CEPC Workshop

H->bb/cc/gg Analysis With Machine Learning

Gang Li (IHEP)

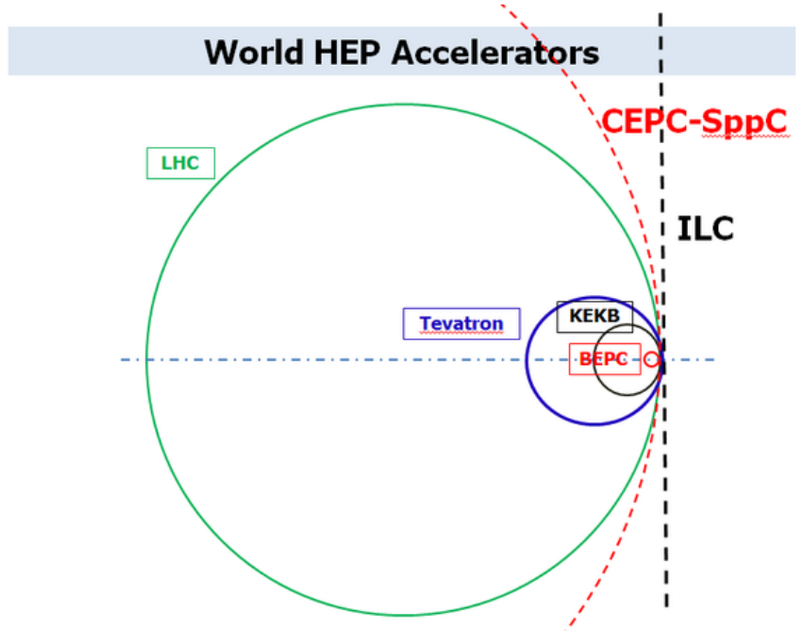
Yu Bai (SEU, Nanjing)

Jianpeng Deng , Jianyu Huang, Weihan Tan,

Heyu Meng, Ke Wang (all from SEU)

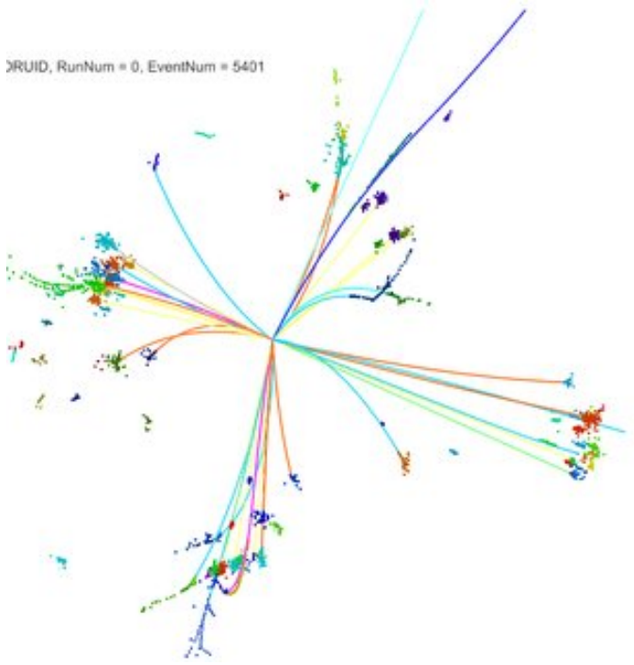
Study Group

November 8, 2021

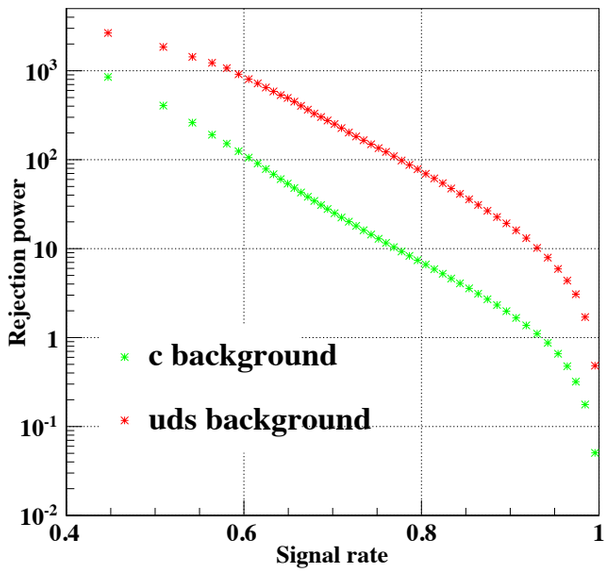


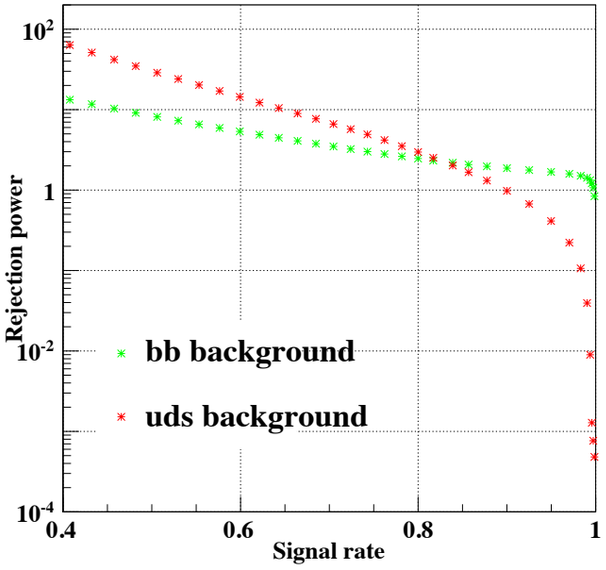
Introduction: Hadronic FS

A hadronic event



Flavor tagging performance





$$L_{qq} = \frac{qq\ pair}{qq\ pair\ +\ neither\ is\ q} = \frac{x_q^1\ x_q^2}{x_q^1 x_q^2 + (1-x_q^1)(1-x_q^1)}\ (qq = bb, cc)$$

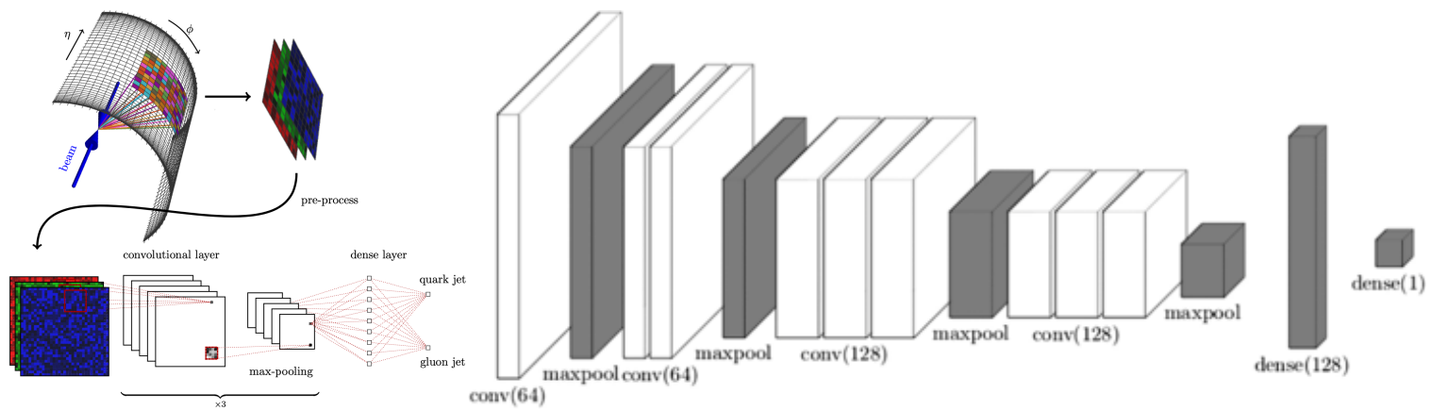
A high performance algorithm to deal with the hadronic final states identification is import to Higgs hadronic decay study

