CEPC JOI

Kaili Zhang

IHEP

zhangkl@ihep.ac.cn

Machine Learnings on Jets

- P-CNN
  - https://scipost.org/10.21468/SciPostPhys.7.1.014
- Particle Flow Network
  - https://arxiv.org/abs/1810.05165
  - CEPC@Xiaotian : https://arxiv.org/abs/2410.04465v2
- LundNet
  - https://doi.org/10.1007/jhep03(2021)052
- ParticleNet
  - Arxiv:1902.08570
  - https://github.com/hqucms/ParticleNet

ParticleTransformer

- https://arxiv.org/abs/2202.03772
- https://github.com/jet-universe/particle_transformer
- Platforms: https://github.com/hqucms/weaver-core
- Application on CEPC: 2309.13231 , PRL 132, 221802 (2024)
- Tutorial on CEPC: https://github.com/ZHUYFgit/CEPC-Jet-Origin-Identification
- Inputs from CEPCsoft: /cefs/higgs/zhangkl/AI/datasets
- Inputs from LHC, JetClass: /cefs/higgs/zhangkl/AI/jetclass
- Require higgsgpu group. Request on https://ccsinfo.ihep.ac.cn/
- Follow the tutorial, build the env if you are interested.

ParticleTransformer @ CEPC

https://github.com/ZHUYFgit/CEPC-Jet-Origin-Identification

- Variable list in M11origin.cc
  - Under development to CEPCSW
  - Unit as one jet: 4 momentum, M11 id information……
- Train in Weaver: JetClass_full.yaml
- Submit jobs on IHEP: train_JetClass.sh
- Output: Pred.root: Label and score for each jets.
- Application: onnx format

Inputs for JOI

/cefs/higgs/zhangkl/CEPCSW/Analysis/JetOrigin/src

- Jet->Event;
- PFO->Component;
- Length: 200
- Label: M11

- Current training use truth PID information, in application reco PID will be used.

