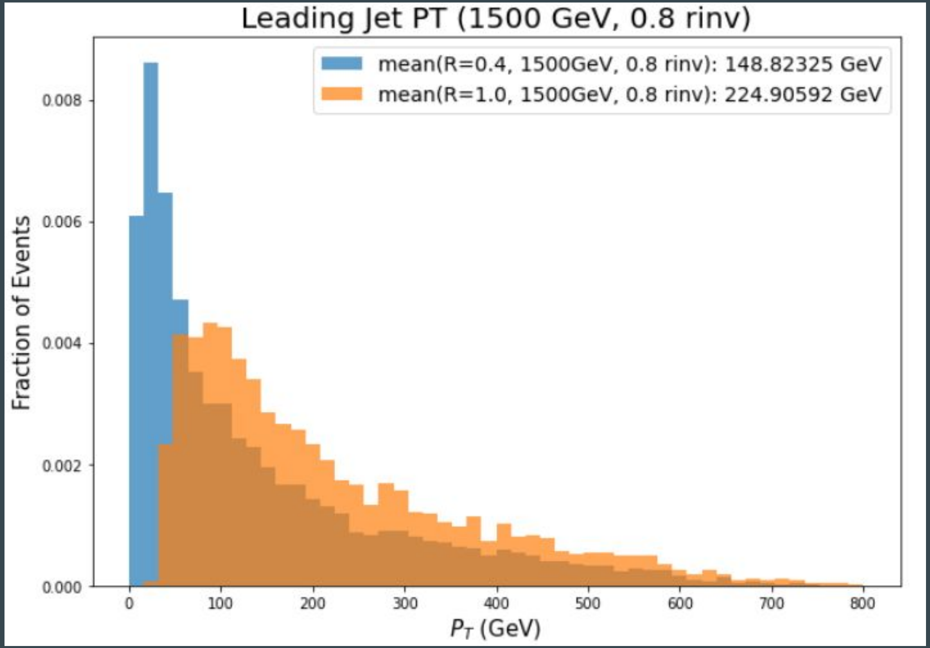
Leading Jet PT



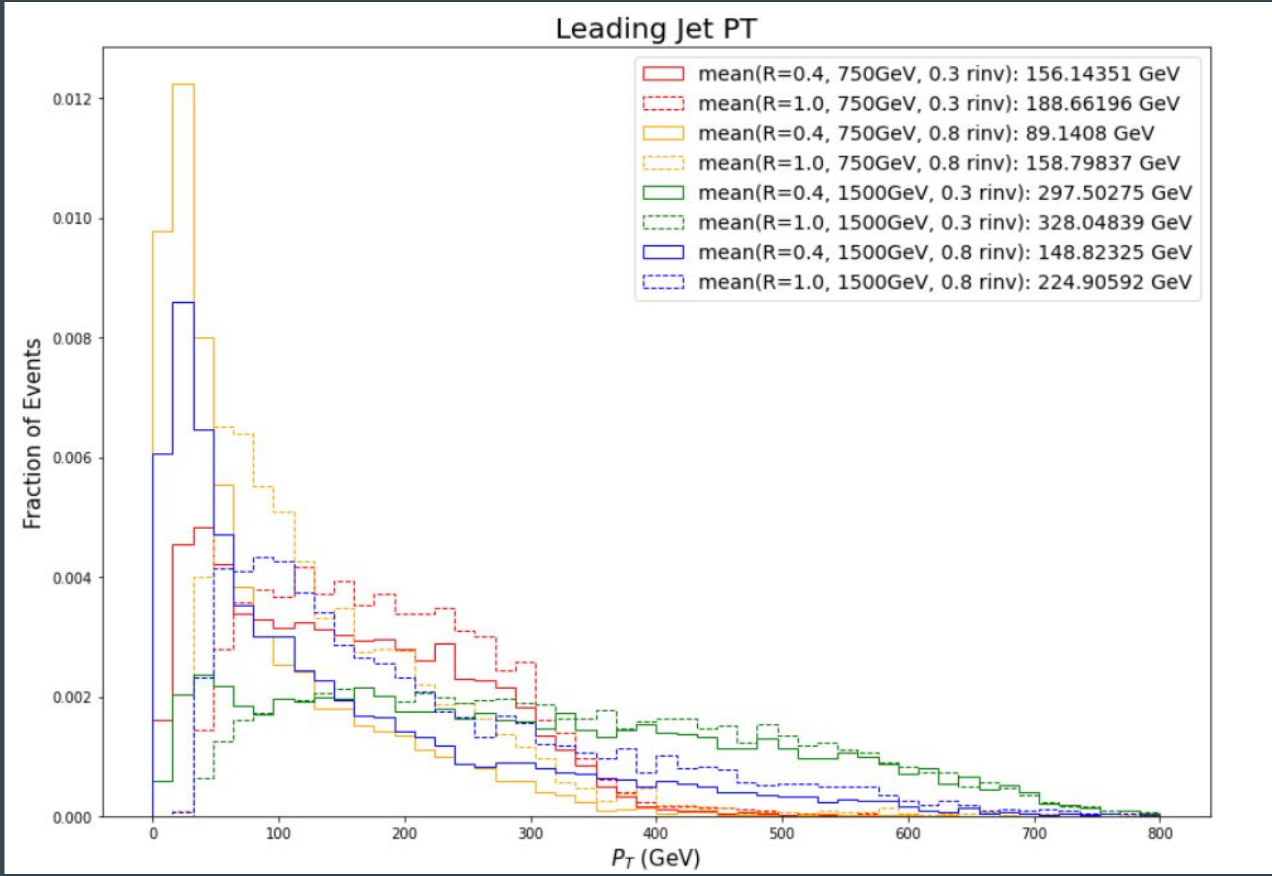


Leading Jet PT

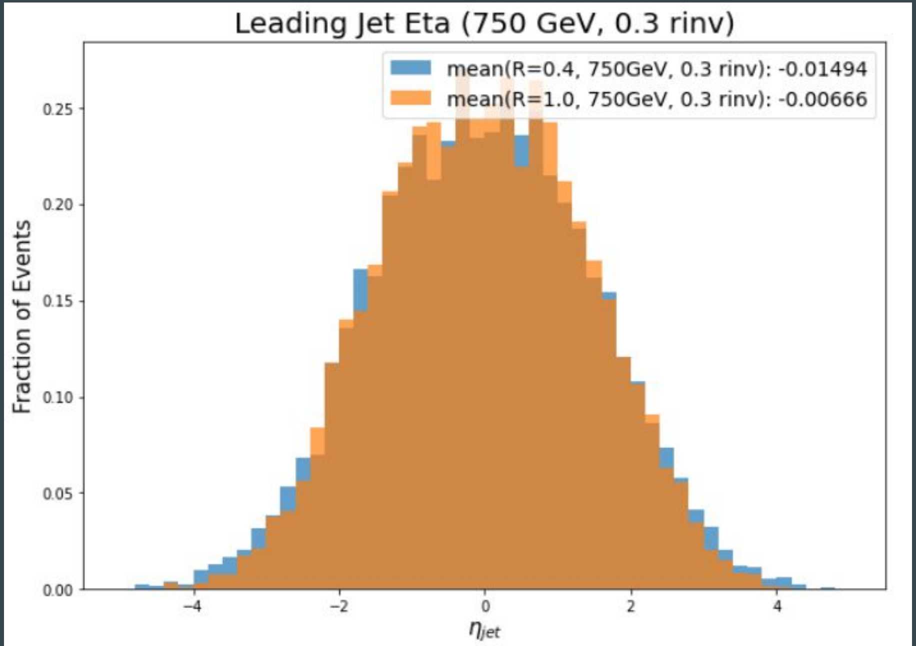


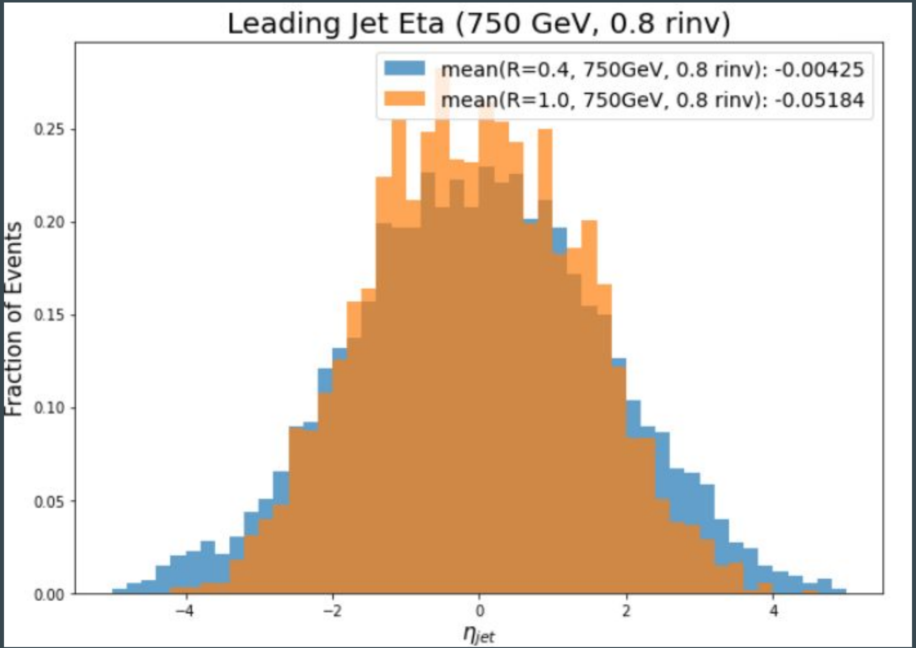


Leading Jet PT

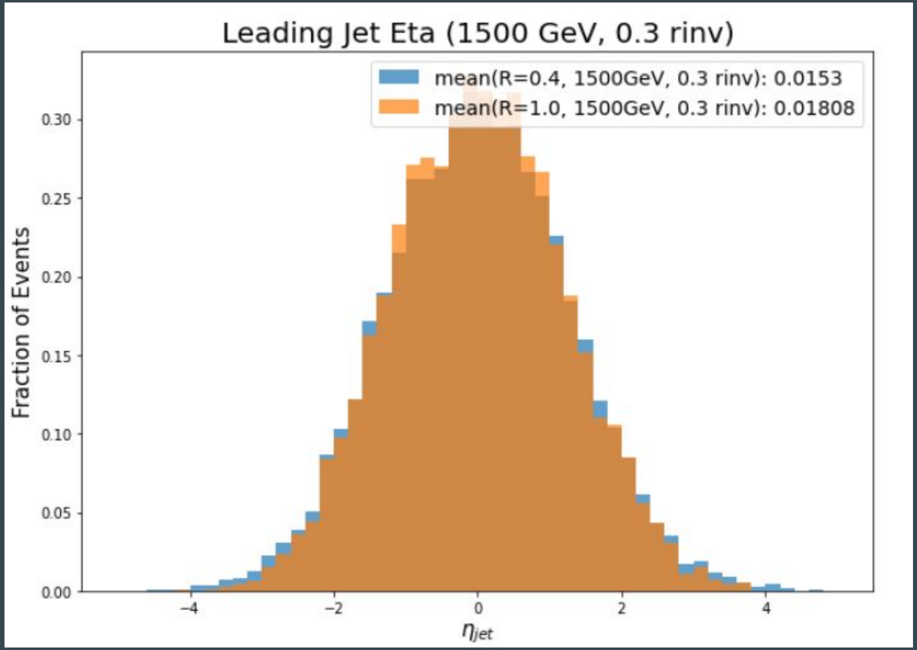


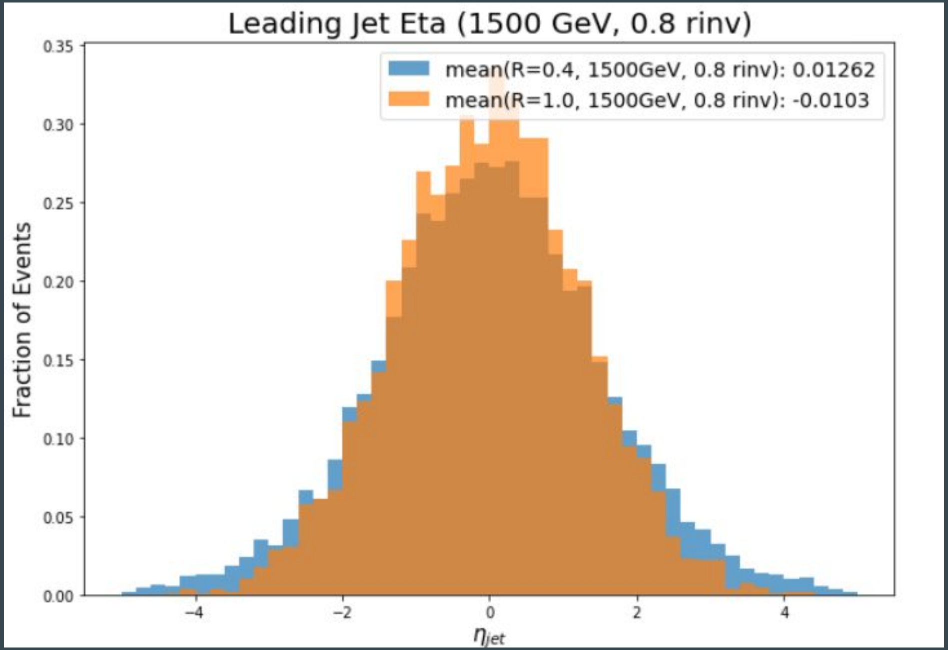
Leading Jet Eta



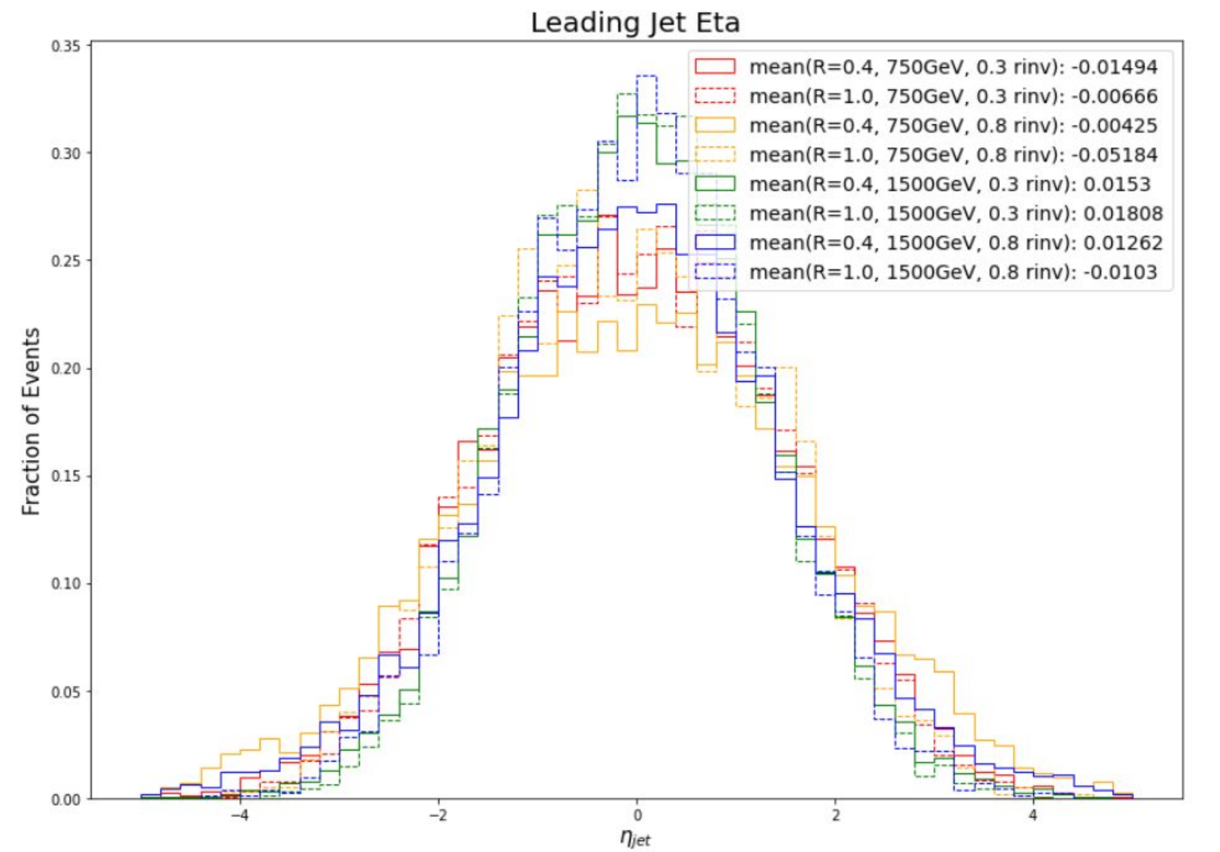


Leading Jet Eta

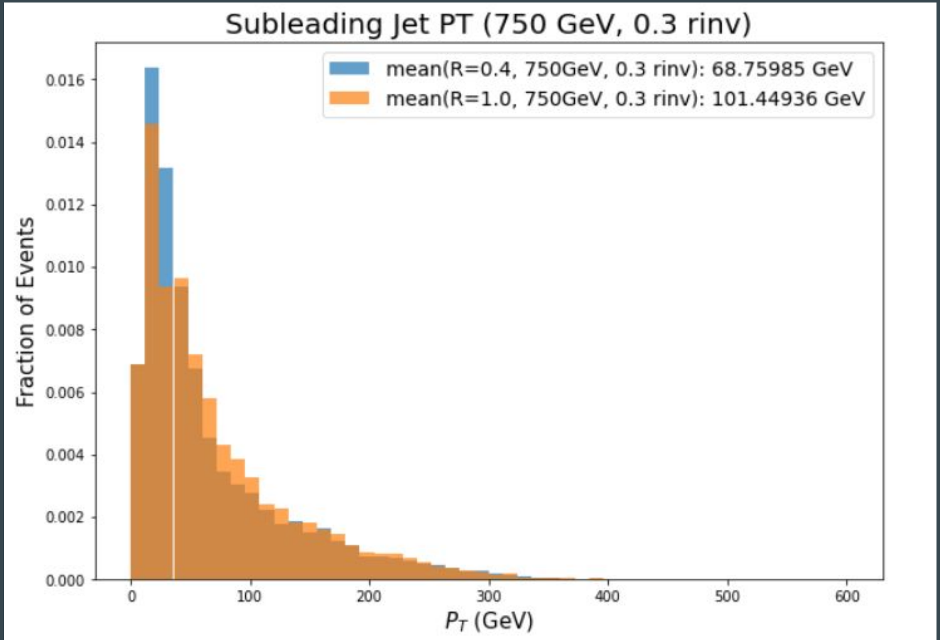


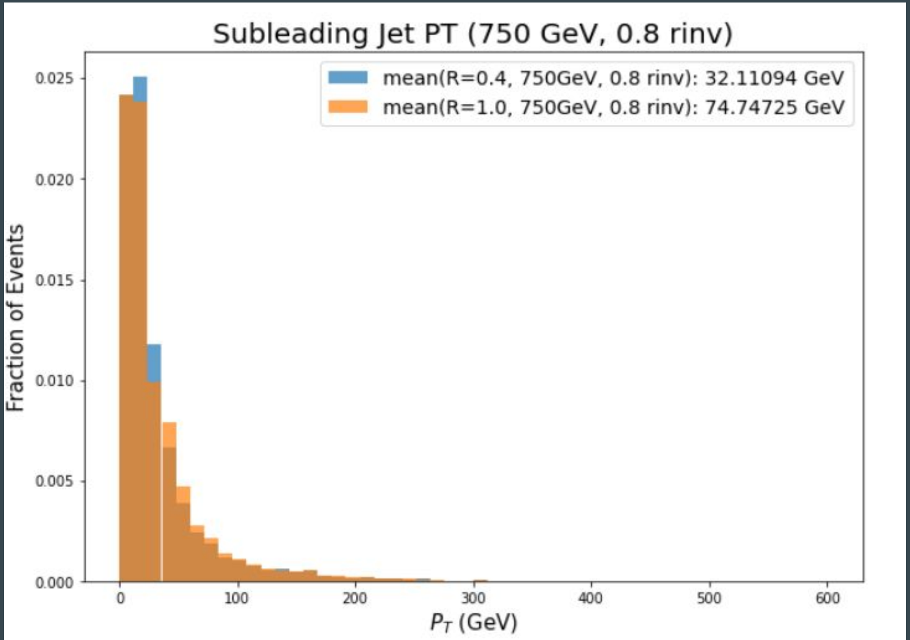


Leading Jet Eta

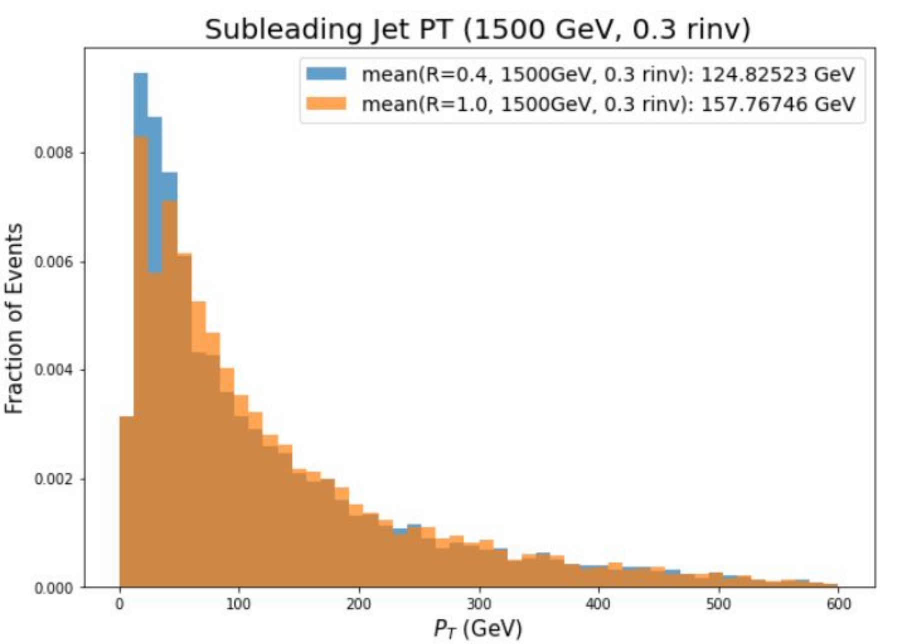


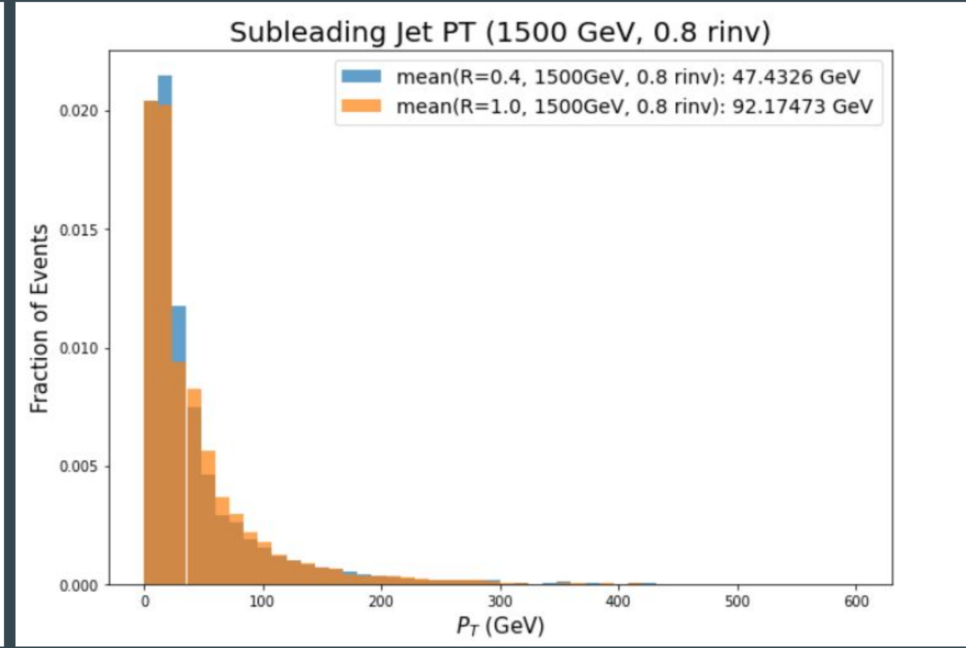