Introduction: Machine Learning in the study of Hadronic FS

Earlier study in gluon-quark separation with CNN:

[JHEP 01(2017) 110]

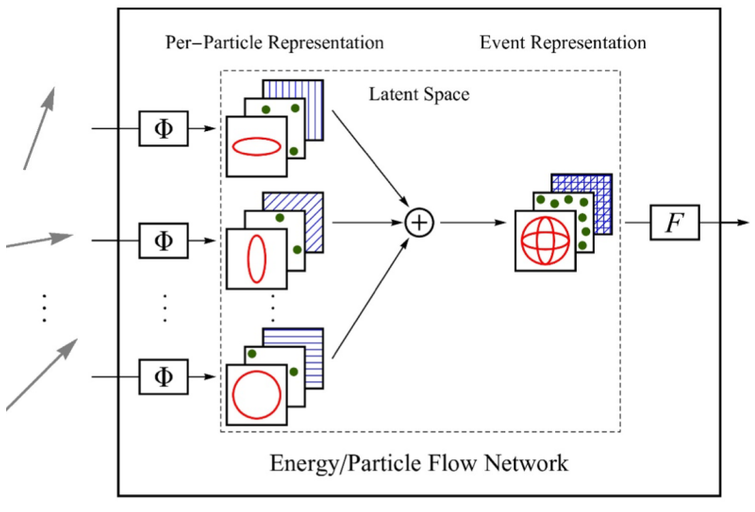
[PRD100, 116013]

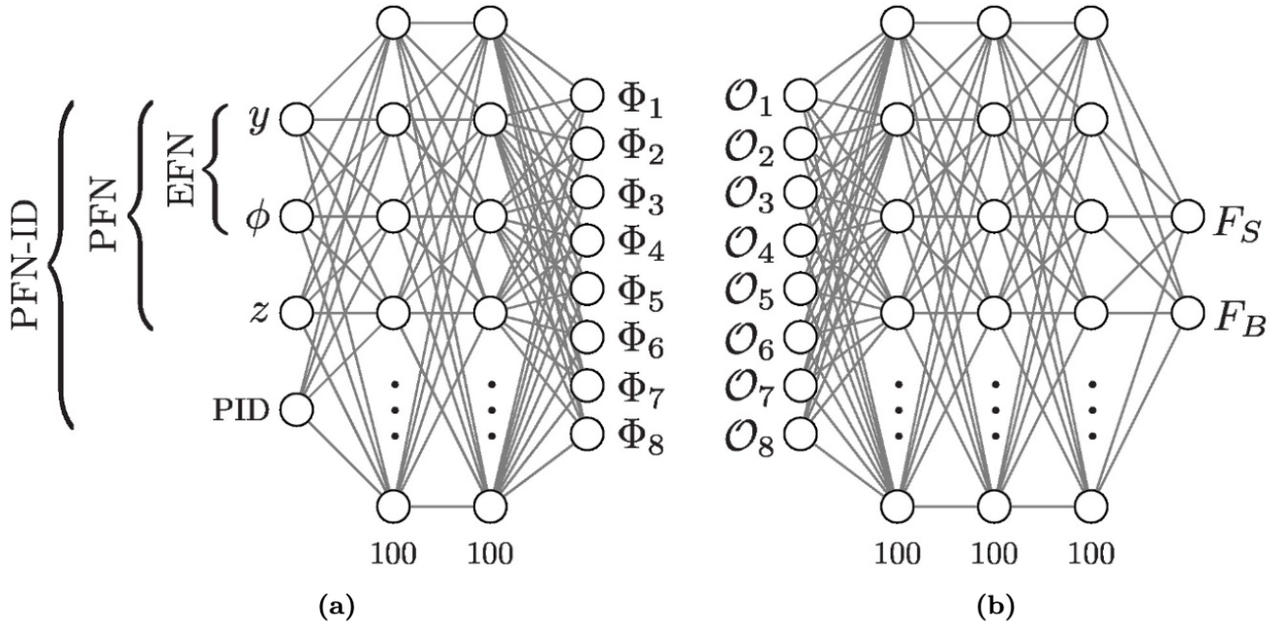


Study with Deep sets:

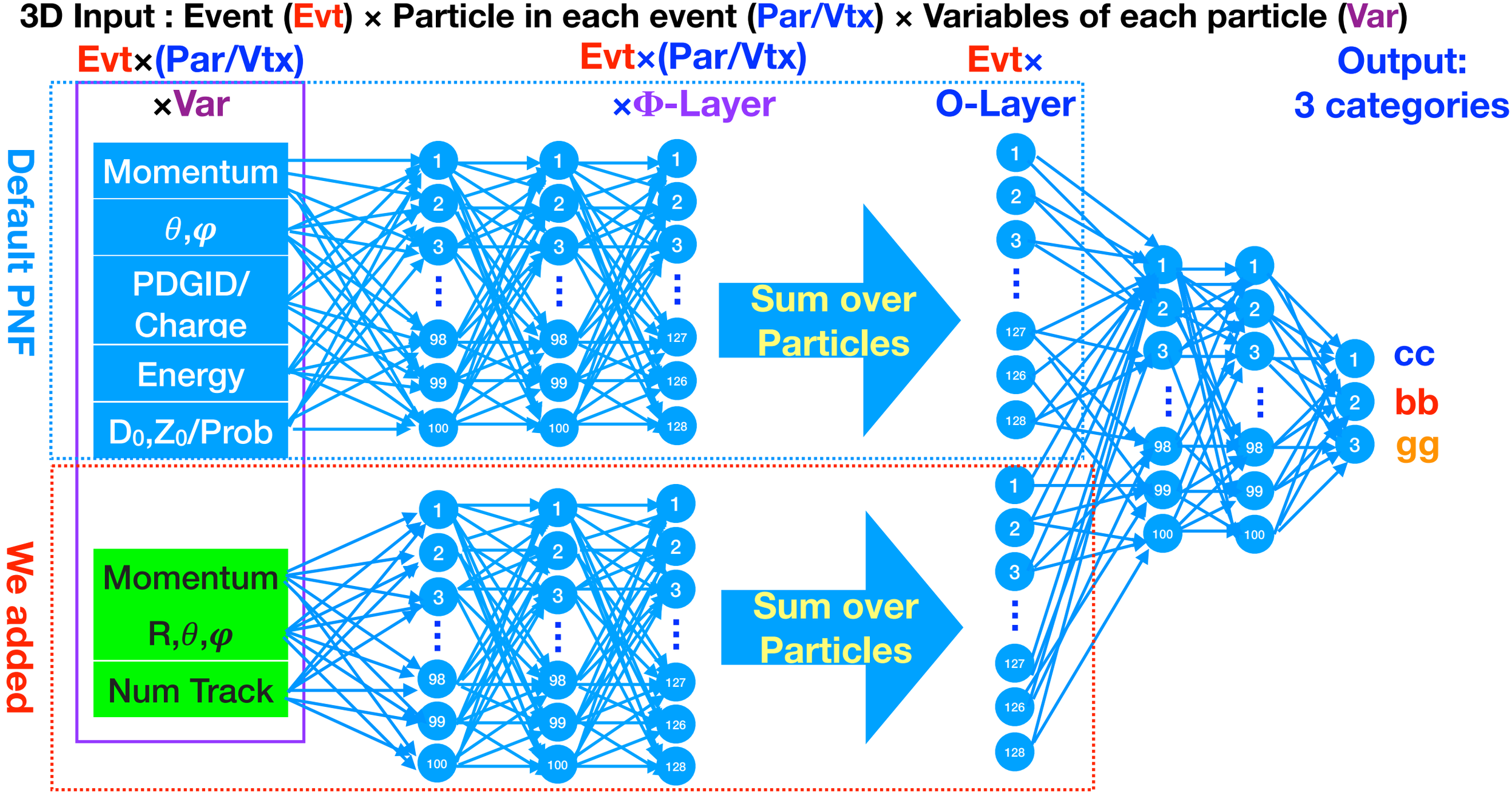
[Deepest theory]