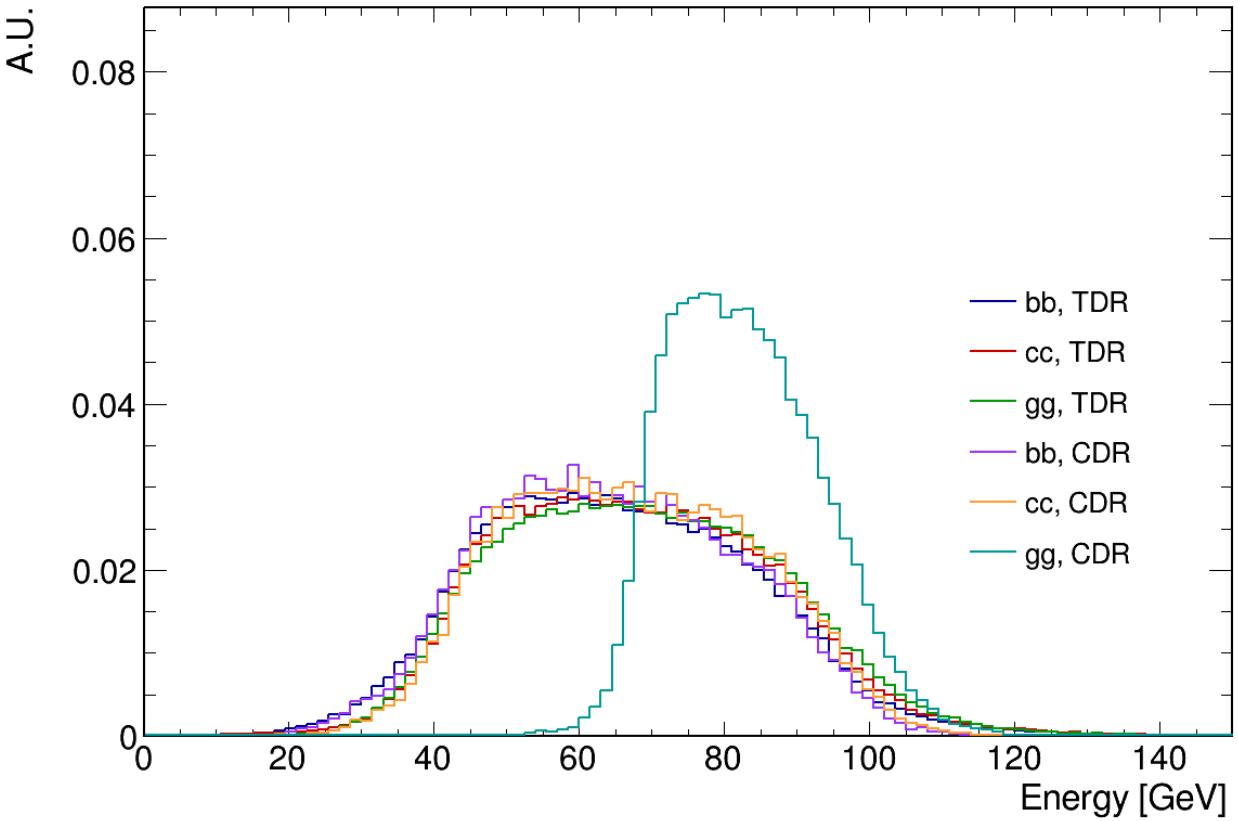
Type	Var	Comment
PFO point distance	$\Delta\phi(pfo, Jet)$	Delta Phi, pfo to jet
	$\Delta\eta(pfo, Jet)$	Delta Eta, pfo to jet
PFO Vector variable	(px, py, pz, E)	4 momentum of PFO
PFO feature variable	P_t^{PFO}, $\log\frac{P_t^{PFO}}{P_t^{jet}}$	Pfo pt and relative pt
	E_t^{PFO}, $\log\frac{E_t^{PFO}}{E_t^{jet}}$	Pfo E and relative E
	$\Delta R(pfo, Jet)$	Delta R, pfo to jet
	N_charge, N_chargeflip	Charge of PFO
	D0, Z0, D0err, Z0err	(if with track) impact parameters
	N_Ecluster, N_Hcluster	
	E_ecal, E_hcal	
	PID	Truth PID type

Variable convention

- Feature variable, Transformer prefer normal distribution with mean ~0, range (-1,1) with cut edge maximum(-5,5).
- (4-momentum vector variable not included)
- Normalization functions like Tanh() used.
-