Subleading Jet PT



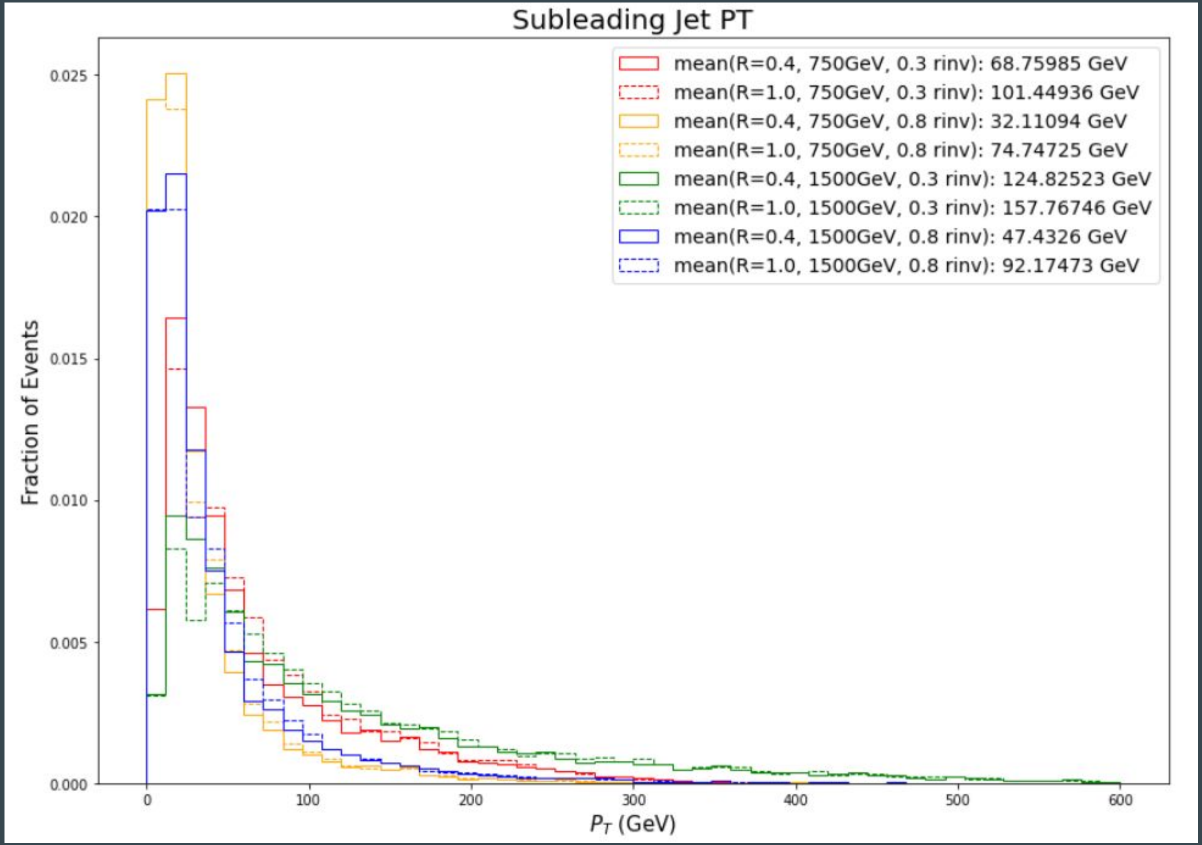


Subleading Jet PT

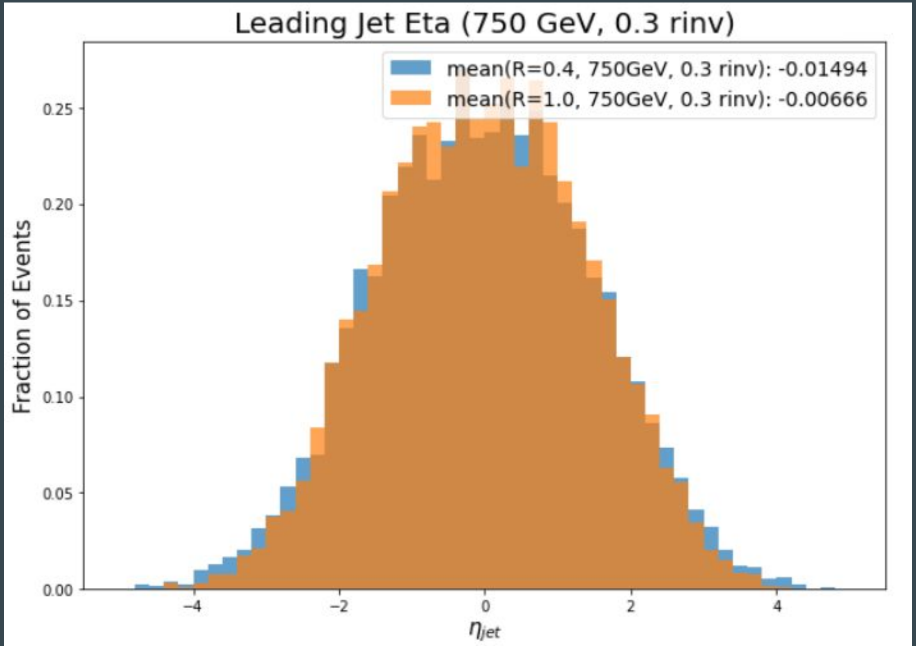


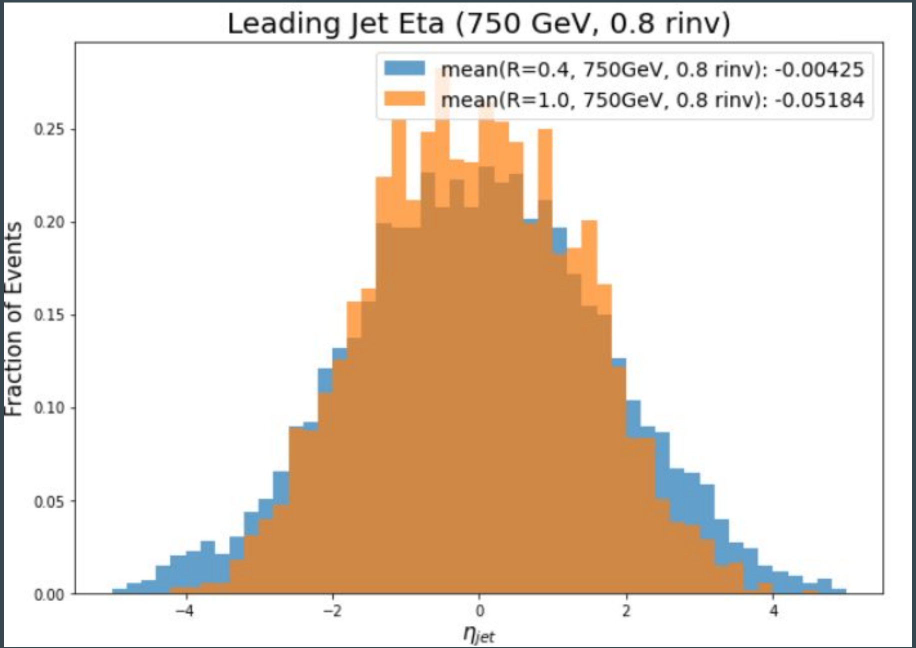


Subleading Jet PT

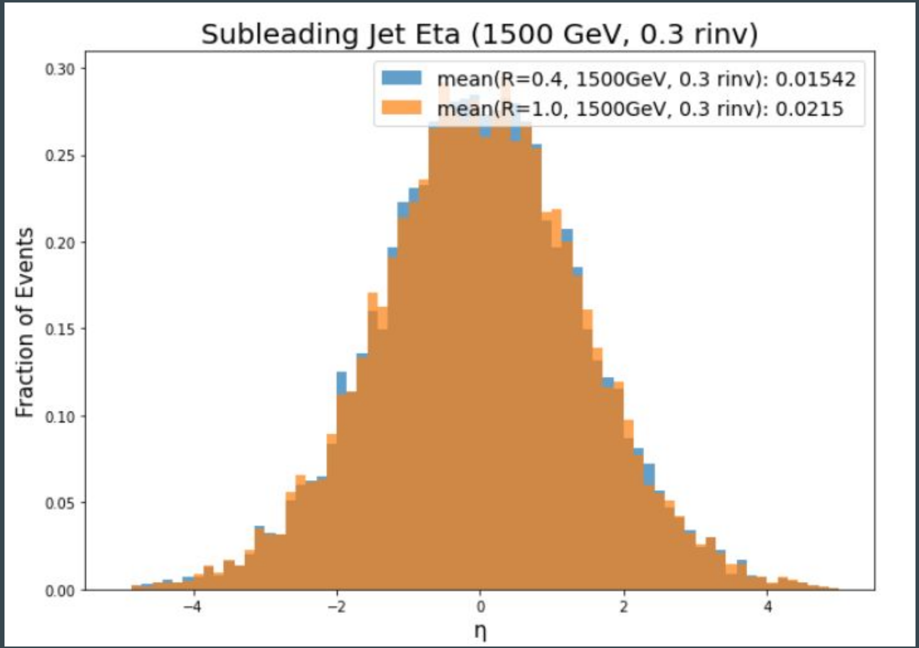


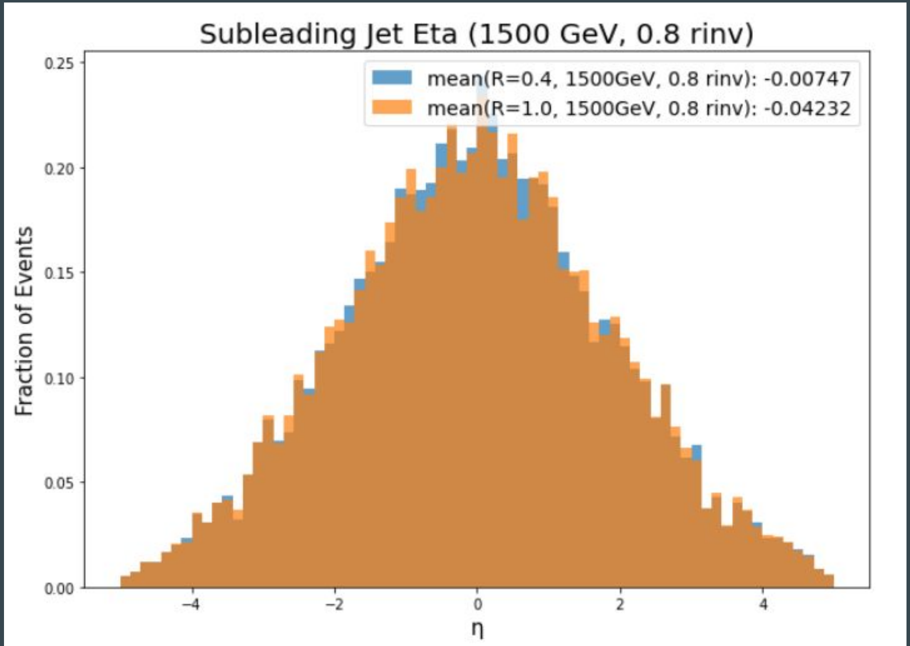
Subleading Jet Eta



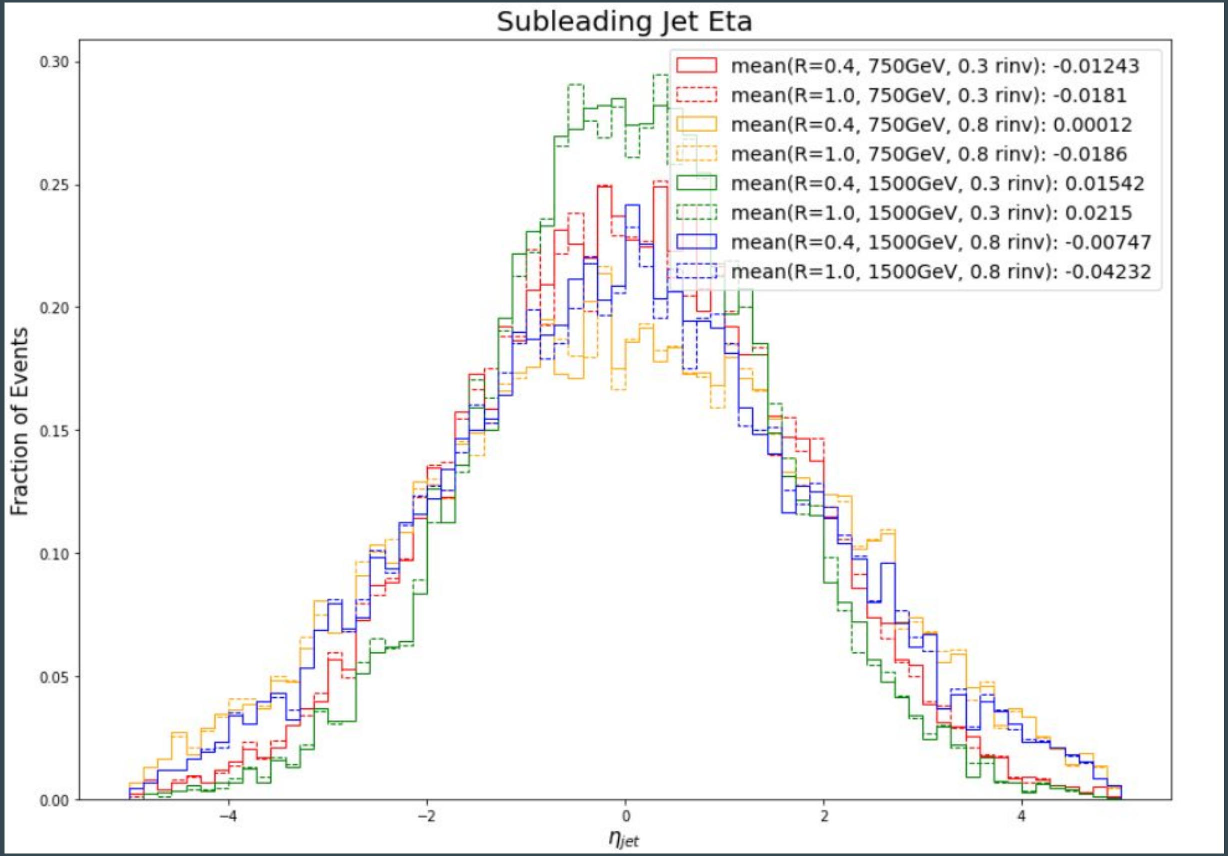


Subleading Jet Eta

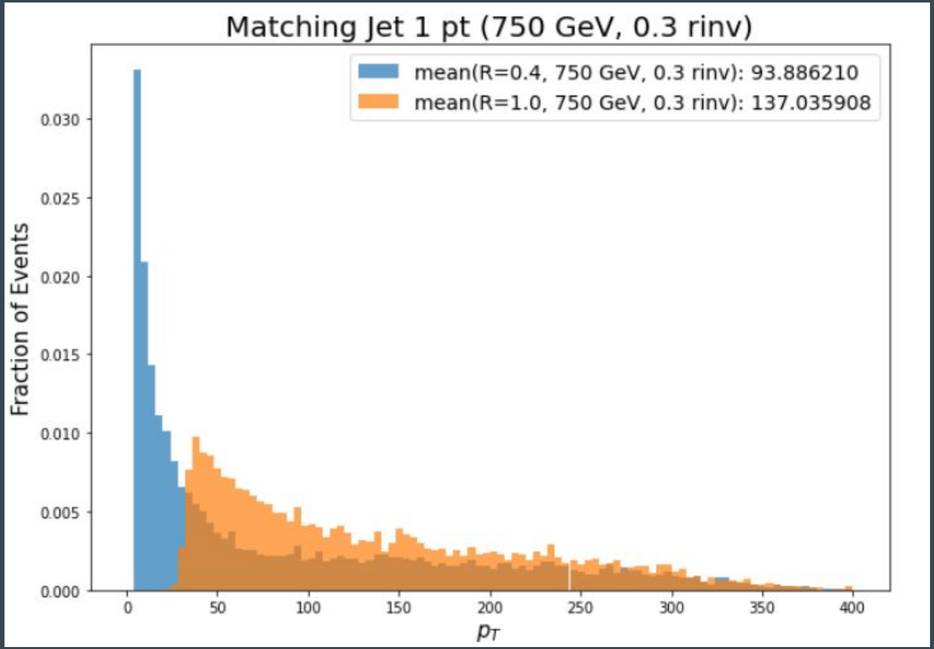


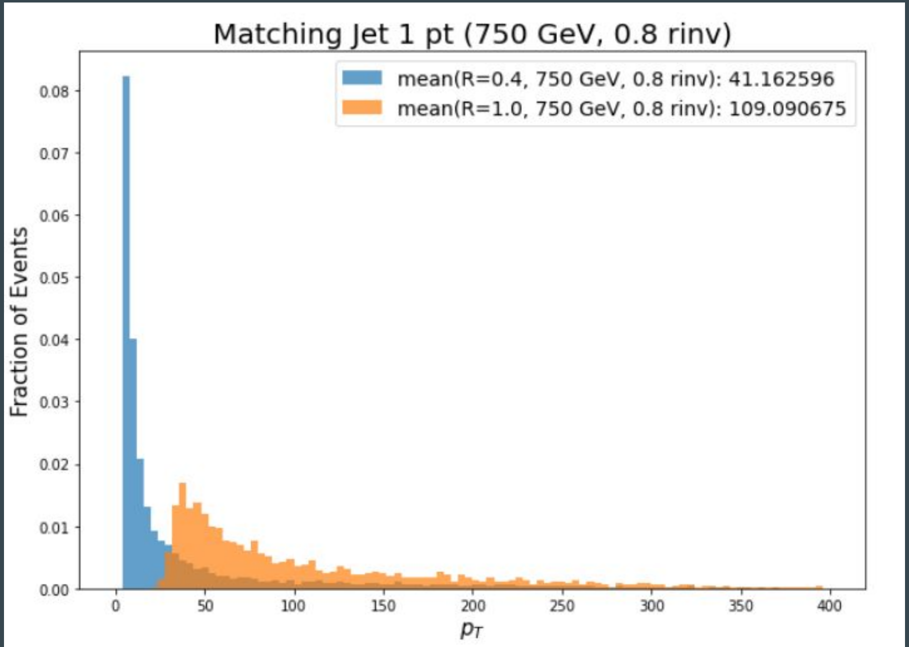


Subleading Jet Eta

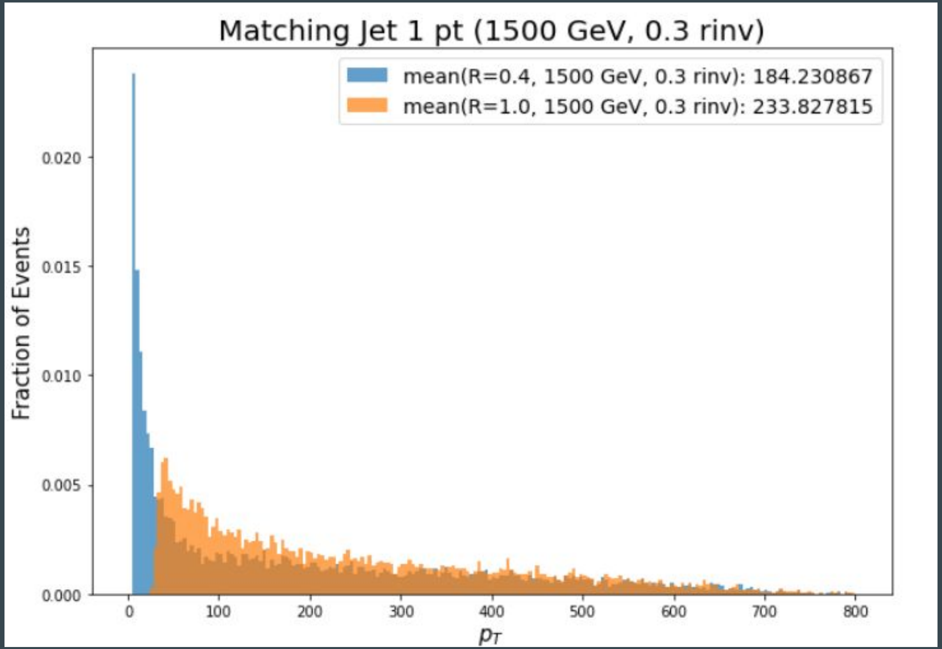


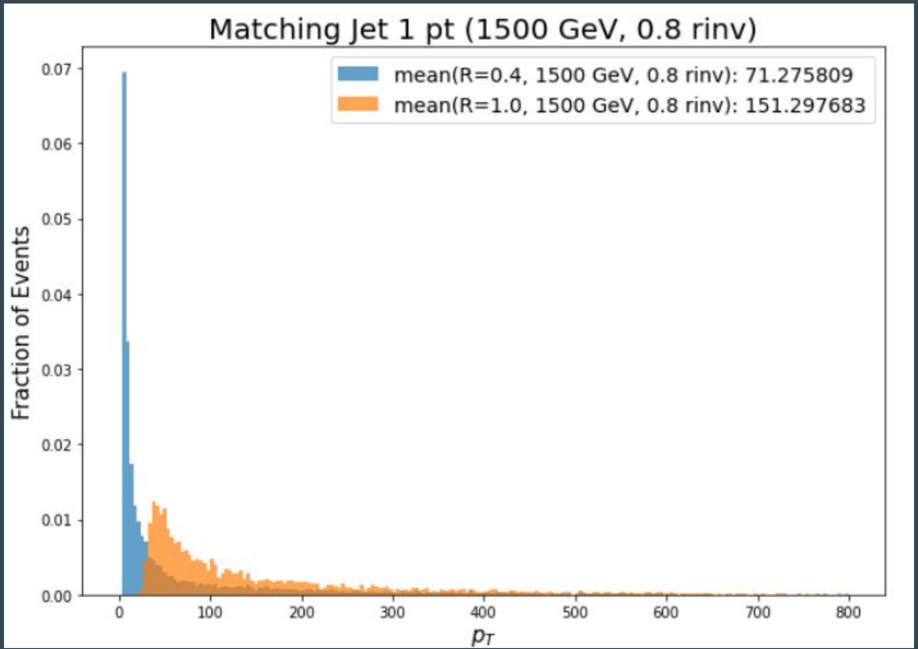