Deep sets for particle jets: [JHEP01(2019)121]





Architecture of Layers



Samples and Selections

- √s = 240 GeV sample: μμH->bb/cc/gg/ww/zz/Everyting full simulation sample:
  1 Million eats /sample,

- √s = 240 GeV: zz->μμ + qq sample

- H->bb/cc/gg for training, H->ww/zz/everything and ZZ-> μμ + qq for evaluation

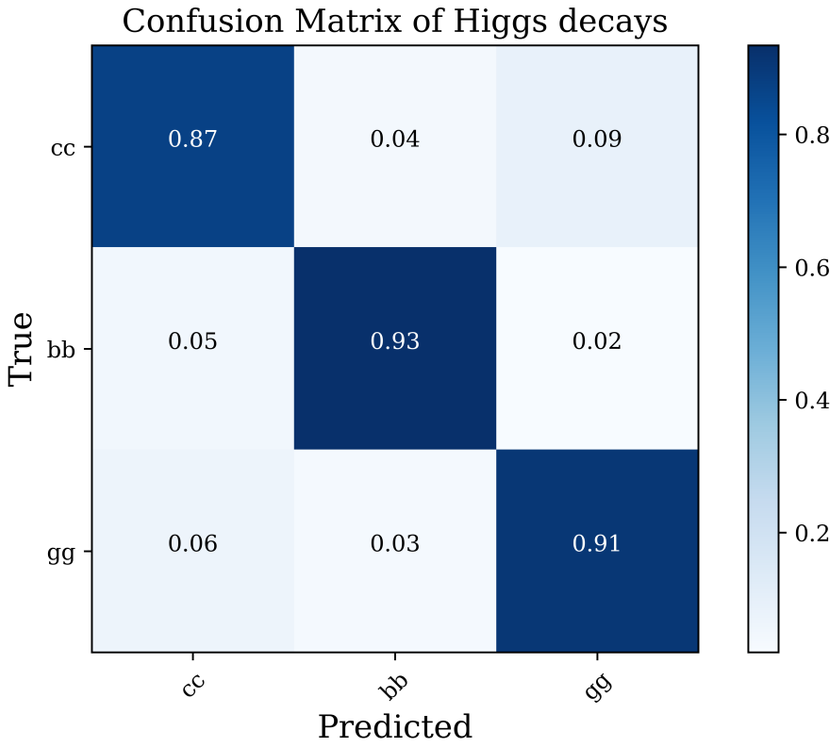
- Loosen cuts applied (higher statistics for comparison)

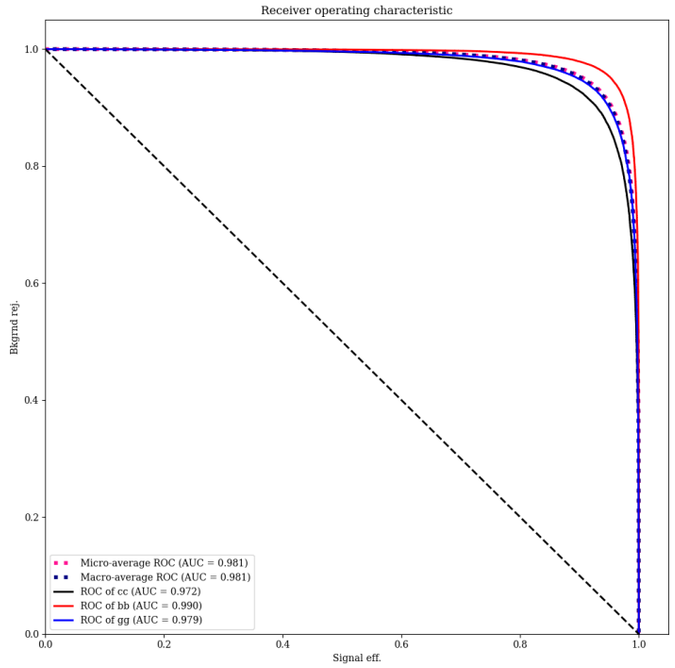
  - a pair of $\mu^+\mu^-$, $\mu^+\mu^-$ recoil mass 120 GeV-140 GeV

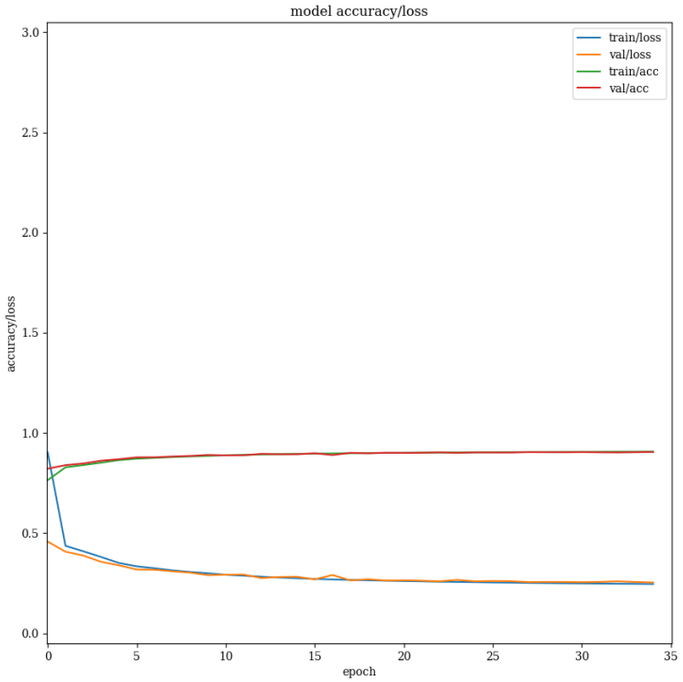
  - jet invariant mass 75 GeV-140 GeV, at least 20 PFOs with energy > 0.4 GeV
    in jet pair

- After selection μμH->bb/cc/gg contain ~700 k events for training+validation+test