Jet Energy

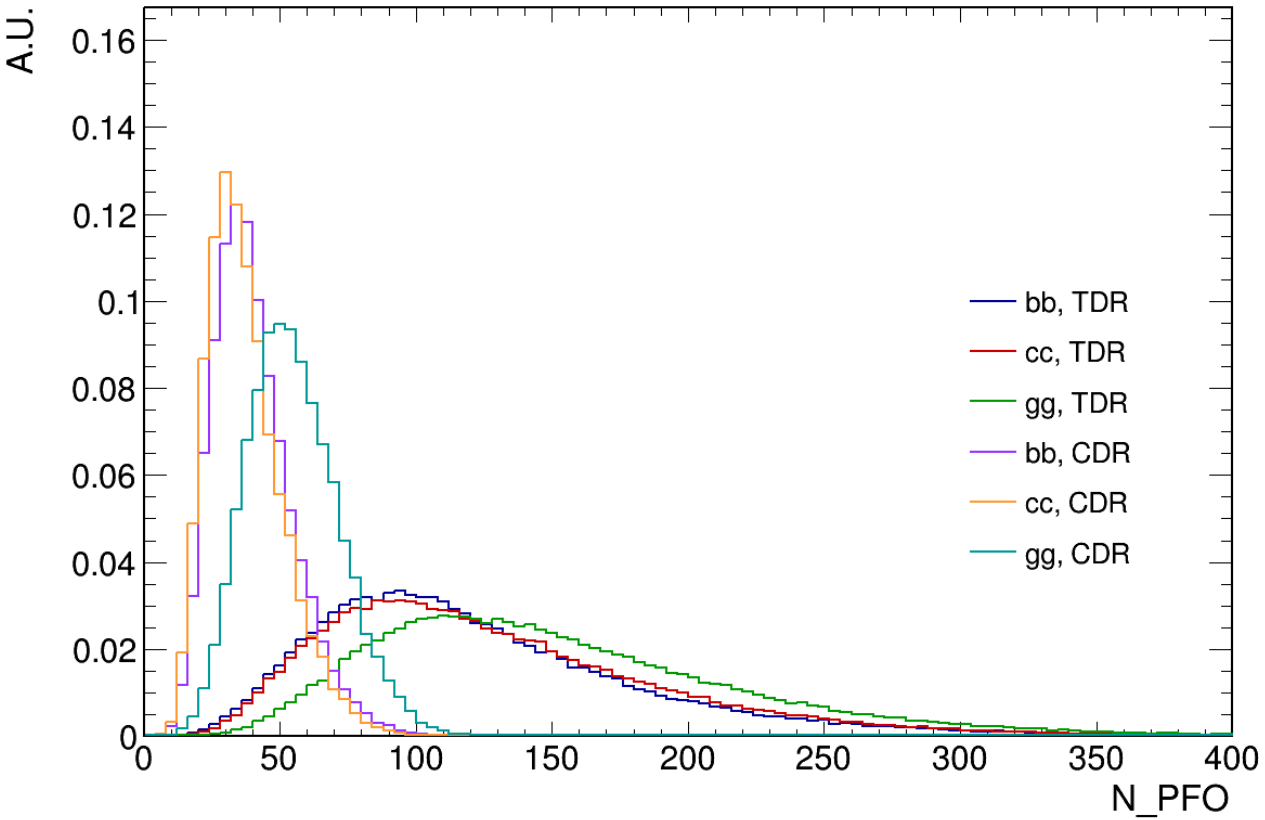
In average, each jet ~66GeV Energy.




```
bbTDR: 65.4526 ±18.8213
ccTDR: 67.7845 ±18.6604
ggTDR: 68.6878 ±18.7631
bbCDR: 64.4942 ±17.0913
ccCDR: 67.0562 ±16.8343
ggCDR: 82.6409 ±10.3267
```

In CDR JOI, one gg event has 2 entries,
choose the leading jet in training. (82GeV) (biased?)