Matching Jet 1 pt



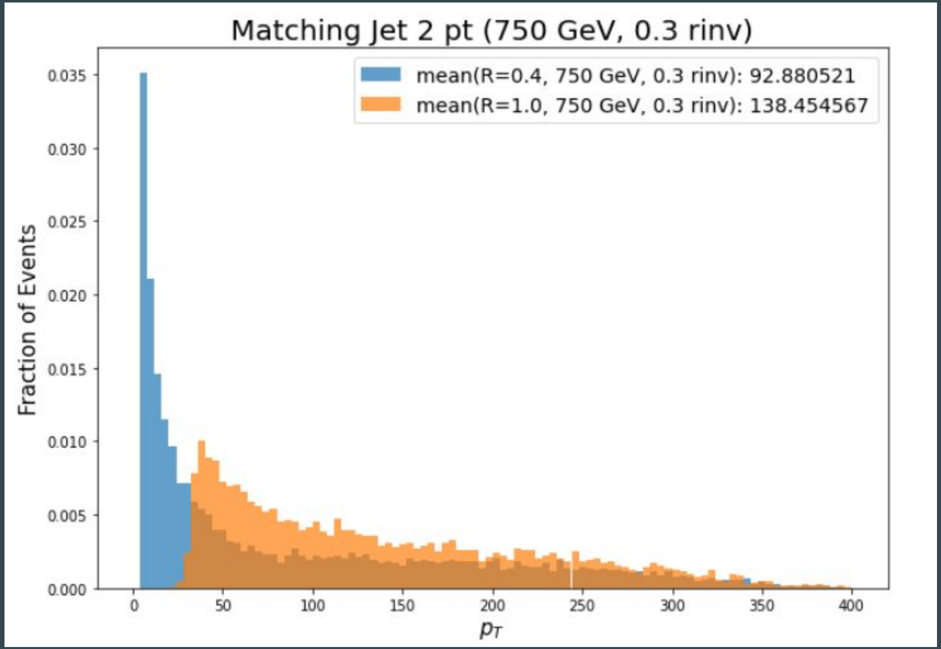


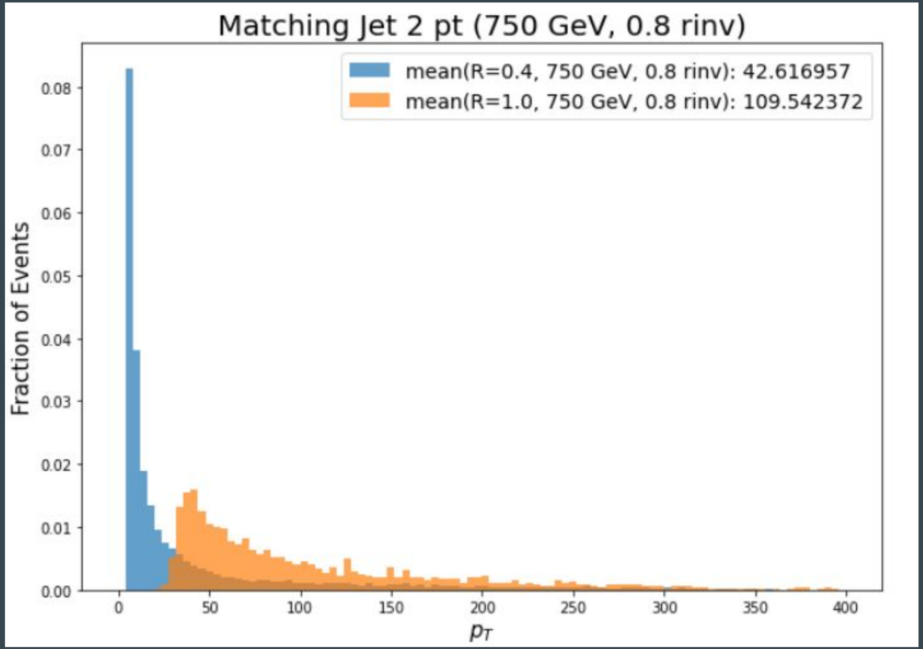
Matching Jet 1 pt



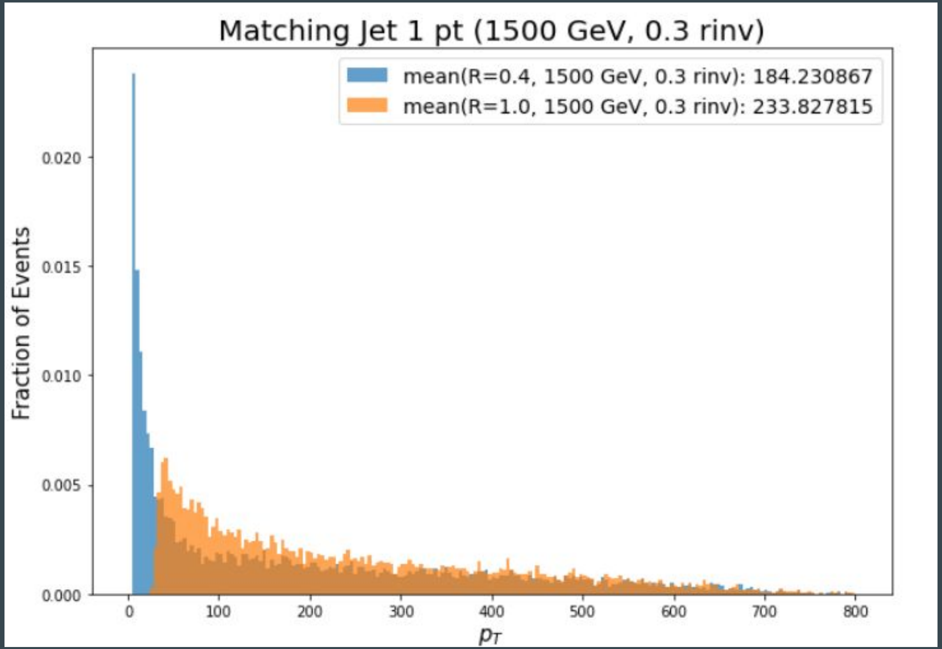


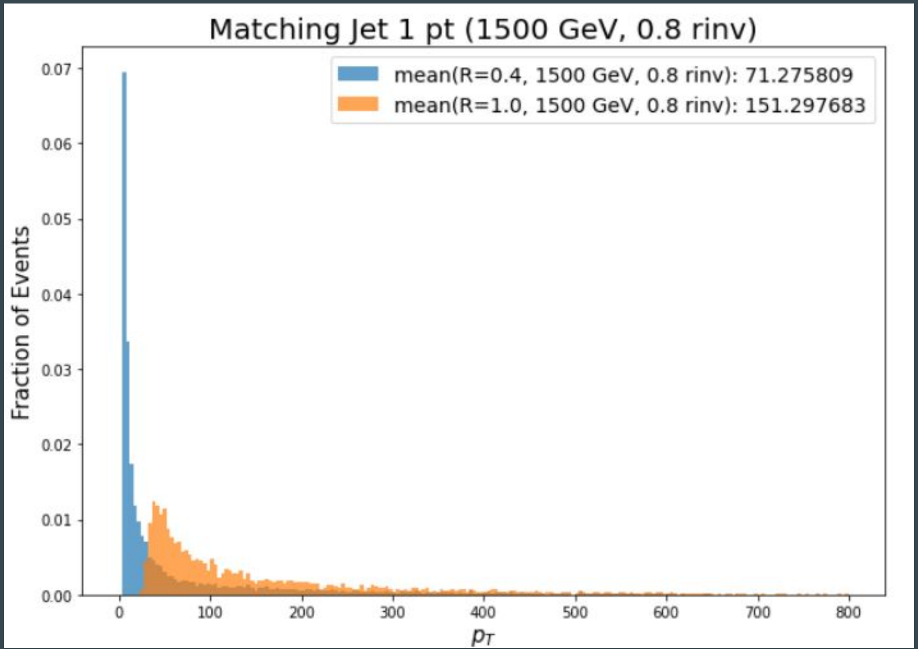
Matching Jet 2 pt



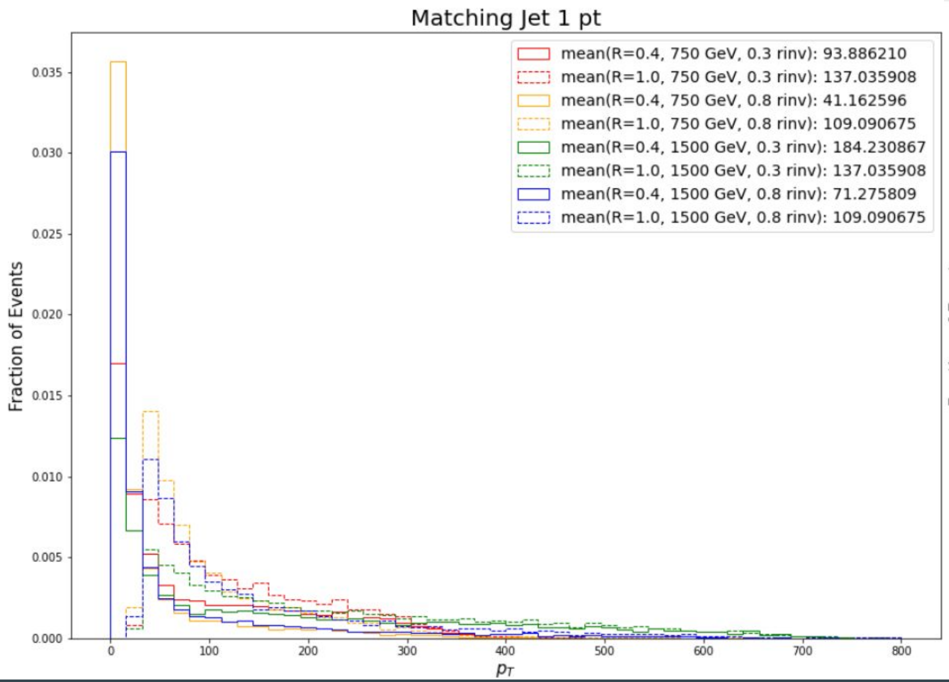


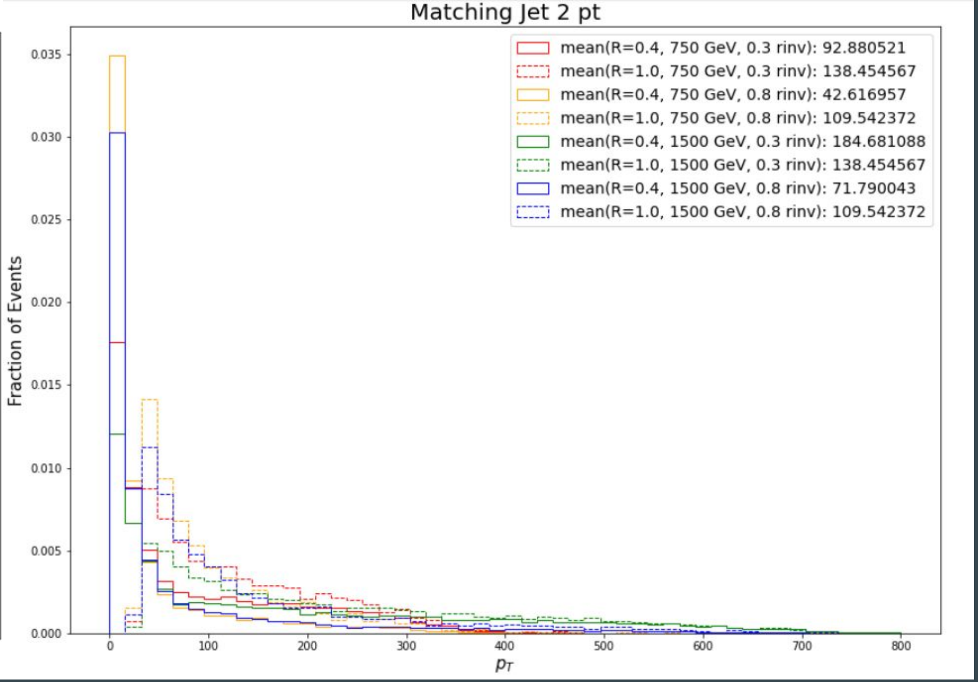
Matching Jet 2 pt



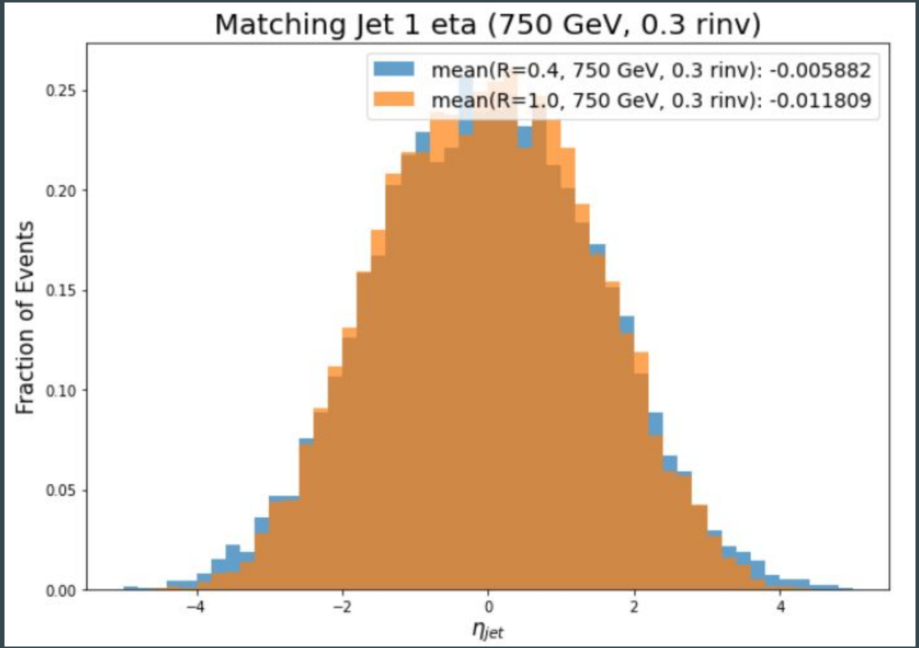


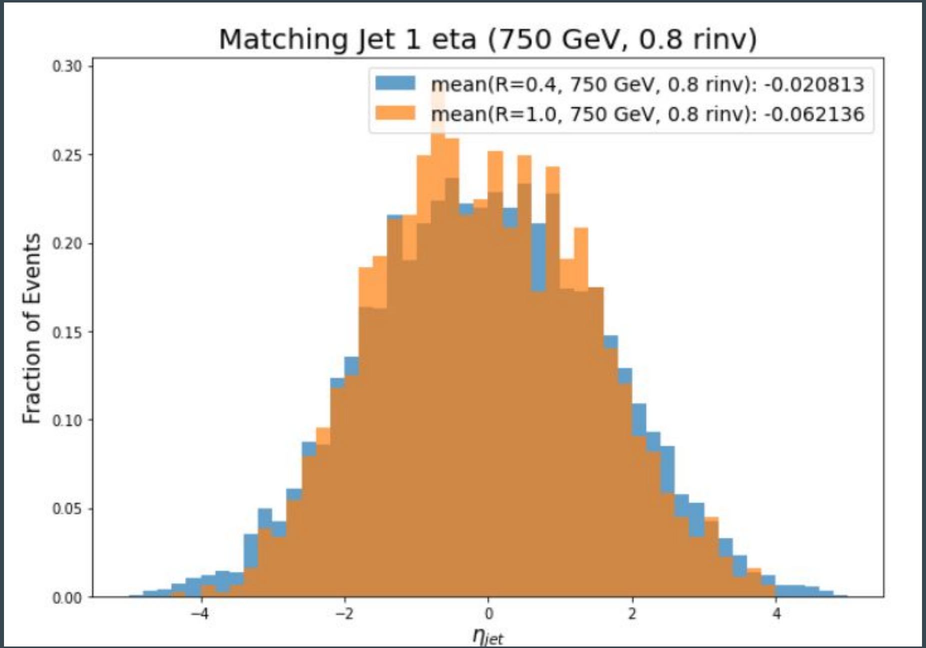
Matching Jet 1 pt & Matching Jet 2 pt



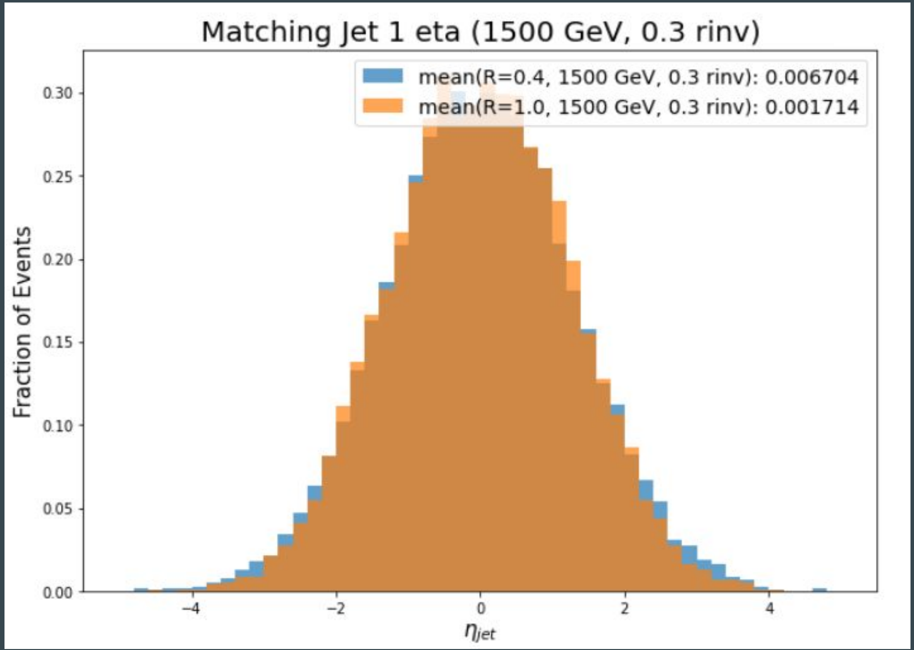


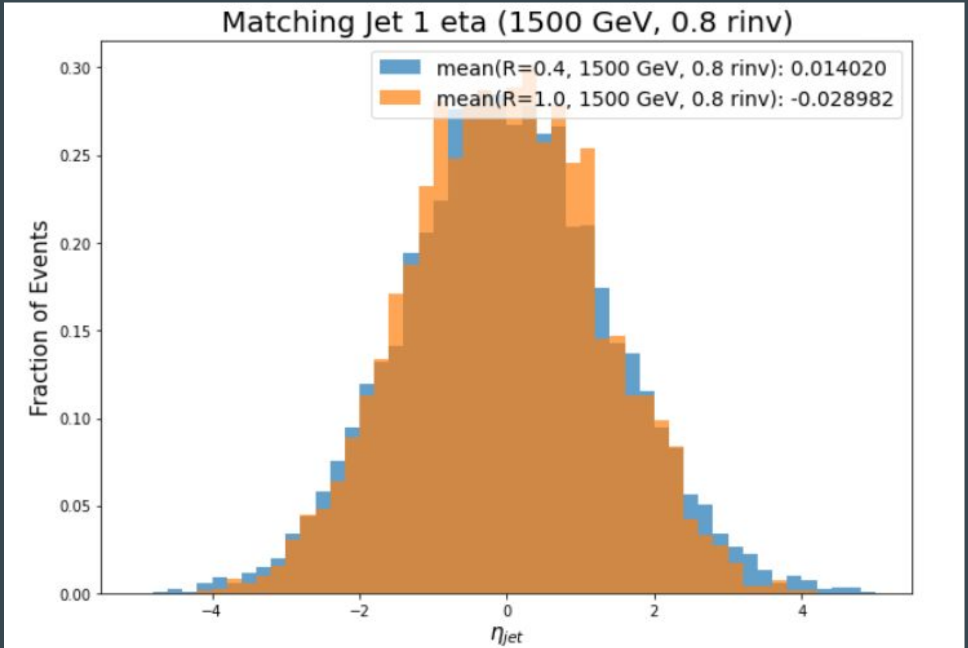