Performance





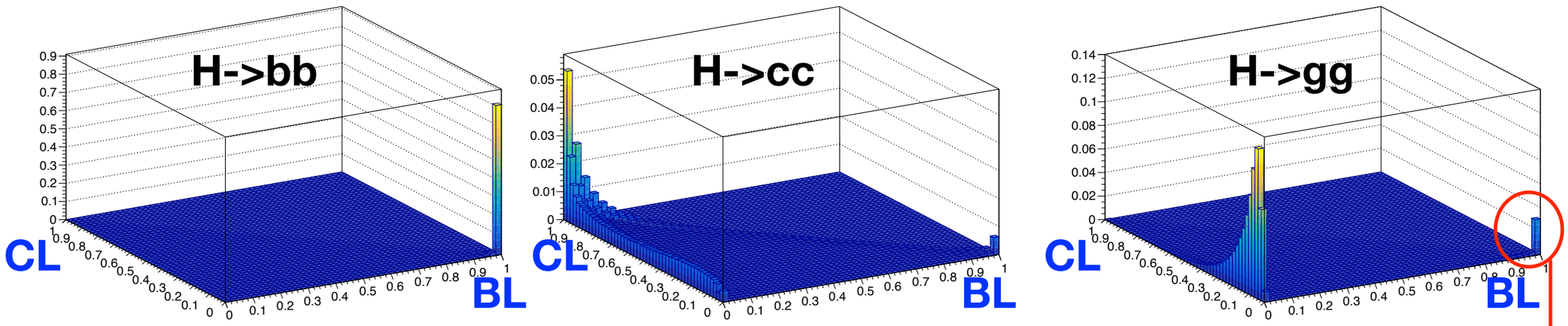


After 35 epochs of training Performance is good

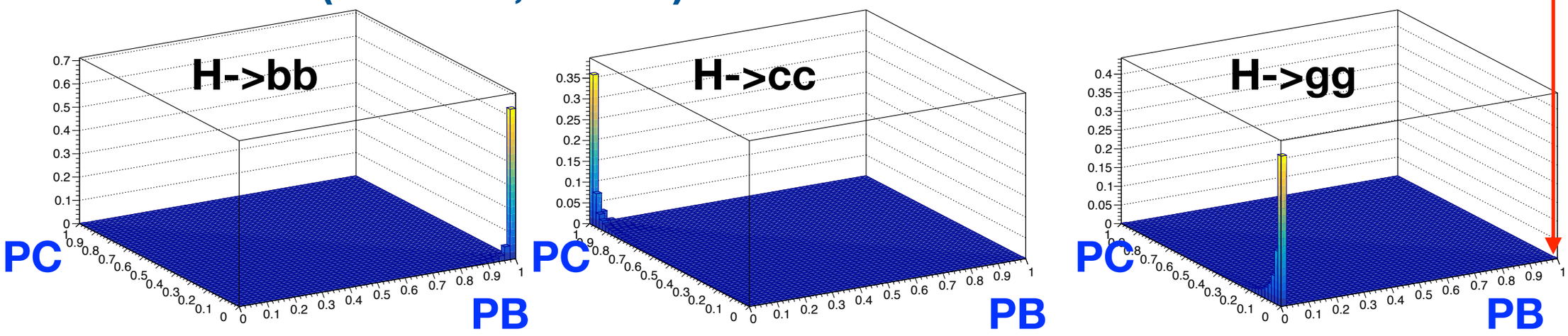
- **Average accuracy over 90%**
- **No over training found**

Comparison between PFN and Conventional Flavor Tagging Variables

Conventional Flavor Tagging: (B/C-likeness, BL/CL)



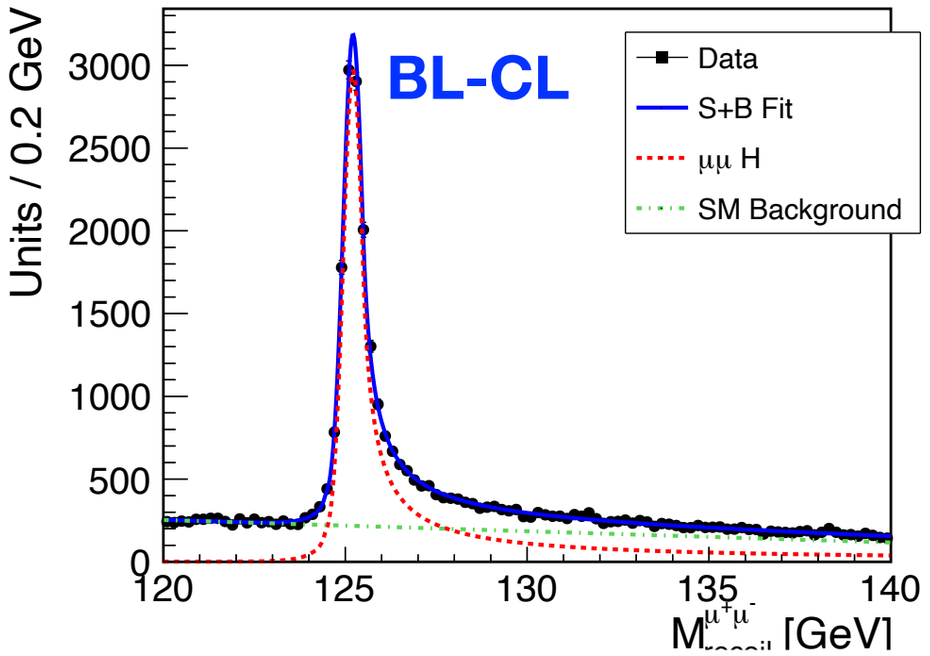
PFN Results: (PFN-B/C, PB-PC)

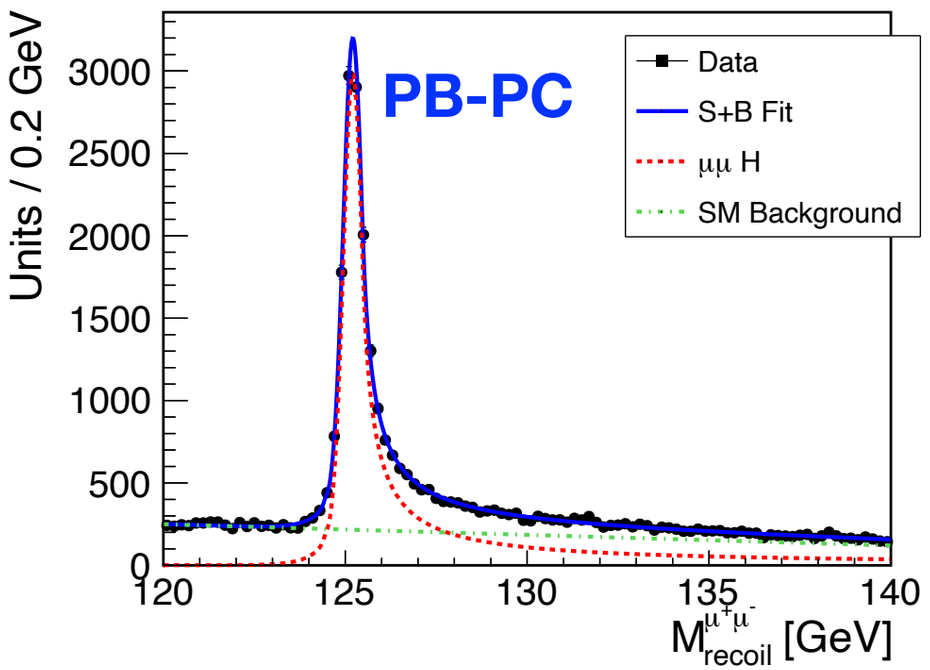


- **PFN provides much better c/g-tag separation**
- **PFN not only learns FT, but learns kinematic information as well**