Jet N_PFOs



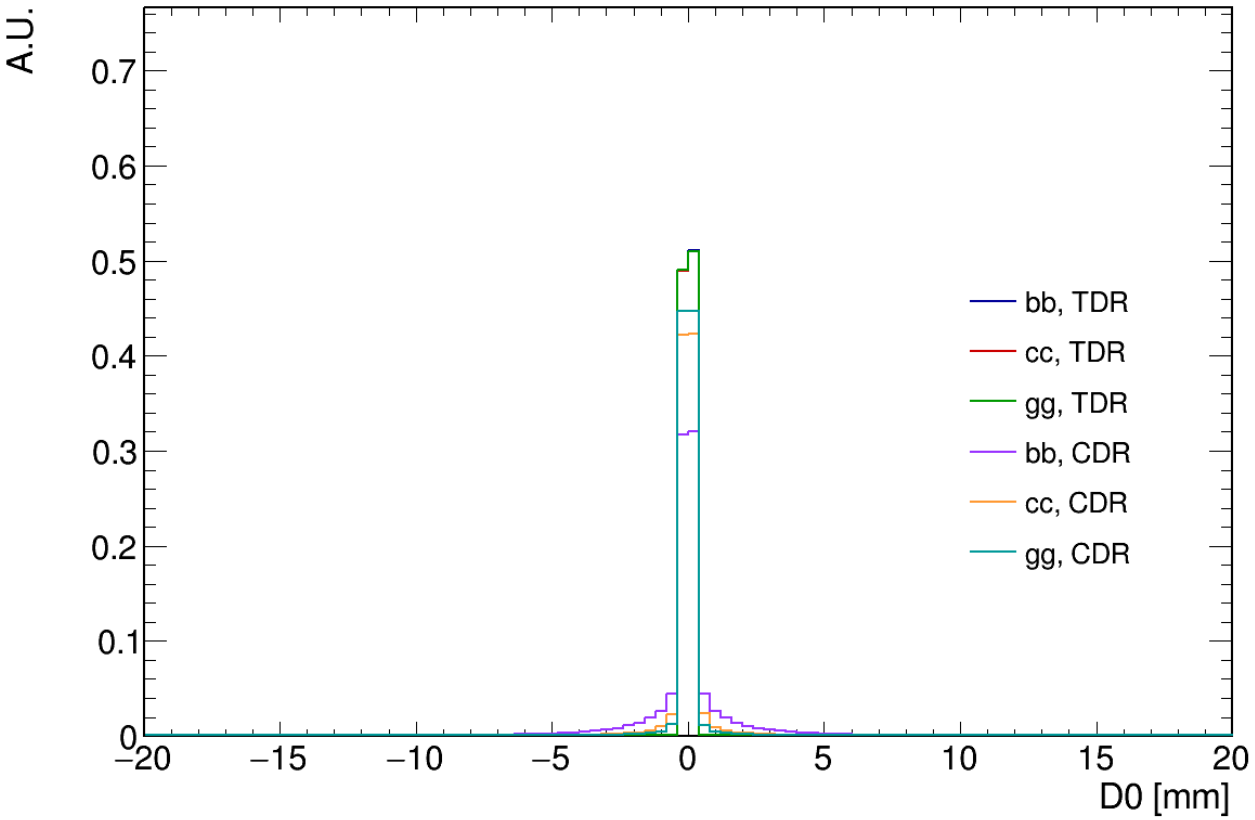
TDR N_PFO with more (broken, nertual)PFOs
TDR cut at 200 (with minimum energy entry ~mev level)
If N_PFO<200, fill with zero.
CDR cut at 128.

```
bbTDR: 122.982 ±59.9375
ccTDR: 127.902 ±61.5878
ggTDR: 150.687 ±65.2885
bbCDR: 39.7391 ±14.6149
ccCDR: 36.9408 ±14.0868
ggCDR: 54.4543 ±16.8207
```

Among these PFOs, ~10-20 PFOs are charged with tracks.

(bug) D0

In the past, wrong trackstate used, not IP but firstHit.
Thus D0 and Z0 with much smaller stddev.
D0 always < 2mm in this case.




```
bbTDR: 0.000138556 ±0.0298339
ccTDR: 9.24658e-05 ±0.0302277
ggTDR: 0.000151907 ±0.028336
bbCDR: 0.000878115 ±3.00354
ccCDR: -0.00749559 ±2.56605
ggCDR: -0.00188047 ±2.49568
```

```
auto trk      = pfo.getTracks(0);
auto trkstate= trk.getTrackStates(0);
Trackmatched = true;
// D0, Z0 mean~0. sigma~0.02. Scale to (-5,5)
D0val = trkstate.D0 ; //
Z0val = trkstate.Z0 ; //
D0err = trkstate.covMatrix[0] ;
Z0err = trkstate.covMatrix[9] ;
```