Matching Jet 1 eta



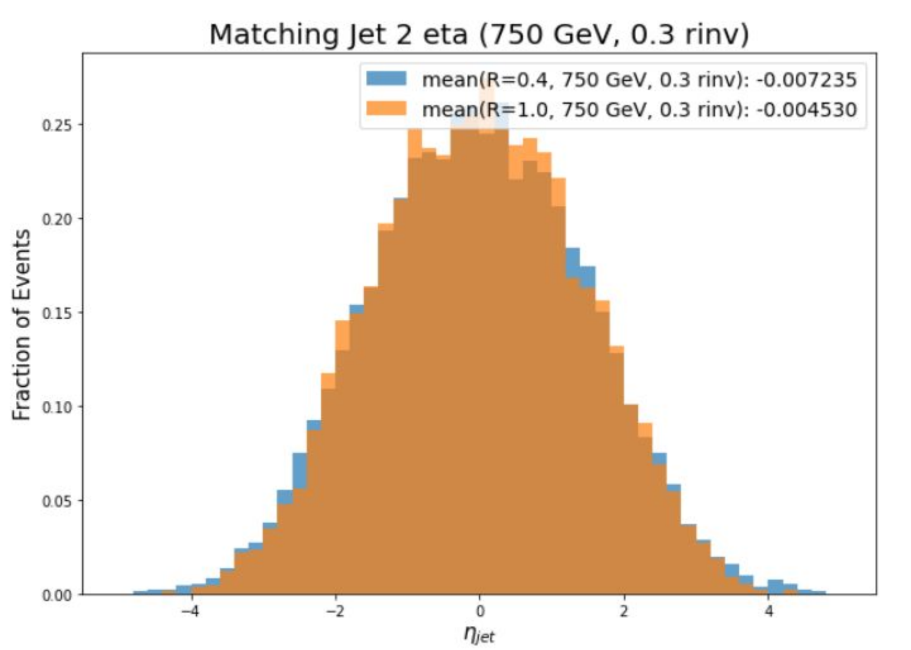


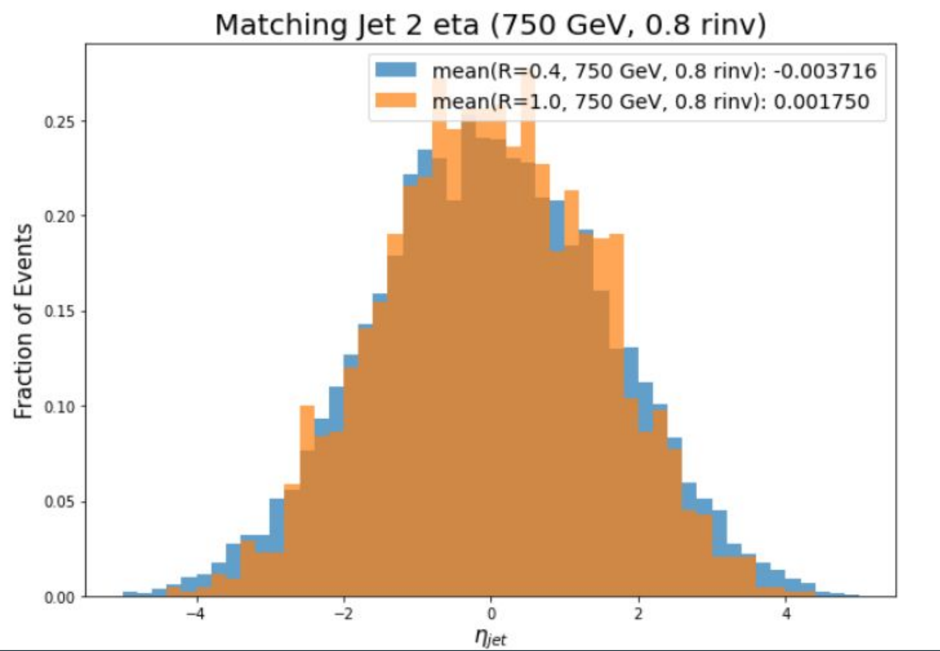
Matching Jet 1 eta



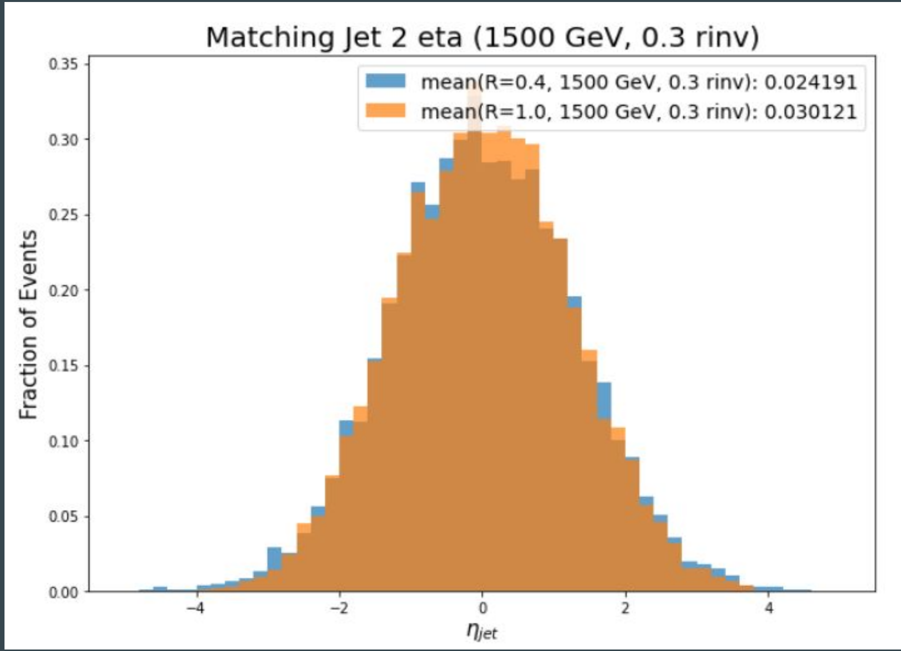


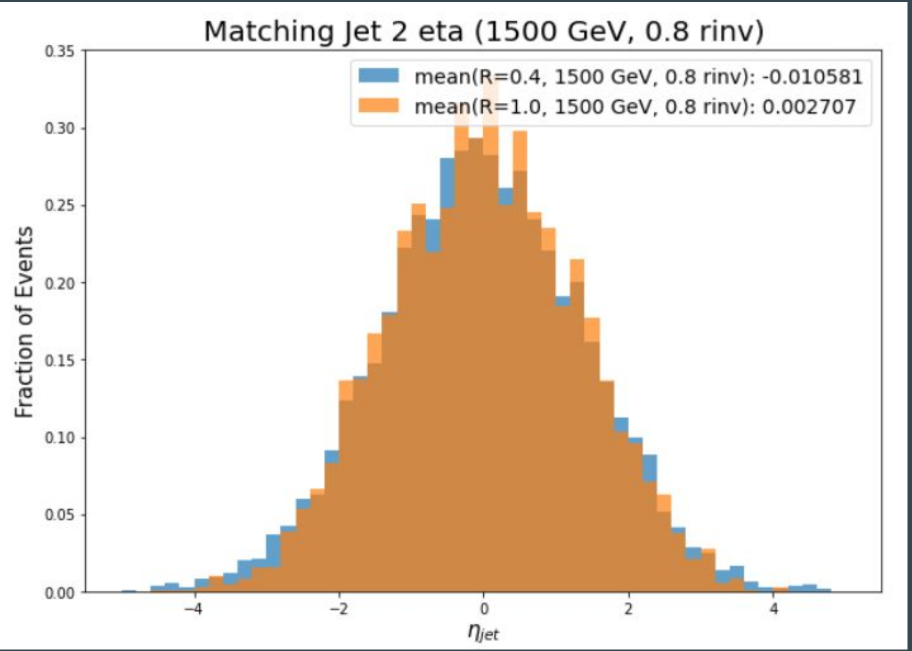
Matching Jet 2 eta



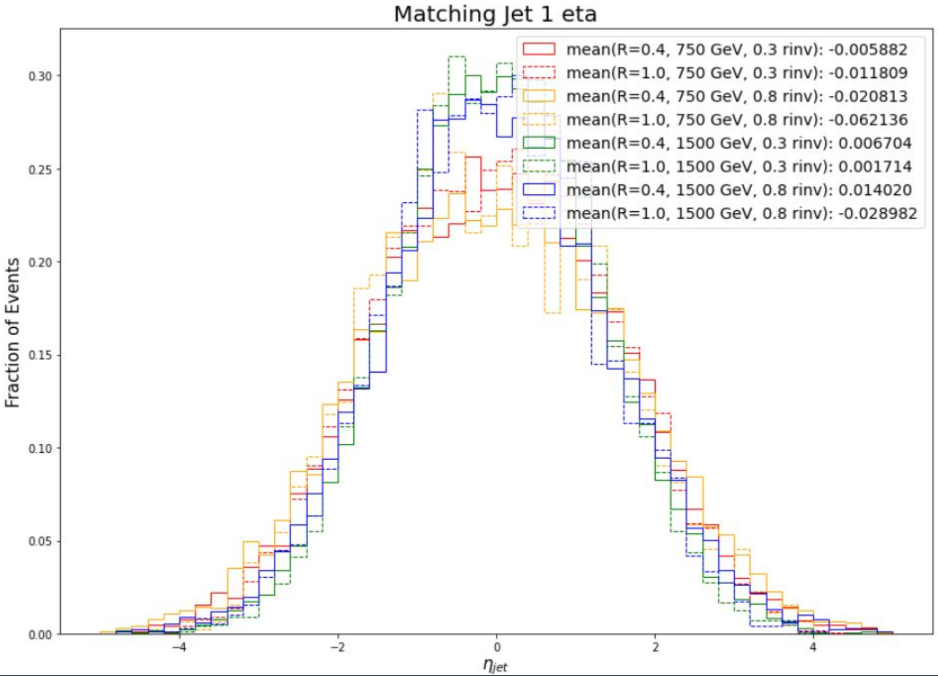


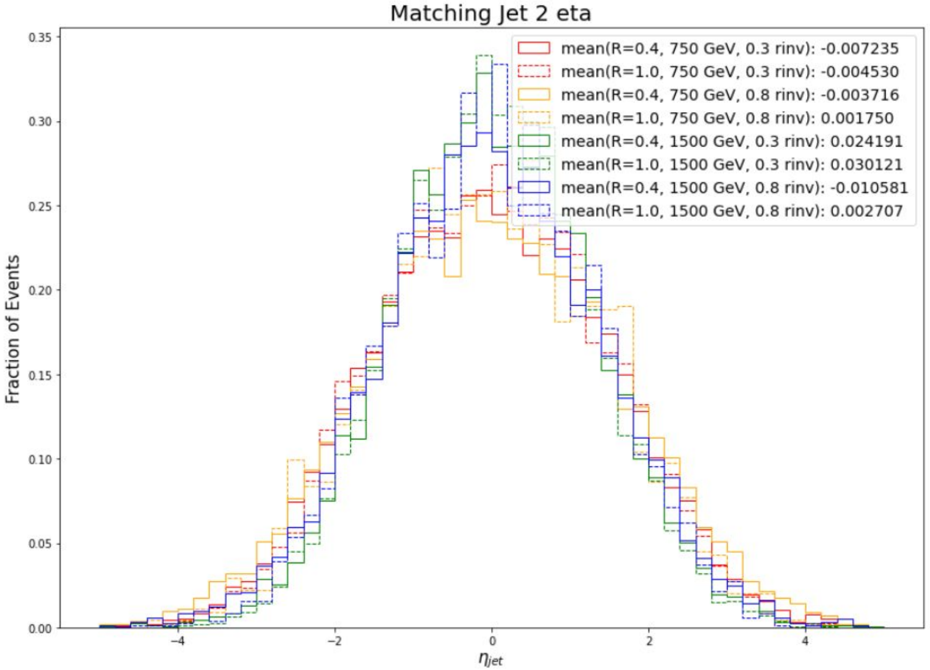
Matching Jet 2 eta



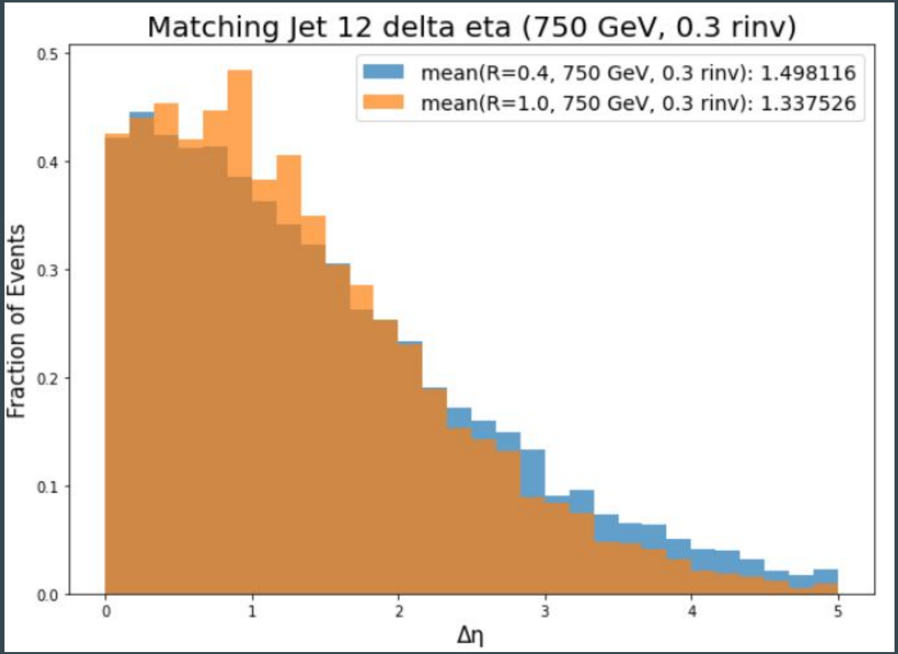


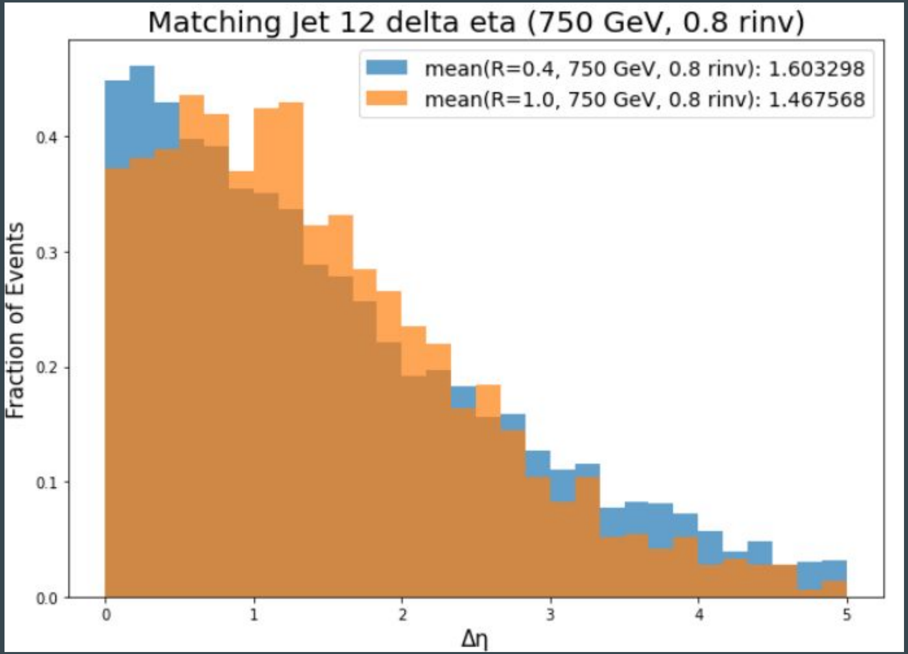
Matching Jet 2 pt



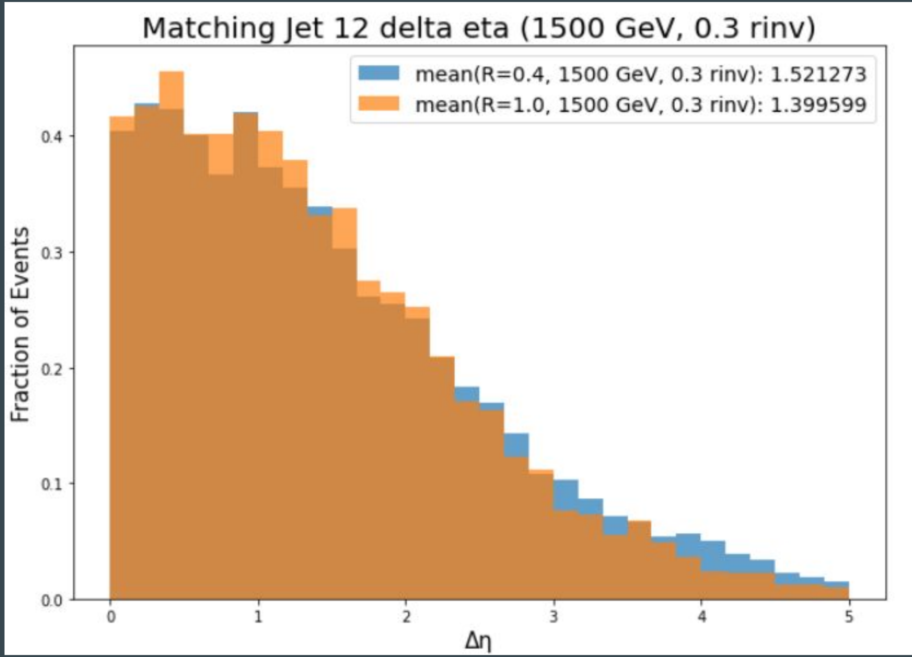


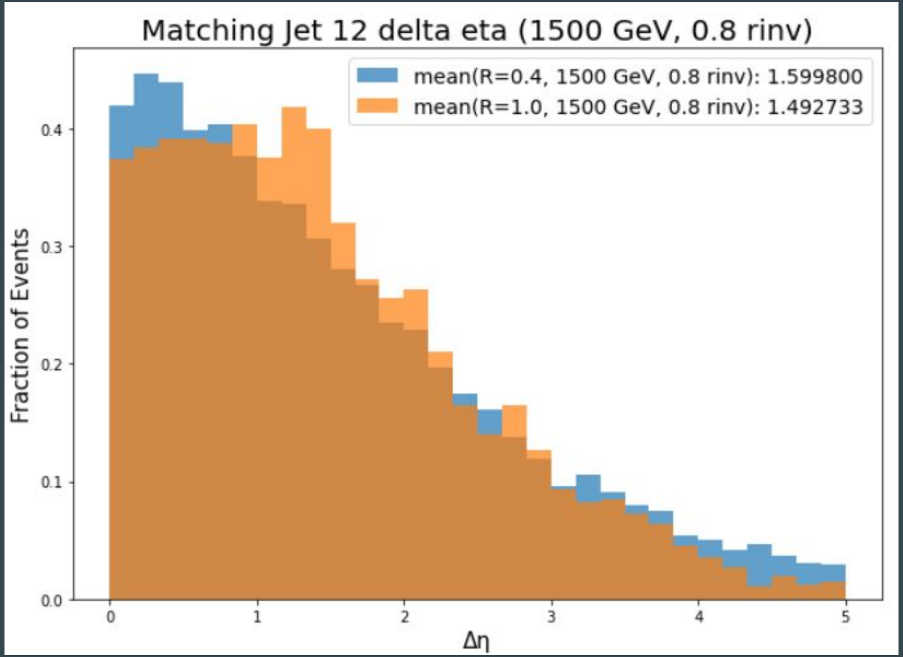
Matching Jet 12 delta eta