Results of Fit:I

	Using BL-CL	Using Pb-Pc
$\mu_{\sigma(H->bb)}$	1.001±0.015	0.994±0.011
$\mu_{\sigma(H->cc)}$	0.986±0.120	1.031±0.070
$\mu_{\sigma(H->gg)}$	0.974±0.063	1.004±0.033
a	-1.005±0.033	-0.973±0.026
n	0.958±0.039	0.986±0.023
σ	0.273±0.006	0.269±0.005
mean	125.2±0.0	125.2±0.0

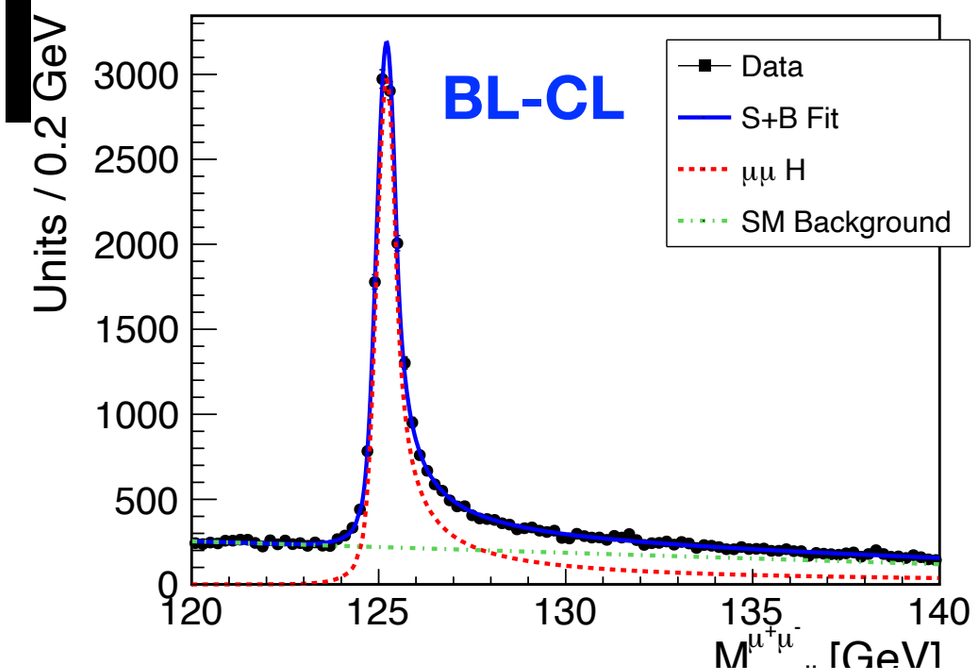


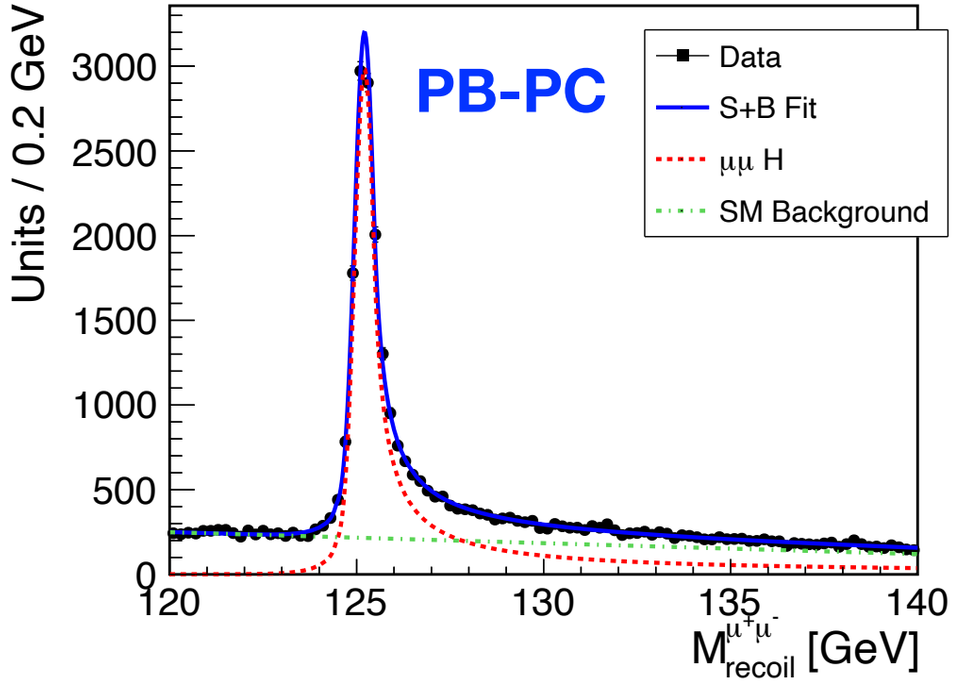


- **Precision improved with Pb-Pc, especially for H->cc/gg**
- **These data are fitted with zz->mumu+qq templates only in fitted M-Recoil region(120 GeV - 140 GeV)**

Results of Fit: II

	Using BL-CL	Using Pb-Pc
$\mu_{\sigma(H->bb)}$	0.995±0.08	0.996±0.011
$\mu_{\sigma(H->cc)}$	0.951±0.125	1.020±0.071
$\mu_{\sigma(H->gg)}$	0.971±0.066	1.016±0.033
a	-1.000±0.035	-0.972±0.026
n	0.963±0.044	0.984±0.024
σ	0.273±0.006	0.268±0.005
mean	125.2±0.0	125.2±0.0





- **Precision improved with Pb-Pc, especially for H->cc/gg**
- **These data are fitted with zz->mumu+qq templates in full M-Recoil region**