- Location=1 -> IP;
- But states() should be 0;
- Correct d0 z0 can be found now.
- Sample under re-generation.

```
JetOrigin          INFO For this track, D0: 1.97829 Z0: -4.47115 D0err: 2.64572 Z0err: 20.7097
JetOrigin          INFO For this track, D0: 0.620478 Z0: 0.852513 D0err: 0.000183923 Z0err: 0.000225732
JetOrigin          INFO For this track, D0: -0.643742 Z0: -0.6305 D0err: 2.26723e-05 Z0err: 2.57715e-05
JetOrigin          INFO For this track, D0: -0.566986 Z0: 0.263159 D0err: 2.48116e-05 Z0err: 2.9748e-05
JetOrigin          INFO For this track, D0: -1.06805 Z0: 1.25654 D0err: 2.78676e-05 Z0err: 3.86399e-05
JetOrigin          INFO For this track, D0: 0.373717 Z0: -0.64295 D0err: 2.77541e-05 Z0err: 3.3744e-05
JetOrigin          INFO For this track, D0: 0.368966 Z0: -0.763947 D0err: 2.95027e-05 Z0err: 3.57912e-05
JetOrigin          INFO For this track, D0: 0.987603 Z0: 1.1879 D0err: 5.5163e-05 Z0err: 9.12332e-05
JetOrigin          INFO For this track, D0: 0.60443 Z0: -0.63053 D0err: 0.0231788 Z0err: 0.0490743
JetOrigin          INFO For this track, D0: 4.36658 Z0: -0.596456 D0err: 0.126764 Z0err: 0.20802
JetOrigin          INFO For this track, D0: -0.164952 Z0: 0.337603 D0err: 2.05198e-05 Z0err: 3.15671e-05
JetOrigin          INFO For this track, D0: 0.377288 Z0: -1.04719 D0err: 2.17551e-05 Z0err: 3.08977e-05
JetOrigin          INFO For this track, D0: 0.359068 Z0: 3.53151 D0err: 0.000226511 Z0err: 0.00076005
JetOrigin          INFO For this track, D0: 0.144355 Z0: -0.27682 D0err: 3.70832e-05 Z0err: 4.03035e-05
JetOrigin          INFO For this track, D0: -2.21977 Z0: -0.0815289 D0err: 0.00311425 Z0err: 0.00431757
JetOrigin          INFO For this track, D0: 12.1517 Z0: -160.117 D0err: 1.75873 Z0err: 3.04779
JetOrigin          INFO For this track, D0: -0.00969746 Z0: 0.124896 D0err: 0.00136036 Z0err: 0.00280031
JetOrigin          INFO For this track, D0: -0.0577078 Z0: -0.152886 D0err: 3.07977e-05 Z0err: 6.40223e-05
JetOrigin          INFO For this track, D0: 19.8924 Z0: 57.1801 D0err: 0.0729774 Z0err: 0.212157
JetOrigin          INFO For this track, D0: -0.00319428 Z0: 0.117056 D0err: 5.36309e-05 Z0err: 6.12757e-05
JetOrigin          INFO For this track, D0: -0.00954015 Z0: 0.164424 D0err: 0.0013619 Z0err: 0.00701892
JetOrigin          INFO For this track, D0: 1.06711 Z0: 0.901302 D0err: 0.00048297 Z0err: 0.00267738
JetOrigin          INFO For this track, D0: 0.525138 Z0: 0.998079 D0err: 4.61002e-05 Z0err: 5.14031e-05
JetOrigin          INFO For this track, D0: -0.532135 Z0: -1.57923 D0err: 0.000226511 Z0err: 0.000238439
JetOrigin          INFO For this track, D0: -0.0306385 Z0: -0.101822 D0err: 0.00105705 Z0err: 0.00260442
JetOrigin          INFO For this track, D0: 0.00689215 Z0: -0.207277 D0err: 1.53168e-05 Z0err: 1.60648e-05
JetOrigin          INFO For this track, D0: 0.0757516 Z0: 0.321308 D0err: 4.63289e-05 Z0err: 5.20079e-05
JetOrigin          INFO For this track, D0: 0.42437 Z0: -0.358423 D0err: 5.02048e-05 Z0err: 5.11278e-05
JetOrigin          INFO For this track, D0: -0.216463 Z0: 0.605704 D0err: 0.000143601 Z0err: 0.000152609
```

Photon reco

In past H->yy, use leading PFO as the photon. With resolution ~0.47%.
But if use all PFO to reconstruct higgs in vvyy: 0.21% resolution.
->Photon p resolution ~0.1%. Comparable with muon tracks.
->Large improvements can be done in photon reco.