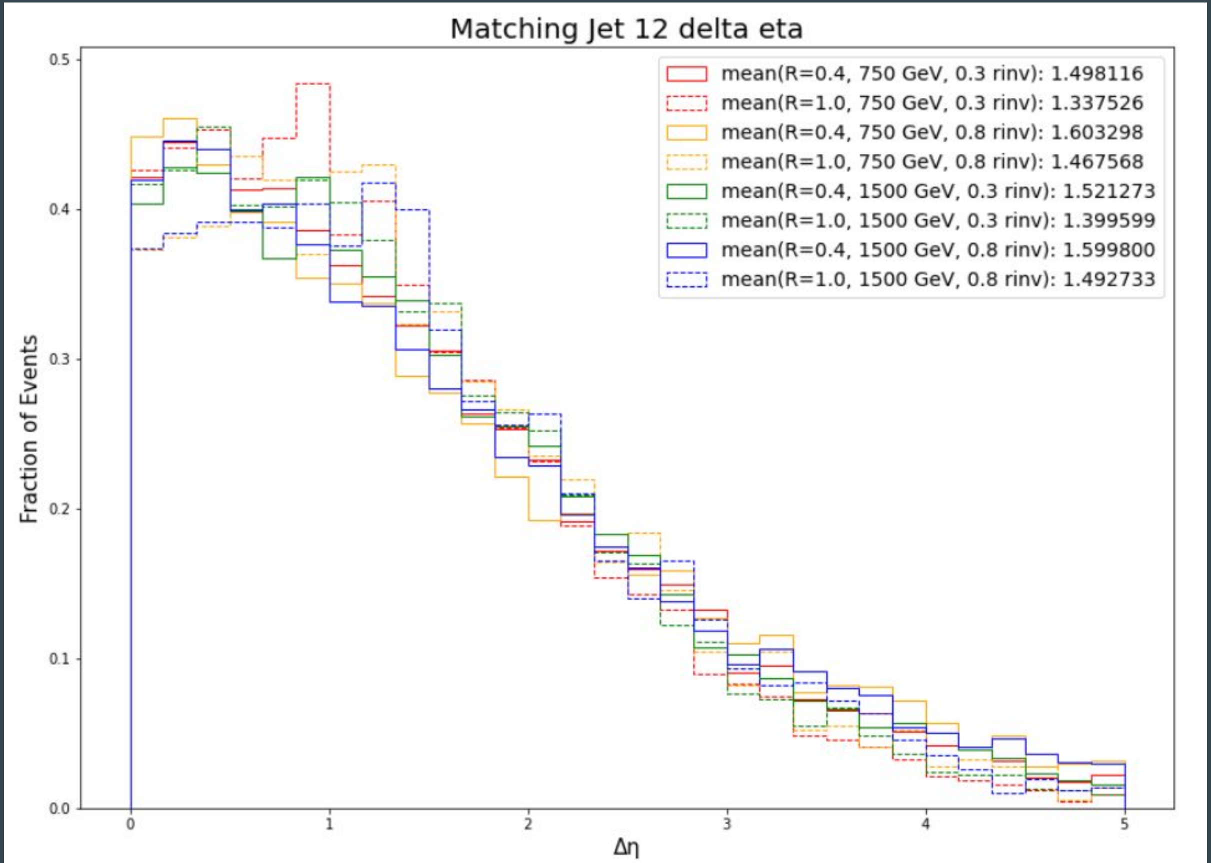


Matching Jet 12 delta eta

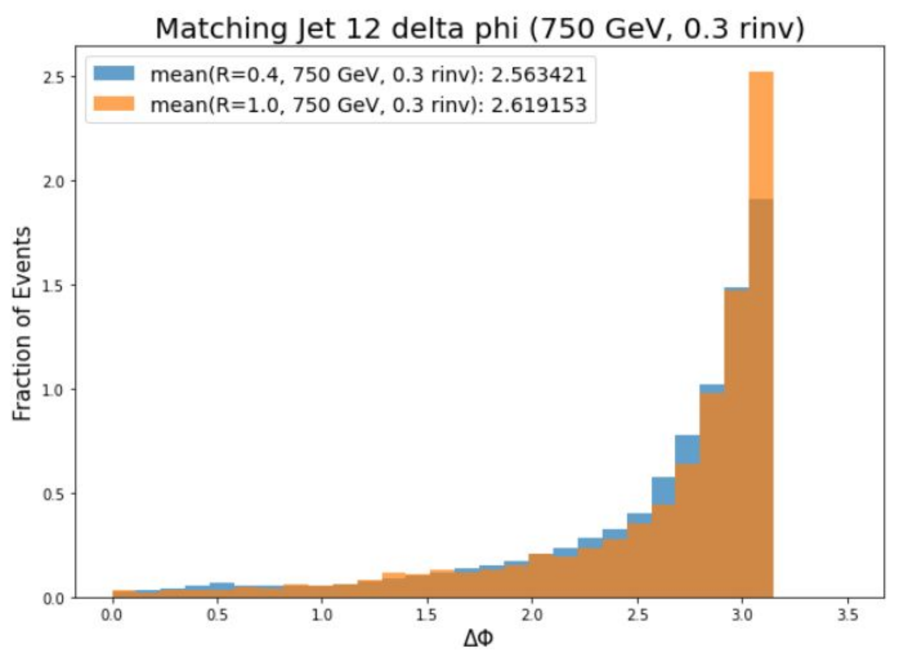


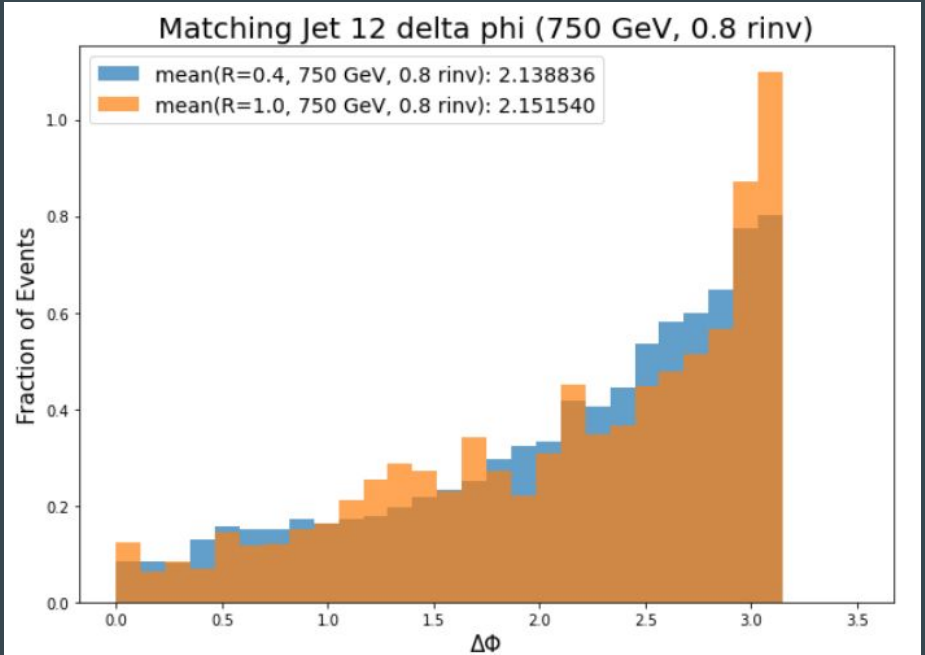


Matching Jet 12 delta eta

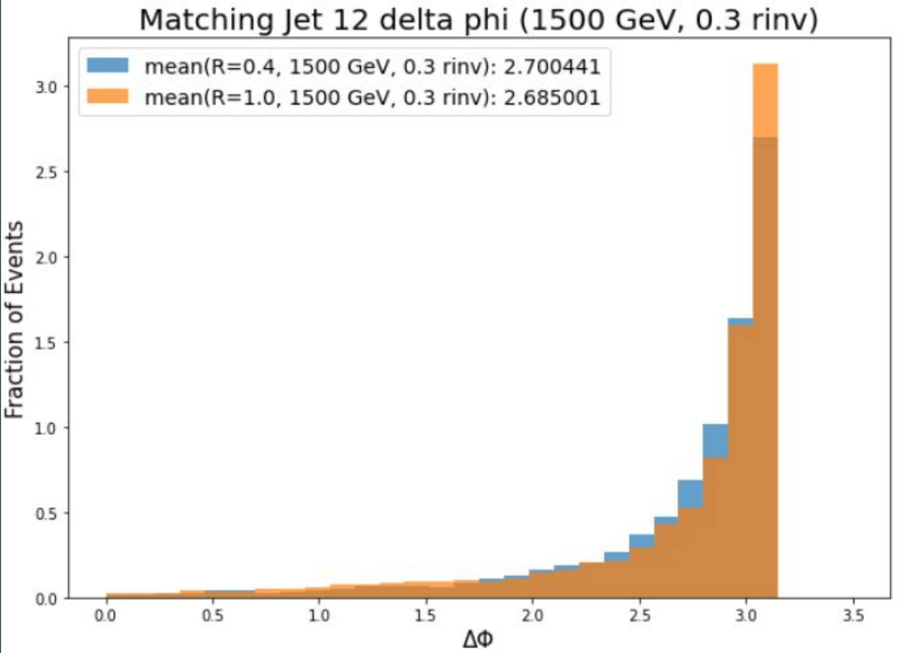


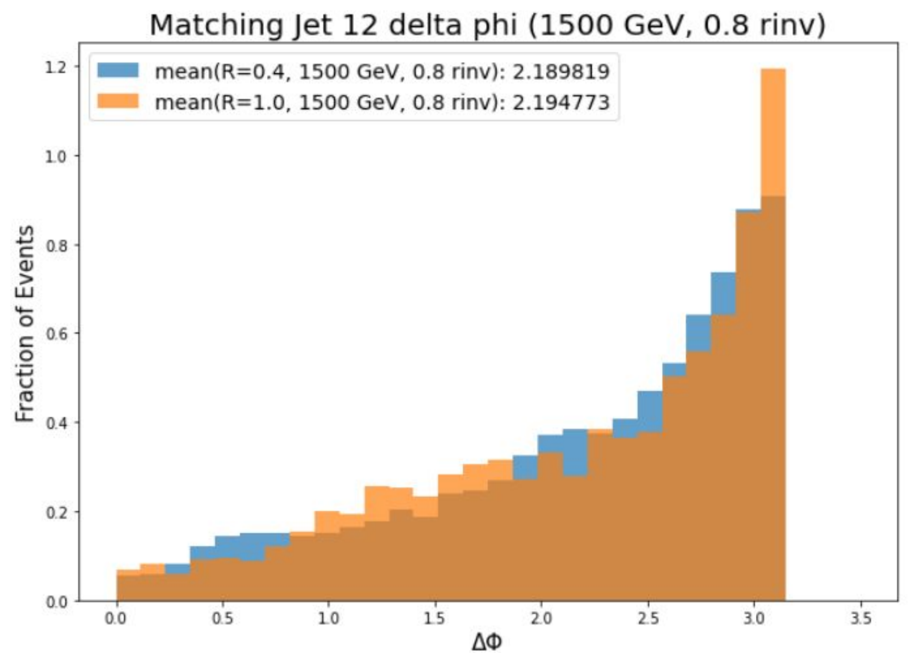
Matching Jet 12 delta phi



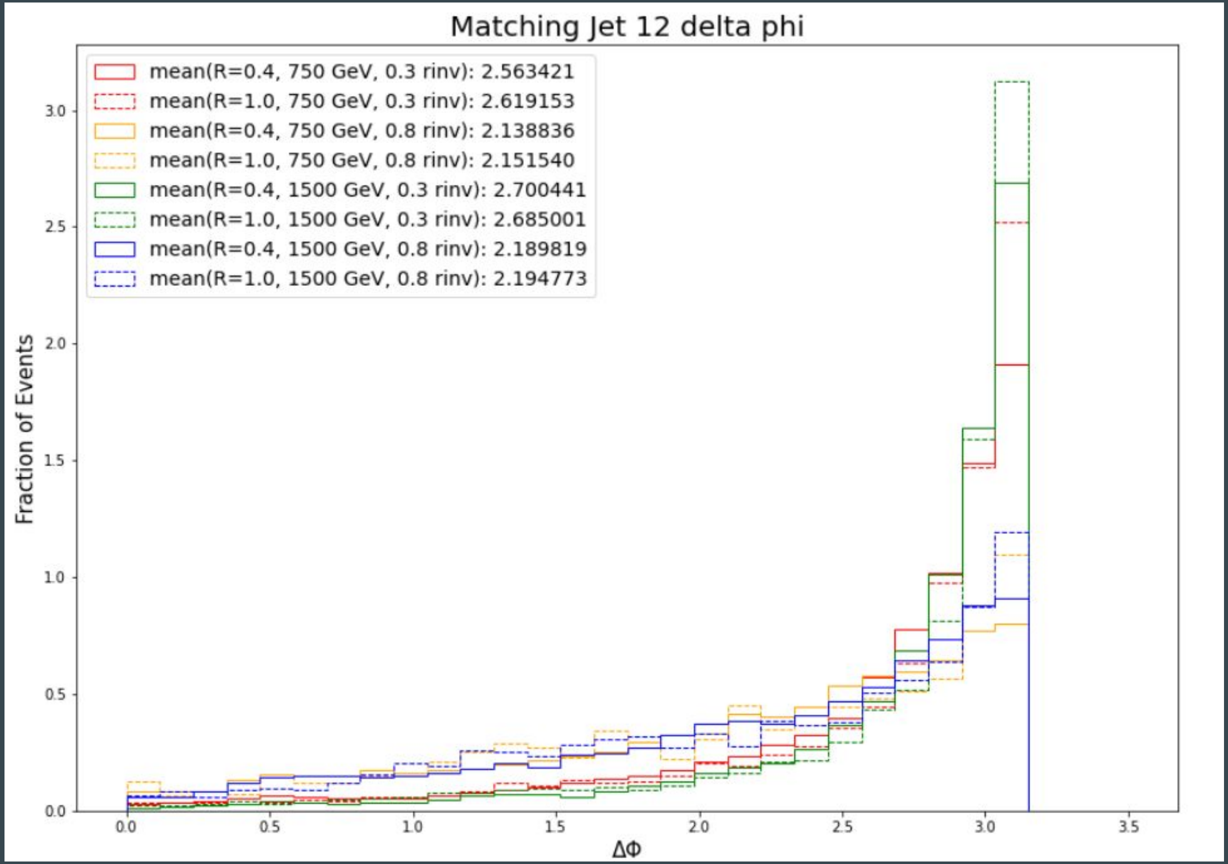


Matching Jet 12 delta phi

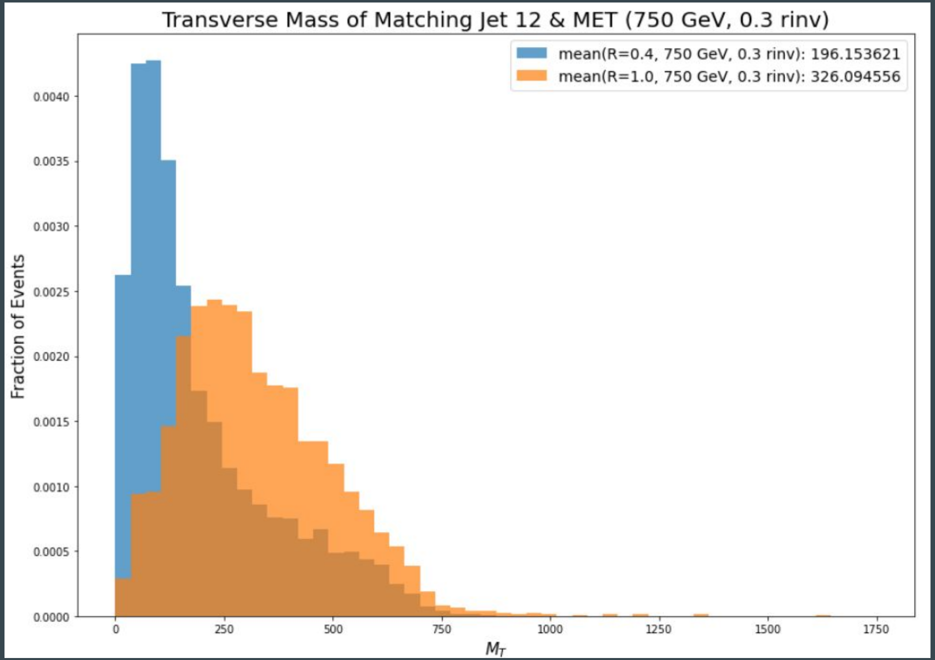


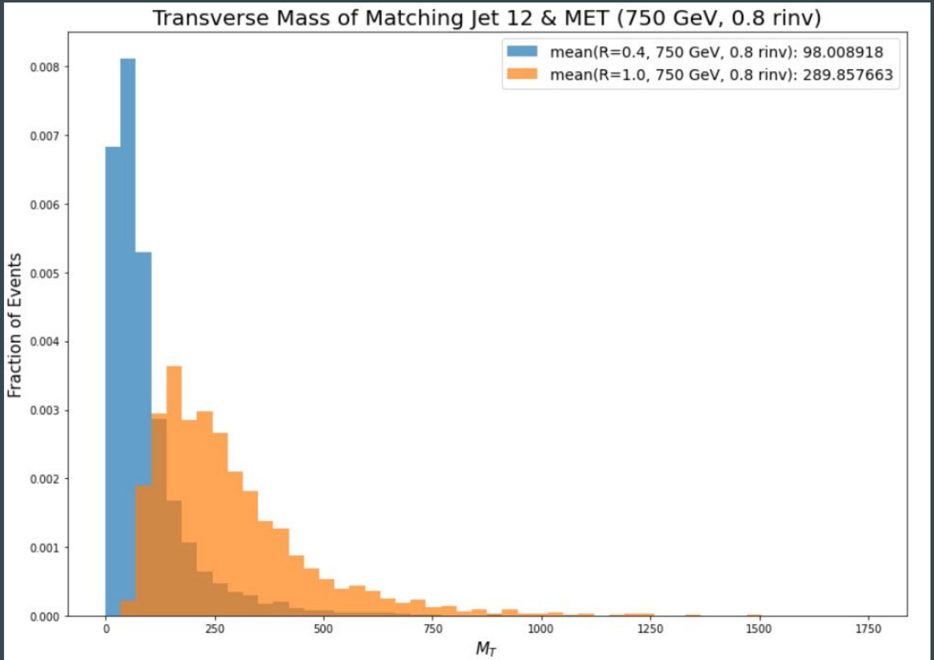


Matching Jet 12 delta phi

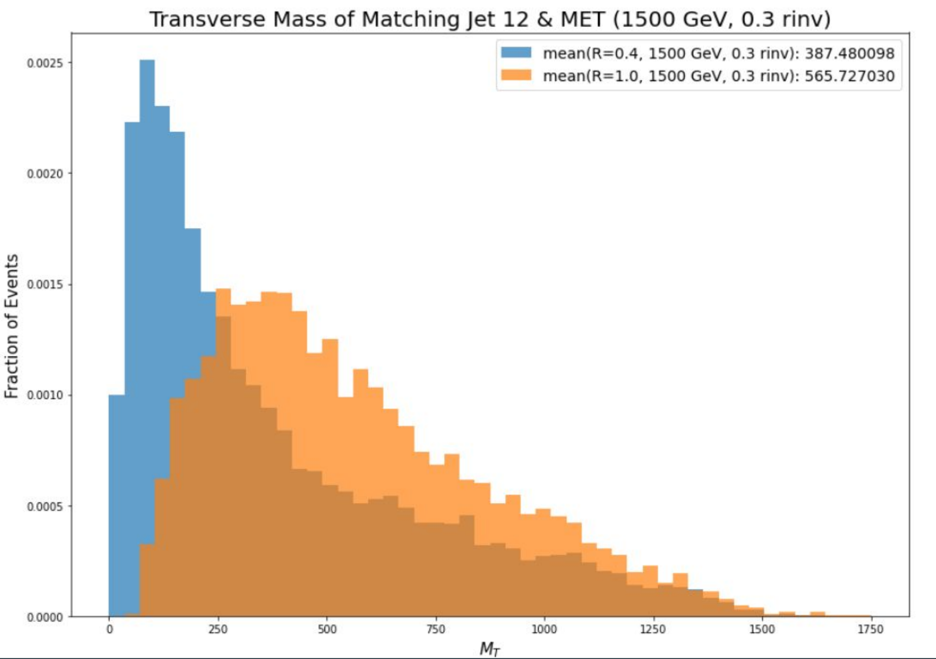