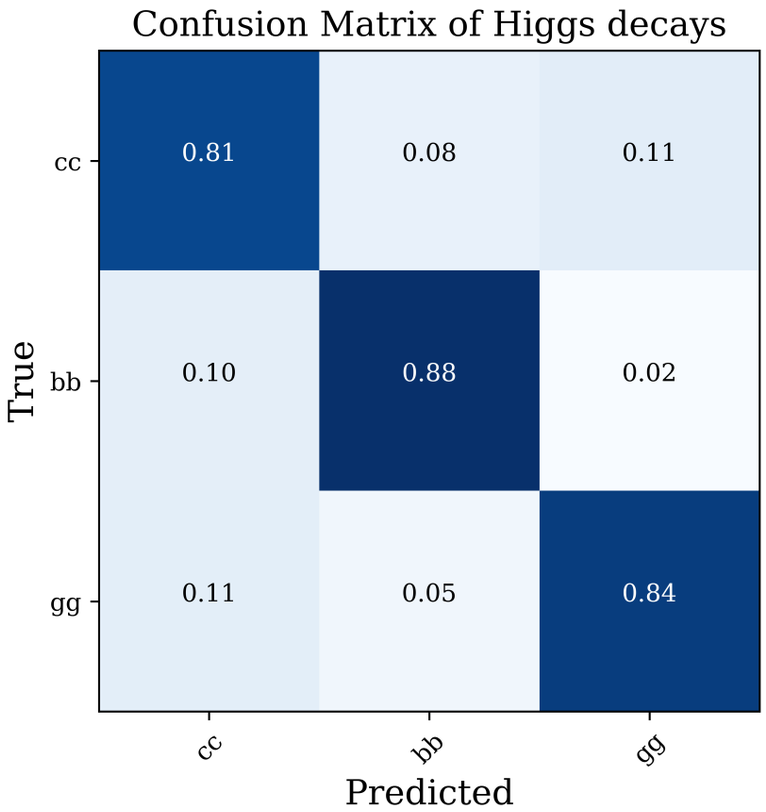
Summary

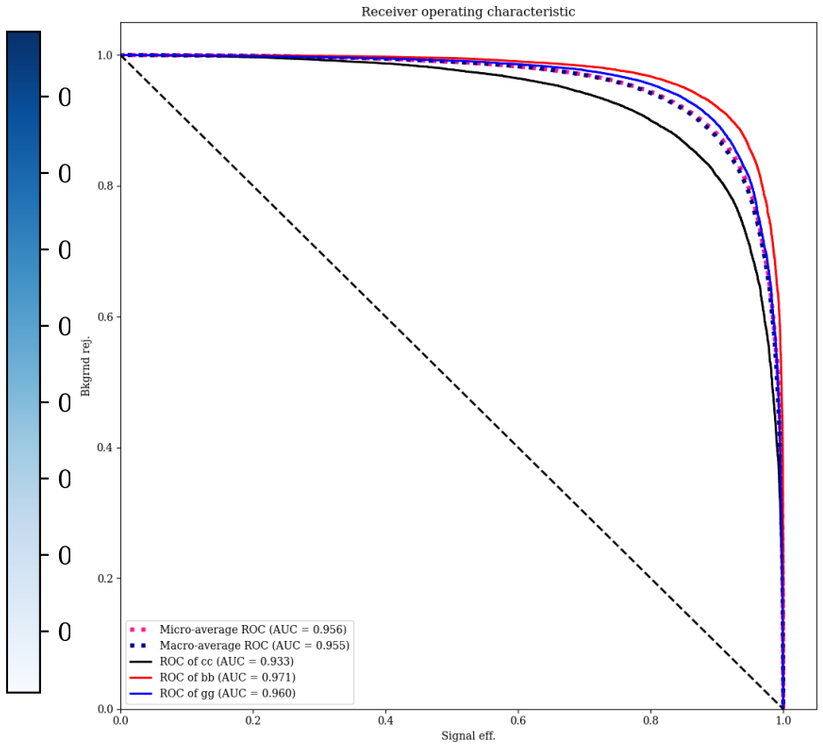
- **Machine learning are useful to deal with complex hadronic final states**
- **In the study presented, we use an architecture based on deep sets/PFN theory**
- **We also add additional structures for vertices**
- **The performance from training is improved, especially for c-tag and c/g separation**
- **The PFN seems learn kinematics deeply, good but also need to be careful**
- **This method can be applied to more complex final states (qqH datasets in prep -aration)**

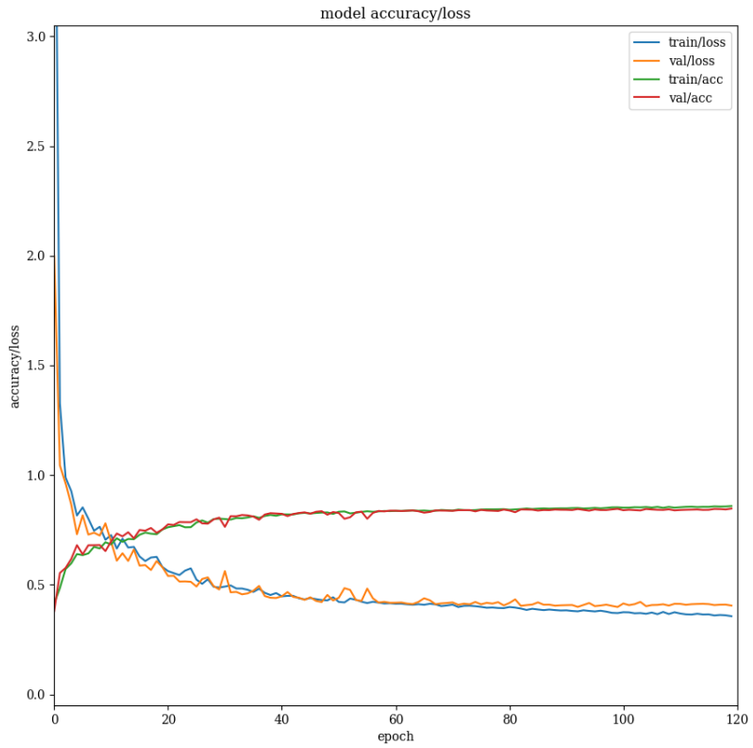
Backup

Some earlier Results

First try on full situation

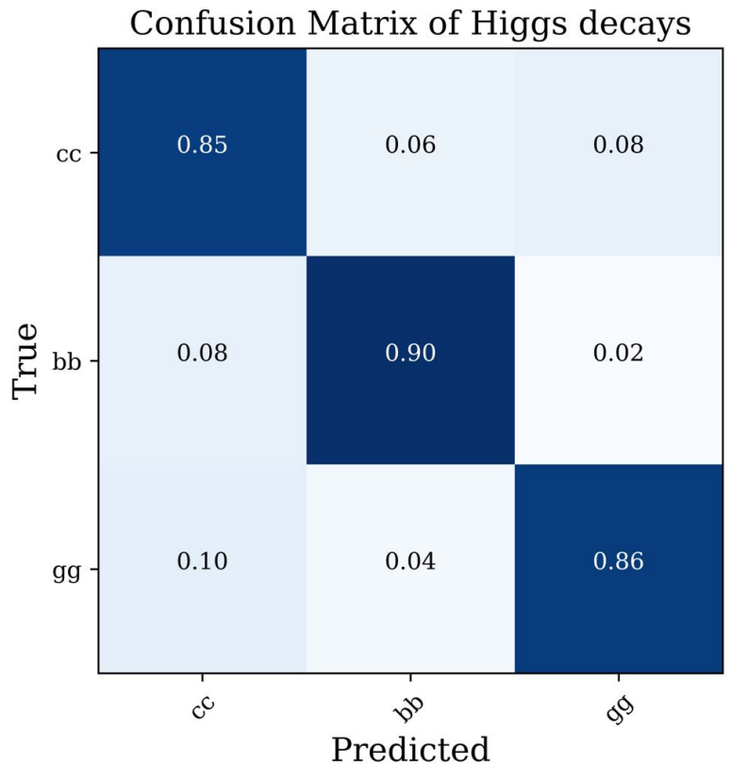


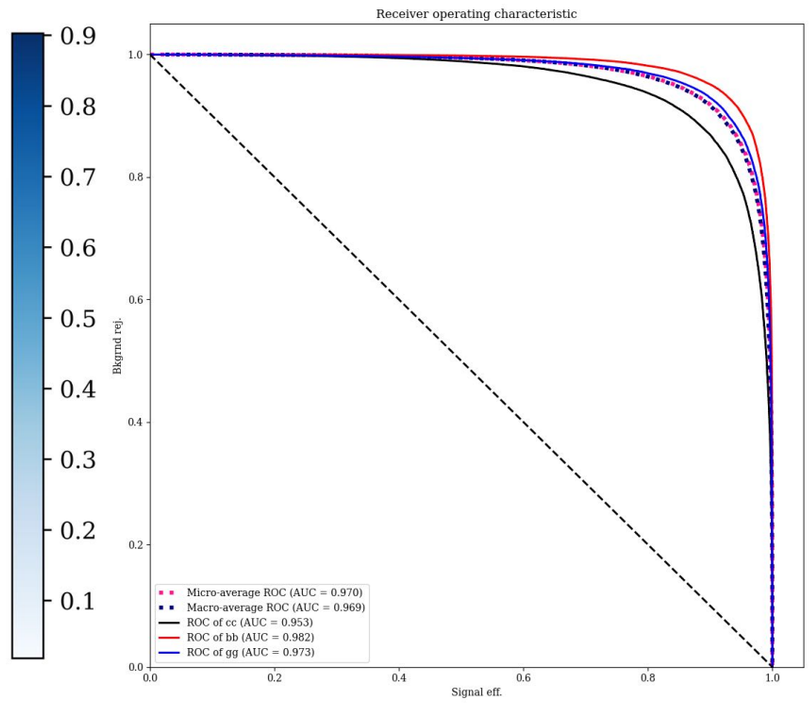


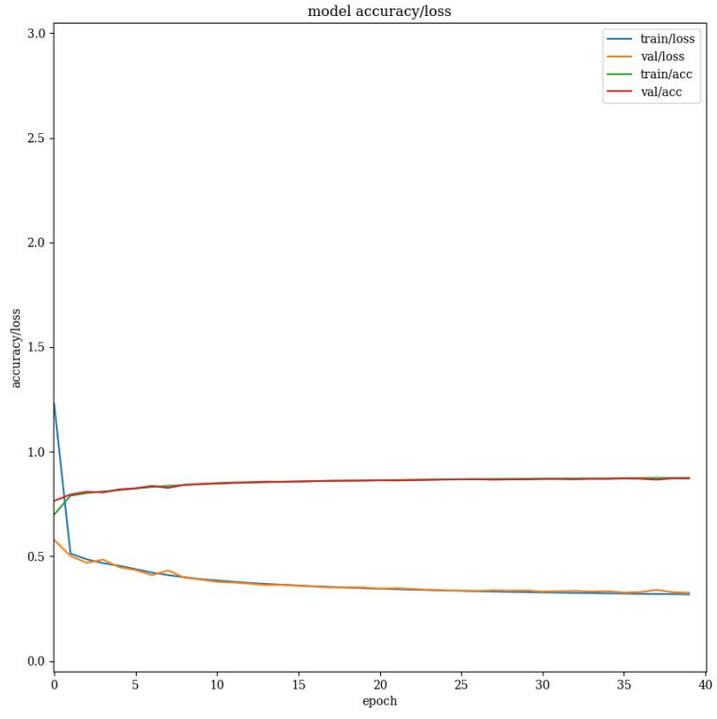


Some earlier Results

After add track probability

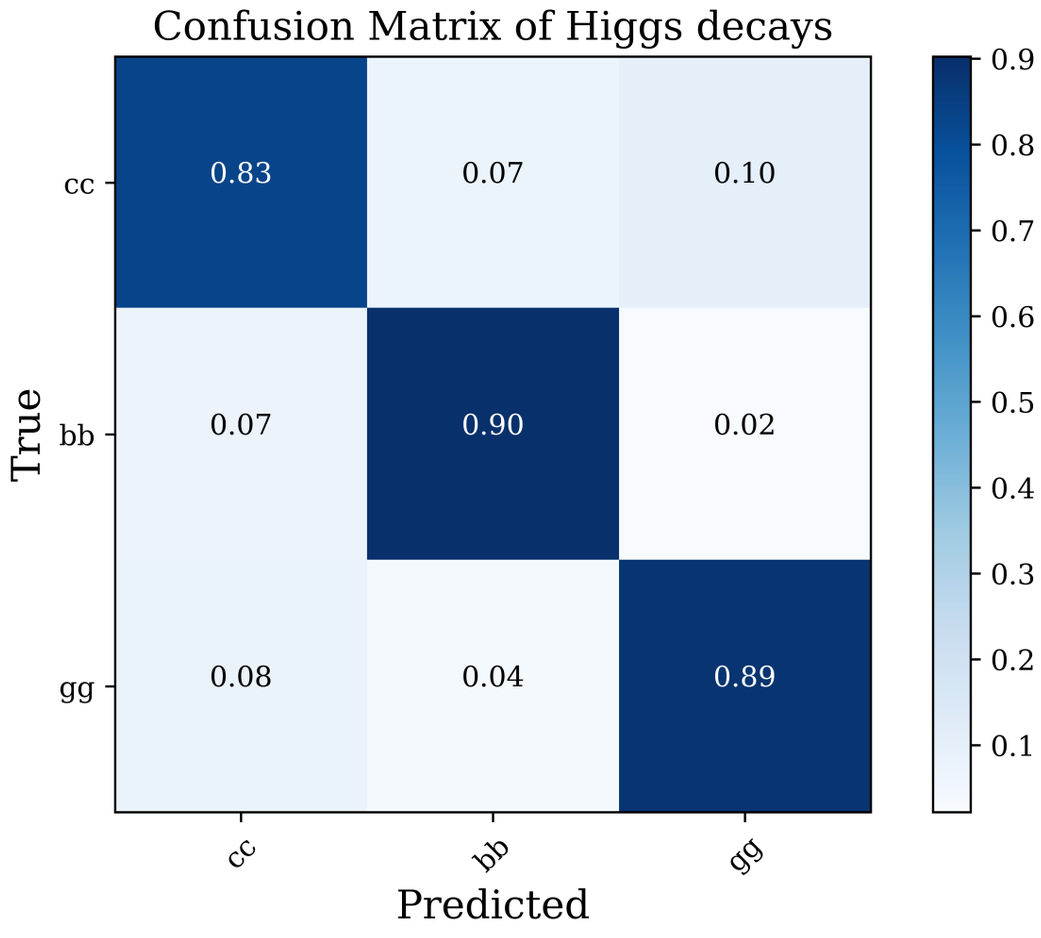


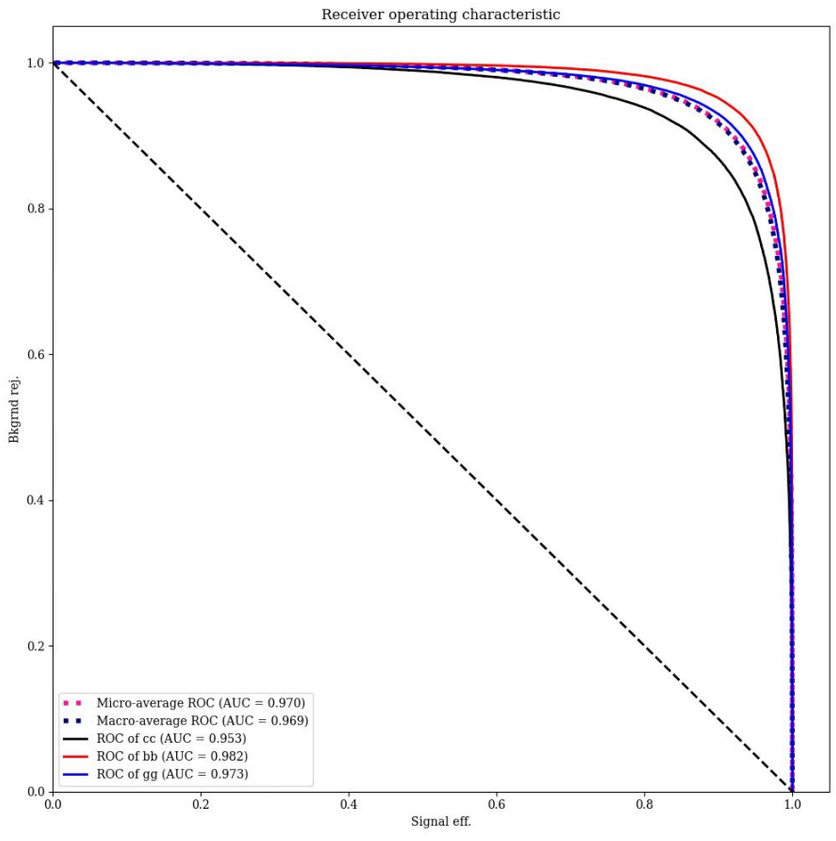