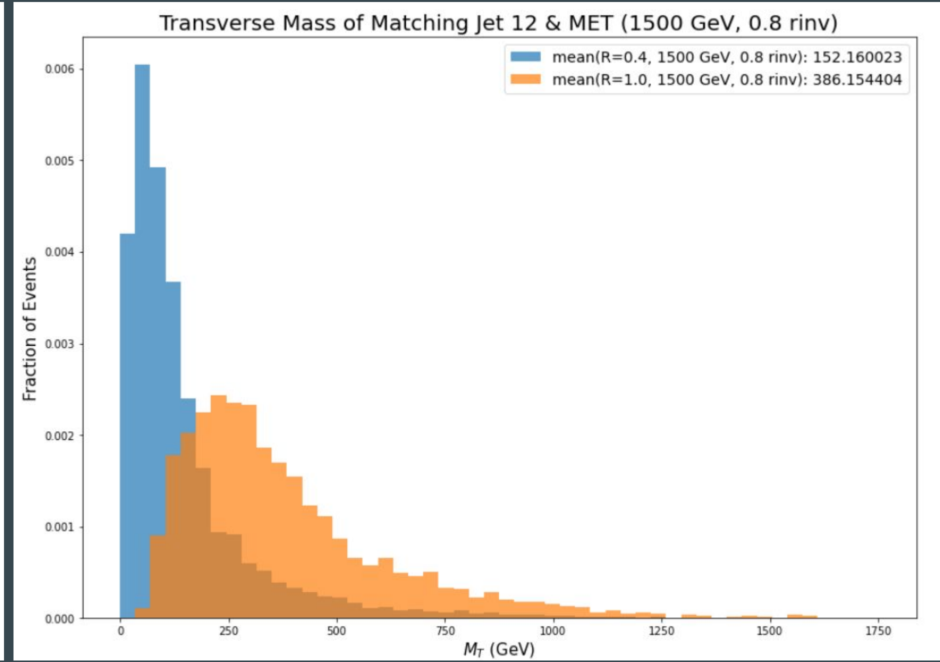
Transverse Mass of Matching Jet 12 & MET



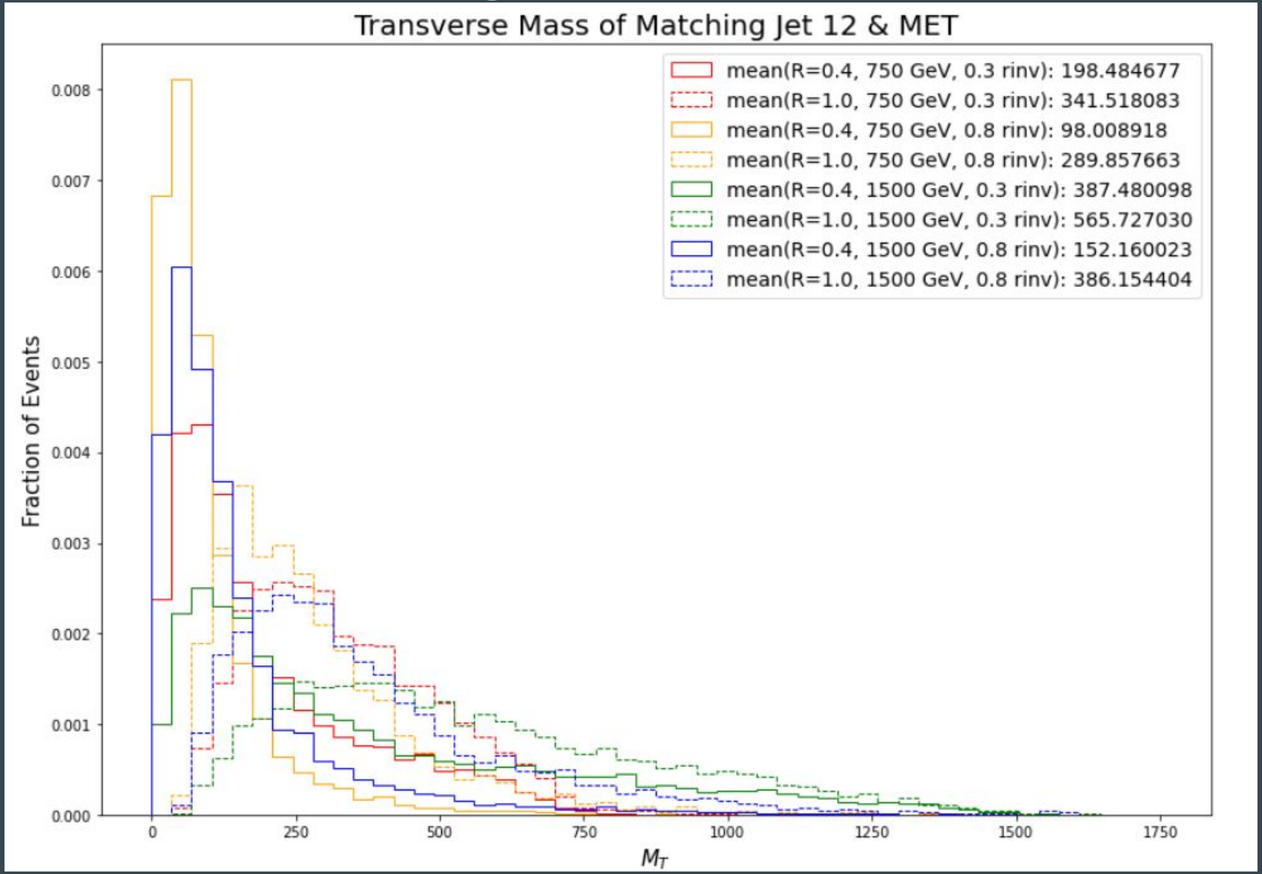


Transverse Mass of Matching Jet 12 & MET

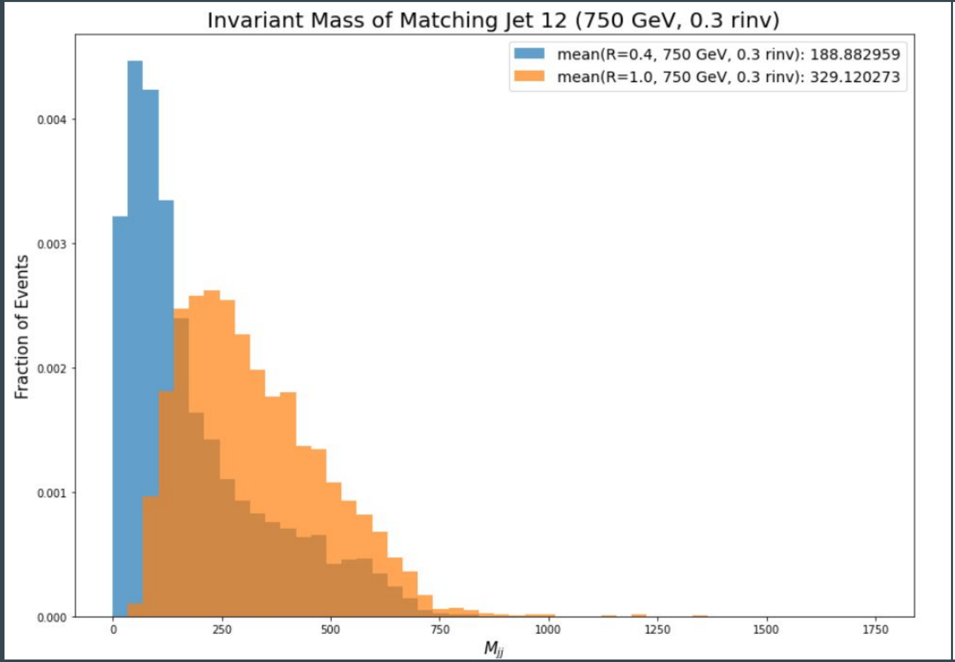


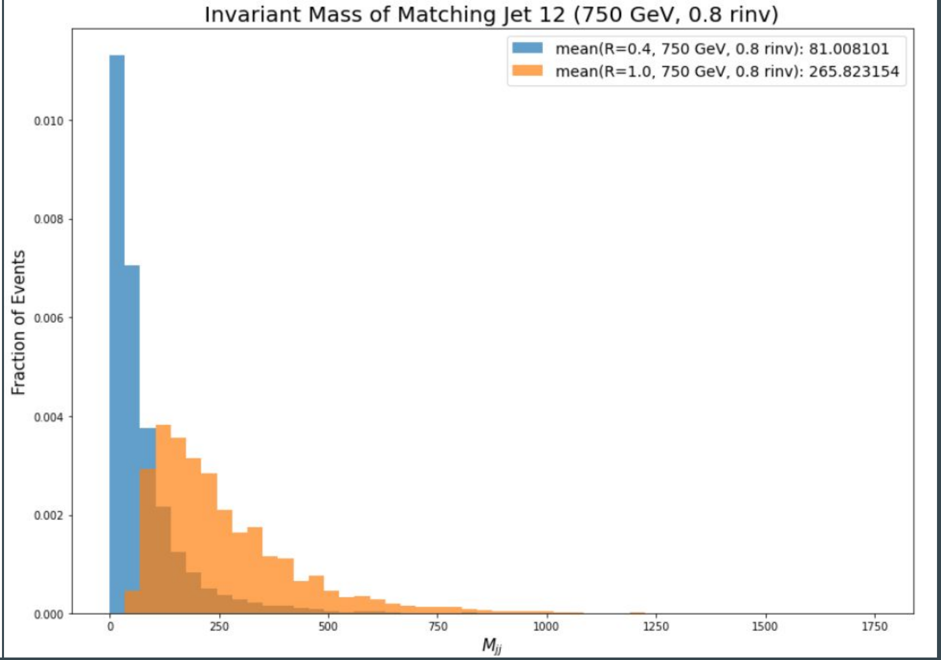


Transverse Mass of Matching Jet 12 & MET

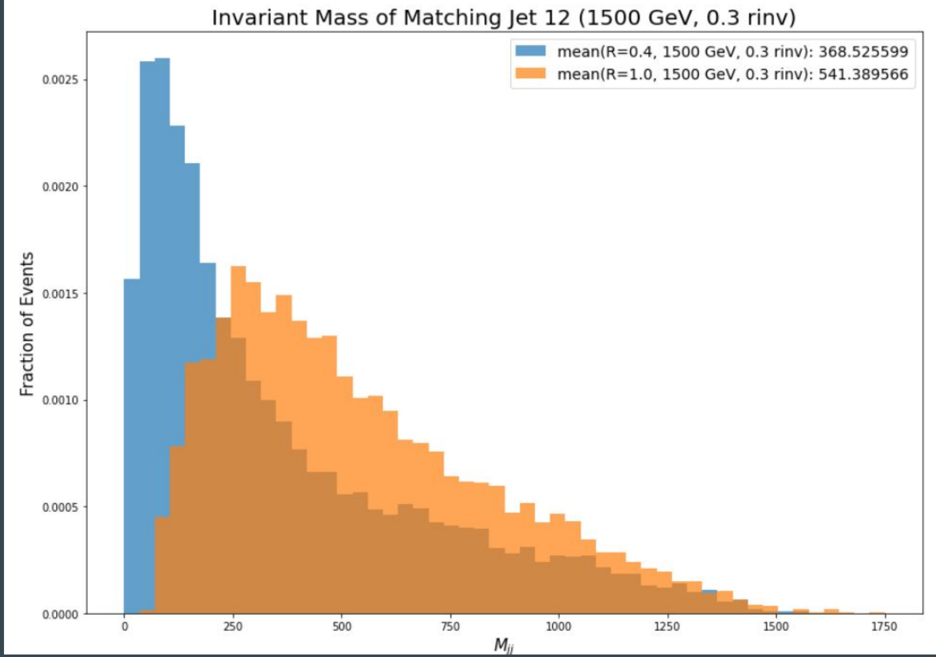


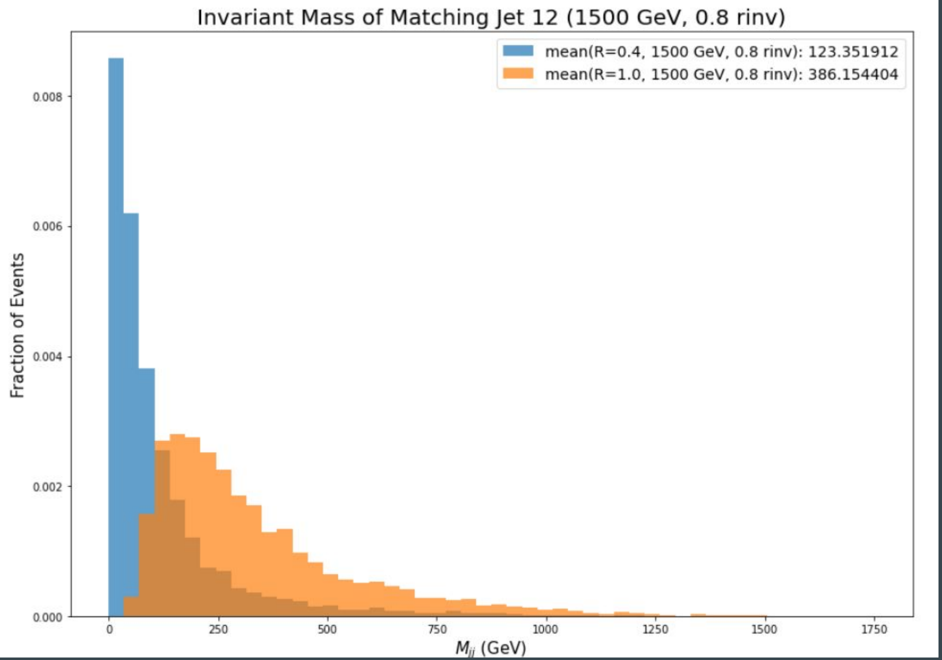
Invariant Mass of Matching Jet 12



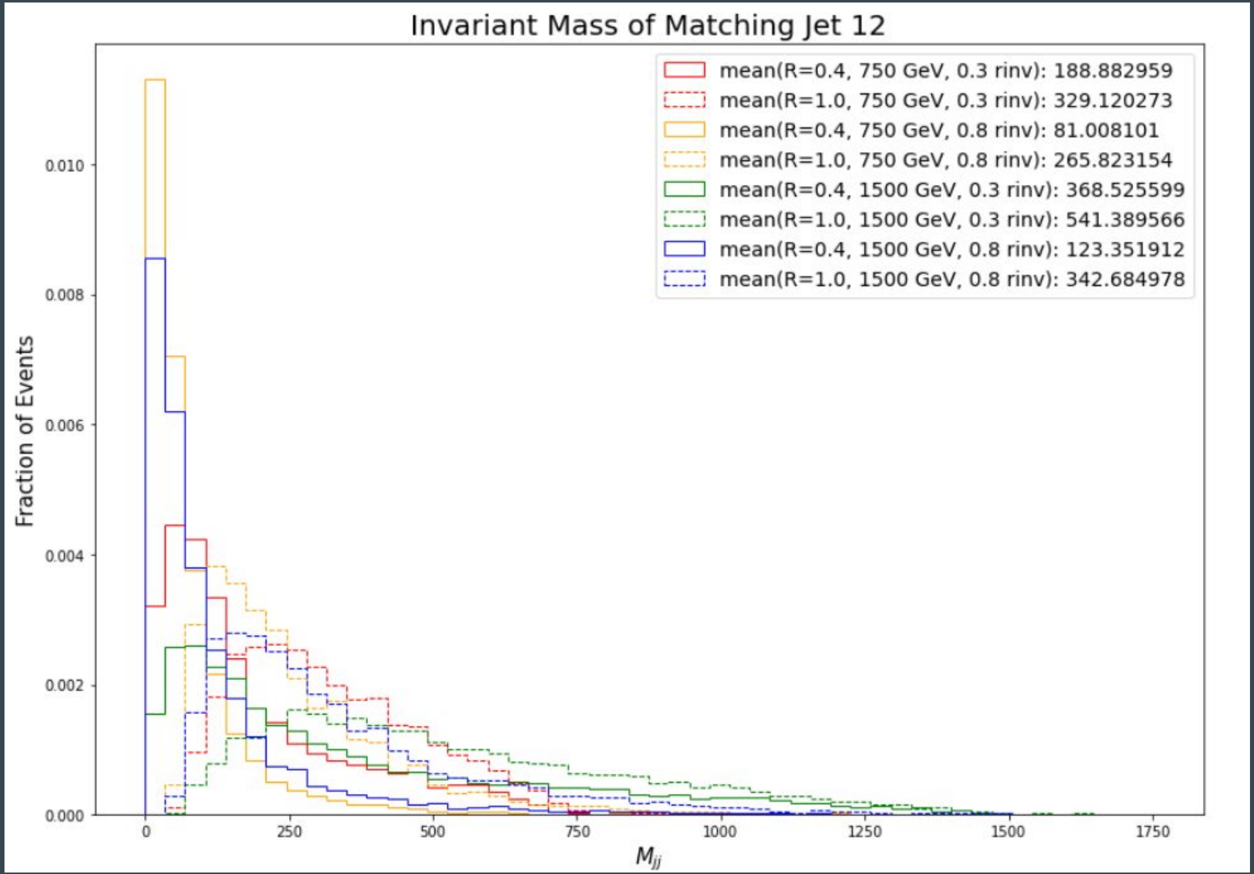


Invariant Mass of Matching Jet 12





Invariant Mass of Matching Jet 12



Problem on Kinematic Plots:

1. large cut on selection cut/ cut flow table
2. energy is approximately equal to momentum in some events.

```
100%|██████████| 9942/9942 [01:09<00:00, 142.80it/s]
  0%|          | 11/9736 [00:00<01:31, 106.56it/s]
Counter({'energy > momentum': 9942, 'same momentum': 9942})
 67%|██████▋   | 6508/9736 [00:46<00:21, 151.10it/s]
problematic energy: 402.8031311035156, momentum:402.80316162109375 in 6615th event
100%|██████████| 9736/9736 [01:09<00:00, 140.26it/s]
  0%|          | 14/9959 [00:00<01:11, 138.28it/s]
Counter({'same momentum': 9736, 'energy > momentum': 9733, 'energy < momentum': 3})
100%|██████████| 9959/9959 [01:08<00:00, 144.71it/s]
  0%|          | 12/9819 [00:00<01:22, 119.17it/s]
Counter({'energy > momentum': 9959, 'same momentum': 9959})
  1%|          | 113/9819 [00:00<01:09, 139.67it/s]
problematic energy: 417.03314208984375, momentum:417.0331726074219 in 86th event
 39%|███▉      | 3875/9819 [00:26<00:48, 122.92it/s]
problematic energy: 372.1103210449219, momentum:372.1103515625 in 3908th event
100%|██████████| 9819/9819 [01:07<00:00, 146.21it/s]
  0%|          | 14/9261 [00:00<01:07, 136.38it/s]
Counter({'same momentum': 9819, 'energy > momentum': 9816, 'energy < momentum': 3})
100%|██████████| 9261/9261 [00:59<00:00, 154.70it/s]
  0%|          | 14/6584 [00:00<00:47, 138.71it/s]
Counter({'same momentum': 9261})
100%|██████████| 6584/6584 [00:43<00:00, 151.99it/s]
  0%|          | 15/9658 [00:00<01:05, 147.03it/s]
Counter({'same momentum': 6584})
100%|██████████| 9658/9658 [01:03<00:00, 151.12it/s]
  0%|          | 14/7754 [00:00<00:57, 135.08it/s]
Counter({'same momentum': 9658})
100%|██████████| 7754/7754 [00:51<00:00, 151.35it/s]
Counter({'same momentum': 7754})
```

Proof: two methods of calculating total momentum are consistent within 10^-4 magnitude. since there is none 'different momentum'.

Proof: there exists jets that has energy approximately equals to momentum (a small difference and it could be python's problem)

Possible Project Idea:

- NN model with Joe
- continue clustering project from Oscar

Autoencoders for Semivisible Jet Detection

- QCD & dark sector sample: Pythia8 and Delphes
  - r_{inv} = {0.3, 0.5, 0.7} & Z' boson mass = {1.5, 2.0, 2.5, 3.0, 3.5, 4.0} TeV
- High-level jet features (total of 9 features)
  - Energy Flow Polynomials (EFPs), Energy Correlation Functions (ECFs) and their ratios: C2 and D2, jet pT dispersion pT D and jet axes, η & φ
- Selection Cut
  - at least 2 jets $|\eta| < 2.4$ and $pT > 200$ GeV
  - for two leading jets: $|\Delta\eta| < 1.5$; $m_T > 1500$ GeV; $E_T/m_T > 0.25$
- η & φ included in the training
  - to allow the network to learn about problematic regions of the detector & avoid tagging noise or other detector failures as anomalous signals.
- pT **not** included in the training to avoid bias.

Input for ML

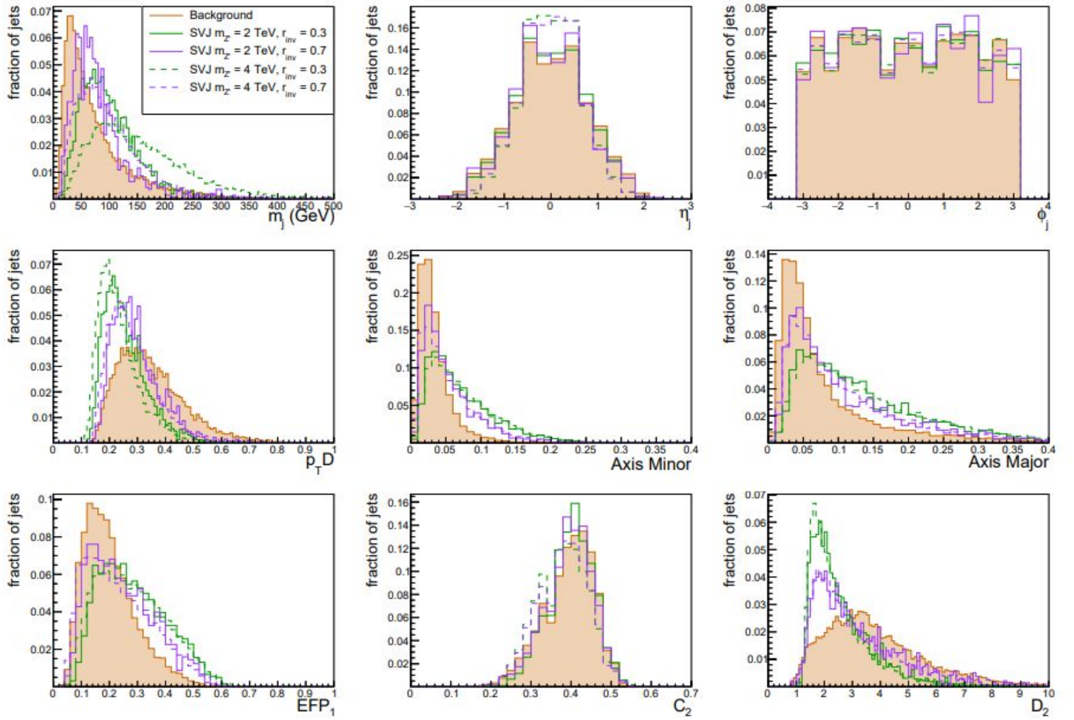


Figure 1: Distributions of input variables for QCD background and selected signal models.

Neural Network

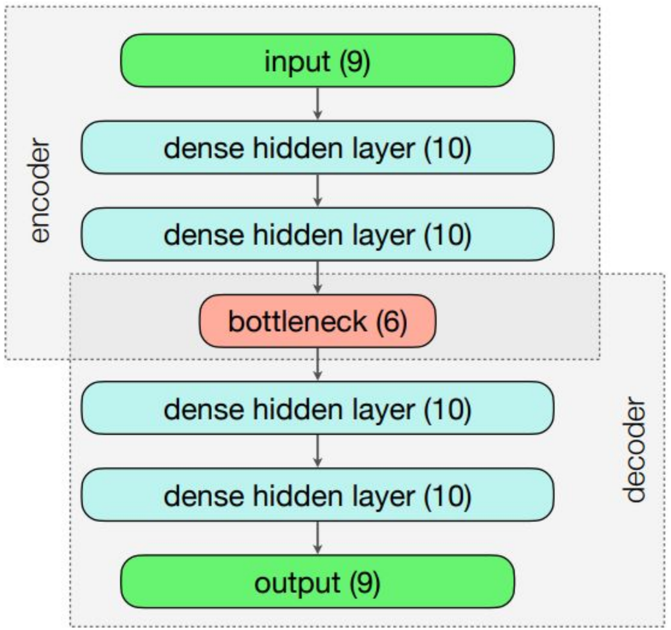


Figure 2: Neural Network architecture.

Sensitivity to SVJ

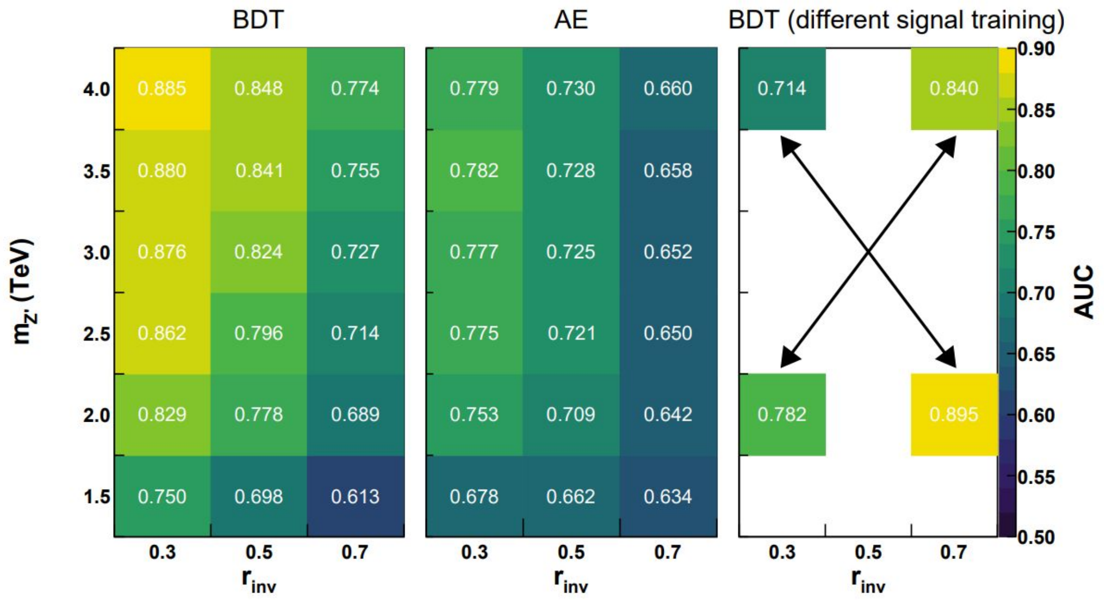


Figure 6: Left and middle panels: comparison of AUC values of the autoencoder and BDT. Right panel: AUC values for a BDT trained on a signal with parameters different from those it was tested on. E.g., the AUC value presented in the top left corner of the table comes from a model trained on the lower right corner sample.

Analysis Flow Chart

Prepare the signal and background

selection cut | QCD background generation

↓

Study high level features (9)

↓

Develop a NN model

↓

Interpret Model

QCD background

- multijet
- W + jets
- ttbar
- Znunu
- single top
- diboson
- ztt

reference to t-channel background:
https://docs.google.com/presentation/d/1KsMlsd9V3JLcYIpgKWlXh76lEtscvWZxD1bz8WMFRXc/edit#slide=id.g106b63b11ab_0_5

High Level Features (9)

- Invariant Mass of Jets (M_j) (complete)
- pseudorapidity (η_j) (complete)
- phi of jets (φ_j) (complete)
- Jet pT dispersion (p_TD) (complete)
- Axis Minor (working)
- Axis Major (working)
- Energy Correlation Functions (ECF1) (complete but have problem)
- C_2 (complete but have problem)
- D_2 (complete but have problem)

ECFs, C_2 and D_2

ECFS and D2

$$E_{CF0}(\beta) = 1,$$

$$E_{CF1}(\beta) = \sum_{i \in J} p_{T_i},$$

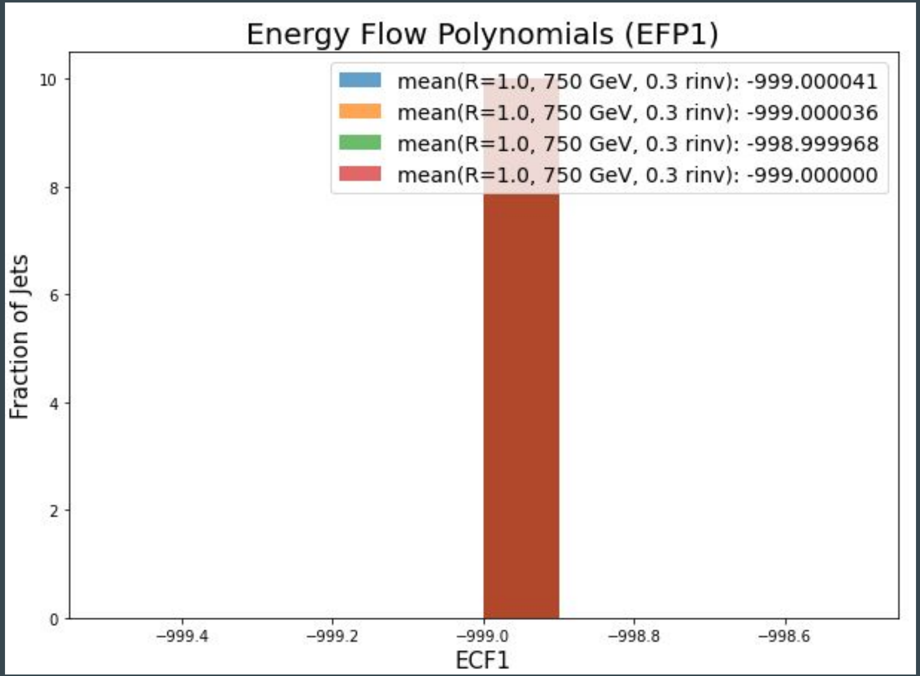
$$E_{CF2}(\beta) = \sum_{i<j \in J} p_{T_i} p_{T_j} \left(\Delta R_{ij}\right)^{\beta},$$

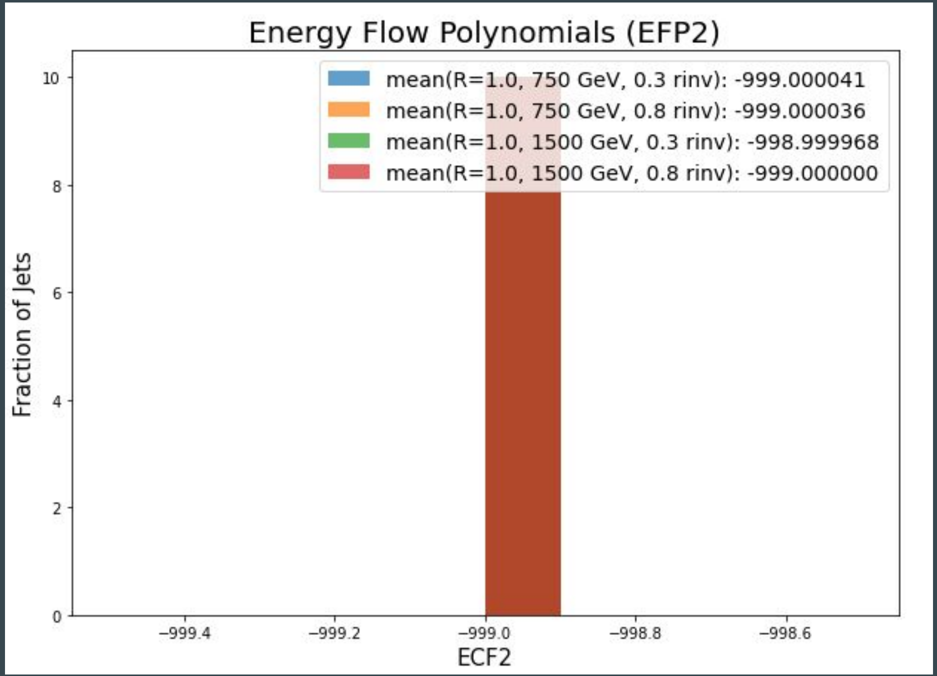
$$E_{CF3}(\beta) = \sum_{i<j<k \in J} p_{T_i} p_{T_j} p_{T_k} \left(\Delta R_{ij} \Delta R_{ik} \Delta R_{jk}\right)^{\beta}$$

$$C_2^{(\beta)} = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^2}$$

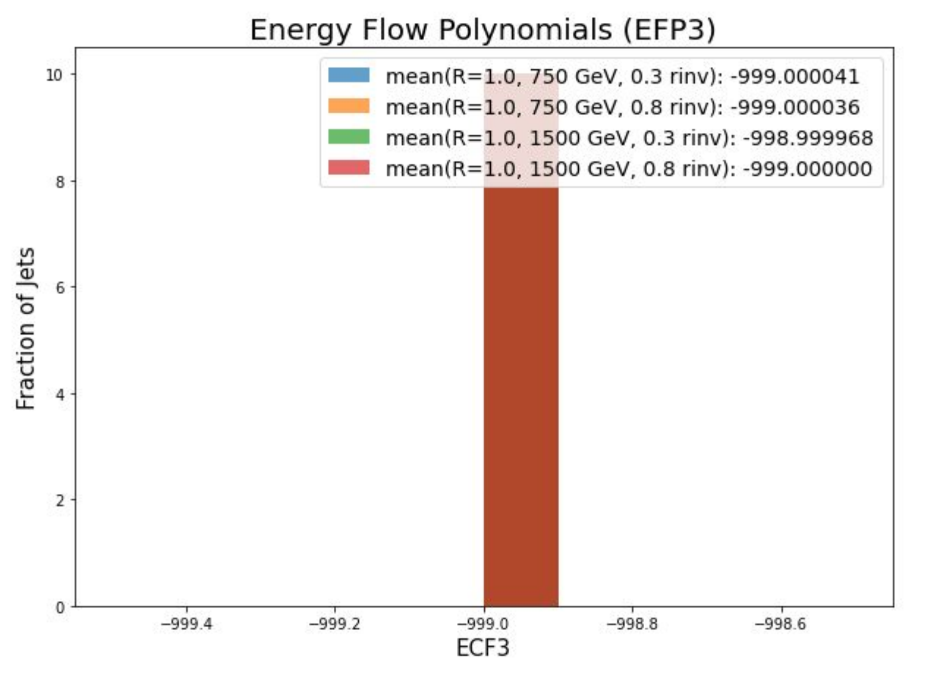
$$D_2^{(\beta)} = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^3}$$

ECFs (there is no data in nTuple files)