Some earlier Results

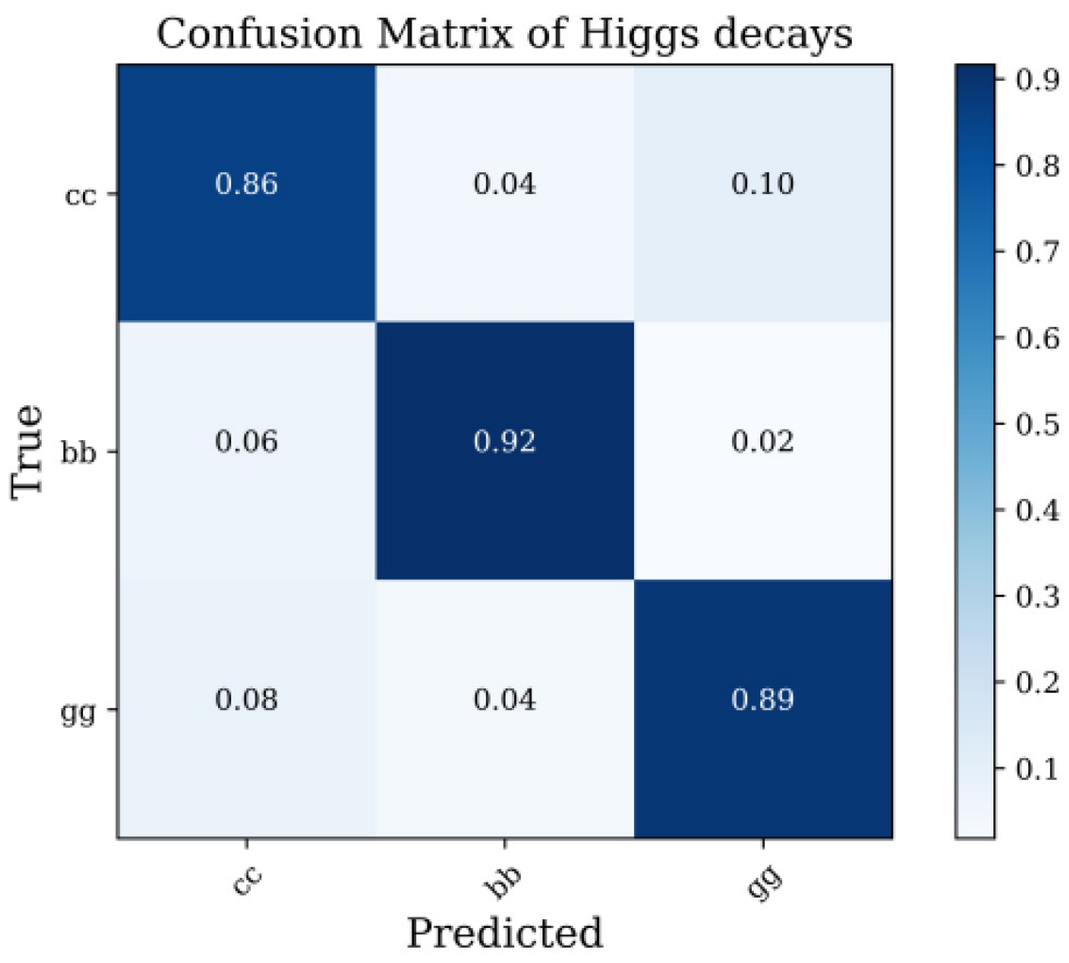
After add track probability, trained in high statistic sample

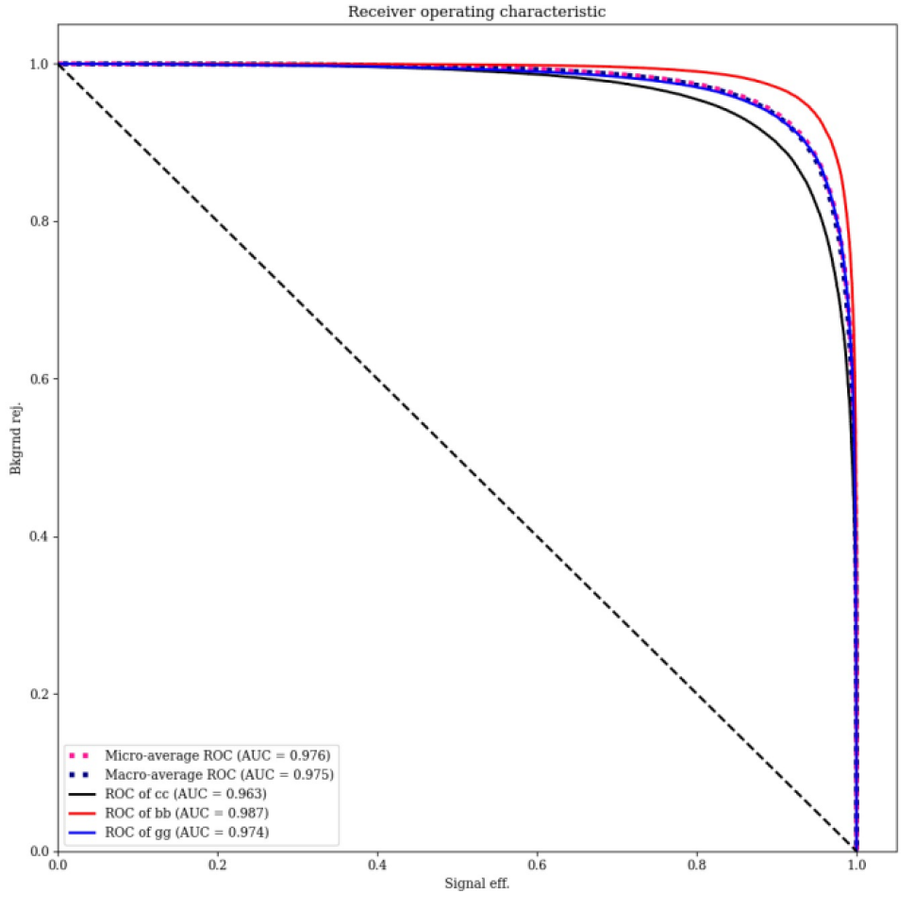




Some earlier Results

After add track secondary vertex(mixed with particles), trained in high statistic sample, no pigid for particles yet





Thank You!