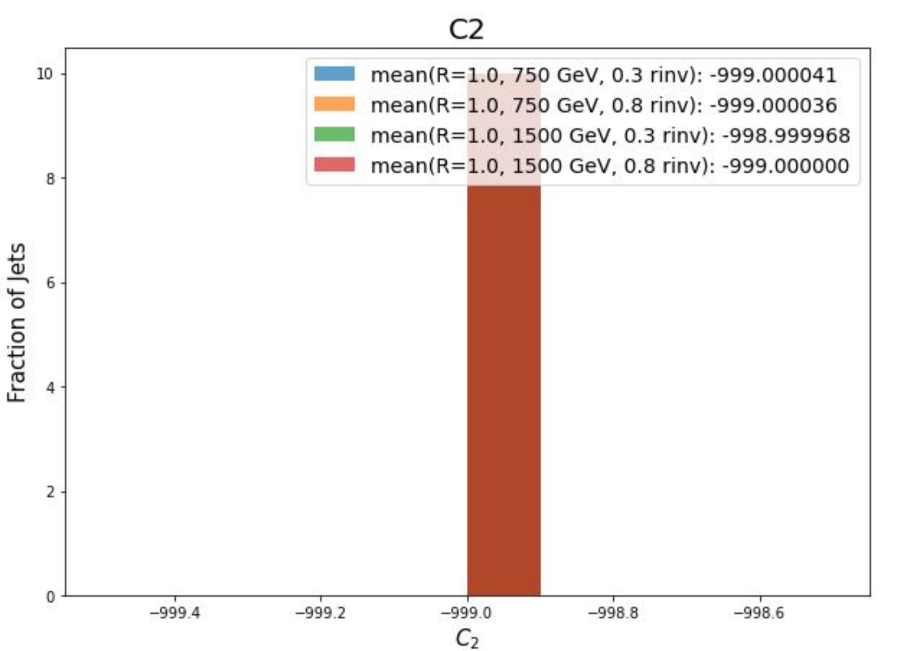


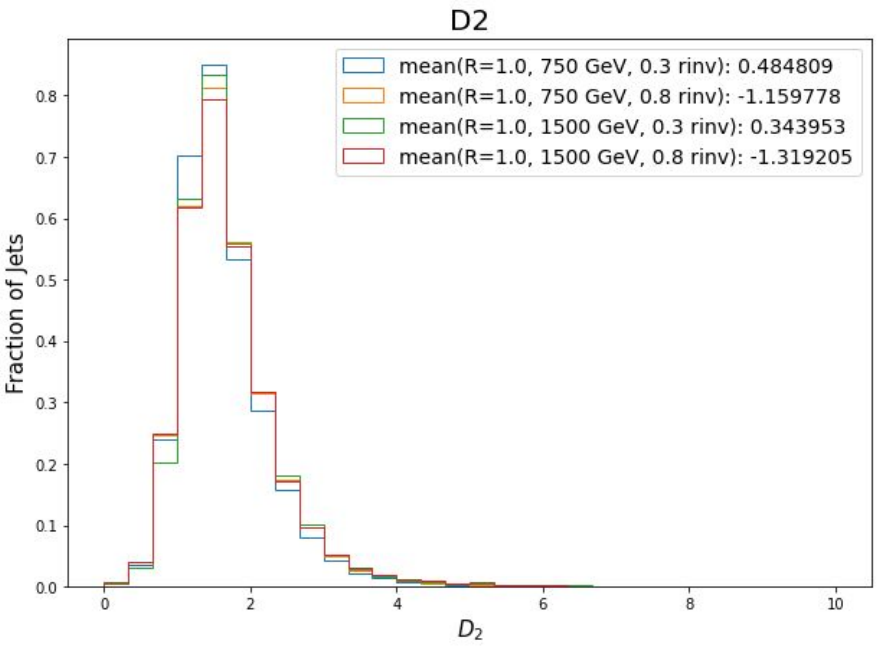
ECFs



Energy Flow Polynomials (EFP3)
mean(R=1.0, 750 GeV, 0.3 rinv): -999.000041
mean(R=1.0, 750 GeV, 0.8 rinv): -999.000036
mean(R=1.0, 1500 GeV, 0.3 rinv): -998.999968
mean(R=1.0, 1500 GeV, 0.8 rinv): -999.000000
Fraction of Jets
ECF3
−999.4
−999.2
−999.0
−998.8
−998.6
0
2
4
6
8
10

C_2 & D_2


C2
mean(R=1.0, 750 GeV, 0.3 rinv): -999.000041
mean(R=1.0, 750 GeV, 0.8 rinv): -999.000036
mean(R=1.0, 1500 GeV, 0.3 rinv): -998.999968
mean(R=1.0, 1500 GeV, 0.8 rinv): -999.000000
Fraction of Jets
C_2
−999.4
−999.2
−999.0
−998.8
−998.6


D2
mean(R=1.0, 750 GeV, 0.3 rinv): 0.484809
mean(R=1.0, 750 GeV, 0.8 rinv): -1.159778
mean(R=1.0, 1500 GeV, 0.3 rinv): 0.343953
mean(R=1.0, 1500 GeV, 0.8 rinv): -1.319205
Fraction of Jets
D_2

Fragmentation Function

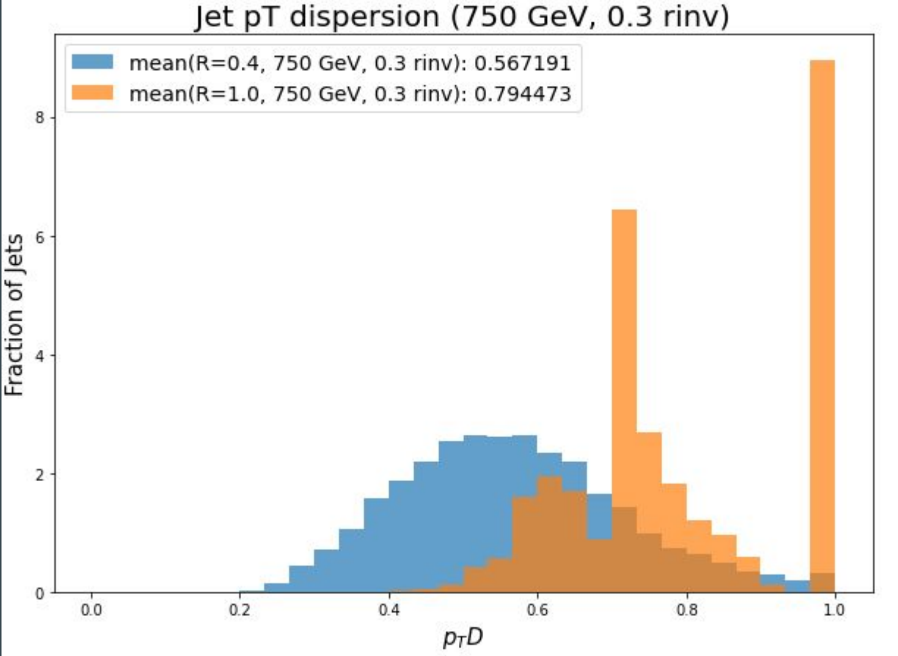
Fragmentation function

Quarks have a harder fragmentation function compared to gluons and are therefore more likely to produce jets with hard constituents that carry a significant fraction of the jet energy. This can be expressed with the $p_{\mathrm{T}}D$ variable, defined as:

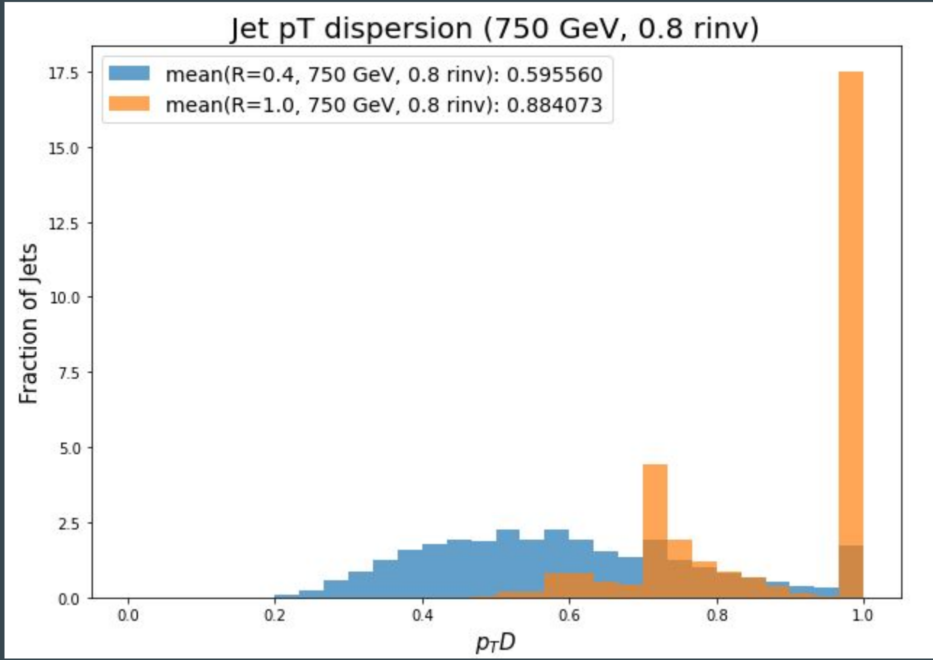
$$p_{\mathrm{T}}D = \frac{\sqrt{\sum_i p_{\mathrm{T},i}^2}}{\sum_i p_{\mathrm{T},i}} \tag{6}$$

where the sum runs over the jet constituents. From its definition, it stems that $p_{\mathrm{T}}D \rightarrow 1$ for jets made of only one particle that carries all of its momentum, and $p_{\mathrm{T}}D \rightarrow 0$ for a jet made of an infinite number of particles.

$p_T D$

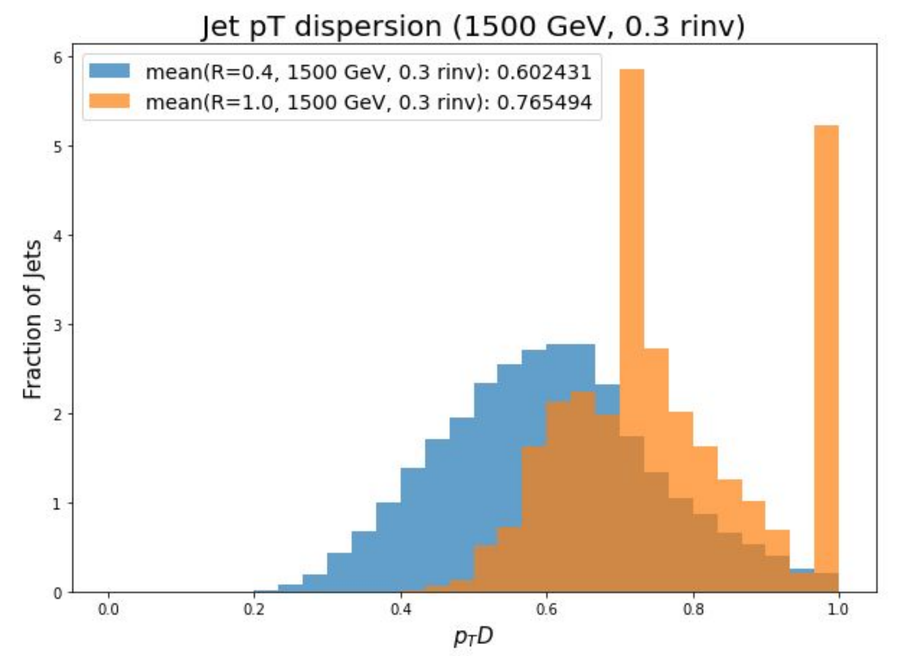


Jet pT dispersion (750 GeV, 0.3 rinv)
mean(R=0.4, 750 GeV, 0.3 rinv): 0.567191
mean(R=1.0, 750 GeV, 0.3 rinv): 0.794473
Fraction of Jets
$p_T D$
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0

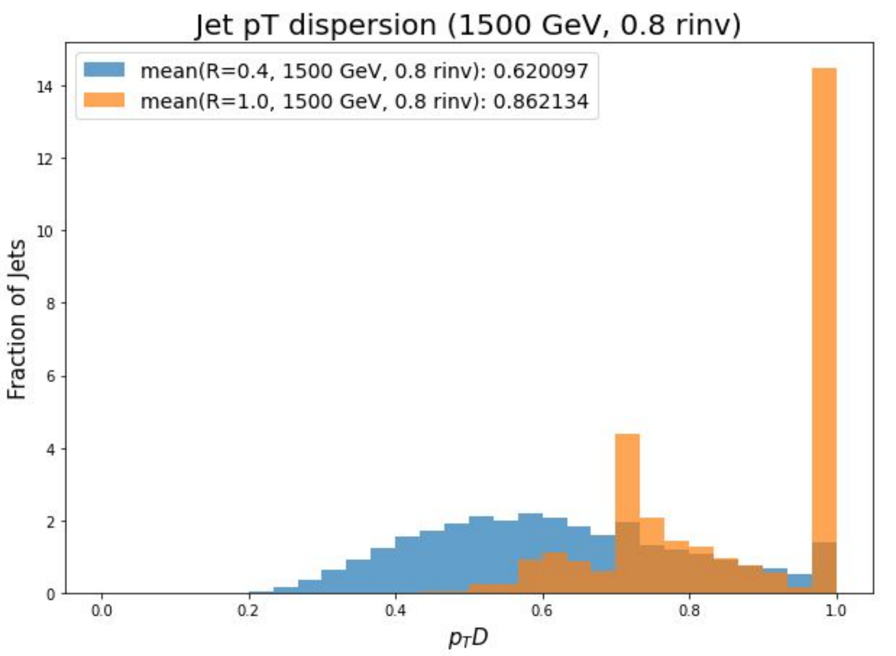


Jet pT dispersion (750 GeV, 0.8 rinv)
mean(R=0.4, 750 GeV, 0.8 rinv): 0.595560
mean(R=1.0, 750 GeV, 0.8 rinv): 0.884073
Fraction of Jets
$p_T D$
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
0.0
0.2
0.4
0.6
0.8
1.0

p_TD



Jet pT dispersion (1500 GeV, 0.3 rinv)
mean(R=0.4, 1500 GeV, 0.3 rinv): 0.602431
mean(R=1.0, 1500 GeV, 0.3 rinv): 0.765494
Fraction of Jets
p_TD
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6



Jet pT dispersion (1500 GeV, 0.8 rinv)
mean(R=0.4, 1500 GeV, 0.8 rinv): 0.620097
mean(R=1.0, 1500 GeV, 0.8 rinv): 0.862134
Fraction of Jets
p_TD
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
12
14

To-do

- arrange a online meeting with Joe
- more background information
- complete high level features study

Personal Notes

•••

succeed [rucio get & DAOD -> nTuple]

```
Your proxy is valid until Thu Mar 10 07:18:02 CET 2022
[jlai@lxplus708 SVJ_Data]$ rucio get mc16_13TeV.508547.MGPy8EG_SVJSChan_1500_8.deriv.DAOD_PHYS.e8357_e7400_s3126_r10724_r10726_p4903
2022-03-09 19:18:14,901 INFO    Processing 1 item(s) for input
2022-03-09 19:18:15,205 INFO    No preferred protocol impl in rucio.cfg: No section: 'download'
2022-03-09 19:18:15,206 INFO    No preferred protocol impl in rucio.cfg: No section: 'download'
2022-03-09 19:18:15,206 INFO    No preferred protocol impl in rucio.cfg: No section: 'download'
2022-03-09 19:18:15,206 INFO    No preferred protocol impl in rucio.cfg: No section: 'download'
2022-03-09 19:18:15,232 INFO    Using 3 threads to download 4 files
2022-03-09 19:18:15,233 INFO    Thread 0/3: Preparing download of mc16_13TeV:DAOD_PHYS.27616103._000001.pool.root.1
2022-03-09 19:18:15,234 INFO    Thread 1/3: Preparing download of mc16_13TeV:DAOD_PHYS.27616103._000002.pool.root.1
2022-03-09 19:18:15,235 INFO    Thread 2/3: Preparing download of mc16_13TeV:DAOD_PHYS.27616103._000003.pool.root.1
2022-03-09 19:18:15,306 INFO    Thread 0/3: Trying to download with root and timeout of 1481s from RAL-LCG2-ECHO_DATADISK: mc16_13TeV:DAOD_PHYS.27616103._000001.pool.root.1
2022-03-09 19:18:15,364 INFO    Thread 2/3: Trying to download with root and timeout of 1476s from RAL-LCG2-ECHO_DATADISK: mc16_13TeV:DAOD_PHYS.27616103._000003.pool.root.1
2022-03-09 19:18:15,366 INFO    Thread 1/3: Trying to download with root and timeout of 1477s from RAL-LCG2-ECHO_DATADISK: mc16_13TeV:DAOD_PHYS.27616103._000002.pool.root.1
2022-03-09 19:18:15,483 INFO    Thread 0/3: Using PFN: root://xrootd.echo.stfc.ac.uk:1094/atlas:datadisk/rucio/mc16_13TeV/83/9a/DAOD_PHYS.27616103._000001.pool.root.1
2022-03-09 19:18:15,484 INFO    Thread 2/3: Using PFN: root://xrootd.echo.stfc.ac.uk:1094/atlas:datadisk/rucio/mc16_13TeV/d8/17/DAOD_PHYS.27616103._000003.pool.root.1
2022-03-09 19:18:15,485 INFO    Thread 1/3: Using PFN: root://xrootd.echo.stfc.ac.uk:1094/atlas:datadisk/rucio/mc16_13TeV/b7/ed/DAOD_PHYS.27616103._000002.pool.root.1
2022-03-09 19:21:16,053 INFO    Thread 2/3: File mc16_13TeV:DAOD_PHYS.27616103._000003.pool.root.1 successfully downloaded. 708.014 MB in 141.36 seconds = 5.01 MBps
2022-03-09 19:21:16,053 INFO    Thread 2/3: Preparing download of mc16_13TeV:DAOD_PHYS.27616103._000004.pool.root.1
2022-03-09 19:21:16,054 INFO    Thread 2/3: Trying to download with root and timeout of 538s from RAL-LCG2-ECHO_DATADISK: mc16_13TeV:DAOD_PHYS.27616103._000004.pool.root.1
2022-03-09 19:21:16,059 INFO    Thread 0/3: File mc16_13TeV:DAOD_PHYS.27616103._000001.pool.root.1 successfully downloaded. 710.977 MB in 141.36 seconds = 5.03 MBps
2022-03-09 19:21:16,064 INFO    Thread 2/3: Using PFN: root://xrootd.echo.stfc.ac.uk:1094/atlas:datadisk/rucio/mc16_13TeV/f0/ce/DAOD_PHYS.27616103._000004.pool.root.1
2022-03-09 19:21:16,073 INFO    Thread 1/3: File mc16_13TeV:DAOD_PHYS.27616103._000002.pool.root.1 successfully downloaded. 708.572 MB in 113.61 seconds = 6.24 MBps
2022-03-09 19:21:36,565 INFO    Thread 2/3: File mc16_13TeV:DAOD_PHYS.27616103._000004.pool.root.1 successfully downloaded. 239.230 MB in 20.05 seconds = 11.93 MBps
----------------------------------
Download summary
----------------------------------------
DID mc16_13TeV:mc16_13TeV.508547.MGPy8EG_SVJSChan_1500_8.deriv.DAOD_PHYS.e8357_e7400_s3126_r10724_r10726_p4903
Total files (DID):                            4
Total files (filtered):                       4
Downloaded files:                             4
Files already found locally:                  0
Files that cannot be downloaded:              0
[jlai@lxplus708 SVJ_Data]$ ls
mc16_13TeV
mc16_13TeV.508547.MGPy8EG_SVJSChan_1500_8.deriv.DAOD_PHYS.e8357_e7400_s3126_r10724_r10726_p4903
mc16_13TeV.508549.MGPy8EG_SVJSChan_750_8.deriv.DAOD_PHYS.e8357_e7400_s3126_r10201_r10210_p4903
[jlai@lxplus708 SVJ_Data]$ |
```

Status Codes

2.1.2 Status codes

When a new particle is added to the event record, it is assigned a positive status code that describes why it has been added, as follows:

code range	explanation
11 – 19	beam particles
21 – 29	particles of the hardest subprocess
31 – 39	particles of subsequent subprocesses in multiparton interactions
41 – 49	particles produced by initial-state-showers
51 – 59	particles produced by final-state-showers
61 – 69	particles produced by beam-remnant treatment
71 – 79	partons in preparation of hadronization process
81 – 89	primary hadrons produced by hadronization process
91 – 99	particles produced in decay process, or by Bose-Einstein effects